In Partnership with

IEEE OpenAI

Smart Algorithms & Models: Your Reach to Advanced Dialogues

Intersection of Artificial Intelligence and Education

BOOK OF PROCEEDINGS

OPEN ACCESS

VOL. 01

5 - 6 October 2026
Seville, Spain & Virtual

Available online at: https://2026.samyrad.com/samyrad-open-access-proceedings/

DOI: HTTPS://DOI.ORG/10.68286/SAMYRAD.2026.PROCEEDINGS

Horizonne Views

INTELLIGENT NUDGE

WEBSITE: HTTPS://2026.SAMYRAD.COM/

S A M Y R A D 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues

Artículos de Acceso Abierto - Edición 2026

La intersección de la inteligencia artificial y la educación

5 y 6 de octubre de 2026
Sevilla, España
Institución anfitriona: Universidad CEU Fernando III

Editora del volumen
Samira Hosseini

Intelligent Nudge
31 de octubre de 2026

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Información de la publicación

Título de la publicación: SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues

Subtítulo: Artículos de Acceso Abierto - Edición 2026

Editora del volumen: Samira Hosseini

Publicado por: Samira Hosseini, bajo el sello editorial Intelligent Nudge

Fecha de publicación: 31 de octubre de 2026

Periodicidad: Anual, Edición: 2026, Artículos: Inglés

Formato: Publicación electrónica en PDF y de acceso abierto

Congreso: 5 y 6 de octubre de 2026, Universidad CEU Fernando III, Sevilla, España

Sitio web de la publicación: https://2026.samyrad.com/samyrad-open-access-proceedings/

Sitio web de la editorial: https://intelligentnudge.com/

Contacto: support@intelligentnudge.com

DOI: https://doi.org/10.68286/samyrad.2026.proceedings

Derechos y condiciones de uso

© 2026 Intelligent Nudge. Los derechos de autor y las condiciones de reutilización de cada contribución se rigen por los acuerdos aplicables con sus autores y por la licencia que, en su caso, se indique en cada artículo. Los materiales atribuidos a terceros están sujetos a los derechos de sus respectivos titulares. El acceso gratuito en línea no otorga, por sí mismo, autorización para reutilizar los contenidos sin restricciones.

Referencia del volumen

Hosseini, S. (Ed.). (2026). SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues: Open Access Articles — 2026 Edition. Intelligent Nudge.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Cintillo legal

SAMYRAD: SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES es una publicación científica de periodicidad anual, edición 2026, en formato electrónico y de acceso abierto, editada y publicada por Samira Hosseini, quien funge como editora responsable, bajo el sello editorial Intelligent Nudge, domicilio de la publicación: Danis núm. 613, colonia Paseo del Vergel, C. P. 64988, Monterrey, Nuevo León, México, teléfono: +52 81 23267576, correo electrónico: support@intelligentnudge.com, sitio web de la editorial: https://intelligentnudge.com/, enlace de acceso a la publicación: https://2026.samyrad.com/samyrad-open-access-proceedings/, Reserva de Derechos al Uso Exclusivo núm. 04-2026-100115181300-102, otorgada por el Instituto Nacional del Derecho de Autor (INDAUTOR), ISSN: pendiente de asignación por el Centro Nacional ISSN México, idioma de los artículos: inglés, páginas preliminares: español e inglés, medio de publicación: electrónico (internet), formato: PDF, fecha de publicación de la edición 2026: 31 de octubre de 2026, responsable de la última actualización de este número: Samira Hosseini, con domicilio en Danis núm. 613, colonia Paseo del Vergel, C. P. 64988, Monterrey, Nuevo León, México, fecha de última modificación: 31 de octubre de 2026, tamaño del archivo: 8.1 MB.

Presentación editorial

Este volumen de SAMYRAD 2026 reúne una selección de artículos completos, revisados por pares, en una colección académica de acceso abierto. Publicado por Intelligent Nudge, el volumen documenta investigaciones y reflexiones críticas sobre la inteligencia artificial en la educación, la formación profesional y el aprendizaje a lo largo de la vida.

La colección incluye trabajos que examinan la práctica educativa y las condiciones institucionales, sociales y éticas en las que se utiliza la tecnología. Invitamos a leer cada contribución a la luz de sus métodos, evidencias y limitaciones declaradas. Su publicación permite someter estos trabajos al análisis, la discusión y la investigación posterior.

Las páginas de cada artículo proporcionan información bibliográfica y enlaces DOI. Para citar un hallazgo o argumento específico, se recomienda utilizar la referencia del artículo correspondiente. Esta publicación anual está concebida para continuar en ediciones sucesivas bajo su título registrado.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Mensaje de la presidenta del congreso

Al concluir un congreso comienza una responsabilidad importante. Las ideas compartidas deben llegar también a quienes no pudieron asistir, y los argumentos que suscitaron debate deben seguir abiertos al análisis. Estos artículos de acceso abierto son nuestra aportación a esa tarea.

En SAMYRAD 2026, celebrado en Sevilla, examinamos cómo la inteligencia artificial está transformando las condiciones en las que aprendemos, enseñamos y nos preparamos para la vida profesional. Una pregunta se mantuvo en el centro de las conversaciones: ¿qué evidencias deben orientar nuestras decisiones? La velocidad del desarrollo tecnológico vuelve urgente esta pregunta, pero no sustituye la investigación rigurosa.

Los artículos reunidos aquí ofrecen distintas formas de abordarla. Su valor reside en los problemas concretos que investigan, los razonamientos que permiten examinar y las preguntas que dejan abiertas. Espero que este volumen sea útil para quienes ponen a prueba nuevas propuestas, para docentes que analizan su práctica y para quienes buscan bases más sólidas para tomar decisiones institucionales.

El acceso abierto es una decisión deliberada de nuestro proyecto editorial. Permite que una conversación iniciada en Sevilla llegue a colegas que trabajan con otros recursos y en contextos educativos distintos. También abre la posibilidad de contrastar resultados, perfeccionar métodos y desarrollar colaboraciones más allá del encuentro original.

Agradezco a las autoras y los autores por confiarnos sus trabajos, así como a quienes participaron en la revisión por pares y en las tareas editoriales por el tiempo y el criterio dedicados a su desarrollo. Agradezco también a nuestra institución anfitriona, la Universidad CEU Fernando III, y a las personas e instituciones reconocidas en estas páginas por hacer posible el primer congreso.

SAMYRAD continuará como un foro anual, y sus publicaciones académicas permitirán construir un registro que crezca con cada edición. Espero que quienes consulten este volumen encuentren ideas para investigar y llevar a la práctica, y que aporten sus propias evidencias y experiencias a las conversaciones que siguen.

Gracias por leer, cuestionar y contribuir a este esfuerzo compartido.

Samira Hosseini, PhD
Presidenta de SAMYRAD 2026

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Agradecimientos

La primera edición del congreso SAMYRAD tuvo como sede la Universidad CEU Fernando III, en Sevilla, España. Agradecemos a la universidad y a su personal las instalaciones, la coordinación y el apoyo que permitieron reunir a los participantes y compartir sus trabajos.

Reconocemos las contribuciones al congreso de IEEE, incluidas la Sección España, la Región 8 y la IEEE Education Society; la Red Interuniversitaria Euroamericana de Investigación sobre Competencias Mediáticas (Alfamed); la Universidad de los Andes; la World Innovation Summit for Education (WISE), iniciativa de Qatar Foundation; la University of Calcutta; Research About Learning in Higher Education (REAL-HE), de Oxford Brookes University; CENTRUM PUCP, de la Pontificia Universidad Católica del Perú; la Universidad Camilo José Cela; y el Instituto para el Futuro de la Educación del Tecnológico de Monterrey.

Agradecemos también a la Association of Economists and Managers of the Balkans (UDEKOM Balkans) y a OpenAI por su apoyo al encuentro inaugural. Estos reconocimientos corresponden a sus contribuciones al congreso; el presente volumen de acceso abierto es publicado por Intelligent Nudge.

Nuestro agradecimiento se extiende a los comités asesor, de programa técnico, editorial y organizador; a quienes promueven nuestra comunidad académica; a ponentes magistrales, facilitadores de talleres y paneles, revisores y responsables de sesión. Sus aportaciones sostuvieron el intercambio científico y el trabajo necesario para realizar este encuentro internacional.

Programa del congreso

El programa del 5 y 6 de octubre combinó conferencias magistrales, actividades prácticas, paneles y presentaciones paralelas, con participación presencial y virtual. Los artículos de este volumen conservan una selección de las contribuciones de investigación de ese intercambio.

Ponentes magistrales

Rose Luckin, PhD, CBE, profesora emérita de Diseño Centrado en el Estudiante, University College London, Reino Unido.

Ignacio Aguaded, PhD, profesor de la Facultad de Educación, Universidad de Huelva, España.

Irena Palamani Xhurxhi, PhD, directora de People Science, Walmart, Estados Unidos.

Diba Kaya, investigadora de Insights, JSTOR (ITHAKA), Estados Unidos.

Soribel Feliz, especialista sénior en Política y Gobernanza de la IA, Departamento de Seguridad Nacional de Estados Unidos.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Talleres y alcance académico

Carlotta Reviglio, de OpenAI, y Julius Nganji, de la University of Toronto, impartieron un taller práctico sobre inteligencia artificial en la enseñanza, el aprendizaje y la investigación. En el panel relacionado participaron May Iliana Portuguez Castro, de la Pontificia Universidad Católica del Perú, y Natalia Lara Nieto Márquez, de la Universidad Camilo José Cela.

El panel plenario de WISE, dirigido por Selma Talha-Jebril, abordó la adaptación a las nuevas demandas de competencias vinculadas con la IA mediante dos estudios de caso. Natalia Lara Nieto Márquez y Carola Hernández Hernández aportaron las perspectivas de la Universidad Camilo José Cela y de la Universidad de los Andes, respectivamente.

Áreas de investigación

El congreso convocó investigaciones de distintas disciplinas y contextos educativos. Su alcance incluyó entornos de aprendizaje apoyados por IA, diseño didáctico, desarrollo y evaluación de herramientas educativas, y los cambios en las responsabilidades de docentes y autoridades institucionales.

Las cuestiones éticas y culturales constituyeron otro eje del programa: equidad, privacidad, accesibilidad y efectos de los sistemas automatizados en el juicio humano. También se abordaron la infraestructura institucional, las políticas, la comunicación académica y los retos prácticos de implementar la IA de manera responsable.

Las perspectivas internacionales e interdisciplinarias relacionaron estos temas con el desarrollo de competencias laborales, la educación continua y el desarrollo sostenible. Esta colección presenta una selección de artículos dentro de ese alcance; no reproduce el programa completo ni la colección independiente de resúmenes.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Dirección y comités del congreso

Presidenta del congreso

Samira Hosseini, PhD - Presidenta del congreso; fundadora de Autonomous Academic Accelerator; editora en jefe de Intelligent Nudge in Education.

Comité asesor

Dušanka Bošković, PhD, University of Sarajevo, Bosnia y Herzegovina

Maria Potes Barbas, PhD, Polytechnic Institute of Santarém (Instituto Politécnico de Santarém), Portugal

Selma Talha-Jebril, World Innovation Summit for Education (WISE), Qatar Foundation, Qatar

Ashutosh Ghosh, PhD, University of Calcutta, India

Comité de programa técnico

Teófilo Ramos, PhD, Tecnológico de Monterrey (emérito), México

Nadine Page, PhD, Hult University, Reino Unido

Edmundo Tovar, PhD, Universidad Politécnica de Madrid, España

Viktoria Dalko, PhD, Hult International Business School, Estados Unidos

Manuel Castro, PhD, Universidad Nacional de Educación a Distancia (UNED), España

Louise Taylor, PhD, Oxford Brookes University; dirección de REAL-HE, Reino Unido

Martín Llamas Nistal, PhD, Universidade de Vigo (University of Vigo), España

Nadine El Agha, PhD, Ministry of Education, Emiratos Árabes Unidos

Elio San Cristóbal, PhD, Universidad Nacional de Educación a Distancia (UNED), España

Katharina von Knop, PhD, Global Applied Ethics Institute; Digital Trust Analytics GmbH, Alemania

Stamatios Papadakis, PhD, University of Crete, Grecia

Francisco Roca Rodríguez, Universidad de Jaén, España

Drissia Chouit, PhD, Moulay Ismail University, Marruecos

Amlan Chakrabarti, PhD, University of Calcutta, India

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Comité editorial

Ángel Hernando-Gómez, PhD, Universidad de Huelva, España
Diego Hernán Peluffo-Ordóñez, PhD, Yachay Tech University, Ecuador
Haylen Alejandra Perines Véliz, PhD, Universidad Complutense de Madrid, España
May Iliana Portuguez Castro, PhD, Pontificia Universidad Católica del Perú, Perú
José Luis Martin Núñez, PhD, Universidad Politécnica de Madrid, España
Patricia Caratozzolo, PhD, Tecnológico de Monterrey, México
Gwanggil Jeon, PhD, Hanyang University, Corea del Sur
Julius T. Nganji, PhD, University of Toronto, Canadá
Francisco Pérez-Fernández, PhD, Universidad CEU Fernando III, España
Yerzhan Kerimbekov, PhD, Zhanibekov University, Kazajistán
Kingsley Okoye, PhD, Tecnológico de Monterrey, México
Patricia de Casas Moreno, PhD, Universidad de Extremadura, España
Asad Abbas, PhD, Tecnológico de Monterrey, México
Claudia Camacho-Zúñiga, PhD, Tecnológico de Monterrey, México
Iván Sánchez-López, PhD, Universitat Oberta de Catalunya, España
Fulya Kula, PhD, University of Twente, Países Bajos
Paula Marcelo Martínez, PhD, Universidad de Huelva, España
Wilson Cheong Hin Hong, PhD, Macao University of Tourism, China (Macao SAR)
Sara Abdullah Alharbi, PhD, Imam Mohammad Ibn Saud Islamic University, Arabia Saudita
Benita Olivier, PhD, Oxford Brookes University, Reino Unido
Manuel Pardo, PhD, Universidad Católica San Antonio de Murcia, España
Ali Hindi, PhD, University of Manchester, Reino Unido
Samar Aad, PhD, Lebanese American University, Líbano
Paras Karmacharya, MD MS, Vanderbilt University Medical Center, Estados Unidos
Lilly Yazdanpanah, PhD, La Trobe University, Australia
Ritesh Chugh, PhD, Central Queensland University, Australia
William Jimenez Leal, PhD, Universidad de los Andes, Colombia
Natalia Lara Nieto Marquez, PhD, Universidad Camilo José Cela, España
Claudia Lizette Garay-Rondero, PhD, Tecnológico de Monterrey, México
Balqis Banjo, PhD, University of the Witwatersrand, Sudáfrica
Carola Hernández Hernández, PhD, Universidad de los Andes, Colombia
Letetia Mary Addison, PhD, The University of the West Indies, Trinidad y Tobago
Catalina Olivas, PhD, Tecnológico de Monterrey, México
Eman Elbealy, PhD, King Khalid University, Arabia Saudita
Sebastian Novomisky, PhD, Universidad Nacional de La Plata, Argentina
Divya R. Balachandran, Clark College, Estados Unidos
Sharon K. Gechure, PhD, Girne American University, Chipre del Norte
Ana Pavković, PhD, University North, Croacia

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • **OPEN ACCESS SERIES**

Syed Naeem Ahmed, PhD, Royal Commission Colleges and Institutes, Arabia Saudita
Asma Khattala, PhD, Université Mohamed Lamine Debaghine Sétif 2, Argelia; Al Yamamah University, Arabia Saudita
Sara Abdalla, PhD, International Islamic University Malaysia, Malasia; University of Khartoum, Sudán
Yogesh Beeharry, PhD, University of Mauritius, Mauricio
Najah Al-Shanableh, PhD, Al al-Bayt University, Jordania
Juan Sebastián Fernández-Prados, PhD, Universidad de Almería, España
Pedro Garcia, PhD, Universidad Católica Andrés Bello, Venezuela

Promotores de la comunidad académica

Priya Singh, PhD, fundadora de Research Made Clear, Reino Unido
Kajs Hadžić, Bosnia & Herzegovina Futures Foundation
Maria Stefanidi, fundadora de Your PhD Coach
Caroline Dale, fundadora de The Thesis Coach (Pty) Ltd, Sudáfrica

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Comité organizador

Francisco Pérez-Fernández, PhD, profesor de Tecnología Educativa, Universidad CEU Fernando III, España; editor en jefe de Escuela Abierta

Rodolfo Zambrano Villarreal, cofundador del Congreso

Samira Hosseini, PhD, editora en jefe de Intelligent Nudge in Education; fundadora de Autonomous Academic Accelerator; cofundadora de Horizonne Views

Amir Hossein Hosseini, responsable de Mercadotecnia y Comunicación, Horizonne Views

Martha Manríquez, responsable de Operación y Coordinación del Congreso; responsable de Éxito del Cliente, Horizonne Views

Nikolina Vrcelj, PhD, cofundadora y directora general de Association of Economists and Managers of the Balkans (UDEKOM Balkans), Serbia

Michelle Espinosa Hernández, directora de Alianzas, Horizonne Views

Aida Rodriguez, PhD, Odontología; Odontopediatría; Biotecnología

Consulta y citación de los artículos

Cada artículo cuenta con título, lista de autores y DOI propios. Para citar una contribución, consulte su registro individual y utilice el intervalo de páginas indicado en el PDF publicado.

Formato sugerido: Apellido, Iniciales. (2026). Título del artículo. SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues: Open Access Articles — 2026 Edition (pp. primera–última). Intelligent Nudge. https://doi.org/[DOI del artículo]

El volumen completo y los registros individuales de los artículos están disponibles en: https://2026.samyrad.com/samyrad-open-access-proceedings/.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues

Open Access Articles - 2026 Edition

The Intersection of Artificial Intelligence and Education

October 5–6, 2026
Seville, Spain
Hosted by Universidad CEU Fernando III

Volume editor
Samira Hosseini

Intelligent Nudge
October 31, 2026

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Publication information

Publication title: SAMYRAD: SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES

Subtitle: Open Access Articles - 2026 Edition

Volume editor: Samira Hosseini

Publisher: Samira Hosseini, publishing under the Intelligent Nudge imprint

Publication date: October 31, 2026

Frequency: Annual | Edition: 2026 | Articles: English

Format: Electronic publication, PDF, open access

Conference: October 5–6, 2026, Universidad CEU Fernando III, Seville, Spain

Publication website: https://2026.samyrad.com/samyrad-open-access-proceedings/

Publisher website: https://intelligentnudge.com/

Contact: support@intelligentnudge.com

DOI: https://doi.org/10.68286/samyrad.2026.proceedings

Rights and use

© 2026 Intelligent Nudge. Copyright and reuse conditions for individual contributions are governed by the applicable author agreements and any license stated with each article. Material credited to third parties remains subject to the rights of its respective owners. Free online access does not, by itself, grant permission for unrestricted reuse.

Citation of this volume

Hosseini, S. (Ed.). (2026). SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues: Open Access Articles — 2026 Edition. Intelligent Nudge.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Legal notice

SAMYRAD: SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES is an annual scientific publication, 2026 edition, in electronic and open-access format, edited and published by Samira Hosseini, who serves as the responsible editor, under the Intelligent Nudge publishing imprint, publication address: Danis No. 613, Paseo del Vergel neighborhood, postal code 64988, Monterrey, Nuevo León, Mexico, telephone: +52 81 23267576, email: support@intelligentnudge.com, publisher's website: https://intelligentnudge.com/, publication access link: https://2026.samyrad.com/samyrad-open-access-proceedings/, Reservation of Rights to Exclusive Use No. 04-2026-100115181300-102, granted by the National Copyright Institute (INDAUTOR), ISSN: pending assignment by the Mexican National ISSN Center, language of the articles: English, preliminary pages: Spanish and English, publication medium: electronic (internet), format: PDF, publication date of the 2026 edition: October 31, 2026, person responsible for the latest update of this issue: Samira Hosseini, at Danis No. 613, Paseo del Vergel neighborhood, postal code 64988, Monterrey, Nuevo León, Mexico, last modification date: October 31, 2026, file size: 8.1 MB.

Editorial statement

The SAMYRAD 2026 Open-Access Proceedings bring selected full-length, peer-reviewed papers into a freely accessible scholarly collection. Published by Intelligent Nudge, the volume documents research and critical inquiry into artificial intelligence in education, professional preparation, and lifelong learning.

The collection includes work that examines educational practice alongside the institutional, social, and ethical conditions in which technology is used. Readers are encouraged to consider each contribution in relation to its methods, evidence, and stated limitations. The purpose of publication is to make that work available for scrutiny, discussion, and further investigation.

Individual article records provide bibliographic information and DOI links. When citing a particular finding or argument, readers should cite the relevant article rather than the complete volume. The annual publication is intended to continue through successive editions under its permanent registered title.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Message from the Conference President

An important responsibility begins when a conference ends. Ideas shared in a room must become available to people who could not be there, and the arguments that prompted discussion must remain open to examination. These open-access proceedings are our contribution to that continuing work.

At SAMYRAD 2026 in Seville, we considered how artificial intelligence is changing the conditions under which people learn, teach, and prepare for professional life. Across these discussions, one question remained central: what evidence should guide the choices we make? The pace of technological development gives that question urgency, but it does not remove the need for careful inquiry.

The papers collected here offer different ways of approaching it. Their value lies in the specific problems they investigate, the reasoning they make visible, and the further questions they leave for readers. I hope this volume will be used by researchers testing new propositions, by educators examining their practice, and by leaders seeking a more informed basis for institutional decisions.

Making this work openly accessible is a deliberate part of our publishing effort. Access allows a conversation that began in Seville to reach colleagues working with different resources and in different educational settings. It also creates an opportunity for findings to be challenged, methods to be refined, and collaborations to develop beyond the original event.

I am grateful to the authors for entrusting us with their work and to the reviewers and editorial colleagues whose time and judgment supported its development. I also thank Universidad CEU Fernando III, our host, and the institutions and individuals acknowledged in these pages for helping make the inaugural conference possible.

SAMYRAD will continue as an annual forum, with its scholarly publications providing a record that can grow across editions. My hope is that readers will return to this volume for ideas they can investigate and use, and that they will bring their own evidence and experience into the discussions ahead.

Thank you for reading, questioning, and contributing to this shared undertaking.

Samira Hosseini, PhD
Conference President, SAMYRAD 2026

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Acknowledgments

The inaugural SAMYRAD conference was hosted by Universidad CEU Fernando III in Seville, Spain. We thank the university and its staff for the facilities, coordination, and practical support that enabled participants to meet and exchange their work.

We acknowledge the contributions to the conference of IEEE, including the IEEE Spain Section, IEEE Region 8, and IEEE Education Society; the Alfamed Euro-American Research Network; Universidad de los Andes; the World Innovation Summit for Education (WISE), an initiative of Qatar Foundation; the University of Calcutta; Research About Learning in Higher Education (REAL-HE) at Oxford Brookes University; CENTRUM PUCP at the Pontificia Universidad Católica del Perú; Universidad Camilo José Cela; and the Institute for the Future of Education at Tecnológico de Monterrey.

We also thank the Association of Economists and Managers of the Balkans (UDEKOM Balkans) and OpenAI for their support of the inaugural event. These acknowledgments recognize conference contributions; this open-access volume is published by Intelligent Nudge.

Our appreciation extends to the advisory, technical program, editorial, and organizing committees; academic advocates; keynote speakers; workshop and panel facilitators; reviewers; and session chairs. Their contributions supported the scientific discussion and the work required to bring an international meeting into being.

Conference program

The October 5–6 program combined keynote sessions, practical learning, panel discussions, and parallel presentations, with opportunities for both on-site and virtual participation. The full papers in this volume preserve selected research contributions from that broader exchange.

Keynote speakers

Rose Luckin, PhD, CBE, Professor Emerita of Learner Centred Design, University College London, United Kingdom.

Ignacio Aguaded, PhD, Professor, Faculty of Education, University of Huelva, Spain.

Irena Palamani Xhurxhi, PhD, Director of People Science, Walmart, United States.

Diba Kaya, Staff Insights Researcher, JSTOR (ITHAKA), United States.

Soribel Feliz, Senior AI Policy and Governance Specialist, United States Department of Homeland Security, United States.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Workshops and scholarly scope

Carlotta Reviglio of OpenAI and Julius Nganji of the University of Toronto led a hands-on workshop on AI in teaching, learning, and research. A related panel included May Iliana Portuguez Castro of Pontificia Universidad Católica del Perú and Natalia Lara Nieto Márquez of Universidad Camilo José Cela.

The WISE plenary panel, led by Selma Talha-Jebril, examined adaptation to changing AI-related skill demands through two case studies. Natalia Lara Nieto Márquez and Carola Hernández Hernández contributed perspectives from Universidad Camilo José Cela and Universidad de los Andes, respectively.

Research scope

The conference invited inquiry across disciplines and educational settings. Its scope encompassed AI-supported learning environments, instructional design, the development and assessment of educational tools, and the changing responsibilities of educators and institutional leaders.

Ethical and cultural questions formed a further strand of the program, including fairness, privacy, accessibility, and the effects of automated systems on human judgment. Contributions also addressed institutional infrastructure, policy, scholarly communication, and the practical demands of implementing AI responsibly.

International and interdisciplinary perspectives connected these themes with workforce development, continuing education, and sustainable development. The open-access collection presents selected papers within this wider scope; it does not reproduce the complete conference program or the separate abstracts collection.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Conference leadership and committees

Conference president

Samira Hosseini, PhD - Conference President; Founder, Autonomous Academic Accelerator; Editor-in-Chief, Intelligent Nudge in Education.

Advisory committee

Dušanka Bošković, PhD, University of Sarajevo, Bosnia and Herzegovina

Maria Potes Barbas, PhD, Polytechnic Institute of Santarém (Instituto Politécnico de Santarém), Portugal

Selma Talha-Jebril, World Innovation Summit for Education (WISE), Qatar Foundation, Qatar

Ashutosh Ghosh, PhD, University of Calcutta, India

Technical program committee

Teófilo Ramos, PhD, Tecnológico de Monterrey (emeritus), Mexico

Nadine Page, PhD, Hult University, United Kingdom

Edmundo Tovar, PhD, Universidad Politécnica de Madrid, Spain

Viktoria Dalko, PhD, Hult International Business School, United States

Manuel Castro, PhD, Universidad Nacional de Educación a Distancia (UNED), Spain

Louise Taylor, PhD, Oxford Brookes University; Director, REAL-HE, United Kingdom

Martín Llamas Nistal, PhD, Universidade de Vigo (University of Vigo), Spain

Nadine El Agha, PhD, Ministry of Education, United Arab Emirates

Elio San Cristóbal, PhD, Universidad Nacional de Educación a Distancia (UNED), Spain

Katharina von Knop, PhD, Global Applied Ethics Institute; Digital Trust Analytics GmbH, Germany

Stamatios Papadakis, PhD, University of Crete, Greece

Francisco Roca Rodríguez, Universidad de Jaén, Spain

Drissia Chouit, PhD, Moulay Ismail University, Morocco

Amlan Chakrabarti, PhD, University of Calcutta, India

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Editorial committee

Ángel Hernando-Gómez, PhD, Universidad de Huelva, Spain
Diego Hernán Peluffo-Ordóñez, PhD, Yachay Tech University, Ecuador
Haylen Alejandra Perines Véliz, PhD, Universidad Complutense de Madrid, Spain
May Iliana Portuguez Castro, PhD, Pontificia Universidad Católica del Perú, Peru
José Luis Martin Núñez, PhD, Universidad Politécnica de Madrid, Spain
Patricia Caratozzolo, PhD, Tecnológico de Monterrey, Mexico
Gwanggil Jeon, PhD, Hanyang University, South Korea
Julius T. Nganji, PhD, University of Toronto, Canada
Francisco Pérez-Fernández, PhD, Universidad CEU Fernando III, Spain
Yerzhan Kerimbekov, PhD, Zhanibekov University, Kazakhstan
Kingsley Okoye, PhD, Tecnológico de Monterrey, Mexico
Patricia de Casas Moreno, PhD, Universidad de Extremadura, Spain
Asad Abbas, PhD, Tecnológico de Monterrey, Mexico
Claudia Camacho-Zúñiga, PhD, Tecnológico de Monterrey, Mexico
Iván Sánchez-López, PhD, Universitat Oberta de Catalunya, Spain
Fulya Kula, PhD, University of Twente, Netherlands
Paula Marcelo Martínez, PhD, Universidad de Huelva, Spain
Wilson Cheong Hin Hong, PhD, Macao University of Tourism, China (Macao SAR)
Sara Abdullah Alharbi, PhD, Imam Mohammad Ibn Saud Islamic University, Saudi Arabia
Benita Olivier, PhD, Oxford Brookes University, United Kingdom
Manuel Pardo, PhD, Universidad Católica San Antonio de Murcia, Spain
Ali Hindi, PhD, University of Manchester, United Kingdom
Samar Aad, PhD, Lebanese American University, Lebanon
Paras Karmacharya, MD MS, Vanderbilt University Medical Center, United States
Lilly Yazdanpanah, PhD, La Trobe University, Australia
Ritesh Chugh, PhD, Central Queensland University, Australia
William Jimenez Leal, PhD, Universidad de los Andes, Colombia
Natalia Lara Nieto Marquez, PhD, Universidad Camilo José Cela, Spain
Claudia Lizette Garay-Rondero, PhD, Tecnológico de Monterrey, Mexico
Balqis Banjo, PhD, University of the Witwatersrand, South Africa
Carola Hernández Hernández, PhD, Universidad de los Andes, Colombia
Letetia Mary Addison, PhD, The University of the West Indies, Trinidad and Tobago
Catalina Olivas, PhD, Tecnológico de Monterrey, Mexico
Eman Elbealy, PhD, King Khalid University, Saudi Arabia
Sebastian Novomisky, PhD, Universidad Nacional de La Plata, Argentina
Divya R. Balachandran, Clark College, United States
Sharon K. Gechure, PhD, Girne American University, Northern Cyprus
Ana Pavković, PhD, University North, Croatia

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • **OPEN ACCESS SERIES**

Syed Naeem Ahmed, PhD, Royal Commission Colleges and Institutes, Saudi Arabia
Asma Khattala, PhD, Université Mohamed Lamine Debaghine Sétif 2, Algeria; Al Yamamah University, Saudi Arabia
Sara Abdalla, PhD, International Islamic University Malaysia, Malaysia; University of Khartoum, Sudan
Yogesh Beeharry, PhD, University of Mauritius, Mauritius
Najah Al-Shanableh, PhD, Al al-Bayt University, Jordan
Juan Sebastián Fernández-Prados, PhD, Universidad de Almería, Spain
Pedro Garcia, PhD, Universidad Católica Andrés Bello, Venezuela

Academic advocates

Priya Singh, PhD, Founder, Research Made Clear, UK
Kajs Hadžić, Bosnia & Herzegovina Futures Foundation
Maria Stefanidi, Founder, Your PhD Coach
Caroline Dale, Founder, The Thesis Coach (Pty) Ltd, South Africa

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

Organizing committee

Francisco Pérez-Fernández, PhD, Professor of Educational Technology, Universidad CEU Fernando III, Spain; Editor-in-Chief, Escuela Abierta

Rodolfo Zambrano Villarreal, Co-founder of the conference

Samira Hosseini, PhD, Editor-in-Chief, Intelligent Nudge in Education; Founder, Autonomous Academic Accelerator; Co-founder, Horizonne Views

Amir Hossein Hosseini, Marketing and Communications Leader, Horizonne Views

Martha Manríquez, Conference Operations and Coordination Manager; Client Success Manager, Horizonne Views

Nikolina Vrcelj, PhD, Co-founder and CEO, Association of Economists and Managers of the Balkans (UDEKOM Balkans), Serbia

Michelle Espinosa Hernández, Director of Partnerships, Horizonne Views

Aida Rodriguez, PhD, Dentistry; Pediatric Dentistry; Biotechnology

Using and citing the proceedings

Each article has its own title, author list, and DOI. Use the individual article record and the page range shown in the published PDF when citing a contribution.

Suggested article citation: Author surname, Initials. (2026). Article title. SAMYRAD: Smart Algorithms & Models: Your Reach to Advanced Dialogues: Open Access Articles — 2026 Edition (pp. first–last). Intelligent Nudge. https://doi.org/[article DOI]

The complete volume and individual article records are available at: https://2026.samyrad.com/samyrad-open-access-proceedings/.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • OPEN ACCESS SERIES

TABLE OF CONTENTS

No.	Article Title	Page
1	From Adoption to Human Judgement: Responsible Generative AI Integration in Higher Education	1
2	Inclusive Adaptation to AI: Adult Education, Digital Skills, and Older Workers' Employment in the EU	11
3	Socio-Technical Perspective on Artificial Intelligence in Secondary Mathematics Education: A Systematic Review	18
4	Artificial Intelligence and the Future of Education in Croatia: Student Professor Co-Creation in a Structured Narrative Review	30
5	A Research-Embedded Implementation Model for Design of AI Applications in Higher Education	38
6	Transforming University Teaching through Generative AI: An Exploratory Pilot Study and a Multidisciplinary Scale-Up Design	48
7	Incorporating Domain Knowledge into Student Risk Prediction: An Ablation Study Using Higher Education Enrollment Records	56
8	Enhancing geometry exemplification through integration of digital technology tools in South African higher education mathematics classrooms	64
9	Beyond Prompting: Embedding Ethical and Equity-Oriented Artificial Intelligence Literacy within the Canadian Dental Public Health Competency Framework	70
10	Human-Curated AI-Assisted Scenario Design for Game-Based Sustainability Learning: A Design-Oriented Case Study	79
11	From Static LMS to Autonomous Agents: Reimagining Moodle-Based Open and Distance e-Learning Delivery for the Agentic AI Era in Kenya	86
12	AI Literacy, Digital Capabilities and Intellectual Property Awareness: Evidence for Preparing Future Professionals in the Digital Age	95
13	Cultivating Faculty's AI Literacy through the Use of Open-Access Platform	104
14	Skill (Mis)Alignments: A case study of a recent educational program against industry needs	111
15	The GenAI Roundabout: From Enforcing Control to Building Capability	119
16	From Automation to Judgment: Evaluating an AI-Augmented Monte Carlo Instructional Architecture for Responsible Managerial Decision-Making	126
17	When Immersive Learning Works: AI- and VR-Supported Feedback for Competency-Sensitive Skill Development in Higher Education	134
18	AI-Supported Learning and Job Readiness in Vocational Tourism Education: A Comparative Study of Two Albanian Schools	143
19	Comparing AI Interaction Logs and Classroom Analytics for Detecting Conceptual Difficulties in Statistics	150
20	An XAI Pipeline for Gender Bias Detection in Advertising Images	157

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

FRONT MATTER • **OPEN ACCESS SERIES**

No.	Article Title	Page
21	Toward AI-Assisted Early Academic Risk Detection in Introductory Physics Courses: Interpretable Profiling among First-Year Engineering Students	166
22	Comparing AI-Generated and Teacher-Generated Lessons in Moodle	175
23	Digital Participation to AI Competence: Challenges for Equitable GenAI Adoption in Mexican Higher Education	188
24	The Illusion of Prompting: A Discourse Analysis of Epistemic Engagement in Student–AI Dialogue	193
25	From AI Adoption to AI Pedagogical Transformation: Developing a Faculty AI Literacy Maturity Model for Higher Education	206
26	Enhancing Inclusive Education Through AI: Supporting Neurodivergent Students and Staff in UK Higher Education	214

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

From Adoption to Human Judgement: Responsible Generative AI Integration in Higher Education

Kurt Naicker[1*], Katlego Thipe[2]

*Corresponding author: kurtn@uj.ac.za

[1]Department of Commercial Accounting, University of Johannesburg, Johannesburg, South Africa
[2]Department of Accountancy, University of Johannesburg, Johannesburg, South Africa

Article DOI: https://doi.org/10.68286/samyrad2026.292

ABSTRACT

Generative artificial intelligence (GenAI) is transforming higher education faster than institutional pedagogy and governance can adapt. Although recent scholarship documents opportunities and risks, the relationships among pedagogical use, personalisation, governance and human cognitive capability remain fragmented. This paper presents a structured qualitative evidence synthesis using directed content analysis of 28 publications from 2024 to 2026, comprising systematic reviews, narrative or conceptual reviews, and selected empirical studies. Four predefined domains guided the analysis: AI-enhanced teaching and learning, personalised and adaptive learning, ethical governance, and AI literacy and critical thinking. The selected literature consistently reports potential benefits for content generation, formative feedback and personalisation, while also identifying fragmented governance, persistent equity constraints and limited empirical testing of safeguards against cognitive overreliance. The paper proposes a human-centred conceptual framework in which pedagogical applications are enabled by AI literacy, bounded by ethical governance and conditioned by institutional context. Critical thinking, professional judgement and graduate readiness are positioned as central human outcomes. This is a conceptual and theoretical contribution rather than an empirical evaluation of GenAI's effects, and the proposed framework has not yet been empirically tested. The framework is conceptual rather than causal and provides propositions for future empirical testing and practical guidance for universities, lecturers, students, policymakers and curriculum developers.

KEYWORDS Generative artificial intelligence, higher education, directed content analysis, responsible AI, AI literacy, critical thinking, educational governance.

How to cite: Naicker, K., & Thipe, K. (2026). From adoption to human judgement: Responsible generative AI integration in higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.292

Introduction

AI Is Transforming Higher Education

Generative artificial intelligence (GenAI) has moved from a peripheral technology to a central feature of the higher-education landscape within a remarkably short period. The public release of large language models such as ChatGPT precipitated what several of the studies reviewed here describe as a rapid, largely unplanned wave of institutional and individual adoption [1], [2]. Universities now encounter GenAI simultaneously as a teaching aid, an assessment risk, a research tool, and an object of policy concern — often before any of these roles has been deliberately defined. This convergence of opportunity and disruption forms the starting point for the present review: GenAI is not a single, bounded innovation to be evaluated once, but a diffuse capability whose implications for pedagogy, equity, ethics and cognition are still being worked out across a fast-growing and fragmented body of literature.

The opportunities are considerable. Reviewed studies report that GenAI can generate differentiated instructional content, provide immediate formative feedback, model learner needs dynamically, and free educator time from routine tasks [1], [3], [4]. Yet the same literature documents a parallel set of concerns around privacy, bias, academic integrity, governance capacity, and the risk that convenience gradually substitutes for independent thought [5], [6], [7]. The result is a literature that is expanding faster than it is being synthesised.

Existing Literature

Recent reviews have begun to map this terrain, typically clustering around one of three concerns. A first cluster addresses teaching and assessment, documenting how GenAI reshapes content creation, feedback and the educator's role in the classroom [8], [9]. A second cluster addresses ethics and governance, cataloguing the privacy, bias, integrity and regulatory risks associated with institutional adoption [10], [11], [12]. A third, smaller cluster addresses AI literacy, examining whether students and staff are equipped to use these tools critically rather than uncritically [13], [14]. Each cluster has produced valuable, largely self-contained syntheses. What has received comparatively little attention is how these three conversations — pedagogy, governance, and cognition — relate to one another as GenAI is integrated into everyday academic practice.

The Gap

Existing reviews have catalogued opportunities and risks associated with AI adoption in higher education. However, comparatively less attention has been devoted to synthesising how these issues collectively reshape the relationship between pedagogy, governance, and human cognitive capability. Studies of AI-enhanced teaching rarely engage with the governance literature that constrains its safe deployment; studies of ethical governance rarely engage with the pedagogical evidence on whether generated content actually reaches higher-order learning outcomes; and studies of AI literacy are frequently proposed as a standalone remedy without being connected back to either the pedagogical or governance conversations that

create the need for it in the first place. This fragmentation matters because responsible AI integration is, by definition, a cross-cutting institutional problem rather than a series of independent technical or ethical questions.

Research Objective

This paper conducts a directed qualitative content analysis of recent literature to identify the dominant themes shaping responsible GenAI integration in higher education and to develop an integrated conceptual account of their relationships. Rather than treating pedagogy, personalisation, governance and literacy as independent concerns, the paper examines how these domains interact in shaping critical thinking, professional judgement and graduate readiness. The study addresses the following research questions:

RQ1: What recurring pedagogical, personalisation, governance and literacy-related themes characterise contemporary literature on responsible GenAI integration in higher education?

RQ2: How do these themes interact in shaping critical thinking, professional judgement and graduate readiness within AI-enabled learning environments?

Literature review

This section situates the analysis conceptually. It does not analyse the 28 reviewed studies individually — that task is undertaken in Section IV — but establishes the definitions and conceptual boundaries the analysis subsequently applies.

Artificial Intelligence in Higher Education

Artificial intelligence in higher education encompasses a broad family of technologies, from earlier rule-based intelligent tutoring systems and predictive learning analytics through to contemporary large language models [1], [15]. What distinguishes the current period from earlier waves of educational technology is scale and accessibility: AI capability that was once confined to specialist systems is now available directly to students and staff through consumer-facing interfaces, collapsing the distance between institutional adoption decisions and individual, often unsanctioned, use [2].

Generative AI

Generative AI (GenAI) refers specifically to models capable of producing novel text, code, images or other content in response to a prompt, as distinct from earlier, predominantly classificatory or predictive applications of AI in education [3]. In the higher-education context, this generative capability is what enables GenAI to function simultaneously as a content-creation tool for educators, a study aid for students, and an assessment risk for institutions — three roles that earlier, narrower AI applications rarely combined [9], [16].

Responsible AI

Responsible AI, as used across the reviewed governance literature, refers to the deliberate design of institutional structures policy, oversight, training and technical safeguards intended to ensure that AI benefits are realised without disproportionate harm to privacy, fairness, or academic integrity [12], [17]. The concept is normative rather than purely technical: it presumes that responsibility is distributed across developers, institutions, educators and students, and that governance mechanisms such as human-in-the-loop review or risk-tiering are only meaningful when embedded in institutional practice rather than left as abstract principle [11].

AI Literacy

AI literacy denotes the capacity of learners and educators to understand, evaluate and use AI tools critically — including recognising their limitations, biases and appropriate boundaries of use — rather than treating AI output as authoritative by default [13], [10]. Across the literature, AI literacy is positioned simultaneously as a curricular outcome in its own right and as an enabling condition for the safe use of GenAI elsewhere in the curriculum, which is part of why it recurs across multiple thematic clusters in the analysis that follows.

Critical Thinking

Critical thinking, in this literature, is generally treated as the higher-order cognitive capacity that GenAI use may either scaffold or erode, depending on how it is deployed [18], [19]. Where GenAI is used to support reflection, iteration and evaluation of ideas, several studies report a positive relationship with metacognitive development; where it is used as a substitute for independent reasoning, the same literature reports a risk of cognitive offloading and de-skilling [6], [7]. This unresolved tension is a central concern of the present review and is returned to directly in the Results and Discussion.

Human Agency, Professional Judgement and Graduate Readiness

Human agency refers to the capacity of educators and learners to retain meaningful control over the purposes, processes and consequences of AI-supported activity. In this paper, professional judgement is treated as the disciplined application of knowledge, critical evaluation and contextual reasoning for which human actors remain accountable. Graduate readiness extends beyond operational familiarity with AI tools to include the capacity to question outputs, recognise uncertainty, justify decisions and act responsibly within disciplinary and professional settings. These concepts provide the human-centred outcomes against which the responsible integration of GenAI is assessed.

Methodology

Research Philosophy

The study adopted an interpretivist orientation because its purpose was to synthesise how contemporary publications construct and connect the pedagogical, governance and cognitive dimensions of GenAI use. A structured qualitative evidence-synthesis design was used, with directed content analysis providing the analytical method.

Review Design

The review is a structured qualitative synthesis of publication-level evidence rather than a systematic review of primary studies. The corpus comprises systematic reviews, narrative or conceptual reviews and selected empirical studies. The publication, rather than the participant sample within an underlying primary study, was the unit of analysis. This design was selected to integrate themes across an already substantial review-level literature and to identify relationships that are often examined separately. The review does not claim exhaustive retrieval. Several included publications synthesise overlapping bodies of primary research; consequently, recurrence across publications indicates thematic consistency in the selected corpus rather than the number of independent empirical datasets supporting a claim.

Literature Search Strategy and Study Selection

A structured, iterative search was conducted through the University of Johannesburg Library environment using Scopus, Web of Science, SpringerLink, ScienceDirect and Google Scholar. Search activity focused on English-language scholarship published from 2022 to 2026 and used combinations of terms relating to artificial intelligence, generative artificial intelligence, higher education, teaching and learning, personalisation, governance, AI literacy and critical thinking. The principal search expression retained in the review audit trail was:

("Artificial Intelligence" OR "AI" OR "Generative Artificial Intelligence" OR "GenAI" OR "ChatGPT") AND ("Higher Education" OR universit OR college*) AND (teaching OR learning OR assessment OR pedagogy OR education OR "AI literacy" OR "critical thinking")*

Additional searches were undertaken around the four analytical domains: AI-enhanced teaching and learning; personalised and adaptive learning; ethical and governance considerations; and AI literacy and critical thinking. Because database-specific search dates, field restrictions, result counts and all supplementary search strings were not recorded contemporaneously, the process cannot be represented as a fully reproducible systematic search. The paper therefore reports the review transparently as a structured and purposive evidence synthesis rather than retrospectively reconstructing unrecorded methodological details.

Publications were retained when they addressed AI or GenAI in higher education and contributed substantively to at least one analytical domain. Peer-reviewed empirical studies, systematic reviews, narrative reviews and conceptual studies were eligible.

Publications were excluded when they focused exclusively on primary or secondary education, addressed technical model development without educational application, were non-scholarly commentary, were not available in English or fell outside the conceptual scope.

The working literature matrix contained 28 publications retained for synthesis, published between 2024 and 2026. Selection and eligibility assessment were completed by the first author using the stated criteria and a structured matrix. A PRISMA flow diagram is not presented because complete database-level identification, deduplication and exclusion-reason records were not maintained. This limitation is acknowledged to avoid presenting retrospective estimates as observed review data.

Data Extraction

The 28 publications were extracted into a structured literature matrix capturing authorship, year, geographical scope, journal, research aim, design, sample, AI technology, primary theme, findings, limitations, practical implications, future-research recommendations, quality-control observations and a provisional evidence-weight classification.

The evidence-weight classification was used as a sensitivity aid rather than a formal risk-of-bias score. Publications were considered higher confidence when they reported transparent retrieval and eligibility procedures, independent screening or coding, and study-level appraisal. Publications with incomplete corpus reporting, internal inconsistencies, miscitation or weak source integrity were treated as lower-confidence or contextual evidence. The synthesis prioritised patterns supported by stronger sources and used lower-confidence publications only for corroboration or context.

Directed Qualitative Content Analysis

Directed qualitative content analysis was applied in accordance with Hsieh and Shannon [20]. The four predefined domains were derived from the research objective and preliminary scoping of the literature: (1) AI-enhanced teaching and learning, (2) personalised and adaptive learning, (3) ethical and governance considerations, and (4) AI literacy and critical thinking.

Analysis proceeded in three stages. First, relevant meaning units were deductively assigned to one or more predefined domains. Second, descriptive codes were generated from the research aims, findings, limitations and recommendations recorded in the matrix. Third, semantically related codes were grouped inductively into subcategories and higher-order interpretive sub-themes within each domain. A publication was counted as contributing to a category when its extracted findings or recommendations contained a substantive meaning unit relevant to that category.

Frequency (f) indicates the number of publications contributing to a category, not the number of code occurrences, participants or independent primary datasets. Frequencies are descriptive indicators of recurrence within the selected corpus and should not be interpreted as prevalence estimates, effect sizes or evidence-quality scores.

Trustworthiness

Trustworthiness was supported through a structured extraction matrix, explicit category definitions and an audit trail linking each reported category to contributing publications. Coding was conducted by the first author and reviewed conceptually with the second author during manuscript development; however, independent double-coding and a formal inter-coder reliability statistic were not completed.

Interpretive findings should therefore be understood as analytically transparent but not independently reliability-tested.

The review has four additional limitations. First, the search was structured and purposive rather than fully systematic, and complete database-level search logs were not retained. Second, the corpus combines heterogeneous publication types and methodological quality. Third, several included reviews may rely on overlapping primary studies, which can amplify apparent convergence. Fourth, English-language and review-level evidence may underrepresent local, non-English and emerging Global South scholarship. These limitations constrain claims of exhaustiveness and causality but do not prevent the review from identifying recurring conceptual relationships within the selected corpus.

These four limitations are interrelated and should be read together when interpreting the Results: because selection was completed by a single researcher rather than through independent double screening, because overlapping primary-study corpora across included reviews can inflate the apparent number of studies “agreeing” on a pattern without adding independent evidence, and because the corpus mixes review types of differing rigour, the frequency counts reported in Section IV should be read as indicators of thematic consistency within this particular, purposively curated corpus — not as independent replication counts, effect-size evidence, or a claim generalisable beyond the reviewed literature.

Results

The following sub-sections report the axial coding summary and narrative pattern (agreements, contradictions and gaps) for each of the four predefined themes.

Theme 1: AI-Enhanced Teaching and Learning

Fourteen of the 28 records address how GenAI reshapes instructional delivery, feedback and the educator's role. Open coding of aims and findings yielded six axial categories (Table 1), synthesised into three sub-themes: GenAI as a content-and-feedback engine; the evolving educator role; and framework-guided versus ad hoc implementation. There is strong agreement that GenAI functions as a generative co-producer of instructional content and a first layer of formative feedback: curricula, quizzes, simulations and differentiated materials [1], [3], automated marking and feedback that supports self-regulated learning [4], and reduced routine

workload freeing educator time for higher-order tasks [9], [21]. This convergence extends across review types — conceptual, PRISMA-systematic, and mixed-method — suggesting the finding is robust to methodological variation. A contradiction emerges around pedagogical depth. Several studies report measurable gains in engagement, motivation and performance attributable to GenAI-enhanced teaching [2], [4], yet Mere [3] explicitly flags weak pedagogical depth in generated content, and Rodríguez-Ortiz et al. [15] find that GenAI-supported learning analytics show weak pedagogical grounding and limited educator involvement. Belkina et al. [8] is the clearest counter-example, demonstrating through Laurillard's Conversational Framework, SAMR and TPACK that GenAI use can progress from mere substitution to genuine pedagogical transformation — but only when implementation is explicitly framework-guided. The tension, therefore, is not whether GenAI can enhance teaching, but whether most reported enhancement is technology-led rather than pedagogy-led. The clearest gap is the scarcity of studies anchoring GenAI integration in an explicit instructional-design framework: only one of fourteen contributing studies [8] does so directly. Domain-specific applications beyond writing and art, and classroom-embedded evaluations of teacher-AI collaboration, remain rare relative to the volume of review-level synthesis.

Theme 2: Personalized and Adaptive Learning

Ten records centre on adaptive and personalised learning mechanisms, their empirical effectiveness, and the equity conditions under which benefits are realised. Six axial categories (Table 2) synthesise into three sub-themes: mechanisms of personalisation; evidence of effectiveness, unevenly theorised; and persistent equity, quality and implementation gaps.

Table 1. Axial coding summary — AI enhanced teaching and learning

Axial Category	Illustrative Open Codes	Contributing Studies	f
Automated content, curriculum & assessment generation	curriculum drafting; quiz/simulation generation; differentiated materials; process-based assessment design	[1], [3], [9], [18], [21], [22]	**6**
Automated & blended feedback mechanisms	AI-generated feedback; instructor–AI feedback integration; formative feedback loops; feedback for metacognition	[1], [4], [9], [15], [16], [21]	**6**
Educator role transformation & workload reduction	facilitation over transmission; reduced routine workload; human-AI collaboration; teacher optimisation	[1], [2], [4], [9], [13], [17], [18], [21]	**8**
Learning analytics for prediction & engagement	dropout prediction; engagement/sentiment analytics; scaffolding via analytics	[15]	**1**
Pedagogical-framework-guided implementation	SAMR substitution-to-transformation; TPACK alignment; Laurillard's Conversational Framework	[8]	**1**
Domain-specific applications (art; academic writing)	creative scaffolding; cultural-heritage content; ideation and outlining support; grammar/revision assistance	[16], [22]	**2**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

Table 2. Axial coding summary — personalized and adaptive learning

Axial Category	Illustrative Open Codes	Contributing Studies	f
Adaptive pathways, pacing & difficulty adjustment	dynamic sequencing; difficulty calibration; individualised learning pathways; early-warning triggers	[1], [3], [23], [24], [25]	**5**
Learner modelling & profiling	dynamic learner profiles; multimodal learner modelling; emotion recognition; algorithmic personalisation engines	[3], [15], [24], [25]	**4**
Empirical effectiveness evidence	performance gains (15–25%); engagement gains (30–40%); retention and accessibility improvements	[23], [24], [26]	**3**
Cognitive-level & theoretical alignment	Bloom's taxonomy mapping; constructivist grounding; PBL/STEAM/UDL alignment	[19], [26]	**2**
Quality, multimodality & workload of generated resources	expert-review evaluation; mixed-methods quality checks; educator workload of resource curation	[15], [27]	**2**
Equity, access & implementation barriers	infrastructure cost; digital-divide access gaps; institutional readiness; cultural adaptation	[3], [23], [24], [26], [28]	**5**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

The selected publications show their clearest agreement on mechanism and reported outcome: adaptive pathway sequencing, pacing and learner profiling are consistently

described as the operational core of AI personalisation [1], [3], [25], and where effectiveness is quantified, the direction of effect is uniformly positive — one included review reports ranges of 15–25% for performance and 30–40% for engagement across its selected studies [23]; these ranges are not pooled effect estimates. Fortuna et al. [24] similarly reports favourable patterns for retention and accessibility.

A contradiction surfaces between the maturity claimed for multimodal personalisation and the maturity observed in practice. Chakraborty [1] and Mere [3] present multimodal learner modelling (including emotion recognition) as an active capability, whereas Yuan et al. [27], examining GenAI-generated personalised resources directly, find that dynamic multimodal resources are comparatively rare and that immediate textual feedback and language support dominate actual implementations. A second tension concerns theoretical grounding: Peng and Li [26] find that constructivist theory is frequently implicit rather than explicit, and that when learning theory is applied deliberately it measurably improves pedagogical alignment. Equity is the most consistently under-operationalised construct in this theme: Peng and Li [26] note it is poorly operationalised even where nominally addressed, and geographic concentration is pronounced — 82.8% of that review's evidence base originates from Asia. Combined with barriers reported by Mere [3], Yuensook et al. [23] and Coman et al. [28], this indicates a gap in equity-focused, Global-South-inclusive evaluation of adaptive learning systems.

Theme 3: Ethical and Governance Considerations

Fourteen records — the largest cluster in the matrix — address ethical risk and governance, reflecting the field's current preoccupation with responsible deployment. Seven axial categories (Table 3) synthesise into three sub-themes: core, recurrent ethical risk domains; governance fragmentation and the institutional response gap; and emerging and underexamined costs.

The selected publications show substantial thematic agreement that privacy, algorithmic bias and academic integrity constitute the dominant, recurring ethical risk triad, appearing independently across conceptual reviews [1], systematic reviews [5], [11], and regionally focused studies [10]. This recurrence across geographically and methodologically diverse publications indicates a stable pattern in the selected corpus: the risk domains themselves are not contested, even though their proposed remedies differ.

Table 3. Axial coding summary — ethical and governance considerations

Axial Category	Illustrative Open Codes	Contributing Studies	f
Privacy & data governance	data protection; surveillance risk; consent; data minimisation	[1], [3], [5], [6], [11], [12], [14], [17]	**8**
Bias, fairness & algorithmic equity	algorithmic bias; cultural/language bias; unequal access to benefits; fairness auditing	[1], [3], [5], [12], [14], [17], [18], [28]	**8**
Academic integrity, authorship & misconduct	plagiarism; ghostwriting; authorship ambiguity; contract-cheating risk	[5], [6], [7], [9], [10], [11], [14], [16], [18]	**9**
Transparency, accountability & human oversight	explainability (XAI); human-in-the-loop review; risk-tiering; accountability structures	[1], [12], [14], [17]	**4**
Regulatory & institutional policy fragmentation	fragmented/reactive policy; absent institutional frameworks; participatory governance gaps	[5], [7], [10], [11], [13], [28]	**6**
Contextual & regional readiness constraints	infrastructure gaps; Global South readiness; policy-capacity constraints	[10]	**1**
Environmental & sustainability costs	compute/energy footprint; underexamined environmental cost of GenAI use	[7]	**1**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

A contradiction exists between the sophistication of proposed governance mechanisms and their reported operational reality. Alfiras et al. [12] articulate concrete mechanisms — risk-tiering, human-in-the-loop review, audit trails and authentic assessment redesign — yet the same study concedes that existing institutional governance is often abstract and lacks operational detail, a tension echoed by Pireci Sejdiu et al. [7], who find that adoption outpaces governance, and by Kaşarcı et al. [11], who describe governance as reactive rather than proactive. A second, contextual contradiction is that readiness constraints reported by Owidi et al. [10] for Kenyan and wider African institutions are largely invisible in the Global-North-centred governance literature.

The most striking gap is environmental cost: of fourteen contributing studies, only Pireci Sejdiu et al. [7] raises GenAI's computational footprint as a governance concern, despite privacy and bias each drawing attention from eight studies. Empirical evaluation of governance-mechanism effectiveness is likewise absent from the corpus.

Theme 4: AI Literacy and Critical Thinking

Eleven records address the cognitive and literacy dimensions of GenAI use — the capacity of educators and learners to engage with these tools critically and productively. Four axial

categories (Table 4) synthesise into three sub-themes: dual potential (cultivation versus erosion of thinking); literacy as the primary mitigation mechanism; and underdeveloped higher-order engagement.

Table 4. Axial coding summary — AI literacy and critical thinking

Axial Category	Illustrative Open Codes	Contributing Studies	f
Critical thinking & metacognition cultivation	metacognitive scaffolding; reflective feedback loops; process-based assessment supporting reasoning	[1], [8], [13], [18]	**4**
Overreliance, cognitive offloading & de-skilling risk	automation bias; reduced critical reflection; dependence on AI output; erosion of independent reasoning	[6], [7], [9], [18]	**4**
Literacy development for students & educators	AI-literacy curricula; ethical-literacy training; educator upskilling; transparent co-writing practices	[3], [6], [10], [11], [12], [13], [14]	**7**
Higher-order cognitive skill gaps	limited evidence for analysis/evaluation/creation; concentration on lower-order recall and comprehension	[19]	**1**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

Agreement centres on the dual-edged nature of GenAI's cognitive impact. Several studies find that, when purposefully scaffolded, GenAI use can support metacognition and reflective reasoning [1], [18], [13], while a comparably sized set of studies documents overreliance and reduced critical reflection as a realised risk rather than a hypothetical one [6], [7], [9].A contradiction worth surfacing is that the overreliance/de-skilling literature largely assumes GenAI tools are cognitively capable enough to be worth offloading to, whereas Latif et al. [19] — the only study in the corpus to map AI personalisation explicitly against Bloom's taxonomy — finds that most tools concentrate on remembering, understanding and applying, with limited evidence of engagement at the analysis, evaluation or creation levels. The offloading risk documented elsewhere may therefore apply more to foundational tasks than to complex reasoning, a distinction the reviewed literature does not yet make explicit.

Literacy training is the most frequently recommended mitigation across the matrix, cited by seven of eleven contributing studies — yet none report an empirical evaluation of its effect on critical-thinking outcomes. This is the theme's central gap, and higher-order cognitive skill development remains markedly under-researched.

Discussion

The Results reported in the Result Section are organised by coding theme; the Discussion instead organises the same evidence around three overarching arguments, moving from what the literature shows toward what it means for responsible AI integration in higher education.

Technology Is Advancing Faster Than Pedagogy

Themes 1 and 2 together (Results Section, Themes 1 and 2) point to a single underlying pattern: the field evaluates GenAI primarily on technological and behavioural outcomes — did engagement rise, did output improve — rather than on whether its use is pedagogically designed. Only Belkina et al. [8] demonstrates, through an explicit instructional-design framework, that GenAI use can progress to genuine transformation rather than substitution Adoption is therefore running ahead of instructional-design maturity, and the multimodal and equity claims noted in the Result Section, Theme 2: Personalized and Adaptive Learning, remain more aspirational than realised.

Governance Remains Reactive

Theme 3's central finding (Result Section, Theme 3: Ethical and Governance Considerations) is not disagreement about risk but a persistent gap between proposed governance mechanisms [12] and their operational reality, described elsewhere in the corpus as abstract, fragmented and reactive [7], [11]. This is consistent with policy formation generally lagging diffusion of a general-purpose technology, particularly where individual adoption can outpace institutional decision-making. Readiness constraints in under-resourced systems [10] being largely absent from the dominant governance literature further indicates that reactive governance is compounded, not uniform, where institutional capacity is already limited.

The Future Challenge Is Preserving Human Judgement

If pedagogy is not yet keeping pace with technology (V-A) and governance is not yet keeping pace with risk (V-B), the residual safeguard in both cases is the individual's own critical judgement — precisely the capacity Results Section, Theme 4, reports as most uncertain. Literacy training is proposed almost universally as the remedy, yet no study empirically evaluates whether it protects critical-thinking outcomes, and current tools rarely engage higher-order cognition in the first place [19]. The field has arrived at a proposed solution without establishing that it addresses the actual mechanism of risk; preserving human judgement, not merely expanding AI capability or policy, is the more durable challenge.

Theoretical contribution

Existing literature frequently treats AI-enhanced teaching, personalised learning, ethical governance and AI literacy as separate areas of inquiry. The present synthesis instead proposes that responsible GenAI integration in higher education should be understood as an interacting, human-centred system. This represents the paper's principal theoretical contribution.

The novelty of this contribution lies not in the four domains themselves, each of which is already documented separately in the reviewed literature, but in the explicit, directional specification of how they relate. Prior syntheses typically treat pedagogy, personalisation, governance and literacy as parallel concerns; this framework instead specifies that AI literacy and critical thinking function as a necessary enabling condition between pedagogical application and human-centred outcome (Propositions 1 and 3), that governance operates as a moderator rather than a downstream compliance step (Proposition 2), and that graduate readiness is contingent on retained human responsibility rather than an automatic by-product of adoption (Proposition 4). It is this set of testable, directional relationships not the underlying themes in isolation — that constitutes the paper's original contribution.

Within the proposed framework, AI-enhanced teaching and personalised learning constitute the main pedagogical applications of GenAI. AI literacy and critical thinking function as enabling human capabilities that influence whether students and educators engage with GenAI reflectively or become excessively dependent on its outputs. Ethical governance, including privacy, fairness, transparency, accountability and meaningful human oversight, surrounds and moderates the use of GenAI across the educational system. The effectiveness and equity of these relationships are further conditioned by infrastructure, institutional capacity, disciplinary requirements and sociocultural context.

The intended outcomes of responsible integration are not technology adoption or efficiency alone. The framework positions professional judgement and graduate readiness as central human outcomes. These outcomes are not automatic consequences of GenAI use; they depend on whether institutions design learning environments in which people remain responsible for evaluating information, recognising uncertainty, explaining decisions and exercising context-sensitive judgement.

The framework is conceptual and does not represent a validated causal sequence. Instead, the synthesis generates five propositions for future empirical investigation:

Proposition 1 (Pedagogical alignment): GenAI is more likely to enhance learning when its use is aligned with explicit learning outcomes, disciplinary requirements and appropriate instructional design.

Proposition 2 (Governance moderation): Ethical governance moderates the relationship between GenAI adoption and educational benefit by constraining unacceptable risks, clarifying accountability and preserving meaningful human oversight.

Proposition 3 (AI literacy and cognitive engagement): AI literacy influences whether engagement with GenAI strengthens reflective learning and critical thinking or encourages cognitive offloading and excessive dependence on generated outputs.

Proposition 4 (Human responsibility and graduate readiness): Graduate readiness depends on retaining human responsibility for evaluating evidence, recognising uncertainty, justifying conclusions and making context-sensitive decisions.

Proposition 5 (Contextual conditions): Infrastructure, institutional capacity, disciplinary context and sociocultural conditions shape the effectiveness, accessibility and equity of the relationships proposed in the framework.

Ethical governance surrounds and moderates the pedagogical use of GenAI, recognising that principles of privacy, fairness, transparency, accountability and meaningful human oversight permeate all stages of AI adoption rather than emerging only after implementation. AI literacy and critical thinking are positioned as enabling human capabilities that support reflective engagement with GenAI and reduce the risk of cognitive offloading and excessive dependence on AI-generated outputs. Professional judgement and graduate readiness represent the intended human-centred outcomes, while infrastructure, institutional capacity, disciplinary context and sociocultural conditions determine the feasibility, effectiveness and equity of responsible GenAI integration across higher education settings. Critically, this influence runs in both directions: observed outcomes for professional judgement and graduate readiness are expected to feed back into and recalibrate ethical-governance mechanisms over time, consistent with the framework's characterisation as an interacting rather than linear system (Fig. 1).

Practical implications

Universities

Institutions should treat GenAI governance as an operational, not merely declarative, function — resourcing risk-tiering, human-in-the-loop review and audit mechanisms rather than issuing policy statements alone [11], [12].

Institutions in resource-constrained contexts should be supported through sector-level or regional coordination, given that readiness constraints are compounded, not merely repeated, where infrastructure and policy capacity are already limited [10].

Lecturers

Educators are best placed to close the pedagogy gap identified by adopting explicit instructional-design rather than relying on ad hoc or tool-led adoption [8]. Assessment design should assume GenAI availability and shift toward process-based and authentic assessment that rewards demonstrated reasoning over polished output [12].

Students

Students require explicit AI-literacy instruction that goes beyond tool operation to include critical evaluation of GenAI output, recognition of hallucination and bias, and awareness of when independent reasoning — rather than AI assistance — is the more appropriate response to a task [13], [14].

Policy Makers

National and sector regulators should prioritise closing the gap between proposed governance mechanisms and their operational detail identified in the Discussion Section of this article, and should extend governance guidance to explicitly address environmental and sustainability costs, currently

addressed by only one study in this corpus [7].

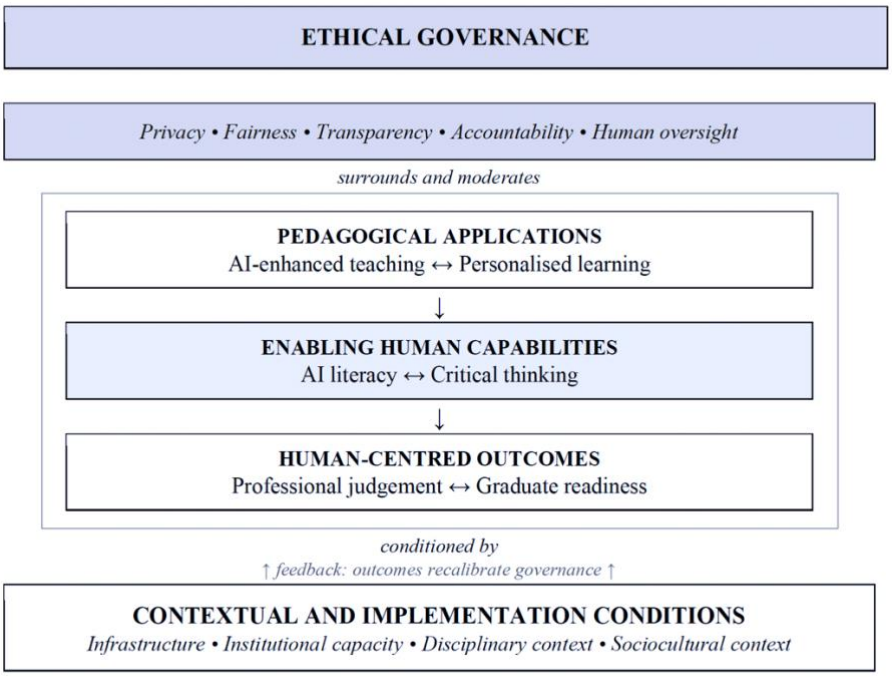


Fig. 1. visually represents the proposed human-centred framework and illustrates how the five propositions collectively conceptualise responsible GenAI integration as an interacting system rather than a linear process.

Curriculum Developers

Curriculum design should treat AI literacy and critical-thinking development as integrated rather than separate outcomes, embedding literacy instruction within disciplinary content rather than as a standalone module, and should build in evaluation mechanisms so that, unlike the studies reviewed here, literacy interventions can be assessed for their actual effect on critical-thinking outcomes rather than assumed to be effective.

Operationalising context: future work should treat "sociocultural context" in Proposition 5 as a measurable rather than background variable — for example, through indicators such as institutional GenAI-access rates, the language of instruction relative to the language(s) supported by deployed tools, and locally negotiated norms around acceptable AI assistance in assessment. Embedding such indicators into future empirical tests of the framework would allow context to function as an explanatory variable rather than a residual category.

Conclusion

This review synthesised 28 contemporary publications into an integrated conceptual account of responsible GenAI integration in higher education, addressing RQ1 (recurring themes) and RQ2 (their interaction). This matters because higher education's central claim to value the development of graduates capable of independent, disciplined professional judgement is precisely the capacity most exposed by the gaps this review identifies. Treating AI-enhanced teaching, personalised learning, ethical governance and AI literacy as four independent concerns, as much of the existing literature currently does, risks addressing each symptom without addressing the underlying system that connects them.

Future research should prioritise three directions: empirical, rather than conceptual, evaluation of governance-mechanism and literacy-intervention effectiveness; equity-focused study of adaptive and generative AI use in under-resourced and Global South institutional contexts; and longitudinal investigation of whether current patterns of GenAI use in higher education strengthen or erode graduates' higher-order cognitive and professional judgement over time.

These conclusions are interpretive rather than causal, for the reasons set out in the discussions around the trustworthiness. Future research should test the proposed framework through longitudinal, cross-institutional and Global South-inclusive designs.

Acknowledgment

The authors thank the University of Johannesburg for institutional support during the preparation of this manuscript.

Author contributions

In accordance with the CRediT (Contributor Roles) taxonomy, author contributions were as follows:

- Kurt Naicker: Conceptualisation; writing – review and editing; project correspondence (conference submission).
- Katlego Thipe: Conceptualisation; literature search and

screening; data curation (literature matrix); formal analysis (directed qualitative content analysis); writing – original draft; writing – review and editing.

Both authors approved the final version of the manuscript.

Refences

[1] S. Chakraborty, "Researching how AI can personalise learning paths, automate content creation, and provide intelligent feedback: Exploring generative AI in learning material creation and AI-powered assessment," International Journal on Science and Technology, 16(3), 1–6, 2025.

[2] J. Garzón, E. Patiño, and C. Marulanda, "Trends, benefits, and challenges of AI integration in education: A systematic review," Multimodal Technologies and Interaction, 9(6), Article 84, 2025. doi: 10.3390/mti9060084.

[3] K. Mere, "Exploring the role of generative AI and deep learning in personalized education content: A systematic literature review," AL-ISHLAH: Jurnal Pendidikan, 17(4), 2025. doi: 10.35445/alishlah.v17i4.8285.

[4] S. S. Lee and R. L. Moore, "Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review," Online Learning, 28(3), 82–104, 2024. doi: 10.24059/olj.v28i3.4593.

[5] I. M. García-López and L. Trujillo-Liñán, "Ethical and regulatory challenges of generative AI in education: A systematic review," Frontiers in Education, 10, Article 1565938, 2025. doi: 10.3389/feduc.2025.1565938.

[6] K. S. N. Rahmawati, O. A. Suciptaningsih, and A. E. Anggraini, "Generative AI in educational research: Opportunities, integrity risks, ethical challenges—A PRISMA-guided systematic review," Journal of Innovation and Research in Primary Education, 5(1), 1522–1536, 2026. doi: 10.56916/jirpe.v5i1.3106.

[7] N. Pireci Sejdiu, S. Grassini, and S. Sejdiu, "Generative AI in education: Process-aware pedagogy, assessment integrity, and institutional governance [Version 1; peer review: 1 approved, 2 approved with reservations]," Open Research Europe, 6, Article 161, 2026. doi: 10.12688/openreseurope.23343.1.

[8] M. Belkina, S. Daniel, S. Nikolic, R. Haque, S. Lyden, P. Neal, S. Grundy, and G. M. Hassan, "Generative artificial intelligence in higher education: A systematic review of empirical case studies and practical guidance," Computers and Education: Artificial Intelligence, 8, Article 100475, 2025. doi: 10.1016/j.caeai.2025.100475.

[9] X. Wang, Z. Zainuddin, and C. H. Leng, "Generative artificial intelligence in pedagogical practices: A systematic review of empirical studies (2022–2024)," Cogent Education, 12(1), Article 2485499, 2025. doi: 10.1080/2331186X.2025.2485499.

[10] S. O. Owidi, J. N. Lyanda, E. O. Okono, and E. M. Micheni, "The dark side of generative AI in higher education: Challenges, ethical concerns, and implications," East African Journal of Arts and Social Sciences, 8(3), 526–543, 2025. doi: 10.37284/eajass.8.3.3690.

[11] I. Kaşarcı, Z. Akın Demircan, G. Çeliker Ercan, and T. İnci, "Managing artificial intelligence ethics in higher education: A systematic framework for issues and policy recommendations," International Journal of Current Educational Studies, 4(2), 112–137, 2025. doi: 10.46328/ijces.223.

[12] M. I. I. Alfiras, A. Q. Emran, and A. M. Mohamed, "Ethics and governance of generative AI in education: A systematic review on responsible adoption," Discover Education, 5, Article 37, 2026. doi: 10.1007/s44217-025-01051-y.

[13] T. Tillmanns, A. Salomão Filho, S. Rudra, P. Weber, J. Dawitz, E. Wiersma, D. Dudenaite, and S. Reynolds, "Mapping tomorrow's teaching and learning spaces: A systematic review on GenAI in higher education," Trends in Higher Education, 4(1), Article 2, 2025. doi: 10.3390/higheredu4010002.

[14] M. Sharma, "The ethical considerations of using Gen AI and AI tools in academic writing in higher education: A systematic review," Kalika Journal of Multidisciplinary Research, 3(1), 82–93, 2025.

[15] M. Á. Rodríguez-Ortiz, P. C. Santana-Mancilla, and L. E. Anido-Rifón, "Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges," Applied Sciences, 15(15), Article 8679, 2025. doi: 10.3390/app15158679.

[16] T. Chanpradit, "Generative artificial intelligence in academic writing in higher education: A systematic review," Edelweiss Applied Science and Technology, 9(4), 889–906, 2025. doi: 10.55214/25768484.v9i4.6128.

[17] R. Chinoracky and N. Stalmasekova, "Ethical problems in the use of artificial intelligence by university educators," Education Sciences, 15(10), Article 1322, 2025. doi: 10.3390/educsci15101322.

[18] G. A. Thomas, S. Abu, A. B. Ajayi, P. A. Frempong, D. Mensah, J. O. Badmus, and A. H. Etim, "Navigating generative AI in education: A narrative review of teaching practices and ethical challenges," Asian Journal of Education and Social Studies, 51(11), 507–521, 2025. doi: 10.9734/ajess/2025/v51i112630.

[19] M. S. Latif, A. S. Bhutta, N. N. Gondal, and S. Yasin, "AI-driven personalized learning in higher education: A systematic review through the lens of Bloom's taxonomy," ACADEMIA International Journal for Social Sciences, 4(4b), 277–300, 2025. doi: 10.63056/academia.4.4(b).2025.1618.

[20] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," Qualitative Health Research, 15(9), 1277–1288, 2005. doi: 10.1177/1049732305276687.

[21] A. B. S. Nayak and S. N. Radhika, "Generative artificial intelligence in higher education: Applications, challenges and future directions," International Journal of Science and Research, 15(5), 1440–1441, 2026. doi: 10.21275/SR26522155915.

[22] Y. Jiang, Y. Fan, and Z. Liu, "Generative AI in art education: A systematic review of research trends, tool applications, and outcomes (2019–2025)," Education Sciences, 16(1), Article 47, 2026. doi: 10.3390/educsci16010047.

[23] T. Yuensook, T. Jantakoon, and P. Limpinan, "AI-driven adaptive learning systems in higher education: A systematic review," Journal of Education and Learning, 15(2), 117–132, 2026. doi: 10.5539/jel.v15n2p117.

[24] A. Fortuna, F. Prasetya, A. D. Samala, S. Rawas, S. Criollo-C, D. Kaya, M. Raihan, W. Andriani, D. Safitri, and R. A. Nabawi, "Artificial intelligence in personalized learning: A systematic literature review," Social Sciences & Humanities Open, 12, Article 101896, 2025. doi: 10.1016/j.ssaho.2025.101896.

[25] H. Farhood, M. Nyden, A. Beheshti, and S. Muller, "Artificial intelligence-based personalised learning in education: A systematic literature review," Discover Artificial Intelligence, 5, Article 324, 2025. doi: 10.1007/s44163-025-00598-x.

[26] J. Peng and Y. Li, "Frontiers of artificial intelligence for personalized learning in higher education: A systematic review of leading articles," Applied Sciences, 15(18), Article 10096, 2025. doi: 10.3390/app151810096.

[27] X. Yuan, Z. Ding, Z. Zhang, Y. Zhou, L. Zhou, and Z. Ge, "Personalized learning resources created by generative artificial intelligence: A systematic review," In Proceedings of the 2025 11th International Conference on Education and Training Technologies (pp. 84–89). IEEE, 2025. doi: 10.1109/ICETT66247.2025.11137092.

[28] C. Coman, V. Gherheș, A. Bucs, E. Coman, V.-R. Cioca, D.-C. Bódi, and R.-M. Șerbănescu, "Navigating opportunities and challenges of generative AI in higher education: Towards responsible, equitable, and human-centered integration," Frontiers in Artificial Intelligence, 9, Article 1750978, 2026. doi: 10.3389/frai.2026.1750978.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Inclusive Adaptation to AI: Adult Education, Digital Skills, and Older Workers' Employment in the EU

Ana Pavković[1], Petar Mišević[1], and Nikolina Pejović[2]
(corresponding author, e-mail: apavkovic@unin.hr), (e-mail: pmisevic@unin.hr), (e-mail: nikolina.pejovic@t.ht.hr)

[1]Department of Economics, University North, Varaždin, Croatia
[2]Croatian Telecom, Zagreb, Croatia

ARTICLE DOI: https://doi.org/10.68286/samyrad2026.301

ABSTRACT

This study examines how artificial intelligence adoption, adult education participation, digital skills, and tertiary educational attainment are associated with older workers' employment in the European Union. Driven by population aging and technological change, the analysis examines whether national labor-market systems across 27 EU member states provide conditions associated with stronger aggregate employment among people aged 55–64. The study evaluates employment in relation to combined technological, educational, digital, and human-capital conditions rather than artificial intelligence adoption alone. Using the employment rate of individuals aged 55–64 as the main outcome, fuzzy-set Qualitative Comparative Analysis (fsQCA) identifies configurations associated with high and low employment. Calibration sensitivity and a temporal comparison with 2023 assess the robustness of the employment-based findings. No single condition is necessary for high or low employment. No configuration meets the predefined criteria for high employment in 2025, whereas four pathways are associated with low employment. The findings identify three descriptive patterns: weak foundations, a technology–learning mismatch, and strong worker-side resources combined with limited AI diffusion. They suggest considering technological diffusion, learning opportunities, and digital inclusion together.

KEYWORDS AI adoption, fuzzy-set Qualitative Comparative Analysis, adult education participation, digital inclusion, active aging

How to cite: Pavković, A., Mišević, P., & Pejović, N. (2026). Inclusive Adaptation to AI: Adult Education, Digital Skills, and Older Workers' Employment in the EU. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.301

Introduction

Artificial intelligence is changing European labor markets by affecting tasks, work organization, and skill requirements. Although AI adoption can support productivity growth and new forms of work organization, its effects are not equally distributed among workers and across countries [1]–[3]. This issue is particularly important in the context of population aging, because European economies increasingly need to extend working lives and support employment at older ages. Therefore, technological change should be considered not only as a source of productivity gains but also as a potential source of exclusion for workers with limited adaptive capacity.

Demographic trends, employment policy, and the inclusiveness of digital and educational systems shape older workers' employment. As AI adoption expands, remaining active and employed at older ages may depend on whether national systems provide the conditions needed for adaptation. Evidence on adult learning is context-dependent. Formal educational upgrading after age 40 was associated with subsequent labor-market participation in Norway [4], whereas a Swedish study found no effect of adult education on retirement timing [5]. Employer-side evidence from Poland

also indicates that computer skills and training requirements can reduce the recruitment prospects of workers aged 50 and over [6].

Existing research has examined the labor market effects of AI, the role of adult education in later working life, and the importance of digital skills for employability. However, these research areas are usually analyzed separately. This represents an important limitation because older workers' employment may depend on the joint presence or absence of several conditions, rather than on a single condition. Less attention has been paid to how technological adoption, adult education participation, digital skills, and human capital combine at the country level. For example, adult education may be less effective when digital skills are limited, and AI adoption may not support employment if older workers lack the foundational skills needed to adapt.

The paper addresses this gap by examining how AI adoption, adult education participation, digital skills, and tertiary educational attainment relate to older workers' employment in the European Union. The empirical analysis uses country-level data for the 27 EU member states. The employment rate of individuals aged 55–64 operationalizes aggregate employment attachment at the country level; it does not measure individual job retention or employment transitions. The AI adoption indicator is interpreted as a country-level measure of the technological environment rather than as direct AI exposure or use among older workers. The applied methodology is fuzzy-set Qualitative Comparative Analysis (fsQCA), which identifies configurations of conditions associated with stronger or weaker outcomes.

The paper's scientific contribution lies in examining older workers' employment in relation to combined national adaptation conditions rather than AI adoption alone. By identifying multiple pathways to low employment, the analysis shows how similar employment outcomes can be associated with different combinations of adult learning, digital inclusion, and accumulated human capital.

Background and Conceptual Framework

The literature on AI and labor markets shows that the labor-market effects of technological change depend not only on adoption, but on how new technologies interact with task structures, skill requirements, and work organization [1], [3]. However, these effects are not equally distributed across workers and countries [2]. This is particularly relevant for older workers because population aging increases the need for longer working lives, while technological change can increase the risk of skill obsolescence.

The effects of AI, therefore, cannot be assessed only through the level of technological adoption. AI technologies include language and image processing, speech recognition, machine learning, and workflow automation. Their effects also depend on workers' and institutions' capacity to adapt to new skill requirements [1], [3], [7]. AI adoption may increase the demand for analytical, digital, and complementary skills, while reducing the value of routine tasks. For this reason, adaptation to AI requires upskilling, reskilling, and organizational support for learning. Countries with higher AI adoption are not necessarily better positioned to support older workers' employment, as technological change delivers such benefits only when accompanied by adequate learning opportunities and digital inclusion.

Adult education matters because it helps workers update their skills throughout their life course. Training participation was positively associated with retirement-eligible workers' decisions to remain employed in Dutch samples, particularly where organizational climates supported older-worker development [8]. However, participation and its effects are not uniform. Evidence indicates that age stereotypes, job skill requirements, time, organizational learning cultures, and workers' perceived growth needs shape recruitment and learning opportunities [6], [8], [9]. These findings suggest that continued learning depends on both individual motivation and supportive organizational conditions.

Digital skills are also important for older workers' employment in an AI-intensive labor market. They include the ability to find and evaluate information, communicate online, create digital content, protect personal data, and solve problems using digital tools. Even when workers do not directly use advanced AI tools, they may still need basic digital skills to adapt to changing work processes, access training, and remain employable. Digital inclusion can be understood as a foundation for broader adaptation to technological change. Tertiary educational attainment is also relevant because it reflects accumulated human capital and the capacity to adjust to more complex and changing work environments.

Based on these arguments, older workers' employment results from a combination of technological, educational, digital, and human capital conditions. The study, therefore, does not assume that any single condition is sufficient on its own. Instead, it assumes that different combinations of conditions may produce similar labor market outcomes. For example, adult education participation may support employment only when paired with adequate digital skills, whereas AI adoption may not strengthen employment if older workers lack the digital or educational foundations needed to adapt.

The four conditions capture distinct but complementary dimensions of national adaptation capacity, consistent with evidence that the labor-market implications of technological change depend on skills, learning opportunities, and organizational adaptation [1], [3], [7]–[9]. AI adoption represents firm-level technological diffusion; adult education participation captures continued learning opportunities at older ages; digital skills reflect the foundational capability required to participate in increasingly digital work and training; and tertiary educational attainment represents accumulated formal human capital. Together, these conditions allow the analysis to assess whether technological diffusion is accompanied by learning opportunities and workforce capabilities.

The framework is deliberately bounded and does not capture all determinants of older workers' employment. It excludes health, pension incentives, sectoral employment structures, and employer age-management practices, while the AI indicator does not measure direct exposure to or use of AI among older workers. These boundaries define the scope of the configurational analysis presented in the following section.

The four conditions also correspond to complementary elements of the current European policy architecture. The Digital Decade combines targets for digital skills and business digitalization, the European Pillar of Social Rights recognizes adult learning as a social and labor-market priority, and the Union of Skills places lifelong learning and workforce adaptability within the EU competitiveness agenda [10]–[12]. Together, these frameworks imply that technological diffusion, learning provision, and foundational capabilities should be treated as an integrated policy package rather than as separate interventions.

Recent evidence cautions against simple causal claims about skills development and extended working lives. A systematic review of prospective cohort studies reports mixed findings: several included studies associated skills development or opportunities for skills development with longer working lives, but the overall certainty of evidence was very low [13]. A systematic review of digital vocational education for older workers identifies flexible delivery and age-appropriate learning design as important for participation and engagement [14]. Together, these findings support examining training within its wider digital and organizational context.

Data and Method

The study uses a country-level dataset for the 27 European Union member states. All indicators are taken from Eurostat datasets to ensure cross-country comparability [15]. The main empirical analysis focuses on 2025, when all core indicators are available, while 2023 serves as a temporal robustness check. Employment among individuals aged 55–64 is the principal outcome; activity is retained only as a secondary outcome check.

The explanatory conditions describe four dimensions of national adaptation capacity. Artificial intelligence adoption is measured by the share of enterprises with 10 or more employees using at least one AI technology. These technologies include text mining, natural language generation, speech and image recognition, machine learning for data analysis, and AI-based workflow automation or decision support [15]. The indicator is interpreted as a country-level measure of the technological environment and does not capture direct AI exposure or use among older workers.

Adult education participation is measured by the participation rate in education and training among individuals aged 55–64. Digital inclusion is measured by the share of individuals aged 55–64 with basic or above-basic overall digital skills. The digital-skills indicator covers information and data literacy, communication and collaboration, digital content creation, safety, and problem-solving [15]. Human capital is measured by the share of individuals aged 55–64 with tertiary educational attainment.

The model therefore combines an enterprise-level AI indicator with individual-level indicators aggregated to the country level; it does not link the same workers to the enterprises reporting AI use. Table 1 reports descriptive statistics. Eurostat dataset codes are lfsa_ergan (EMP), isoc_eb_ai (AI), trng_lfse_01 (AET), isoc_sk_dskl_i21 (DIG), and edat_lfse_03 (TER).

Table 1. Descriptive Statistics

Variable	Mean	Median	Min	Max
EMP	66.8	69.4	49.5	78.3
ACT	69.9	71.8	51.9	83.7
AI	20.7	19.6	5.2	42.0
AET	10.3	8.5	2.1	31.2
DIG	49.2	46.6	18.0	80.0
TER	28.7	29.8	11.4	43.4

Note: EMP = employment 55–64; ACT = activity 55–64; AI = enterprises using at least one AI technology; AET = adult education/training 55–64; DIG = basic or above-basic digital skills 55–64; TER = tertiary attainment 55–64. All variables are expressed as percentages. N = 27 EU member states.

Descriptive statistics show clear cross-country variation in older workers' employment and adaptation conditions. Employment among individuals aged 55–64 ranges from 49.5% to 78.3%, while AI adoption ranges from 5.2% to 42.0%. Variation is also visible in adult education participation, digital skills, and tertiary attainment. This supports a configurational approach because EU member states differ not only in individual indicators but also in combinations of conditions.

The variables are calibrated into fuzzy-set membership scores using direct calibration, with the 10th percentile representing full non-membership, the median representing the crossover point, and the 90th percentile representing full membership (Table 2) [16], [17]. Because externally validated substantive thresholds are not available consistently across all indicators, these distribution-based anchors provide a transparent comparative calibration for the full population of 27 EU member states. They distinguish relatively low, intermediate, and high positions within the EU distribution rather than compliance with absolute substantive benchmarks.

Because the 10th and 90th percentiles are located near the distribution tails in a small sample, an additional check uses the 20th percentile, median, and 80th percentile to assess sensitivity to the placement of the outer anchors. Exact crossover memberships are assigned to 0.501 in the main analysis, while a 0.499 specification evaluates sensitivity to the opposite assignment. The 2023 sets are recalibrated using their own annual distributions.

The empirical strategy relies on fuzzy-set Qualitative Comparative Analysis (fsQCA), which identifies configurations associated with an outcome rather than isolated net effects [16], [17]. The analysis uses QCA version 3.25 and conservative minimization. Consistency assesses how closely the observed set memberships conform to the tested necessity or sufficiency relation. Proportional reduction in inconsistency (PRI) assesses whether a sufficient configuration is associated with the outcome rather than also with its negation. Coverage indicates the empirical relevance of a sufficient solution by showing how much of the outcome membership it covers. Necessity is evaluated at a consistency threshold of 0.90.

Table 2. Variables and Calibration Anchors

Set	Calibration Anchors		
	Out	Cross	In
EMP	57.4	69.4	76.3
ACT	59.3	71.8	79.5
AI	8.78	19.6	34.7
AET	3.6	8.5	19.1
DIG	31.4	46.6	72.8
TER	19.6	29.8	38.9

Note: EMP = employment 55–64; ACT = activity 55–64; AI = enterprises using at least one AI technology; AET = adult education/training 55–64; DIG = basic or above-basic digital skills 55–64; TER = tertiary attainment 55–64. Out/Cross/In are the p10/p50/p90 empirical calibration anchors and represent relatively low, intermediate, and high positions within the annual EU distribution. Values are percentages.

For the sufficiency analysis, each truth-table row represents a combination of the presence or absence of the four conditions. A row is included in the minimization when it contains at least one case and meets the consistency threshold of 0.80 and the PRI threshold of 0.50. Limited diversity occurs when not all logically possible combinations are observed. Conservative minimization uses only observed truth-table rows and does not use unobserved logical remainders to simplify the solution. Prime implicants and complete solutions must also reach 0.80 consistency.

The configurational model is expressed as:

$$Y_i = f(AI_i, AET_i, DIG_i, TER_i) \quad (1)$$

where *i* denotes an EU member state, and *Y* denotes the calibrated outcome set. Older workers' employment is evaluated as EMP and its complement LOW_EMP. This notation preserves set-theoretic asymmetry: configurations associated with low employment need not be the inverse of configurations associated with high employment.

The analysis is conducted in two steps. First, a necessity analysis evaluates whether any condition, or its absence, is required for the outcome. Second, truth tables and Boolean minimization identify sufficient configurations [16], [17]. In fsQCA, sufficiency means that membership in a configuration is consistently associated with membership in the outcome set; it does not establish causation.

Robustness is assessed through the alternative 0.499 crossover assignment, the recalibrated 2023 model, and a secondary activity-outcome check. Three additional checks examine the sensitivity of the 2025 LOW_EMP solution to alternative p20/p50/p80 calibration anchors, a stricter PRI threshold of 0.60, and a frequency threshold of two cases. The stricter PRI threshold requires a clearer distinction between the outcome and its negation, while the higher frequency threshold tests whether the solution depends on truth-table rows represented by a single country.

Results and Discussion

The necessity analysis shows that no individual condition, or its absence, meets the conventional consistency threshold of 0.90 for high or low employment. The conclusion is virtually unchanged under the 0.501 and 0.499 crossover assignments. Older-worker labor-market disadvantage therefore cannot be reduced to a single technological, educational, or digital condition.

At the predefined thresholds, no sufficient configuration was identified for high employment in 2025. This finding is consistent with set-theoretic asymmetry because the conditions associated with low employment need not be the inverse of those associated with high employment. However, it may also reflect the limited diversity of the 27-country sample, the predefined consistency and PRI thresholds, and conservative minimization. It should therefore not be interpreted as proof of causal asymmetry or as evidence that pathways to high employment do not exist. By contrast, LOW_EMP produces four conservative pathways (Table 3), with a solution consistency of 0.821, a PRI of 0.734, and a coverage of 0.699. The four pathways are obtained by conservatively minimizing truth-table rows that meet the predefined frequency, consistency, and PRI thresholds. They are interpreted below as alternative patterns associated with LOW_EMP.

Table 3. Final fsQCA Solution for Low Employment

Condition	Pathway			
	P1	P2	P3	P4
AI	⊗	●		⊗
AET		⊗	●	●
DIG	⊗		⊗	●
TER	⊗	●	⊗	●
Cons.	0.833	0.852	0.887	0.888
PRI	0.741	0.697	0.763	0.679
Raw cov.	0.510	0.261	0.330	0.199
Uniq. cov.	0.224	0.075	0.046	0.019

Note: ● indicates the presence of a condition, ⊗ indicates the absence of a condition, and blank cells indicate that the condition is not part of the minimized pathway. AI = enterprise AI adoption; AET = adult education and training participation among people aged 55–64; DIG = basic or above-basic digital skills among people aged 55–64; TER = tertiary attainment among people aged 55–64. Cons. = consistency; PRI = proportional reduction in inconsistency; Raw cov. = raw coverage; Uniq. cov. = unique coverage. Solution consistency = 0.821; PRI = 0.734; solution coverage = 0.699. Cases: P1 = BG, EL, HU, IT, PL, RO, SK; PT; P2 = LT; DE; P3 = PT; MT, SI; P4 = FR. Semicolons separate distinct observed truth-table configurations covered by the same minimized expression. The main analysis assigns exact crossover memberships to 0.501.

Figure 1 shows the sufficiency relation between membership in the combined 2025 LOW_EMP solution and membership in the LOW_EMP outcome.

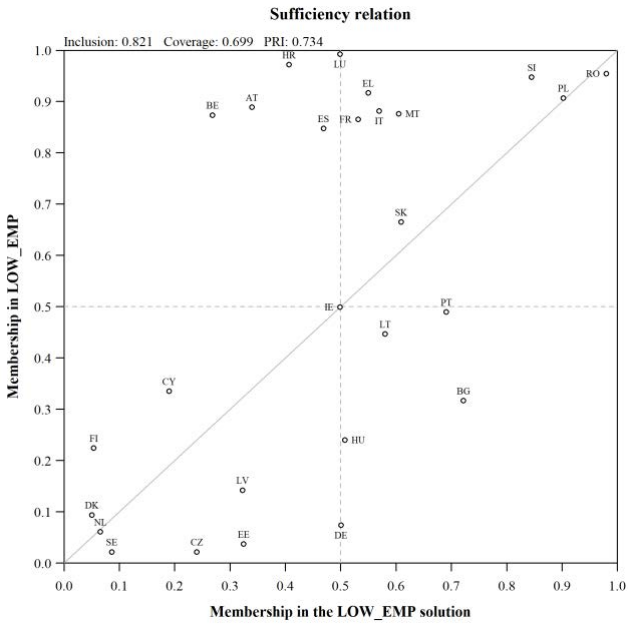


Fig. 1. Sufficiency plot for the 2025 LOW_EMP solution

To make the country distribution of the verified main solution transparent, Table 4 reports the cases covered by each 2025 LOW_EMP pathway. Semicolons separate cases that belong to distinct observed truth-table configurations covered by the same minimized expression.

Table 4. Country Membership in the Low-Employment Solution

Outcome	Pathway	Country membership	Main pattern
LOW_EMP	P1	BG, EL, HU, IT, PL, RO, SK; PT	Low AI, low DIG, and low TER
LOW_EMP	P2	LT; DE	High AI, low AET, and high TER
LOW_EMP	P3	PT; MT, SI	High AET, low DIG, and low TER
LOW_EMP	P4	FR	Low AI, high AET, high DIG, and high TER

Note: The table reports cases covered by the four pathways in the main 2025 LOW_EMP solution; membership is not exclusive. LOW_EMP = low employment among people aged 55–64; AI = enterprise AI adoption; AET = adult education and training participation among people aged 55–64; DIG = basic or above-basic digital skills among people aged 55–64; TER = tertiary attainment among people aged 55–64. "Low" and "high" denote calibrated fuzzy-set membership. BG = Bulgaria; DE = Germany; EL = Greece; FR = France; HU = Hungary; IT = Italy; LT = Lithuania; MT = Malta; PL = Poland; PT = Portugal; RO = Romania; SI = Slovenia; SK = Slovakia. Additional country codes shown in Figure 1 are: AT = Austria; BE = Belgium; HR = Croatia; CY = Cyprus; CZ = Czechia; DK = Denmark; EE = Estonia; FI = Finland; IE = Ireland; LV = Latvia; LU = Luxembourg; NL = Netherlands; ES = Spain; and SE = Sweden. Semicolons separate distinct observed truth-table configurations covered by the same minimized expression.

The pathway labels used below are descriptive and do not imply causal relationships. The first and broadest pathway combines low AI adoption, weak digital skills, and low tertiary attainment (~AI*~DIG*~TER). It accounts for the largest share of LOW_EMP and describes a configuration in which low employment coexists with limited technological diffusion, basic digital capability, and accumulated human capital. The same combination of weak digital skills and low tertiary attainment also appears in the 2023 comparison. Poland clearly illustrates P1: low AI adoption, weak digital skills, and low tertiary attainment coexist with high membership in LOW_EMP.

The second pathway combines high AI adoption, low adult education participation, and high tertiary attainment (AI*~AET*TER). In this configuration, accumulated education and technological diffusion coexist with limited continued learning and weak employment. P2 covers Lithuania and Germany and should therefore be interpreted as a case-specific pattern rather than a general EU pattern. The third pathway (AET*~DIG*~TER) describes a configuration in which high adult learning coexists with low employment and weak digital and educational foundations. P3 is calibration-sensitive because its exact form changes under the alternative crossover assignment and should therefore be interpreted with caution.

The fourth pathway combines low AI adoption with high adult learning, digital skills, and tertiary attainment (~AI*AET*DIG*TER). In this configuration, favorable worker-side resources coexist with weak employment, while firm-level AI diffusion remains limited. P4 is represented only by France and should therefore be interpreted cautiously. It remains stable across the two 2025 crossover assignments but does not recur in 2023 and is therefore treated as period-specific rather than as a general EU pattern.

Table 5 summarizes the robustness checks. In addition to the alternative crossover assignment, the 2023 comparison, and the secondary activity outcome, the additional checks vary the outer calibration anchors, the PRI threshold, and the minimum number of cases required for a truth-table row.

Table 5. Analytical Robustness Checks

Check	Solution consistency	PRI	Coverage
2025, crossover 0.501	0.821	0.734	0.699
2025, crossover 0.499	0.805	0.710	0.686
2025, alternative anchors	0.805	0.739	0.661
2025, PRI threshold 0.60	0.821	0.734	0.699
2025, frequency threshold 2	0.854	0.777	0.524
2023 LOW_EMP	0.818	0.741	0.767
2025 LOW_ACT	No solution ≥ 0.80	-	-

Note: The main 2025 model uses p10/p50/p90 calibration anchors, a 0.501 crossover assignment, a consistency threshold of 0.80, a PRI threshold of 0.50, and a frequency threshold of one case. Each additional 2025 check changes only the stated specification. The alternative-anchor check uses p20/p50/p80. The 2023 model is recalibrated using its annual distribution. LOW_ACT produced no conservative solution at the predefined solution-consistency threshold of 0.80.

The p20/p50/p80 calibration retains the same four pathways, with a solution consistency of 0.805, a PRI of 0.739, and a coverage of 0.661. Raising the PRI threshold to 0.60 reproduces the main solution and its parameters. When the frequency threshold is raised to two cases, the solution narrows to two more specific configurations nested within P1 and P3. Consistency increases to 0.854 and PRI to 0.777,

while coverage decreases to 0.524. This result shows that the multi-case configurations are more consistent but cover a smaller share of LOW_EMP. It also supports the cautious interpretation of P2 and P4, which depend on truth-table rows represented by single countries. Across the remaining checks, weak digital and human-capital foundations and the technology–learning mismatch remain the most relevant recurring patterns.

Overall, inclusive adaptation to AI cannot be reduced to technological adoption alone. The results identify different combinations of technological diffusion, learning participation, digital inclusion, and accumulated human capital. These patterns point to different policy priorities.

Rather than providing a single ranking of national performance, Table 6 links each policy priority to the conditions present or absent in the corresponding 2025 LOW_EMP pathway. These implications are diagnostic and configuration-sensitive; they do not estimate causal treatment effects or support universal prescriptions.

Table 6. Configuration-Specific Policy Implications

Pathway	Policy bottleneck	Priority response
P1	Weak digital and educational foundations	Combine basic digital-skills provision with educational upgrading before or alongside AI diffusion.
P2	Technology-learning mismatch	Strengthen accessible mid- and late-career learning and employer-supported adaptation to AI-enabled work.
P3	Learning without foundation; calibration-sensitive	Integrate adult learning with foundational digital and educational support; avoid stand-alone training measures.
P4	Strong worker-side resources with limited AI diffusion	Support responsible AI adoption and work redesign so existing skills translate into age-inclusive employment.

Note: Implications are derived from the verified 2025 LOW_EMP pathways. Pathway labels are descriptive and do not imply causal effects. P3 is calibration sensitive.

For the weak-foundations pathway, digital-skills objectives and adult-learning systems need to operate together, consistent with the combined orientation of current EU skills and digital-policy frameworks [10]–[12]. For the technology–learning mismatch, policy should emphasize age-appropriate and flexible training design rather than access alone [14]. Because evidence on whether skills development extends working lives remains uncertain [13], these recommendations should be treated as pathway-specific priorities rather than universal causal prescriptions.

Croatia illustrates why this distinction matters. In 2025, its basic-or-above-basic digital-skills rate among people aged 55–64 was 50.0%, slightly above the EU-27 median of 46.6%, whereas adult education participation was 2.4% compared with 8.5%, tertiary attainment was 19.6% compared with 29.8%, and enterprise AI adoption was 15.2% compared with 19.6%. This profile therefore highlights not only digital access, but also continued learning, accumulated human capital, and organizational technology adoption. This descriptive profile does not, by itself, assign Croatia to a sufficient pathway.

Conclusion

This paper examined how AI adoption, adult education participation, digital skills, and tertiary educational attainment relate to older workers' employment in EU member states. No individual condition is necessary for high or low employment. In 2025, no sufficient configuration was identified for high employment, whereas four were identified for low employment. This difference is consistent with set-theoretic asymmetry, but it may also reflect the limited diversity of the 27-country sample, the predefined consistency and PRI thresholds, and conservative minimization. It does not demonstrate causal asymmetry.

The main contribution is to distinguish persistent and period-specific forms of adaptation disadvantage. The most stable cross-time pattern combines weak digital skills and low tertiary attainment. A second recurring pattern combines high AI adoption with limited continued learning despite stronger accumulated human capital. A third, 2025-specific configuration combines strong learning and skill resources with limited AI diffusion and weak employment.

The findings point to different policy priorities. For countries matching the weak-foundations pattern, these include integrated digital and educational capacity building. For those matching the technology–learning mismatch, they include access to relevant mid- and late-career training. These are configuration-specific priorities rather than universal prescriptions.

The pattern combining strong learning and skill resources with limited AI diffusion points to responsible AI diffusion, work redesign, and organizational support as possible priorities. Because only France represents this pathway and it does not recur in 2023, its interpretation remains case-specific and period-specific.

Pathway-based interpretation provides a configuration-sensitive approach to implementing EU and national policy instruments [10]–[12]. The four-path solution remains unchanged under the alternative calibration anchors and the stricter PRI threshold. Requiring at least two countries per truth-table row yields a narrower solution with lower coverage, reinforcing the need for caution when interpreting single-country configurations.

Limitations include the small number of national cases, reliance on empirical calibration, the cross-sectional design, use of aggregate indicators, and a mismatch

between enterprise-level AI adoption and individual-level skill measures. Configurations supported by only one or two countries are case-specific and should not be generalized without further evidence. In particular, the AI indicator does not capture older workers' direct exposure to or use of AI. Employment rates do not track individual job retention or transitions. Future research should use longitudinal microdata and sector-level AI exposure to test whether the identified configurations predict continued employment.

Acknowledgment

Dr. Ana Pavković acknowledges financial support from University North, Croatia, under its institutional funding scheme for young researchers.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Lane and A. Saint-Martin, "The impact of artificial intelligence on the labour market: What do we know so far?," OECD Social, Employment and Migration Working Papers, no. 256, OECD Publishing, Paris, 2021.
[2] S. Albanesi, A. Dias da Silva, J. F. Jimeno, A. Lamo, and A. Wabitsch, "New technologies and jobs in Europe," Economic Policy, vol. 40, no. 121, pp. 71–139, 2025.
[3] R. Cramarenco, M. I. Burcă-Voicu, and D.-C. Dabija, "The impact of artificial intelligence (AI) on employees' skills and well-being in global labor markets: A systematic review," Oeconomia Copernicana, vol. 14, no. 3, pp. 731–767, 2023.
[4] T. Midtsundstad and R. A. Nielsen, "Lifelong learning and the continued participation of older Norwegian adults in employment," European Journal of Education, vol. 54, no. 1, pp. 48–59, 2019.
[5] A. Stenberg, X. de Luna, and O. Westerlund, "Can adult education delay retirement from the labour market?," Journal of Population Economics, vol. 25, pp. 677–696, 2012.
[6] K. Turek and K. Henkens, "How skill requirements affect the likelihood of recruitment of older workers in Poland: The indirect role of age stereotypes," Work, Employment and Society, vol. 34, no. 4, pp. 550–570, 2020.
[7] S. Morandini, F. Fraboni, M. De Angelis, G. Puzzo, D. Giusino, and L. Pietrantoni, "The impact of artificial intelligence on workers' skills: Upskilling and reskilling in organisations," Informing Science: The International Journal of an Emerging Transdiscipline, vol. 26, pp. 39–68, 2023.
[8] Y. Li, K. Turek, K. Henkens, and M. Wang, "Retaining retirement-eligible older workers through training participation: The joint implications of individual growth need and organizational climates," Journal of Applied Psychology, vol. 108, no. 6, pp. 954–976, 2023.
[9] A. L. Wulff and A. J. Lassen, "Capacity for competence development: Unlocking potential for lifelong learning in later working life," Journal of Aging & Social Policy, vol. 37, no. 6, pp. 1125–1146, 2025.
[10] European Parliament and Council of the European Union, "Decision (EU) 2022/2481 establishing the Digital Decade Policy Programme 2030," Official Journal of the European Union, L 323, pp. 4–26, 2022.
[11] European Commission, The European Pillar of Social Rights Action Plan. Luxembourg: Publications Office of the European Union, 2021.
[12] European Commission, "Union of Skills," 2025. [Online]. Available: https://commission.europa.eu/topics/competitiveness/union-skills_en. Accessed: Jun. 25, 2026.
[13] K. G. V. Seeberg, M. N. Stilling, J. Vinstrup, E. Sundstrup, O. S. Mortensen, and L. L. Andersen, "Importance of skills development on labour market participation among workers aged 35 or older—A systematic review of prospective cohort studies," BMC Public Health, vol. 25, Art. no. 3125, 2025.
[14] S. Staniczek, L. Naegele, and W. Schmitz, "Digital learning for old(er) workers: A systematic review of age-appropriate vocational education and training," Educational Gerontology, pp. 1–29, 2026.
[15] Eurostat, "Eurostat database," European Commission, 2026. [Online]. Available: https://ec.europa.eu/eurostat. Accessed: May 15, 2026.
[16] C. C. Ragin, Redesigning Social Inquiry: Fuzzy Sets and Beyond. Chicago, IL, USA: University of Chicago Press, 2008.
[17] I. O. Pappas and A. G. Woodside, "Fuzzy-set qualitative comparative analysis (fsQCA): Guidelines for research practice in information systems and marketing," International Journal of Information Management, vol. 58, Art. no. 102310, 2021.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Socio-Technical Perspective on Artificial Intelligence in Secondary Mathematics Education: A Systematic Review

Balqis Olawumi Banjo

(e-mail: balqis.banjo@wits.ac.za)

Mathematics Education Division, University of the Witwatersrand, Johannesburg, South Africa

Article DOI: https://doi.org/10.68286/samyrad2026.304

ABSTRACT

Artificial intelligence (AI) is increasingly integrated into mathematics education; however, research on its application in secondary mathematics education remains limited. This systematic review examines AI adoption in secondary mathematics from a socio-technical systems perspective, aiming to clarify opportunities and challenges and identify technologies that align with educational needs and practices. The review follows the PRISMA methodology and employs a hybrid deductive-inductive qualitative content analysis. The review included 32 studies predominantly from Asia. The findings indicate that approximately 47% of reviewed studies involved ChatGPT, with adaptive scaffolding and teacher mediation identified as primary mechanisms. AI provides personalized learning and additional benefits, although socio-technical challenges persist. This review offers insights into effective AI adoption in secondary mathematics education and delineates socio-technical implications for practice, policy, and future research.

KEYWORDS AI adoption constraints, Secondary mathematics education, Pedagogical mechanisms, Socio-technical systems, ChatGPT

How to cite: Banjo, B. O. (2026). Socio-technical perspective on artificial intelligence in secondary mathematics education: A systematic review. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.304

Introduction

Research has shown that artificial intelligence (AI) enables machines to perform tasks that require human intelligence, including reasoning, learning, problem-solving, and decision-making (Yanar & Ergene, 2025; Zawacki-Richter et al., 2019). AI encompasses methods such as machine learning, natural language processing, and neural networks. In contrast, digital technologies such as calculators, spreadsheets, geometry software, and whiteboards automate processes, provide content, or supply static resources but operate on pre-set rules and cannot learn or adapt (Zawacki-Richter et al., 2019). AI technologies, including intelligent tutoring systems (ITS), perform tasks that require human-like intelligence, such as learning from data, reasoning, prediction, and response customization (Zawacki-Richter et al., 2019).

Furthermore, the literature indicates that AI enhances mathematics teaching and learning by using technologies

that simulate human behaviors, including adaptive instruction, feedback provision, and personalized learning support (Yi et al., 2025). According to Yi et al. (2025), AI applications such as ITS, adaptive platforms, and generative tools improve achievement, motivation, and engagement in mathematics. AI also supports teachers in planning, resource development, and assessment, enabling greater focus on instruction and student support (Colen, 2025; Yanar & Ergene, 2025). Despite AI's potential to enhance mathematics teaching and learning, its adoption in secondary schools remains limited (Colen, 2025; Pardede et al., 2026). Findings suggest that this problem stems from insufficient professional development, inadequate institutional support, and infrastructure challenges (Lima et al., 2026; Marienko et al., 2026; Yu Tarasova & Doroshko, 2025). Yu Tarasova and Doroshko (2025) outline a four-part remedial foundation: (a) understanding essential AI concepts and skills, (b) implementing interactive and project-based teaching methods, (c) establishing guidelines for curriculum and resource development, and (d) providing recommendations for teacher professional development. The authors present structured, practical solutions and aim to make AI education in secondary schools more comprehensive, engaging, and effective.

Constraints to adopting AI in secondary schools extend beyond technical challenges. Accordingly, this review examines AI adoption in secondary mathematics education from a socio-technical systems (STS) perspective. The STS framework clarifies the relationship between technology and human or social factors, enabling a comprehensive assessment of AI's impact, adoption, and effectiveness in educational contexts (Li & Manzari, 2025; Zha et al., 2025). This review is guided by the following questions: (1) What types of AI technologies are used in secondary mathematics education, and through which pedagogical mechanisms are they implemented? (2) What socio-technical challenges hinder AI adoption in secondary mathematics education? Applying the STS perspective can help teachers and policymakers design, implement, and sustain AI adoption in secondary mathematics education by ensuring they consider both technical and social factors. This review is significant because it clarifies the opportunities and challenges of AI adoption in secondary schools and how to align AI technologies with educational needs and practices.

Socio-Technical Systems Perspective

STS emphasizes the interdependence of social and technical components within organizations (Ali et al., 2023; Imanghaliyeva et al., 2020). Effective AI integration requires considering environmental and institutional contexts alongside technological factors (Ali et al., 2024). STS advocates that stakeholders integrate these elements to achieve effective, ethical, and sustainable outcomes. UNESCO (United Nations Educational, Scientific and Cultural Organization, 2023) further provides guidelines for ethical AI use, prioritizing human agency, inclusion, equity, and protective safeguards. The framework focuses on ensuring that AI empowers both teachers and students, enables equitable access, particularly for marginalized groups, while upholding data privacy, transparency, and age-appropriate protections (Ali et al., 2024; UNESCO, 2023).

Methods or analytical approach

Research Design

This study followed the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Figure 1) (Page et al., 2021).

Search Strategy

This study searched Scopus, Web of Science, and ERIC in May 2026 to identify relevant studies on AI integration in secondary mathematics education. The search included keywords related to "artificial intelligence," "mathematics education," and "secondary school," including their alternative spellings and conceptually related terms combined with Boolean operators. Filters limited results to empirical, journal-source research articles published from 2022 to 2026, in the social sciences and arts and humanities, at the final publication stage, and written in English (Appendix).

Eligibility Criteria

The initial database search identified 75 articles (Scopus n=11, Web of Science n=25, ERIC n=39). The study excluded 14 records before screening: 10 were duplicates, and the other four were inaccessible (Figure 1). Thus, this study screened only 61 studies for eligibility. The author conducted the screening using the quality appraisal protocol (Table 1), and a senior researcher independently verified it. Disagreements were resolved by consensus. This study included studies that focused on AI technologies in secondary mathematics education and were published in peer-reviewed journals. Otherwise, they were excluded. The study excluded 29 studies. As a result, this study systematically reviewed 32 eligible studies.

Data Analysis

This review employed a hybrid deductive–inductive qualitative content analysis to quantify and describe content from the included studies (Saldaña, 2018). The study used CIMO (Context, Intervention, Mechanism, and Outcome) to organize the data (Denyer et al., 2008). This study organized the data into the following sub-categories: study ID, year, study title, study location, methodology, participants, AI name, mechanisms, findings on educational benefits, and findings on AI adoption constraints.

This study coded the data manually using a descriptive process in Microsoft Word. To enhance the trustworthiness of the findings, this study conducted peer debriefing by engaging two independent researchers to review the coding and themes. The study developed inductive codes based on the terminology, descriptions, and functions reported in the included studies. The study further refined the categories into themes. The AI technology themes followed Hammad (2025). The framework categorizes AI in education into four groups: personalized and adaptive learning (PAL), intelligent tutoring systems (ITS), generative AI (GenAI), and multimodal learning analytics (MMLA).

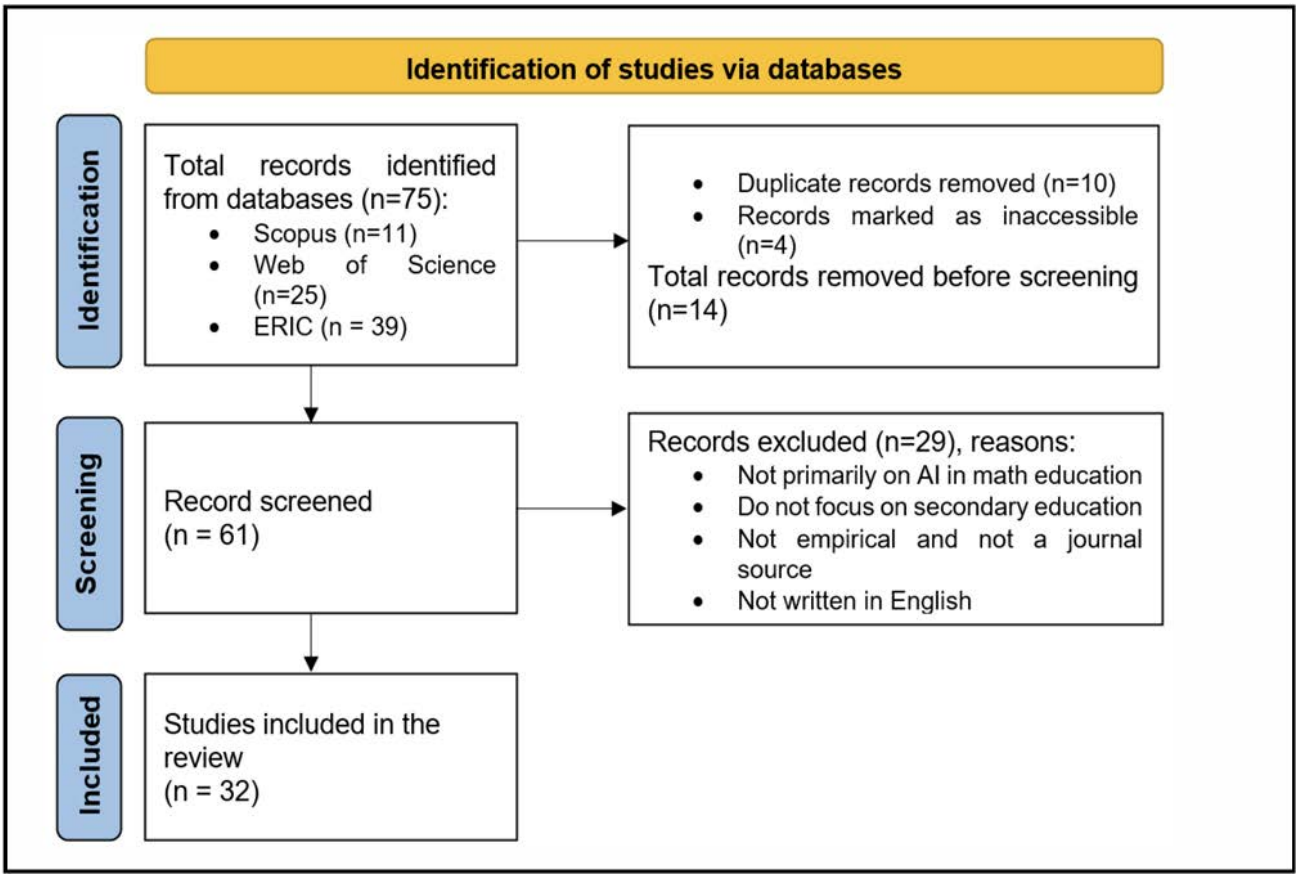


Fig. 1. PRISMA Diagram

Table 1. Quality Appraisal Protocol

Inclusion Criteria	Exclusion Criteria
Relevance to AI in mathematics education. Must address mathematics only.	Irrelevant to AI in mathematics education. Excluded even if they report the use of AI. Not relevant to math only if focused on STEM or STEAM, even if primarily about AI.
Relevance to secondary education and secondary education alongside other educational levels. Including Grades 5-8, Grades 9 12, Grades 6-9, Junior secondary, K-12, Lower secondary, Middle and High school mathematics, Middle school, Secondary school, Secondary mathematics teachers, and Senior Secondary School.	Irrelevant if the focus is not on secondary education; even if it focuses on teacher preparation or references secondary mathematics preservice.
Publication period: from January 2022 to July 2026	Irrelevant if outside of January 2022 and July 2026.
Subject areas: Social Sciences & Arts and Humanities	Other subject areas apart from Social Sciences & Arts and Humanities are irrelevant.
Source & Document type: Journal Article	Irrelevant if they are non -empirical or not a journal source.
Language: English	Irrelevant if written in any language other than English.

PAL encompasses AI-driven systems that customize educational content, pacing, and feedback according to each student's needs, preferences, and progress (Akçay & Altıner, 2025; Cho & Kim, 2025; Hammad, 2025). This approach departs from traditional one-size-fits-all models to improve engagement and learning outcomes. ITS are AI-powered platforms that model student knowledge and behavior to provide individualized instruction, guidance, and assessment (Akçay & Altıner, 2025; Hammad, 2025; Manto et al., 2026). These systems adapt teaching strategies and content based on real-time student data. GenAI, including large language models, produces new content such as text, images, or code in response to prompts (Hammad, 2025). This technology supports both students and teachers in tasks such as idea generation, feedback provision, and resource creation. MMLA utilizes advanced AI analytics to integrate and analyze data

from multiple sources, such as learning management systems and sensors (Hammad, 2025; Manto et al., 2026). MMLA aims to understand learning processes, monitor progress, and inform interventions across diverse educational contexts.

Results and findings

Characteristics of the Included Studies

The 32 studies cover Asia (Indonesia, Turkey, Philippines, Singapore, Nepal, China, Taiwan), Europe (Germany, Greece, Serbia, Portugal, Ukraine), Africa (Nigeria), North America (USA, Mexico), and South America (Argentina, Peru). Study 9 involved ten Southeast Asian countries, and Study 23 combined Nigeria, Ghana, and South Africa. Quantitative methods were the most common methodology used in the studies (n=19), followed by mixed methods (n=9), while qualitative methods were the least common (n=4).

Of the 32 studies, 24 focused on students only, five on teachers only, two studies involved both students and teachers, and one study involved no humans. Sample sizes ranged from 2 to 53298; qualitative (Study 13) and quantitative (Study 23) studies, respectively, both involved only students.

Table 2. The Studies' Years and Locations

Study ID	Year	Country
1	2025	Turkey
2	2024	Greece
3	2026	Greece
4	2026	Indonesia
5	2023	Nigeria
6	2026	Turkey
7	2025	Turkey
8	2026	Ukraine
9	2025	10 Southeast Asia: Indonesia, Thailand, Malaysia, Singapore, Philippines, Vietnam, Cambodia, Laos, Myanmar, and Brunei Darussalam
10	2025	USA
11	2024	Germany
12	2025	Germany
13	2024	Greece
14	2025	Serbia
15	2024	Indonesia
16	2025	Germany
17	2024	Nigeria
18	2025	Germany
19	2022	Indonesia
20	2025	Indonesia
21	2025	Singapore
22	2026	Portugal
23	2025	Nigeria, Ghana, and South Africa
24	2022	Mexico
25	2025	Argentina
26	2023	Philippines
27	2026	Nepal
28	2024	Philippines
29	2026	USA
30	2025	China
31	2025	Taiwan
32	2026	Peru

AI technologies involved in Included Studies

Some included studies evaluated AI's capabilities, while others assessed different aspects of AI; these distinctions informed the categorization. For example, both Studies 15 and 22 involved ChatGPT and GeoGebra, but the purpose of ChatGPT use differed. Study 15 utilizes ChatGPT to simplify GeoGebra tasks for students, whereas Study 22 investigated how ChatGPT supports learning. Consequently, these studies were categorized as GenAI and 'others', respectively (Table 3).

The review found that AI platforms such as Fractions Lab, Smart Learning Environments (SLE), Magic School, and Alberta AI also provided teachers with resources such as lesson planning, assignments, and instructional materials, while supporting students with explanations and tailored feedback. Their functionality aligned with ITS, while they often integrated GenAI. Similarly, this review identified some enabler models, including Photomath, Optical Character Recognition (OCR), and RoBERTa, as AI-based methods. These methods often use ChatGPT (Study 8 and Study 21) to help OCR systems handle complex tasks such as recognizing characters in various fonts, languages, and handwritten notes.

Some studies did not fit into Hammad's (2005) categories and were therefore categorized as 'others'. These include studies examining AI perceptions and acceptance, measuring AI's impact on mathematics outcomes, or exploring psychological and behavioral implications (Table 3). 'Others' was the dominant category with 12 counts (37.5%). The categories used in this study are not mutually exclusive. As a result, this review selected the category that best reflects each study's primary purpose.

Pedagogical Mechanisms of AI Adoption

The reviewed studies indicate that AI's impact in mathematics education depends on the pedagogical mechanisms through which these tools influence learning, teaching, feedback, assessment, and engagement. For example, Study 3 utilized DreamBox Learning, Fractions Lab, ChatGPT, and Mathia within a theory-driven intervention focused on fraction learning. Study 20 incorporated AI-driven scaffolding into realistic mathematics education and gamification. In both cases, AI served as an integral component of a purposefully designed instructional system rather than as an isolated tool. Conversely, when AI primarily generated solutions, as in Study 1 (chatbot-generated responses for e-textbook problems) and Study 14 (o1 and o3-mini solutions for competition problems), results indicated that AI-generated answers might be beneficial but still required critical evaluation and expert oversight. Key pedagogical approaches identified in the included studies include adaptive scaffolding, teacher mediation, dialogic reasoning, output evaluation, and representation mediation (Table 4).

Table 3. Categories of AI Technologies Adapted from Hammad (2025)

AI Category	Example	Study ID
ITS	ASSISTments, Smart Learning Environment (SLE), ALTER-Math, ADA C-GPT, Photomath, Multi-target classification model (LLaMA 3 + LoRA), and Cooperative conversation-based intelligent tutor system (CCITS)	2; 5; 10; 12; 24; 26; 29; 31 (n=8)
MMLA	ChatGPT on Google Sites & Google Forms, ChatGPT on MRIIA+ ecosystem & OCR, k nearest neighbors (KNN) & Random Forest & Support vector machine (SVM), Roberta powered by ChatGPT, and Chain of-thought (CoT) & GPT-4 on AMMORE Dataset	4; 8; 18; 21; 23 (n=5)
GenAI	ChatGPT, MathGPT, Gemini, OpenAI o1, and OpenAI o3-mini	1; 13; 14; 15 (n=4)
PAL	DreamBox Learning, Fractions Lab, ChatGPT, Mathia, Semantha® (AIEDN), and AI-based Gamification	3; 11; 20 (n=3)
Others	ChatGPT, AI in General, Machine-learning paradigm, ChatGPT-based Questionnaire (MTAPACQ), and Chatbots (ChatGPT, Gemini & Copilot)	6; 7; 9; 16; 17; 19; 22; 25; 27; 28; 30; 32 (n=12)

Table 4. Pedagogical Mechanisms Revealed in the Studies

Mechanism	Description
Adaptive scaffolding	Immediate feedback, hints, adaptive pathways, personalized worksheets, and scaffolding.
Teacher mediation	Select tools, design tasks, verify accuracy, customize resources, and address equity or infrastructure issues.
Dialogic reasoning	Encourage students to ask questions, explain their reasoning, justify their answers, or engage in mathematical dialogue.
Evaluation of outputs	Evaluate outputs, correct errors, refine prompts, debug, interpret feedback, or teach AI.
Representation mediation	Visual models, videos, GeoGebra, Photomath, decision trees, artifacts, and multimodal tools.

Adaptive scaffolding. Study 3 used DreamBox and Mathia to adjust the difficulty and feedback during fraction lessons, while ChatGPT provided questions and prompts that elicited explanations. Study 20 presented a clearer scaffolding model with tiered AI hints that guide students from action-level help to reasoning at the process and schema levels, without direct answers. Similarly, Study 31 employed a conversation-based ITS that offered hints, prompts, and cues during Pythagorean argumentation tasks, and Study 8 employed MRIIA+ with ChatGPT and OCR to analyze handwritten solutions, classify errors, and generate personalized feedback. These findings suggest that AI feedback is most effective when it focuses on students' reasoning processes rather than just confirming answers.

Teacher mediation. This was common across several studies. The mechanism showed that AI was more effective when teachers selected appropriate tools, designed tasks, monitored outputs, helped students interpret feedback, and aligned AI activities with curriculum goals. For example, Study 13 found that prompt quality influenced the usefulness of ChatGPT-created worksheets for students with special needs. Study 25 showed that teachers produced more specific, pedagogically sound fractal lesson plans than chatbots did. At the same time, Study 12 noted that the ADA chatbot still required teacher oversight because of potential hallucinations and confusing guidance. These findings demonstrate that successful AI integration depends heavily on teacher judgment.

Dialogic reasoning. This mechanism involves dialogic interaction. GenAI chatbots and teachable agents provided students with opportunities to ask questions, explain concepts, justify solutions, correct AI-generated errors, and engage in iterative reasoning. In Study 10, ALTER-Math positioned students as teachers of an AI-powered teachable agent, requiring them to explain mathematical concepts and correct the agent's errors. Study 29 adopted a similar learning-by-teaching approach, in which students identified errors in AI-generated solutions and justified their corrections. Study 22 demonstrated students using ChatGPT alongside GeoGebra to ask questions about quadratic functions and to compare textual explanations with graphical representations. These examples indicate that dialogue is most effective when students articulate and test their mathematical reasoning rather than simply requesting

answers. However, Study 15 found that students often struggled to get usable GeoGebra commands from ChatGPT because their prompts were vague. Study 12 showed that the ADA chatbot could facilitate Socratic questioning but also posed risks, such as confusing terminology, hallucinations, and increased cognitive load when used with paper-based materials. These findings suggest that AI dialogue does not replace mathematical expertise but instead shifts part of the learning challenge to questioning, interpreting, and validating machine-generated explanations.

Evaluation of outputs. Learning through critique emerged as a central mechanism across the reviewed studies. Multiple studies demonstrated that AI enhances pedagogical value when students actively inspect, challenge, revise, or improve AI-generated responses. For instance, Study 15 reported that students tested ChatGPT-generated GeoGebra commands, identified syntax or language errors, and refined prompts until the output was usable. Similarly, Study 4 described a culturally grounded geometry task in which students revised prompts when ChatGPT's representations of the Sanggah Cucuk did not align with the intended geometric structure. Studies 10 and 29 further developed this approach by using teachable agents that required students to teach, correct, and justify mathematical ideas to AI, shifting the focus from passively receiving solutions to exercising mathematical authority. These mechanisms are significant because they transform AI limitations, such as hallucinations, inconsistent terminology, incorrect

reasoning, or incomplete justifications, into learning opportunities when tasks are carefully designed. Study 14 recommended that advanced students critically analyze each step of AI-generated solutions, as not all were correct. Study 25 compares chatbot- and teacher-generated fractal solutions and finds that chatbots often omit intermediate reasoning steps or provide less detailed classroom proposals. Collectively, these findings indicate that the mechanism extends beyond AI providing explanations; students and teachers learn through questioning AI explanations, under conditions that promote verification and reflection.

Representation mediation. AI has influenced learning by expanding students' access to diverse mathematical representations. For example, Study 3 used Fractions Lab to enable visual manipulation of fraction models, helping students connect equivalence and operations to concrete visuals. Study 16 implemented CODAP with the Arbor plug-in to demonstrate decision-tree construction and misclassification rates, thereby making aspects of machine learning observable. Studies 15 and 22 integrated ChatGPT with GeoGebra, enabling students to combine textual explanations with dynamic geometric or graphical representations. Study 26 showed that Photomath supported pre-calculus learning through step-by-step solutions, while Study 11 linked students' questions to relevant videos using a semantic AI assistant. Collectively, these examples demonstrate that representational mediation can help students connect symbolic procedures with visual, procedural, or contextual understanding.

Nevertheless, several studies indicated that support based on representations was inconsistent. Study 7 reported that ChatGPT could generate numerous mathematics problems but struggled with visual or geometric representations, which could lead to misclassification or gaps in understanding. Similarly, Study 22 found that ChatGPT-3.5 could not produce graphs, so researchers used GeoGebra to represent quadratic functions visually. Study 1 omitted geometry and figure-based problems because certain AI models could not process images. These findings indicate that the most effective approach combines AI with representational tools and teacher expertise.

Several studies also found that GenAI chatbots and teachable agents enable students to explain concepts, justify solutions, correct AI-made mistakes, and reason iteratively. In Study 10, ALTER-Math had students teach an AI agent. Study 29 used a similar approach: students identified errors in AI solutions and justified their corrections. Similarly, Study 22 showed students using ChatGPT and GeoGebra to ask questions about quadratic functions and to compare textual explanations with graphical representations. The students actively inspected, challenged, revised, or improved AI-generated responses.

Reported Outcomes of AI Integration

The following sub-sections present the study's outcomes regarding AI educational benefits and challenges related to AI integration.

Educational benefits of AI adoption

The review categorized AI benefits in secondary mathematics, as reported in the included studies, into six categories (Table 5): Student agency and social learning, which include independent learning, autonomy, self-reliance, collaboration, and learning by teaching (Studies 1, 2, 3, 4, 6, 10, 13, 20, 21, 22, 26, 31); Personalized support and feedback, which include explanations, accessibility, personalization, scaffolding, feedback, error correction, and learning support (Studies 1, 2, 5, 6, 8, 10, 14, 17, 18, 22, 31); Motivational and affective outcomes, which include engagement, interest, motivation, and attitudes (Studies 3, 4, 5, 10, 12, 13, 20, 21, 26, 27); Higher-order mathematical skills, which include critical thinking, reasoning, problem solving, debugging, and argumentation (Studies 4, 15, 16, 20, 21, 22, 31); Cognitive learning outcomes, which include conceptual understanding, procedural fluency, achievement, knowledge transfer, and retention (Studies 3, 4, 10, 11, 13, 24, 26); and Teaching and assessment benefits, which include teaching efficiency, instructional support, and grading (Studies 1, 17, 23).

These results indicated that AI enhanced mathematical thinking and achievement, primarily in studies that embedded AI within structured instruction. A quasi-experimental study showed that students who received AI-supported mathematics instruction using DreamBox Learning, Fractions Lab, ChatGPT, and Mathia for fraction lessons had higher understanding, improved procedural skills, and greater engagement than those using traditional methods (Study 3). Study 20 found similar increases in understanding with the AI-RME-gamification model, which may contribute to the

positive overall learning outcomes observed. Some of the included studies also suggested that AI-driven learning is most effective when aligned with curriculum goals, scaffolding, and active reasoning. Several studies reported improvements in reasoning, problem-solving, critical and computational thinking, and argumentation. One example is Study 20, which showed that AI-supported RME enhanced problem-solving gamification, and tiered hints helped students persist without direct answers. Similarly, Study 31 found that students using an ITS (CCITS) outperformed their peers during the argumentation stages of the Pythagorean theorem. A qualitative study, Study 15, added interpretive depth, noting that ChatGPT and Gemini could handle demanding tasks and improve computational thinking through debugging AI commands.

Regarding AI-supported assessment, Study 8 showed that AI-assisted assessment using MRIIA+ and ChatGPT facilitated error correction, while also noting that traditional feedback was clearer and more engaging. GPT-4-based automated grading, particularly with chain-of-thought prompting, aligned more closely with human scoring than rule-based string-matching and text-processing methods (Study 23). In addition, some studies suggested that AI also boosted affective, engagement, and motivational outcomes (Studies 2, 3, 12, and 26). Notably, Study 2 reported that 94% of students (high) were interested in using ASSISTments; however, more than half of the students also reported that the platform caused them anxiety. This was like Study 27, which linked more time spent on digital resources to enjoyment but also to negative effects on achievement. Students with higher mathematics anxiety relied more on AI, but AI did not reduce anxiety (Study 30).

Some included studies also found that AI promotes equitable and inclusive education, especially through personalized learning (Studies 4, 9, and 13). For example, Study 13 found that personalized worksheets helped students with special needs gain concept mastery and engagement. Similarly, Study 4 used culturally grounded tasks to help students connect geometry to Balinese traditions. Culturally adapted AI technologies, local languages, and familiar examples are key to engagement and relevance (Study 9). In addition, the five teacher-only studies explored teachers' perceptions or knowledge of AI (Studies 9, 17, 25, 28 and 32). Teachers' responses indicated that AI supported teachers and teaching.

Challenges of integrating AI

Six themes emerged as constraints to AI adoption in secondary mathematics education (Table 6). The dominant constraint was AI's technical, output, and accuracy limitations. These include AI-generated reasoning errors, symbolic or notation errors, difficulty with visual representations, hallucinations, technical glitches, lag or system errors, and OCR problems. 15 studies across Asia, Europe, Africa, and America reported AI's technical, output, and accuracy limitations (Table 6).

Table 5. AI Benefits in Secondary Mathematics Education

Benefit	Main contributive AI	Total count
Student agency and social learning	ITS (Studies 2; 10; 26; 31)	12
Personalized support and feedback	ITS (Studies 2; 5; 10; 31)	11
Motivational and affective outcomes	ITS (Studies 5; 10; 12; 26)	10
Higher-order mathematical skills	MMLA (Studies 4; 21)	7
Cognitive learning ITS outcomes	(Studies 10; 24; 26)	7
Teaching and assessment benefits	GenAI & MMLA (Study 1 and Study 23 respectively)	3

Table 6. Constraints of AI Adoption in Secondary Mathematics Education

Constraints	Studies involved
Technical, output & accuracy limitations	1, 3, 4, 6, 7, 8, 10, 11, 14, 15, 17, 21, 22, 23, 25 (n=15)
Human oversight, competency & training	1, 3, 5, 6, 7, 9, 10, 15, 17, 19, 20, 27, 28, 32 (n=14)
Infrastructure, equity & access	1, 5, 11, 15, 17, 20, 22, 23, 27, 28, 31, 32 (n=12)
Student learning & psychological impacts	1, 2, 3, 6, 8, 10, 12, 17, 27, 28, 30 (n=11)
Diversity, culture & ethics	1, 2, 6, 8, 9, 11, 17, 19, 21, 22, 28 (n=11)
Pedagogy, instruction & integration barriers	1, 3, 8, 10, 11, 15, 17, 20, 28 (n=9)

Studies from Indonesia, Southeast Asia, Nigeria, Nepal, Peru, China, and the Philippines indicated that AI adoption needs language support, cultural relevance, infrastructure, teacher preparation, and students' digital literacy. Study 9 highlighted that culturally responsive AI technologies require local examples, multilingual options, and respect for regional norms. Study 4 cautioned that while cultural artifacts can improve mathematical understanding, AI without guidance might distort them. Study 5 and Study 17 noted that limited access to computers hindered AI use in Nigerian public secondary schools, and only 17% of Nigerian teachers were familiar with ChatGPT (Study 17). Similarly, Study 28 found that half of Metro Manila mathematics teachers were

unfamiliar with ChatGPT, and 82.7% cited technical training as a key need.

In addition, the findings indicated that GenAI is less reliable due to hallucinations. In Study 1, GPT-4o and GPT-4 performance accuracy declined as tasks became more complex. Study 14 confirmed this pattern, with OpenAI models excelling in Serbian mathematics contests but struggling with complex questions. Studies 7, 23, and 25 reported that ChatGPT reasoning is unstable: they sometimes accept correct corrections, defends false claims, or produces flawed proofs. Furthermore, AI representational mediation is limited for mathematical visuals, spatial, or handwritten artifacts. Study 1 excluded geometry problems with figures; Studies 7, 8, and 22 noted ChatGPT struggles with visual and geometric representations, especially with triangles. Study 8 indicated that ChatGPT assessments in MRIIA+ depend on OCR accuracy, scan quality, and recognition of mathematical notations and diagrams. Study 22 showed that ChatGPT-3.5 needs GeoGebra to generate dynamic graphs, as it cannot do so independently.

The review further revealed that the quality of AI output depends on users' AI knowledge. Students often failed because ChatGPT provided commands in an incorrect programming language or with syntax errors that GeoGebra could not interpret. In worksheet-generation and narrative-task studies, teachers had to modify ChatGPT outputs to suit students, curriculum levels, and classroom contexts. Meanwhile, some studies revealed that AI limitations often begin before output is generated. For example, Study 8 relied on OCR quality and the structured digitization of handwritten solutions. Study 21 showed that RoBERTa's dialogue-act classification required annotated tutoring data and synthetic examples from ChatGPT to handle rare codes.

ASSISTments, DreamBox, and adaptive scaffolding platforms often faced challenges including limited adaptability, limited emotional responsiveness, fixed interaction formats, and partial modeling of student thinking. Study 10 found that while ALTER-Math enhanced learning, it had uncleared prompts, limited math problem diversity, poor frustration management, and an occasional risk of hallucinations. Study 20 showed that the AI scaffolding tutor intentionally withholds direct answers to promote productive struggle, which can frustrate students who get stuck. Study 31 demonstrated that a cooperative conversation-based intelligent tutor system (CCITS) improved argumentation skills but limited students to responding rather than asking questions. Nevertheless, Studies 6, 17, 26, 27, 28, and 30 identify issues of overdependence, academic dishonesty, diminished independent reasoning, and less social interaction. These issues mostly involved students and related to how technologies like ChatGPT and Photomath provide instant answers and step-by-step solutions. Studies have found that AI hinders learning when students rely on it to do all the thinking. Study 30 showed that students with mathematics anxiety may turn to AI as a coping mechanism, yet frequent AI use does not necessarily lower anxiety.

Discussion

This study found that AI effectiveness in secondary mathematics education depends on socio-technical alignment.

ChatGPT: The Most Common AI Technology

Reviewed studies reported that ITS, GenAI, and PAL are some of the AI technologies adopted in secondary mathematics education. Some studies involved multiple AI technologies. The findings show that conversational AI, especially ChatGPT and similar large language models (LLMs), was widely used in secondary mathematics education. The literature supports this finding that ChatGPT is popular in secondary school mathematics for lesson planning, problem-solving, and activities (Yanar & Ergene, 2025). This may be because it can generate explanations, support reasoning, provide feedback, and create materials. Although ChatGPT was common among the included studies – 15 out of the 32 included studies involved ChatGPT – it did not drive the popularity of GenAI (Table 3). This shows ChatGPT's influence and contributions in other categories. For instance, RoBERTa serves as an enabler model in natural language processing (Moussa et al., 2022). OCR systems often use AI to improve text recognition accuracy and efficiency (Zhang et al., 2023). Among the included studies, RoBERTa and OCR were powered by ChatGPT, and their use aligned more with the MMLA. Other frequently used AI technologies in the included studies include DreamBox; AI-powered teaching agents such as ALTER-Math; automated assessment systems such as MRIIA+; and specialized apps such as Photomath and GeoGebra.

Interconnection Effects of AI's Capacity, Pedagogical Mechanisms, and Educational Benefits

Another key finding is that the educational benefits of AI technologies, AI's technical characteristics, and their pedagogical mechanisms were interconnected. Adaptive scaffolding was the dominant mechanism, implying PAL addresses challenges such as feedback timeliness and individualization for many students. Some included studies suggest that AI breaks tasks into steps, provides hints, and supports struggle without giving answers, aligning technology with teaching goals. However, this mechanism depends on technology quality, student engagement, and teacher integration. Song et al. (2025) confirm that dialogue and scaffolding play a vital role, especially in GenAI, which encourages inquiry-based learning and critical thinking. Existing studies highlight personalized feedback as the most effective mechanism for providing real-time, tailored feedback, thereby increasing student engagement and improving learning outcomes (Mohamed et al., 2022; Yi et al., 2025). This suggests AI is most effective when it acts as a pedagogical mediator, supporting active knowledge construction through questioning.

Challenges of AI Adoption

AI in secondary mathematics education faces technical, pedagogical, and contextual challenges. These issues reflect system-social tensions, and effective outcomes require improvements in the system's social and technical factors (Ali et al., 2024). This review found that five out of the six

categories of AI constraints were socio-environmentally related (Table 6).

Challenges related to the AI adoption environment

Teacher-related challenges were prominent, including inadequate training, limited AI literacy, increased workload, and uncertainty about effective pedagogical integration. The literature confirms that in-service and pre-service teachers often lack structured opportunities to develop AI literacy and pedagogical strategies for effective integration (Açıkyıldız & Şahin, 2025; Ruiz & Yambao, 2026; Shrestha & Yi, 2026). Professional development seems insufficient, as teachers often call for targeted training to improve their skills and confidence in using AI (Atuahene & Boateng, 2026). Teachers also reported that heavy workloads, large class sizes, and administrative duties limited their time to explore new AI tools (Atuahene & Boateng, 2026). This underscores the need for professional development that covers pedagogical, ethical, and curricular aspects of AI use.

Teachers also worry that students' overreliance on AI reduces critical thinking and that AI's incorrect outputs often cause misconceptions and shallow learning. Existing studies confirm these challenges (Pardede et al., 2026; Shrestha & Yi, 2026). Pavlova (2024) adds that overreliance on AI can lead students to submit work that does not reflect their own reasoning, sometimes resulting in incorrect answers and hindering genuine learning. Moreover, easy access to AI-generated solutions can discourage students from developing their own problem-solving and critical thinking skills (Shrestha & Yi, 2026). To use AI effectively, students and teachers alike need digital literacy, which many lack, particularly those with special needs or from under-resourced areas (Atuahene & Boateng, 2026). This aligns with the STS theory that human capabilities shape effectiveness.

The findings also identified infrastructure, internet, device shortages, language issues, and mismatched AI systems as factors influencing AI success. Studies emphasize that unreliable electricity, poor internet connectivity, and limited access to digital devices remain significant barriers, particularly in low-resource or rural areas (Atuahene & Boateng, 2026; Fakoyede et al., 2026; Lessing & Ogbonnaya, 2025). In Ghana, mathematics teachers reported that poor internet connectivity and limited access to devices significantly hindered classroom use of ChatGPT (Atuahene & Boateng, 2026). Similarly, for students with special educational needs (SEN), unequal access to technology and high costs of AI-powered systems were identified as critical barriers (Lessing & Ogbonnaya, 2025). The literature further suggests that AI systems and their training data may embed cultural biases or depict scenarios that do not align with local students' experiences, leading to disengagement (Niño et al., 2026; Walkington, 2025). AI technologies may not support all local languages, limiting access for students unfamiliar with the tool's primary language (Sofroniou et al., 2025). STS places these systems within policies, resources, and support structures.

Challenges caused by the features of the AI system

A key challenge of AI systems is accuracy and reliability issues: included studies reported hallucinations, errors, and inconsistent feedback, especially with complex problems or advanced reasoning. The literature consistently reports that while models like GPT-4 demonstrate high accuracy on many mathematics problems, significant limitations persist, especially with complex, open-ended, or context-dependent tasks (Weinhandl et al., 2026). Teachers and students alike have raised concerns about the dependability of AI solutions. In a study of Ghanaian mathematics teachers, many reported skepticism about ChatGPT, citing misleading or incorrect outputs (Atuahene & Boateng, 2026). Nevertheless, some studies have shown that AI tutoring and assessment tools provide reliable feedback for structured tasks, suggesting that reliability varies with task complexity, context, and human mediation.

Research emphasizes the need for transparent data governance and well-defined privacy protection. Additionally, cultural challenges such as algorithmic bias, lack of transparency, and the potential to deepen educational inequalities arise when AI systems are not thoughtfully developed and supervised (Niño et al., 2026; Walkington, 2025). AI adoption requires supportive leadership, coherent policies, professional development, and resources, yet these are often inadequate. Schools often lack AI guidance, causing inconsistency and uncertainty. STS opines that AI integration depends on organizational support, teacher expertise, student AI literacy, ethical governance, and infrastructure.

This review's findings indicate that although AI increases engagement, provides personalized feedback, supports both students and teachers, promotes reasoning, and improves access, it is also associated with anxiety, inaccurate feedback, the need for substantial teacher mediation, overreliance, and inequity. These contrasting outcomes suggest that AI's benefits in secondary mathematics are conditional rather than inherent. These complexities highlight the need for a socio-technical systems approach, where effective AI integration relies on aligning reliable technology, teacher guidance, equitable infrastructure, ethical safeguards, and purposeful mathematics instruction.

Implications

The findings indicate that teachers should regard AI as a pedagogical partner rather than a replacement, especially when integrated into activities that foster reasoning and reflection. Teachers play a critical role in mediating AI use by teaching students to verify AI outputs and understand these technologies' limitations, thereby incorporating AI literacy into mathematics education. The effectiveness of AI adoption depends on the readiness of both teachers and students, requiring activities to align with their existing knowledge and skills. Further evidence suggests that schools and administrators should approach AI as a socio-technical transformation that requires adequate infrastructure, training, pedagogical support, and preparedness. Ongoing professional development is essential because of limited familiarity with AI and concerns about reliability. Schools should provide continuous training in technical skills, pedagogy, AI literacy, ethics, curriculum, and assessment. A

supportive environment with reliable internet access, devices, and technical assistance is crucial to prevent disparities. Administrators should promote responsible AI use through clear guidelines on integrity, privacy, ethics, and oversight. AI integration in secondary mathematics education should be approached as a socio-technical reform, encompassing policies on teacher training, curriculum development, infrastructure, and equity.

Limitations

A key limitation of this review is the potential for selection bias arising from the search strategy. The databases were restricted to the social sciences and the arts and humanities, and the review did not include backward and forward citation tracking or grey literature searches. These factors may have resulted in the omission of relevant studies. In addition, the evidence is skewed toward Asian and European countries. Educational systems and cultural attitudes toward AI vary widely across nations; results from one context may not apply directly to others. Many included studies also examined specific versions of ChatGPT, Gemini, Photomath, or other AI technologies that are continually updated. As AI capabilities evolve rapidly, some reported features, limitations, and performance results may become outdated. Future research should explore topic-specific uses of AI in secondary mathematics education, as the reviewed studies address topics such as fractions, geometry, calculus, quadratic functions, and problem-solving. It remains unclear whether AI's pedagogical benefits and challenges differ by topic.

Conclusion

This systematic review demonstrates that AI provides substantial benefits for secondary mathematics education, particularly through adaptive scaffolding. However, effective AI adoption requires recognizing AI technologies as specialized pedagogical mediators whose success depends on aligning STS. The reviewed studies indicate that evaluating AI limitations should focus on the degree of alignment between AI mechanisms, mathematical tasks, instructional design, and sociocultural context, rather than simply determining whether AI "works" or "does not work." When this alignment is strong, AI can enhance feedback, student engagement, problem posing, assessment, and personalized practice. In contrast, poor alignment may lead to hallucinations, superficial reasoning, generic outputs, overreliance, inequity, or pedagogical mismatches.

Acknowledgment

Special thanks to Dr Lawan Abdulhamid and Dr Moses Adeleke Adeoye for participating in the peer debriefing exercise for this study.

References

[1] Açıkyıldız, G., & Şahin, A. (2025). Investigation of preservice secondary school mathematics teachers' views on artificial intelligence. Journal of Pedagogical Sociology and Psychology, 7(3), 78–90. https://doi.org/10.33902/jpsp.202536762

[2] Akçay, A. O., & Altıner, E. Ç. (2025). Sustaining Mathematics Education in the AI Era (pp. 263–288). https://doi.org/10.4018/9798-3373-1998-8.ch009

[3] Ali, M., Wood-Harper, T., & Wood, B. (2023). Understanding the technical and social paradoxes of learning management systems usage in higher education: A sociotechnical perspective. Systems Research and Behavioral Science, 41(1), 134–152. https://doi.org/10.1002/sres.2945

[4] Atuahene, E., & Boateng, F. O. (2026). Mathematics teachers' awareness, perceptions, and challenges in using ChatGPT. Discover Education. https://doi.org/10.1007/s44217-025-01083-4

[5] Cho, M. K., & Kim, S. (2025). Analyzing AI-based educational platforms for supporting personalized mathematics learning. International Electronic Journal of Mathematics Education, 20(4), em0847. https://doi.org/10.29333/iejme/16664

[6] Colen, J. (2025). Teacher Candidates' Perceptions of AI Integration in Secondary Education: A Comparative Study of STEM and Non STEM Majors. International Journal of Education in Mathematics, Science and Technology, 13(5), 1097–1112. https://doi.org/10.46328/ijemst.5701

[7] Denyer, D., Tranfield, D., & Van Aken, J. E. (2008). Developing design propositions through research synthesis. Organization Studies, 29(3), 393–413. https://doi.org/10.1177/0170840607088020

[8] Fakoyede, S. J., Babalola, F. E., & Akomolafe, O. D. (2026). Exploring the Influence of AI on the Professional Development of Science Teachers in STEM Education. Canadian Journal of Science, Mathematics and Technology Education, 26(1). https://doi.org/10.1007/s42330-026-00434-z

[9] Hammad, R. (2025). A Framework for Effective Utilisation of Artificial Intelligence in Education. World Journal of Science, Technology and Sustainable Development, 20(1–2), 139–152. https://doi.org/10.47556/J.WJSTSD.20.1-2.2025.9

[10] Imanghaliyeva, A. A., Thompson, P., Salmon, P., & Stanton, N. A. (2020). A Synthesis of Sociotechnical Principles for System Design. Advances in Intelligent Systems and Computing, 955, 665 676. https://doi.org/10.1007/978-3-030-20227-9_63

[11] Lessing, M.-J., & Ogbonnaya, U. I. (2025). ARTIFICIAL INTELLIGENCE IN MATHEMATICS EDUCATION OF STUDENTS WITH SPECIAL EDUCATIONAL NEEDS. Problems of Education in the 21st Century, 83(6), 786. https://doi.org/10.33225/pec/25.83.786

[12] Li, M., & Manzari, E. (2025). AI utilization in primary mathematics education: a case study from a southwestern Chinese city. Education and Information Technologies, 30(9), 11717–11750. https://doi.org/10.1007/s10639-025-13315-z

[13] Lima, A. J. S., Pahuyo, C. A. L., & Tastar, A. P. (2026). Evaluating Teachers' Readiness and Implementation Strategies of Artificial Intelligence in Science Education. 2026 15th International Conference on Educational and Information Technology (ICEIT), 1072–1077. https://doi.org/10.1109/ICEIT68991.2026.11521839

[14] Manto, A., Visitacion, M. A., Yee, A. M., Roble, J., Capuyan, S. N., Hermoso, F., Milano, M. L., Gonzales, R., Hallarte, D. K., & Gonzales, G. (2026). AI integration in mathematics education: A systematic review of implementation challenges and pedagogical implications (2017-2025). Social Sciences & Humanities Open, 14, 103147. https://doi.org/10.1016/j.ssaho.2026.103147

[15] Marienko, M. V., Markova, O. M., & Semerikov, S. O. (2026). AI literacy in secondary education: framework, assessment, and professional development in the Ukrainian context. Computers and Education: Artificial Intelligence, 10. https://doi.org/10.1016/j.caeai.2026.100605

[16] Mohamed, M. Z. bin, Hidayat, R., Suhaizi, N. N. binti, Sabri, N. binti M., Mahmud, M. K. H. bin, & Baharuddin, S. N. binti. (2022). Artificial intelligence in mathematics education: A systematic literature review. International Electronic Journal of Mathematics Education, 17(3), em0694. https://doi.org/10.29333/iejme/12132

[17] Moussa, A., Fournier, S., Mahmoudi, K., Espinasse, B., & Faiz, S. (2022). Mixing Static Word Embeddings and RoBERTa for Spatial Role Labeling. Procedia Computer Science, 207, 2944–2951. https://doi.org/10.1016/j.procs.2022.09.353

[18] Niño, J. R. G., Delgado, L. P. A., Chiappe, A., & Sáez-Delgado, F. (2026). Teachers' conceptions of AI for inclusive mathematics learning: a conceptual analysis. Frontiers in Education, 11. https://doi.org/10.3389/feduc.2026.1778339

[19] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for

reporting systematic reviews. Systematic Reviews, 10(1). https://doi.org/10.1186/s13643-02101626-4
[20] Pardede, S., Simamora, S., & Sinaga, D. (2026). ChatGPT in Secondary Education: A Multi-Theoretical Case Study from a Developing Country. Education as Change, 30. https://doi.org/10.25159/1947-9417/20308
[21] Pavlova, N. H. (2024). Flipped dialogic learning method with ChatGPT: A case study. International Electronic Journal of Mathematics Education, 19(1). https://doi.org/10.29333/iejme/14025
[22] Ruiz, C. C., & Yambao, F. M. P. (2026). AI Literacy and Professional Development Engagement as Predictors of Mathematics Teachers' Readiness for AI-Enhanced Instruction. International Journal of Learning, Teaching and Educational Research, 25(2), 655–675. https://doi.org/10.26803/ijlter.25.2.29
[23] Saldaña, J. (2018). Writing qualitatively: The selected works of Johnny Saldaña.
[24] Shrestha, S., & Yi, J. (2026). TPACK-based professional development for the AI era: fostering pre-service teachers' acceptance of generative AI in mathematics classrooms. Journal of Mathematics Teacher Education. https://doi.org/10.1007/s10857-026-09747-0
[25] Sofroniou, A., Patel, M. H., Premnath, B., & Wall, J. (2025). Advancing Conceptual Understanding: A Meta-Analysis on the Impact of Digital Technologies in Higher Education Mathematics. In Education Sciences (Vol. 15, Number 11). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/educsci15111544
[26] Song, X., Mak, J., & Chen, H. (2025). Teachers and Learners' Perceptions about Implementation of AI Tools in Elementary Mathematics Classes. SAGE Open, 15(2). https://doi.org/10.1177/21582440251334545
[27] United Nations Educational Scientific and Cultural Organization. (2023). UNESCO Guidance for AI in Education and Research.
[28] Walkington, C. (2025). The implications of generative artificial intelligence for mathematics education. School Science and Mathematics. https://doi.org/10.1111/ssm.18356
[29] Weinhandl, R., Andic, B., Wijaya, T. T., Bleckenwegner, V., Riegler, V., Baldinger, S., Mayrhofer, J., & Helm, C. (2026). Beyond calculators: Decoding students' usage behaviour regarding AI chatbots for mathematics using UTAUT. Journal of Open Innovation: Technology, Market, and Complexity, 12(2). https://doi.org/10.1016/j.joitmc.2026.100773
[30] Yanar, A. N., & Ergene, Ö. (2025). Integrating artificial intelligence in education: How pre-service mathematics teachers use ChatGPT for 5E lesson plan design. Journal of Pedagogical Research, 9(2), 158–176. https://doi.org/10.33902/JPR.202533163
[31] Yi, L., Liu, D., Jiang, T., & Xian, Y. (2025). The Effectiveness of AI on K-12 Students' Mathematics Learning: A Systematic Review and Meta-Analysis. International Journal of Science and Mathematics Education, 23(4), 1105–1126. https://doi.org/10.1007/s10763-024-10499-7
[32] Yu Tarasova, O., & Doroshko, V. S. (2025). Methodological foundations of teaching the basics of artificial intelligence to lyceum students. 7th InternationalWorkshop on Augmented Reality in Education, 240–249. https://kdpu.edu.ua/personal/oyutarasova.html
[33] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? In International Journal of Educational Technology in Higher Education (Vol. 16, Number 1). Springer Netherlands. https://doi.org/10.1186/s41239-019-0171-0
[34] Zha, H., Li, W., Wang, W., & Xiao, J. (2025). The Paradox of AI Empowerment in Primary School Physical Education: Why Technology May Hinder, Not Help, Teaching Efficiency. Behavioral Sciences, 15(2). https://doi.org/10.3390/bs15020240
[35] Zhang, X., Bai, W., & Cui, H. (2023). Review of Optical Character Recognition for Power System Image Based on Artificial Intelligence Algorithm. In Energy Engineering: Journal of the Association of Energy Engineering (Vol. 120, Number 3, pp. 665-679). Tech Science Press. https://doi.org/10.32604/ee.2023.020342

Appendix

Databases used for the study and their respective query codes.

Database	String/Link	Filters
Scopus	TITLE-ABS-KEY ((artificial intelligence OR AI OR ChatGPT OR generative AI) AND (mathematics education OR mathematics teaching OR mathematics learning) AND (secondary school OR high school)) AND PUBYEAR > 2021 AND PUBYEAR < 2027 AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "ARTS")) AND (LIMIT-TO (LANGUAGE , "English"))	AI, math education, secondary school, 2022-2026, article, social sciences & humanities, English
Web of Science	https://www.webofscience. com/wos/woscc/summary/ b44924de-2382-4f8d-a15a-a987f4f11f94-01b6454469/relevance/1	AI, math education, secondary school, 2022-2026, article, social sciences & humanities, English
ERIC	https://eric.ed.gov/?q=(%2 2artificial+intelligence%22 +OR+AI+OR+ChatGPT+ OR+%22generative+AI%2 2)+AND+(%22mathematics+education%22+OR+%22mathematics+teaching%2 2)&pr=on&ff1=dtySince_ 2022&ff2=eduSecondary+ Education&ff3=pubJourna l+Articles	AI, math education, secondary school, 2022-2026, article, social sciences & humanities, English

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Artificial Intelligence and the Future of Education in Croatia: Student Professor Co-Creation in a Structured Narrative Review

Tamara Drezner

(e-mail: tdrezner@gmail.com)

Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

Article DOI: https://doi.org/10.68286/samyrad2026.307

ABSTRACT

This structured narrative review examines student–professor co-creation in AI-supported higher education in Croatia. It distinguishes routine AI use from shared educational decision-making and synthesizes Croatian higher-education evidence with international research, using K–12 evidence only as adjacent context. The review organizes the literature around five readiness domains: stakeholder participation, institutional capacity, AI and digital competencies, responsible technical conditions, and external relevance and sustainability. Across the evidence reviewed, AI use and preparation are more visible than sustained student-professor co-creation. The paper therefore proposes an unvalidated AI Co-Creation Readiness Framework and a mixed-method research agenda for future Croatian evaluation. As a selective narrative synthesis, it does not establish causal effects, national prevalence, comparative effectiveness, or financial return.

KEYWORDS artificial intelligence, student–professor co-creation, Croatia, higher education, responsible AI

How to cite: Drezner, T. (2026). Artificial Intelligence and the Future of Education in Croatia: Student Professor Co-Creation in a Structured Narrative Review. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.307

Introduction

Artificial intelligence (AI) now features across higher education teaching, individualized learning activities, and administrative processes [1]. Rapid growth in generative and conversational AI has intensified debate about pedagogy, assessment, and institutional practice [2]. Evidence from Romania and Serbia demonstrates both the opportunities and challenges associated with implementation [3]. These developments underscore the significance of educational relationships, ethical responsibilities, and institutional capacity, which are now as critical as technical access [4], [5]. This review examines the ways in which students and professors collaboratively influence AI-supported learning. In this context, AI adoption denotes the use of artificial intelligence in educational activities. Student–professor co creation is defined as the active involvement of both groups in designing, evaluating, or revising learning tasks, assessments, feedback, or guidelines for AI use. This distinction is significant, as frequent AI use alone does not indicate shared educational decision-making. The Croatian context is analyzed in conjunction with international evidence to identify design principles and outstanding questions.

Human-centered approaches maintain human judgment and responsibility, even when artificial intelligence is involved in collaborative activities [6]. Studies on AI assisted academic writing and teacher–AI collaboration emphasize the importance of coordination, shared control, transparency, and student agency [7]–[9].

These forms of human–AI interaction are relevant to educational design, but they should not be treated as evidence that students and professors co-create the rules, tasks, or assessments through which AI is used.

The review asks one central question: What conditions and evidence gaps shape student–professor co-creation in AI-supported higher education in Croatia? The analysis therefore treats AI use and AI co-creation as distinct constructs and examines the conditions under which students and professors can jointly design, evaluate, and revise AI-supported learning. Existing Croatian research primarily documents AI use, attitudes, literacy, and implementation, while evidence examining students and professors as joint decision-makers in the design of AI supported learning remains limited. This review addresses that gap by distinguishing AI adoption from educational co creation, synthesizing the Croatian studies identified in this review alongside relevant international research, and proposing a context-specific readiness framework for future empirical testing. The framework is intended not as a validated measurement model, but as a structured basis for future empirical examination of the conditions under which student–professor co-creation can occur in AI-supported higher education.

Conceptual Background and Evidence

AI Integration and Collaborative Learning

Existing reviews of artificial intelligence in education identify applications that include instructional support, adaptive learning, and administrative functions [10]. Comprehensive syntheses emphasize the need for enhanced educational foundations, increased ethical consideration, interdisciplinary collaboration, and greater methodological rigor [11], [12].

For the present review, the key implication is that access to an AI tool neither establishes educational value nor determines who participates in decisions about its use.

Co-creation research in higher education places relationships, participation, and shared design at the center of educational practice [13]. Research on AI and value co creation extends this perspective to interactions among students, educators, institutions, and technology [14]. Related work distinguishes using AI outputs as finished products from involving learners actively in creating, examining, and evaluating AI-supported content [15], [16]. Together, these perspectives provide a conceptual basis for separating AI use from shared educational design.

Teacher-focused studies examine learning design and collaboration with AI [17], [18], active learning with generative systems [19], and AI-supported feedback and collaborative work [20], [21]. These studies identify processes relevant to co-creation, but teacher–AI and student–AI interaction remain analytically distinct from student–professor co-creation.

Stakeholders and Institutional Conditions

Student-focused studies emphasize usefulness for learning, feedback, and efficiency while also reporting concerns about accuracy, dependence, and academic integrity [22]–[24]. Educators similarly identify potential benefits alongside needs for professional knowledge and institutional support [25], [26]. Research comparing teacher and student perspectives further highlights questions of support, replacement, and digital literacy [27]. These differences make shared expectations and role negotiation central to co-creation research.

Institutional integration also depends on ethical, social, and organizational conditions [28]. Studies of academics identify implementation opportunities and constraints [29], [30], while institutionally mediated AI initiatives illustrate how organizational structures can support engagement [31]. For co-creation, the relevant issues are whether students and professors have sufficient time, skills, guidance, authority, and accountability to make shared educational decisions.

School-level evidence is used only as adjacent context. Studies involving parents, teachers, or pupils can inform questions of trust, skills, and responsible use, but their findings cannot be transferred directly to university co creation because adult learner autonomy, assessment responsibilities, and governance arrangements differ [32].

Workforce Preparation and Sustainability

AI-supported entrepreneurship education and studies of student perceptions provide context for how learners connect AI with practical and professional development [33], [34]–[36]. These indicators assist in identifying expectations regarding future employment trends; however, direct evidence from employers is necessary before such expectations can be recognized as established labor-market requirements.

Economic sustainability is significant only to the extent that it influences the feasibility of co-created practices. Potential efficiency gains must be distinguished from demonstrated financial returns, which require evidence of attributable benefits and consideration of all implementation costs, including training, maintenance, staff time, and technical support. Educational value should be evaluated directly based on learning quality, learner agency, accessibility, and responsible participation [5], [29], [37].

Methods or analytical approach

Review Approach and Scope

This paper uses a structured narrative review to organize Croatian and international literature on student–professor co-creation in AI-supported higher education. Croatian higher-education studies are treated as direct local context, international higher-education studies as comparative evidence, and K–12 studies as adjacent evidence requiring adaptation. The objective is interpretive synthesis and framework development rather than exhaustive coverage, a protocol-based scoping review, or a pooled estimate of effects.

Sources were retained when they addressed stakeholder expectations, Croatian AI use and preparation, collaborative educational design, responsible technical conditions, employer or economic evaluation, or relevant conceptual and measurement boundaries. Literature was identified through targeted searches of Scopus, Web of Science, and IEEE Xplore, supplemented by searches of publisher platforms

including SpringerLink, ScienceDirect, MDPI, Taylor & Francis, Wiley, Frontiers, SAGE, and Nature Portfolio. Search terms combined concepts related to artificial intelligence, higher education, co-creation, student–staff partnership, AI literacy, responsible AI, and Croatia. Priority was given to peer-reviewed higher education research, Croatian empirical evidence, and recent international studies relevant to the conceptual boundaries of student–professor co-creation. K–12 studies were included only where they informed implementation, readiness, or responsible-use considerations not yet represented in the Croatian higher-education evidence identified. Because the review was designed as a structured narrative synthesis rather than a systematic review, no claim of exhaustive retrieval is made. Searches have been conducted during the last six months.

Synthesis and Interpretation

The synthesis distinguishes educational setting, stakeholder group, and form of collaboration. School-level studies are treated as adjacent evidence rather than direct tests of university co-creation. Student attitudes are separated from observed collaboration and learning outcomes, while perceived employability is separated from direct employer evidence. Proposed evaluation criteria are reported as future research recommendations rather than as findings from the reviewed studies.

Consequently, identified gaps refer to the literature synthesized in this review rather than demonstrating that no other relevant studies exist. The review therefore cannot establish causal effects, national prevalence, or comparative effectiveness.

For auditable interpretation, the synthesis distinguishes direct Croatian higher-education evidence, international higher-education evidence, K–12 evidence, and conceptual or review evidence. Each category supports a different level of inference: Croatian studies inform their reported settings, international studies support comparison and design principles, and school-level studies require adaptation to university autonomy, assessment, and governance.

Each substantive claim is interpreted in relation to the cited study's population, setting, design, and outcome. Table I makes these evidentiary roles explicit for sources central to Croatian contextualization and international comparison. It is a narrative evidence map rather than a quality-ranked inventory. Independent duplicate screening and formal risk of-bias assessment were not used, increasing the possibility of selection and interpretation bias.

Results and findings

The synthesis identifies five connected areas relevant to student–professor co-creation: stakeholder perspectives, Croatian AI use and preparation, evidence on collaborative design, employer and economic evaluation, and responsible technical conditions. Across these areas, the recurring distinction is between evidence of AI use or attitudes and evidence of shared educational decision-making. Proposed design and evaluation criteria are therefore reported separately from findings in the cited studies.

Stakeholder Perspectives on AI-Supported Education

Student-focused studies report interest in learning assistance and feedback together with requests for reliable information and guidance [22], [38]. Research across disciplines shows that attitudes and perceptions are important dimensions of AI integration [23], [39], and studies of education and careers indicate why student expectations should inform institutional discussion [33], [35]. However, favorable attitudes or frequent tool use do not establish participation in the design of AI-supported teaching or assessment.

Educator-focused research identifies potential teaching benefits alongside concerns about dependence, integrity, and professional preparation [25]–[27]. Student views about possible teacher replacement and AI–teacher collaboration further highlight questions of educator roles, control, and interpretation of learning activity [8], [40]. Taken together, the evidence suggests that co-creation requires explicit negotiation of roles and responsibilities rather than an assumption of shared expectations.

Multi-stakeholder research shows that AI acceptability depends on context and use case [32]. The implication for co-creation is methodological as well as practical: students' and professors' perspectives should be examined together, and enthusiasm from either group should not substitute for evidence about the fairness, educational quality, or acceptability of a specific collaborative practice.

Croatian Evidence on AI Use and Preparation

A Croatian institutional study by Vrana [41], conducted at FHSS Zagreb, provides an example of student AI use alongside limited formal preparation. Its importance lies in the potential gap between tool access and organized preparation; it does not estimate national prevalence or test student–professor co-creation. Rožman et al. [42] surveyed 471 business students in Slovenia, Croatia, and India and reported differences in AI knowledge and applications. These findings identify curricular questions, including ethical and legal understanding, but do not establish national educational effectiveness or the effect of co-design.

Spanish studies address a different construct. Research on disciplinary attitudes [39] and validation of the four-item Artificial Intelligence Attitude Scale (AIAS-4) in 650 university students [43] inform measurement of AI attitudes, not co-creation. The Croatian institutional study therefore informs local preparation needs, the Slovenia–Croatia–India study informs AI-literacy dimensions, and the Spanish studies inform instrument adaptation. Their different constructs and samples preclude a common country ranking; valid cross-context comparison would require aligned constructs, sampling frames, and measurement checks.

Croatian school-level research addresses teacher adoption, a three-year K–12 pilot, personalized learning and creative integration of AI in teaching [44], [45], [48], [49]. These studies identify issues of skills, task design, and implementation support, but do not establish how university students and professors jointly design curricula, assessment, or governance.

Table 1. Evidence Roles and Comparative Limits

Evidence and setting	Role in this review	Inference boundary
Vrana [41] FHSS Zagreb	Institutional AI use and preparation	One institutional context; no national prevalence estimate
Rožman et al. [42] Slovenia, Croatia, India	AI literacy among business students	Sample and construct differences preclude national effectiveness ranking
Almaraz-López et al. [39]; Luque-Reca et al. [43] Spain	Disciplinary attitudes and attitude measurement	Attitudes are distinct from co-creation and learning effects
Brodarič and Babić [44]; Lišnić et al. [45] Croatian schools	Teacher adoption and K–12 implementation	Requires adaptation to university autonomy and assessment
Martin et al. [46]; Williams and Ingle [47] International higher education	Participatory assessment and student–staff partnership	Examples for local design, not proof of Croatian effectiveness

Research on young entrepreneurs in Croatia provides context for the environment graduates may enter, but is not direct evidence about university teaching or employer requirements [50]. Across the Croatian sources reviewed, AI use, literacy, and implementation are more visible than sustained student–professor co-creation. The principal local evidence gap is therefore not whether AI is being used, but how shared educational design operates and with what outcomes.

Evidence on Co-Creation and Its Limits

International research provides examples of participatory assessment design, student-generated content, and student staff partnership [46], [47], [51]. Work on human–AI co creation and postdigital co-design broadens the conceptual discussion [52], [53]. These studies show that participation can be examined through shared tasks and design decisions rather than attitudes alone, but they do not establish a single model that can be assumed to transfer across Croatian universities.

Related studies examine AI-supported personal and collaborative learning environments [54], coordination in academic writing [7], comparisons of human and AI support in teacher education [55], and human–AI co-creation in programming education [56]. Because populations, interventions, and outcomes differ, these studies should inform design choices rather than be combined into a general causal claim about student–professor co-creation.

The literature is strongest in identifying design questions and examples; transfer across institutions remains uncertain. Reviews of co-creation and educational AI emphasize context, conceptual clarity, and rigorous evaluation [11] [13]. For Croatia, the unresolved question is how shared educational design operates over time, for which participants, and with what consequences for learning, agency, and responsibility.

Employer Perspectives and Economic Evaluation

Student perceptions of professional development and employment illuminate learner expectations [33], [35], but are not equivalent to direct employer assessments of graduate competence. Similarly, broader discussions of economic transformation and entrepreneurship education justify examining workforce relevance without establishing which AI competencies employers require in particular Croatian sectors [5], [34].

The reviewed literature also provides insufficient evidence to estimate the financial return of student professor co-creation in Croatian universities. A credible economic evaluation would require attributable benefits, full implementation costs, and an appropriate time horizon, while pedagogical outcomes should remain visible in their own terms rather than being converted automatically into monetary savings [5], [29], [37].

The evidence need is therefore specific: future research should connect observed shared educational design with learning outcomes, institutional conditions, direct employer input, and longitudinal cost data while keeping these evidence types analytically distinct.

Responsible Technical Conditions for Co-Creation

Educational AI systems can shape shared decisions through the data they use, the outputs they generate, and the mechanisms available for correction. Generative systems used for feedback raise questions about unsupported statements and prompt-sensitive responses, while predictive or adaptive systems raise questions about recommendations, differential error, and the data used to allocate support [10], [11]. These system classes require different evaluation approaches. The present review does not assess architectures or compare technical performance. Responsible-AI literature nevertheless supports attention to fairness, interpretability, privacy, and human oversight [4], [32], [57].

A future co-creation pilot could therefore include a task specific evaluation in which students and professors agree on criteria for acceptable AI feedback and document substantive errors, omissions, corrections, and appeals. Where justified and privacy-preserving, subgroup patterns and accessibility could also be examined. Data minimization, approved access, retention rules, and records of human review should accompany the task. These are proposed evaluation criteria, not technical findings of this review.

Discussion

Evaluating Shared Educational Design

Future Croatian research should directly examine how students and professors share decisions about AI-supported learning tasks, assessment criteria, feedback, and rules for responsible use. Project-based learning, reflective dialogue with AI, and participatory curriculum or assessment teams provide plausible settings for evaluation [46], [47], [58], but should be treated as designs to test rather than established solutions.

A feasible preliminary study would follow one semester across at least two comparable courses, preferably at more than one Croatian institution. Within a common assignment type, students and professors would jointly define assessment criteria, permissible AI assistance, disclosure, verification, and appeal procedures. A usual-practice comparator with the same learning objectives and access to the selected AI tool would help distinguish participatory design from tool access alone. Random allocation should be used only where sufficient independent course sections make it methodologically defensible; otherwise, the study should be framed as an exploratory matched-course pilot.

Baseline and post-task measures could be combined with interviews, design artefacts, observations, and workload logs. Primary pilot outcomes should emphasize feasibility, recruitment, retention, participation fidelity, and usable data; exploratory educational outcomes could include AI literacy, student agency, and task quality assessed with a common rubric. AIAS-4 may capture general attitudes [59], but should not substitute for measures of participation or learning. Any Croatian adaptation would require appropriate linguistic and measurement checks.

The study would also require institutional ethics approval, informed consent, voluntary research participation without grade penalties, and a suitable non-research learning option. With a small number of courses, analysis should emphasize feasibility, descriptive change, uncertainty, and qualitative explanation rather than a definitive causal estimate. The selected AI system, task conditions, errors, correction requests, and privacy incidents should be documented. These procedures describe future research and have not been implemented in this review.

A mixed-method design would connect reported attitudes with observed participation and assessment products. Prior research on shared control and student concerns provides dimensions such as transparency and responsibility [8], [9], while work on AI-assisted content creation illustrates the value of examining both process and product [51]. Agreement in self-reports alone would not establish the quality of a co-creation partnership.

Fig. 1 summarizes the AI Co-Creation Readiness Framework. AI access and use provide the context, but co creation requires readiness across stakeholder participation, institutional capacity, AI and digital competencies, responsible technical conditions, and external relevance and sustainability. These domains support shared educational decisions, which can then be evaluated through participation, learning, agency, responsibility, implementation burden, and costs. The framework is a conceptual synthesis for future testing rather than a validated causal model.

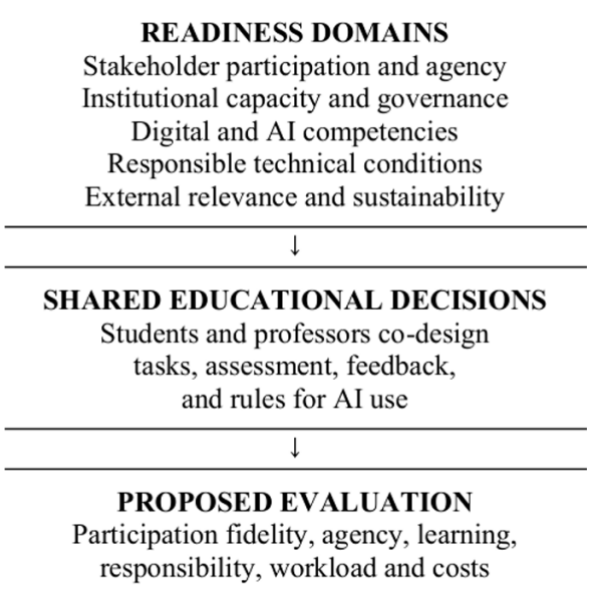


Fig. 1. Proposed AI Co-Creation Readiness Framework linking readiness domains, shared educational decisions, and evaluation outcomes.

Within the framework, readiness refers to conditions that permit responsible co-creation. Stakeholder participation concerns whose proposals shape tasks and assessment; institutional capacity concerns policy clarity, time, support, and accountability; digital and AI competencies concern critical appraisal of system capabilities and limitations; responsible technical conditions concern error handling, fairness, privacy, explanation, and human review; and external relevance and sustainability concern curriculum relevance, employer input, and implementation burden. Candidate indicators include decision logs, documented support arrangements, task-based literacy measures, error and-appeal records, and cost logs. These indicators are proposed operationalizations, not a composite readiness score, and would require content review and psychometric testing before any scale development.

Measuring Attitudes and Learning Processes

Validated instruments can support the measurement of clearly defined constructs. The four-item Artificial Intelligence Attitude Scale (AIAS-4) measures general attitudes toward AI [59], with validation studies in Norwegian and Spanish university populations providing methodological examples [43], [60]. Additional research has examined AI attitudes in other student samples [61]. An attitude scale, however, should not be treated as a direct measure of co-creation, institutional readiness, or learning effectiveness.

For Croatian research, translated instruments should be examined for linguistic suitability and measurement quality in the target population. Attitude measures should be combined with evidence of shared decision-making, task participation, and evaluation of AI outputs. Longitudinal follow-up could help distinguish initial novelty from sustained practice. Comparison groups or interventions may strengthen inference, but causal conclusions would still depend on design quality and its assumptions.

Institutional Capacity and Responsible Participation

Educator development should address educational design and responsible use as well as technical skills. Studies of co creation among innovation champions and open-schooling initiatives offer adjacent examples of participatory competence development [62], [63]. Their transfer to higher education requires adaptation to adult learner autonomy, disciplinary standards, assessed authorship, university grading, data access, accountability, and student representation. Principles may transfer; effectiveness requires separate university-level evaluation.

Institutions should also create structured opportunities for students and professors to discuss acceptable AI assistance, authorship, output verification, and sensitive information. Research on responsible-AI education informs these priorities [4], [57], while discipline-specific evidence, including medical education, reinforces the importance of context [64]. Croatian studies should therefore compare institutional types and resource conditions rather than generalize from a single faculty.

Employer Input and Economic Sustainability

Future research should collect direct employer perspectives through interviews, surveys, or participation in curriculum discussions. Comparing these data with student and professor expectations would identify areas of agreement and disagreement about graduate AI competencies while preventing assumed market needs from substituting for educational judgment.

Economic evaluation should compare co-designed practice with usual practice over a defined period and report implementation costs alongside educational outcomes. Costs may include software, training, maintenance, technical support, and staff time, with student time reported separately. A cost-consequence approach can preserve outcomes such as learning quality, agency, accessibility, workload, and responsible participation rather than forcing them into a single monetary measure. Financial ROI should be estimated only when benefits can be credibly attributed and monetized. The present review estimates neither pedagogical effects nor financial returns.

Limitations

This structured narrative review is selective and does not establish exhaustive coverage. Croatian higher-education evidence is geographically and institutionally limited and cannot establish national representativeness. International comparisons use heterogeneous populations, measures, designs, and policy settings; they support contextual interpretation rather than causal country ranking. K–12 and discipline-specific evidence is adjacent and requires adaptation before application to Croatian universities. No primary participant data, technical algorithm experiment, formal meta-analysis, evidence-certainty grading, or pedagogical or financial ROI estimate is reported. The proposed AI Co-Creation Readiness Framework is conceptual and unvalidated; effectiveness, generalizability, technical performance, and economic value remain questions for empirical testing.

Conclusion

This review distinguishes AI adoption from student professor co-creation in higher education. Across the Croatian evidence identified, research has focused primarily on AI use, literacy, attitudes, and implementation rather than sustained shared educational decision-making. The proposed AI Co-Creation Readiness Framework integrates five conditions for future empirical investigation: stakeholder participation and agency, institutional capacity and governance, digital and AI competencies, responsible technical conditions, and external relevance and sustainability.

Future Croatian research should test these domains through mixed-method co-design studies that directly observe how students and professors jointly shape AI supported learning. The central research challenge is therefore no longer simply whether AI is used in higher education, but how its use can be jointly designed, governed, and evaluated in an educationally meaningful and responsible manner.

Acknowledgment

The author thanks the reviewers for their constructive comments and suggestions, which helped improve the manuscript.

Declarations

Funding: This work was not supported by any university or organization.

Conflicts of interest: The author declares no conflicts of interest.

References

[1] V. Kuleto et al., “Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions,” Sustainability, vol. 13, no. 18, Art. no. 10424, 2021, doi: 10.3390/su131810424.
[2] T. K. F. Chiu, “Future research recommendations for transforming higher education with generative AI,” Comput. Educ. Artif. Intell., vol. 6, Art. no. 100197, 2024, doi: 10.1016/j.caeai.2023.100197.
[3] R. Bucea-Manea-Ţoniş et al., “Artificial intelligence potential in higher education institutions enhanced learning environment in Romania and Serbia,” Sustainability, vol. 14, no. 10, Art. no. 5842, 2022, doi: 10.3390/su14105842.
[4] E. Đerić, D. Frank, and D. Vuković, “Exploring the ethical implications of using generative AI tools in higher education,” Informatics, vol. 12, no. 2, Art. no. 36, 2025, doi: 10.3390/informatics12020036.
[5] Y. Wu and X. Gou, “The impact of AI-driven educational integration on high-quality economic development: An analytical study,” Tehnički vjesnik, vol. 32, no. 6, pp. 2098–2110, 2025, doi: 10.17559/TV 20250220002404.
[6] R. Katsenou, K. Kotsidis, A. Papadopoulou, P. Anastasiadis, and I. Deliyannis, “Beyond assistance: embracing AI as a collaborative co agent in education,” Educ. Sci., vol. 15, no. 8, Art. no. 1006, 2025, doi: 10.3390/educsci15081006.
[7] A. Nguyen, Y. Hong, B. Dang, and X. Huang, “Human-AI collaboration patterns in AI-assisted academic writing,” Stud. High. Educ., vol. 49, no. 5, pp. 847–864, 2024, doi: 10.1080/03075079.2024.2323593.
[8] B. Han, S. Coghlan, G. Buchanan, and D. McKay, “Who is helping whom? student concerns about AI-teacher collaboration in higher education classrooms,” Proceedings of the ACM on Human-Computer Interaction, vol. 9, no. 2, pp. 1–32, 2025, doi: 10.1145/3711104.
[9] L. Lawrence, V. Echeverria, K. Yang, V. Aleven, and N. Rummel, “How teachers conceptualise shared control with an AI co orchestration tool: A multiyear teacher-centred design process,” Br.

J. Educ. Technol., vol. 55, no. 3, pp. 823–844, 2024, doi: 10.1111/bjet.13372.
[10] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," IEEE Access, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/access.2020.2988510.
[11] M. Bond et al., "A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour," Int. J. Educ. Technol. High. Educ., vol. 21, no. 1, Art. no. 4, 2024, doi: 10.1186/s41239-023-00436-z.
[12] M. Y. Mustafa et al., "A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda," Smart Learn. Environ., vol. 11, no. 1, Art. no. 59, 2024, doi: 10.1186/s40561-024-00350-5.
[13] M. Omland et al., "Co-creation in higher education: a conceptual systematic review," High. Educ., vol. 90, no. 4, pp. 1017–1047, 2025, doi: 10.1007/s10734-024-01364-1.
[14] O. Robayo-Pinzon, S. Rojas-Berrio, J. Rincon-Novoa, and A. Ramirez-Barrera, "Artificial intelligence and the value co-creation process in higher education institutions," Int. J. Hum.-Comput. Interact., vol. 40, no. 20, pp. 6659–6675, 2024, doi: 10.1080/10447318.2023.2259722.
[15] M. Romero, "From consumption to co-creation: A systematic review of six levels of AI-enhanced creative engagement in education," Multimodal Technol. Interact., vol. 9, no. 10, Art. no. 110, 2025, doi: 10.3390/mti9100110.
[16] D. Ifenthaler and C. Schumacher, "Reciprocal issues of artificial and human intelligence in education," J. Res. Technol. Educ., vol. 55, no. 1, pp. 1–6, 2023, doi: 10.1080/15391523.2022.2154511.
[17] J. Kim, "Leading teachers' perspective on teacher-AI collaboration in education," Educ. Inf. Technol., vol. 29, no. 7, pp. 8693–8724, 2024, doi: 10.1007/s10639-023-12109-5.
[18] J. Kim, H. Lee, and Y. H. Cho, "Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education," Educ. Inf. Technol., vol. 27, no. 5, pp. 6069–6104, 2022, doi: 10.1007/s10639-021-10831-6.
[19] K. Pahi, S. Hawlader, E. Hicks, A. Zaman, and V. Phan, "Enhancing active learning through collaboration between human teachers and generative AI," Comput. Educ. Open, vol. 6, Art. no. 100183, 2024, doi: 10.1016/j.caeo.2024.100183.
[20] L. Zheng, Y. Fan, B. Chen, Z. Huang, L. Gao, and M. Long, "An AI-enabled feedback-feedforward approach to promoting online collaborative learning," Educ. Inf. Technol., vol. 29, no. 9, pp. 11385 11406, 2024, doi: 10.1007/s10639-023-12292-5.
[21] L. I. Ruiz-Rojas, L. Salvador-Ullauri, and P. Acosta-Vargas, "Collaborative working and critical thinking: adoption of generative artificial intelligence tools in higher education," Sustainability, vol. 16, no. 13, Art. no. 5367, 2024, doi: 10.3390/su16135367.
[22] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: perceptions, benefits, and challenges in higher education," Int. J. Educ. Technol. High. Educ., vol. 20, no. 1, Art. no. 43, 2023, doi: 10.1186/s41239-023-00411-8.
[23] A. Fošner, "University students' attitudes and perceptions towards AI tools: implications for sustainable educational practices," Sustainability, vol. 16, no. 19, Art. no. 8668, 2024, doi: 10.3390/su16198668.
[24] X. Zhou, J. Zhang, and C. Chan, "Unveiling students' experiences and perceptions of artificial intelligence usage in higher education," J. Univ. Teach. Learn. Pract., vol. 21, no. 6, 2024, doi: 10.53761/xzjprb23.
[25] C. K. Y. Chan and L. H. Y. Tsi, "Will generative AI replace teachers in higher education? A study of teacher and student perceptions," Studies in Educational Evaluation, vol. 83, Art. no. 101395, 2024, doi: 10.1016/j.stueduc.2024.101395.
[26] S. Soleimani, M. Farrokhnia, A. van Dijk, and O. Noroozi, "Educators' perceptions of generative AI: investigating attitudes, barriers and learning needs in higher education," Innov. Educ. Teach. Int., vol. 62, no. 5, pp. 1598–1613, 2025, doi: 10.1080/14703297.2025.2530767.
[27] S. Haroud and N. Saqri, "Generative AI in higher education: Teachers' and students' perspectives on support, replacement, and digital literacy," Education Sciences, vol. 15, no. 4, Art. no. 396, 2025, doi: 10.3390/educsci15040396.
[28] A. M. Al-Zahrani and T. M. Alasmari, "Exploring the impact of artificial intelligence on higher education: the dynamics of ethical, social, and educational implications," Humanit. Soc. Sci. Commun., vol. 11, no. 1, Art. no. 912, 2024, doi: 10.1057/s41599-024-03432-4.
[29] A. I. Pisica, T. Edu, R. M. Zaharia, and R. Zaharia, "Implementing artificial intelligence in higher education: pros and cons from the perspectives of academics," Societies, vol. 13, no. 5, Art. no. 118, 2023, doi: 10.3390/soc13050118.
[30] M. Doğan, A. Celik, and H. Arslan, "AI in higher education: risks and opportunities from the academician perspective," Eur. J. Educ., vol. 60, no. 1, Art. no. e12863, 2025, doi: 10.1111/ejed.12863.
[31] M. A. Bulut, "'Aidemics' examined analytically through the connectivist theory: institutional AI-mediation in higher education via an artificial intelligence center (CILT-AI)," Interactive Learning Environments, vol. 34, no. 2, pp. 972–1001, 2026, doi: 10.1080/10494820.2026.2647982.
[32] A. J. Karran et al., "Multi-stakeholder perspective on responsible artificial intelligence and acceptability in education," npj Science of Learning, vol. 10, no. 1, Art. no. 44, 2025, doi: 10.1038/s41539-025 00333-2.
[33] S. R. Thomson, B. A. Pickard-Jones, S. Baines, and P. C. J. Otermans, "The impact of AI on education and careers: what do students think?" Front. Artif. Intell., vol. 7, Art. no. 1457299, 2024, doi: 10.3389/frai.2024.1457299.
[34] M. Vecchiarini and T. Somià, "Redefining entrepreneurship education in the age of artificial intelligence: an explorative analysis," Int. J. Manag. Educ., vol. 21, no. 3, Art. no. 100879, 2023, doi: 10.1016/j.ijme.2023.100879.
[35] J. M. Santos-Jaén, A. León-Gómez, M. C. Valls Martínez, and V. Llorente Muñoz, "University students' perceptions of the impact of artificial intelligence in the business sector on their educational and professional development," Educ. Inf. Technol., vol. 30, no. 17, pp. 24395–24428, 2025, doi: 10.1007/s10639-025-13712-4.
[36] C. K. Y. Chan and W. Zhou, "An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI," Smart Learn. Environ., vol. 10, no. 1, Art. no. 64, 2023, doi: 10.1186/s40561-023-00284-4.
[37] H. U. Rahiman and R. Kodikal, "Revolutionizing education: artificial intelligence empowered learning in higher education," Cogent Education, vol. 11, no. 1, Art. no. 2293431, 2024, doi: 10.1080/2331186x.2023.2293431.
[38] G. M. Idroes et al., "Student perspectives on the role of artificial intelligence in education: A survey-based analysis," Journal of Educational Management and Learning, vol. 1, no. 1, pp. 8–15, 2023, doi: 10.60084/jeml.v1i1.58.
[39] C. Almaraz-López, F. Almaraz-Menéndez, and C. López-Esteban, "Comparative study of the attitudes and perceptions of university students in business administration and management and in education toward artificial intelligence," Educ. Sci., vol. 13, no. 6, Art. no. 609, 2023, doi: 10.3390/educsci13060609.
[40] V. Okulich-Kazarin, A. Artyukhov, Ł. Skowron, N. Artyukhova, O. Dluhopolskyi, and W. Cwynar, "Sustainability of higher education: study of student opinions about the possibility of replacing teachers with AI technologies," Sustainability, vol. 16, no. 1, Art. no. 55, 2024, doi: 10.3390/su16010055.
[41] R. Vrana, "AI as a study companion: an exploration of student experiences at the faculty of Humanities and Social Sciences, Zagreb, Croatia," in Proc. 48th MIPRO ICT Electron. Conv. (MIPRO), pp. 461–466, 2025, doi: 10.1109/mipro65660.2025.11131750.
[42] M. Rožman, D. Oreški, A. A. Elias, M. F. Pynadath, and P. Tominc, "AI literacy among university students: A comparative study of three countries—Slovenia, Croatia, and India," IEEE Access, vol. 13, pp. 110671–110688, 2025, doi: 10.1109/access.2025.3581128.
[43] O. Luque-Reca, C. Peñacoba, P. Catalá, L. Gutiérrez Hermoso, and S. Grassini, "Artificial intelligence attitudes in higher education: Spanish validation of the AIAS-4 and its associations with well-being and academic AI usage," Curr. Psychol., vol. 45, no. 4, Art. no. 370, 2026, doi: 10.1007/s12144-025-08566-5.
[44] L. Brodarič and S. Babić, "The adoption of artificial intelligence tools in education: A case study of primary and secondary school teachers in Pula, Croatia," in Human Systems Engineering and Design (IHSED 2025): Future Trends and Applications, vol. 198. AHFE Open Access, 2025, pp. 299–308, doi: 10.54941/ahfe1006809.
[45] B. Lišnić, G. Zaharija, and S. Mladenović, "Integration of artificial intelligence in K-12: analysis of a three-year pilot study," AI, vol. 6, no. 3, Art. no. 49, 2025, doi: 10.3390/ai6030049.

[46] A. F. Martin, S. Tubaltseva, A. Harrison, and G. J. Rubin, "Participatory co-design and evaluation of a novel approach to generative AI-integrated coursework assessment in higher education," Behav. Sci., vol. 15, no. 6, Art. no. 808, 2025, doi: 10.3390/bs15060808.
[47] A. Williams and E. Ingle, "Assessment design through co-creation: student-staff partnership in evaluating the impact of artificial intelligence," Int. J. Students Partners, vol. 9, no. 1, pp. 214–226, 2025, doi: 10.15173/ijsap.v9i1.5802.
[48] I. Sabic, H. Puljiz, and A. Smoljo, "Personalized learning in the Croatian national education system: A study of AI implementation in the e-Class Register," SN Computer Science, vol. 5, no. 8, Art. no. 1145, 2024, doi: 10.1007/s42979-024-03515-8.
[49] M. Filipov, K. Mikulić, and L. Mustač, "Creative integration of artificial intelligence (AI) tools in effective teaching," Creativity. Theories – Research – Applications, vol. 12, no. 1, pp. 139–156, 2025, doi: 10.2478/ctra-2025-0007.
[50] M. Rajko, "The use of artificial intelligence by young entrepreneurs in the Republic of Croatia," Interdisciplinary Description of Complex Systems, vol. 22, no. 6, pp. 656–665, 2024, doi: 10.7906/indecs.22.6.2.
[51] S. Pozdniakov, J. Brazil, M. Mohammadi, M. Dollinger, S. Sadiq, and H. Khosravi, "AI-assisted co-creation: bridging skill gaps in student- generated content," J. Learn. Anal., vol. 12, no. 1, pp. 129–151, 2025, doi: 10.18608/jla.2025.8601.
[52] J. Zhong and T. Cochrane, "Heutagogy-based human-AI co-creation practice: A framework for enhancing undergraduate creativity," in Proc. ASCILITE 2024, pp. 351–356, 2024, doi: 10.14742/apubs.2024.1083.
[53] S. Zeivots et al., "Reshaping higher education designs and futures: postdigital co-design with generative artificial intelligence," Postdigit. Sci. Educ., vol. 7, no. 4, pp. 1334–1374, 2025, doi: 10.1007/s42438 025-00595-4.
[54] M. M. Msambwa, Z. Wen, and K. Daniel, "The impact of AI on the personal and collaborative learning environments in higher education," Eur. J. Educ., vol. 60, no. 1, Art. no. e12909, 2025, doi: 10.1111/ejed.12909.
[55] T. Li, Y. Ji, and Z. Zhan, "Expert or machine? comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course," Asia Pac. J. Educ., vol. 44, no. 1, pp. 45– 60, 2024, doi: 10.1080/02188791.2024.2305163.
[56] Y. Zhuang, Y. Ji, Z. Huang, Y. Shang, X. Li, and X. Su, "Exploring the integration of artificial intelligence and programming education: A practice-oriented study on the development of computational thinking based on human-AI co-creation," in Proc. IEEE 5th Int. Conf. Softw. Eng. Artif. Intell. (SEAI), pp. 325–329, 2025, doi: 10.1109/seai65851.2025.11108869.
[57] A. Aler Tubella, M. Mora-Cantallops, and J. C. Nieves, "How to teach responsible AI in higher education: challenges and opportunities," Ethics Inf. Technol., vol. 26, no. 1, Art. no. 3, 2024, doi: 10.1007/s10676-023-09733-7.
[58] T. K. F. Chiu, H. Meng, C.-S. Chai, I. King, S. Wong, and Y. Yam, curriculum," IEEE Trans. Educ., vol. 65, no. 1, pp. 30–39, 2022, doi: 10.1109/te.2021.3085878.
[59] S. Grassini, "Development and validation of the AI attitude scale (AIAS-4): a brief measure of general attitude toward artificial intelligence," Front. Psychol., vol. 14, Art. no. 1191628, 2023, doi: 10.3389/fpsyg.2023.1191628.
[60] A. Møgelvang and S. Grassini, "Validating the AI attitude scale (AIAS-4) and exploring attitudinal differences in a large sample of Norwegian university students," Discov. Educ., vol. 4, no. 1, Art. no. 212, 2025, doi: 10.1007/s44217-025-00657-6.
[61] A. Sultana, N. Abdul Latheef, N. Siby, and Z. Ahmad, "Exploring students' attitudes toward artificial intelligence (AI): psychometric validation of AI-attitude scale," Sage Open, vol. 15, no. 4, Art. no. 21582440251378375, 2025, doi: 10.1177/21582440251378375.
[62] E. I. Yuwono, D. Tjondronegoro, C. Riverola, and J. Loy, "Co-creation in action: bridging the knowledge gap in artificial intelligence among innovation champions," Comput. Educ. Artif. Intell., vol. 7, Art. no. 100272, 2024, doi: 10.1016/j.caeai.2024.100272.
[63] A. Okada, T. Sherborne, G. Panselinas, and G. Kolionis, "Fostering transversal skills through open schooling supported by the CARE KNOW-DO pedagogical model and the UNESCO AI competencies framework," Int. J. Artif. Intell. Educ., vol. 35, no. 4, pp. 1953–1998, 2025, doi: 10.1007/s40593-025-00458-w.
[64] E. H. Hernández Rincón, D. Jimenez, L. A. Chavarro Aguilar, J. M. Pérez Flórez, Á. E. Romero Tapia, and C. L. Jaimes Peñuela, "Mapping the use of artificial intelligence in medical education: a scoping review," BMC Medical Education, vol. 25, no. 1, Art. no. 526, 2025, doi: 10.1186/s12909-025-07089-8.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

A Research-Embedded Implementation Model for Design of AI Applications in Higher Education

Hiba Kassab and Nancy Maynes

Nipissing University, Canada

Article DOI: https://doi.org/10.68286/samyrad2026.309

ABSTRACT

In the context of higher education, particularly in the current era of expansive growth of Artificial Intelligence (AI), the challenges that confront students and faculty are numerous and multifaceted. Potential paths for researching and defining effective implementation of this technology are complex. This article provides an emerging model for further discussion of exploring, utilizing, and expanding the implementation of AI in educational contexts, with a focus on higher education. Meta-synthesis has been applied to 1679 studies to identify themes/recurring knowledge and uncover recurring knowledge to inform processes for adopting AI use in education and optimizing its potential to enhance instruction and learning. ChatGPT GenAI synthesis identified 12 recurring meta-themes, whereas Copilot GenAI synthesis identified 13 recurring meta-themes that describe how GenAI technologies are used and experienced in higher education environments. Both AI tools produced frameworks to illustrate their findings. ChatGPT's framework synthesizes the mixed role of GenAI in higher education, highlighting its pedagogical values and associated risks, and tracing its implications for learning processes, student experience, instructional design, and institutional governance, whereas Copilot's framework illustrates how effective GenAI integration relies on coordinated pedagogical design, institutional governance, operational stability, and equity-focused practices. Four main ideas have been identified as consistently displayed in previous studies on use in higher education, including learning, pedagogy, governance, operations, and inclusion/industry, potentially influencing institutional and instructor readiness to use AI effectively in the higher learning environment.

KEYWORDS Meta-synthesis, Artificial Intelligence, ChatGPT, Higher education, Copilot

How to cite: Kassab, H., & Maynes, N. (2026). A research-embedded implementation model for design of AI applications in higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.309

Introduction

In higher education, amid the rapid growth of Artificial Intelligence (AI), the challenges facing students and faculty are numerous and multifaceted. AI and immersive technologies such as virtual and augmented reality are reshaping "virtual classrooms" into "modified learning formats and flexible course modules" [1], while simultaneously redefining the global educational landscape. As Elayyan [2] notes, integrating Industry 4.0 technologies into education requires cultivating "liquid or soft skills [such as] critical thinking, coordinating with others, verbal communication, and time management" alongside traditional knowledge-based skills, a shift accelerated by the rise of Massive Open Online Courses (MOOCs) as a democratizing, disruptive force in education [2].

These advancements also bring challenges: redefining "learning outcomes and curricula" and preparing students for a labour market where traditional roles may give way to emerging professions requiring AI literacy [2]. This article explores these implications, emphasizing the necessity for higher education to

adapt and prepare students for an AI-driven future. It is intended to help post-secondary instructors and institutions adopt a proactive, evidence-informed approach to this reality by synthesizing current qualitative research on the instructional use of AI and how best to use it.

Background

While article length limitations prohibit lengthy definitions of AI and claims of its popularity in our culture [3]; [4], it is valuable to define the different types of AI as they are currently presented in the research literature, while simultaneously recognizing that these classifications are in continuous shift with the developments in technology. DiCarlo [5] categorizes AI into three stages: Artificial narrow intelligence (or ANI), "artificial general intelligence (or AGI), and artificial super intelligence (or ASI)". Currently, most AI applications use ANI, designed to execute "specific tasks and to directly follow its digital command", in applications such as recommendation systems, speech recognition, and automated grading tools, generally purposed to enhance task efficiency.

AGI represents AI systems capable of independent reasoning, "problem-solving, learning, and long-term planning abilities, as well as and, eventually, much better than any human" [5]. AGI can learn, adapt, and generate novel solutions similar to human cognition; however, it also raises concerns regarding ethical decision-making, job displacement, and the potential loss of human oversight in critical sectors. Additionally, users will have experienced the cultural nature of the current "learning" of some AGI tasks. For example, depending on where a program was trained in certain content, a lesson on writing structure might default to the five-paragraph essay format (indicating that it was trained in North American curricula) but exclude oral storytelling traditions or argumentative structures common in other cultures or educational systems.

DiCarlo's third stage, ASI, refers to "superintelligence, ... intelligence that surpasses the cognitive abilities of any individual human or even the combined intellect of all humanity" [5], with the resultant existential risks it entails. Speculation about ASI risks has implications for education, society, and global stability, highlighting the need for informed, responsible AI development and regulation to ensure that technological advancements align with human values and safety in higher education.

Our current applications of AI draw on several disciplines that have contributed to its current structures. Among these machine learning uses [6], are, analysis of algorithms to identify patterns and optimize performance through exposure to extensive datasets [6]; machine-to-machine interactions (also called IoT or the Internet of Things) that extend beyond human-to-machine interactions [7]; [8]; formal representations of algorithmic thinking and computational models [9]; intelligent tutoring systems and frameworks for educational technology systems [10]; human information processing (e.g., memory, attention, and decision-making) applied to computational frameworks and selective attending [11]; formal models of reasoning and problem-solving [12]; neural-network architectures, brain structure and functions and neural network mimicking [13]; computational linguistics foundations for parsing human language into machine-readable structures [14]; and the simulation of human cognitive processes [4].

AI is reshaping education by creating interconnected systems to personalize learning experiences. This shift may optimize educational processes and demonstrate the broader impact of IoT, where devices interact autonomously to enhance efficiency and accessibility across various domains [15]; [16]; [17]. It may also make the post-secondary education world uncertain about adoption steps and its own values and principles as AI emerges into instructional and institutional territory. This revolution underscores the urgency of our understanding of current research on AI in our learning environments.

Methodology

A meta-synthesis is needed to identify the themes that emerge from the interaction among faculty who teach, students who learn, administrators who process, and the AI tools available to enhance this collaboration. Systematic literature reviews and meta-level studies can provide "systematic ways of collecting and synthesizing previous research" [18] to create a theoretical framework for AI adoption [4]; [19] and form the basis for "building a conceptual model or theory" [18].

This study began with a broad inclusion of relevant qualitative studies, gradually applying further criteria to prioritize measures of validity, reliability, generalizability, and relevance, and refining the full range of available studies to those most likely to enrich a meta-synthesis theory. By synthesizing diverse perspectives, there is potential to uncover the nuanced ways in which AI is reshaping higher educational practices, through "gathering rich data to discover underlying patterns and construct theories that are grounded in the data" [20].

Adopting the interpretivist/constructivist framework in meta-synthesis and grounded theory has key implications, including how data are collected, analyzed, and how valid the conclusions may be [21]. The meta-synthesis procedure as outlined by Walsh and Downe [22] was adopted for this study. Six steps of that procedure were followed. including:

Framing a meta-synthesis exercise: defining clear research questions, such as "What are the emergent themes in AI applications in higher education?"

Locating relevant papers: conducting a comprehensive search for qualitative studies about AI in higher education across various databases.

Deciding what to include: selecting all relevant qualitative studies, focusing on those related to AI in higher education, teaching, learning, or administration.

Appraisal of studies: evaluating the quality and relevance of the chosen studies, with an emphasis on inclusivity.

Analytical technique: employing a grounded theory approach

for data analysis, extracting relevant data through comparison, contrast, and thematic synthesis.

Synthesis of translation*:* integrating the findings to reveal refined meanings, exploratory theories, and new concepts that contribute to a comprehensive understanding of the topic.

This was an iterative process, using constant comparative approaches to isolate studies from the original 1679 (see Table 1), to focus analysis on those most closely aligned with strong qualitative methodologies and therefore most reliable in their conclusions. After filtering the initial body of related AI studies using various criteria, 154 studies were selected for detailed analysis of themes and findings. Table 1 shows the original body of studies from the full range of AI search engines that were used.

Table 1. Number of articles generated from the different AI tools and other search engines

AI tool/ Search Engine	Number of articles generated
Undermind (April 04-2025)	598
ChatGPT03-Deep Research	58
ChatGPT03-Search	124
Undermind (July 10-2025)	122
Research Rabbit (July 26-2025)	488
Deepseek and Scholar	201
Other articles	88
Total	**1679**

To investigate AI, AI tools were used: ChatGPT (versions 3 and 3-Deepsearch), Undermind, and ResearchRabbit, to identify relevant studies. In addition, online databases for journal articles and chapters were employed in the broadest stages of the search (e.g., Google Scholar). As searches were initiated, inclusion criteria were made specific to ensure adherence to the standards and the post-positivist, constructivist approach. These inclusion criteria included: relevance to AI in higher education; qualitative study design; peer-reviewed; currency (within the last 10 years); empirical studies with rich thematic insights, and a focus on sustainability and educational outcomes. This study excluded previous research that used only quantitative or mixed methods, was conducted in unrelated educational contexts, was out of date (10 years or more), was published in a language other than English, was not peer-reviewed, or had an obsolete technological focus. The application of both inclusion and exclusion criteria to the initial data set of 1679 articles is displayed in Table 2.

Table 2. Number of articles per criterion for inclusion/exclusion

Criterion	Number of Articles
Duplicates	260
Mixed Methods	130
Quantitative	243
Not English, irrelevant, no record found, retracted, restricted access	299
Date parameter	33
Non-Peer-reviewed	15
Qualitative	792

The remaining 154 studies that met the rigid criteria for this purpose were analyzed using the Walsh and Downe [22] method.

Research Design

This study employed a qualitative meta-synthesis to integrate and interpret findings from published qualitative papers on GenAI in higher education. This method permits researchers to move beyond aggregation by re-analyzing and conceptually extending qualitative results across multiple studies [23]. An analytical strategy was applied for the retrieval of the thematic and interpretive synthesis [24], which involves (1) generating codes that remain faithful to participants' or authors' language; (2) grouping conceptually related codes into higher-order themes; and (3) developing integrative propositions that transcend individual studies. This approach is widely used in qualitative synthesis because it enables researchers to integrate qualitative findings methodically while maintaining transparency and traceability to the original data sources [25]. The analysis focuses on identifying recurring patterns in educational practice, as well as the opportunities and challenges associated with GenAI adoption.

Data Sources and Coding Architecture

The synthesis drew on two distinct outputs generated using the two GenAI tools: Copilot and ChatGPT. Copilot was chosen because it is the AI tool approved for use across Universities and Colleges in Canada, whereas ChatGPT was picked for its global reputation as the go-to GenAI tool. The developed coded datasets ("Coding-Copilot" and "Coding-ChatGPT" workbooks) each contain qualitative excerpts extracted from the same peer-reviewed qualitative studies on GenAI in higher education contexts. Each coded entry retained four elements:

1. The source article from which the excerpt originated;
2. The excerpt itself, representing a relevant segment of qualitative evidence;
3. The initial code assigned during the coding phase; and
4. The corresponding thematic interpretation.

Maintaining this structured coding format ensured that each analytical theme could be traced back to its original textual evidence, thereby preserving methodological transparency and an auditable trail of interpretation. Establishing such traceability is essential in qualitative synthesis research to enhance transparency and accuracy [23]. The coded articles were subsequently consolidated into 13 themes (Copilot) and 12 themes (ChatGPT), which served as the structural basis for the two synthesis results.

Coding and Thematic Consolidation

The coding process followed an iterative thematic synthesis approach, consistent with qualitative meta-synthesis methodologies described by Thomas and Harden [24] Initial coding preserved the verbatim text retrieved from the studies using Copilot and ChatGPT, and the language and meaning of the original excerpts were kept as closely as possible to maintain fidelity to the source material. Two workbooks were created to

conduct the thematic analysis for each GenAI tool.

At the specific-code level, retaining open codes that reflect participants' or authors' language to maintain the semantic integrity of the original language during early coding is a recommended practice because it minimizes interpretive distortion during the initial analytical stage [26].

Following the initial coding phase, codes were grouped into broader conceptual categories through axial coding, a process that organizes related codes according to semantic and conceptual similarities. This stage resulted in the development of 13 thematic categories for Copilot and 12 for ChatGPT, representing key domains of GenAI use and impact in higher education.

To improve thematic precision, the coding framework underwent a two-pass refinement process, including:

Pass 1: Rule-based classification. In the first refinement stage, decisions were applied to ensure consistent thematic assignment. For example:

- Codes referring to hallucinated outputs or inaccurate AI responses were classified under Technical Limitations.
- Codes describing tutoring, explanations, or conceptual clarification were assigned to Learning Support.

Pass 2: Pragmatic thematic adjustment. A second refinement stage adjusted the codes based on the contextual interpretation of the quote. For example:

- Codes referring to time savings or productivity improvements were reassigned to the Efficiency and Productivity category.
- Codes referring to the integration of human instruction and AI systems were grouped under Human-AI Pedagogy.

This iterative consolidation process reduced reliance on overly broad categories such as Other or General while improving the conceptual clarity of the thematic structure.

Thematic Synthesis and Analytical Calibration

The synthesis progressed through three levels of analysis when using the Copilot GenAI tool.

Within-category synthesis: Coded excerpts within each category were examined to identify recurring subthemes. These subthemes were informed by two quantitative indicators, involving:

- Article counts indicating the number of unique studies contributing evidence to a theme; and
- Occurrence counts indicating the frequency with which the theme appeared in the dataset.

Meta-level conceptual synthesis*:* The final stage of analysis involved developing higher-order propositions that connected individual themes into a broader interpretive framework; patterns observed across the dataset were grounded in the coded evidence extracted from the source studies.

Throughout this process, the emphasis in the findings was calibrated to the thematic counts (recorded in Table 3) to reduce the risk of over-representing themes supported by limited evidence, thereby maintaining alignment between analytical claims and the distribution of coded data.

Limitations and trustworthiness

Several strategies were used to enhance the trustworthiness of this synthesis. First, traceability was maintained through structured coding of Excel ensuring an audit trail [26]. Second, unique codes, occurrences, and contributing articles, allowed thematic emphasis to be adjusted according to the distribution of evidence across the dataset and to reduce the tendency to overinterpret themes based on limited sources.

Nevertheless, several limitations should be acknowledged. The coded dataset reflects heterogeneous source coverage, meaning that some thematic categories draw on fewer contributing studies than others. For example, themes related to data privacy and security were supported by fewer articles and should therefore be understood as offering directional insight rather than definitive conclusions. Additionally, the rapidly evolving nature of GenAI technologies introduces the possibility of data deviation, in which technological developments may overtake the evidence available in the literature.

Tool-Specific sensitivity. A further limitation concerns the extent to which the findings reported here are attributable to GenAI-assisted synthesis in general versus the particular behaviour of the two tools used, ChatGPT and Copilot, at the specific versions and points in time this study was conducted. The two tools organized identical coded evidence according to different underlying logics, producing frameworks that diverged in the number of excerpts extracted, the themes, and the structure, despite being subject to the same human validation and adjudication procedure. This suggests that at least part of the study's output reflects each tool's default synthesis behaviour rather than solely the properties of the underlying literature. Because GenAI tools are updated frequently and their synthesis behaviour is not fully documented or controlled by the human using them, this study's specific findings may not replicate identically if repeated with newer versions of these tools, or with other GenAI systems entirely. This limitation does not undermine the validity of the human-adjudicated codes and themes, which were independently verified against source excerpts regardless of which tool proposed them; rather, it constrains the generalizability of the specific frameworks presented in Figs. 1 and 2, which should be read as one exemplification of GenAI-assisted synthesis rather than a stable ground truth. Future work replicating this design across additional GenAI tools, or at multiple points in time with the same tools, would help distinguish which patterns are durable features of the underlying evidence and which are exploratory of a given tool's synthesis at a given point in its development.

Table 3. Thematic findings summary using Copilot GenAI: Limitations and Trustworthiness

Theme	Concise definition	Unique codes	Occurrences	Unique articles
Learning Support & Understanding	On-demand explanations, scaffolding, reassurance	216	286	**20**
Engagement & Participation	AI as conversation starter; critique of outputs	94	127	**5**
Efficiency & Productivity	Time compression for routine tasks	98	112	**7**
Assessment & Feedback	AI for breadth/speed; humans for specificity/affect	63	67	**2**
Accessibility & Inclusion	Language support, captioning, neurodiversity	97	113	**4**
Academic Integrity & Ethics	Concerns: plagiarism, fabricated citations; redesign	123	146	**5**
Data Privacy & Security	Sandboxing, provenance, consent	30	33	**2**
Human–AI Pedagogy	Designed use; role-play; scenario-based learning	88	95	**4**
Technical Limitations & Accuracy	Hallucinations; formulaic style; shallow reasoning	115	122	**6**
Infrastructure & Access	Bandwidth, devices, premium features	43	47	**4**
Training, Policy & AI Literacy	PD, clear policies, norm-setting	107	119	**5**
Accessibility of Data Governance	Privacy, Integrity interface: watermarking, provenance	—	—	**—**
Industry & Workplace Readiness	Task reallocation; AI literacy for work	25	26	2

Methods validation and trustworthiness

Following Lincoln and Guba's [27] framework, the trustworthiness of the synthesis was evaluated through the criteria of credibility, dependability, and confirmability to strengthen transparency, reliability, and interpretive validity [27]; [23]; [28].

Thematic Findings: Process and Results

The thematic synthesis followed a structured multi-stage analytic procedure designed to identify recurring conceptual patterns across the coded dataset [29]; [23]; [24]. The same prompts were used in both Copilot and ChatGPT at each level of the analysis.

Meta-Themes Identified in the Qualitative Meta-Synthesis

The ChatGPT GenAI identified 12 meta-themes, whereas the Copilot GenAI distinguished 13 recurring meta-themes that describe how GenAI technologies are used and experienced in higher education environments.

Resulting Conceptual Framework

After completing the data analysis and retrieving the meta-themes for each GenAI tool, each output was uploaded to the corresponding tool to produce a conceptual framework. Although the datasets were extracted from the same qualitative studies, the two tools produced different themes and conceptual frameworks.

ChatGPT:

Fig. 1 illustrates the interaction between ChatGPT, other large language models, and higher education systems. The model identifies two primary dimensions: educational functions and educational risks. Educational functions include learning support, writing and language development, and research assistance, which enable students to access explanations, develop academic writing, and explore research ideas. In parallel, educational risks include cognitive dependence, ethical concerns related to AI misuse, and reliability issues associated with biased or inaccurate AI-generated outputs. These factors collectively influence learning processes and shape the student learning experience, including engagement, productivity, and learner autonomy. The framework further demonstrates how these dynamics contribute to broader pedagogical transformation, such as assessment redesign and AI-integrated teaching practices, and ultimately inform institutional governance responses, including AI literacy initiatives, policy development, and responsible AI use in higher education.

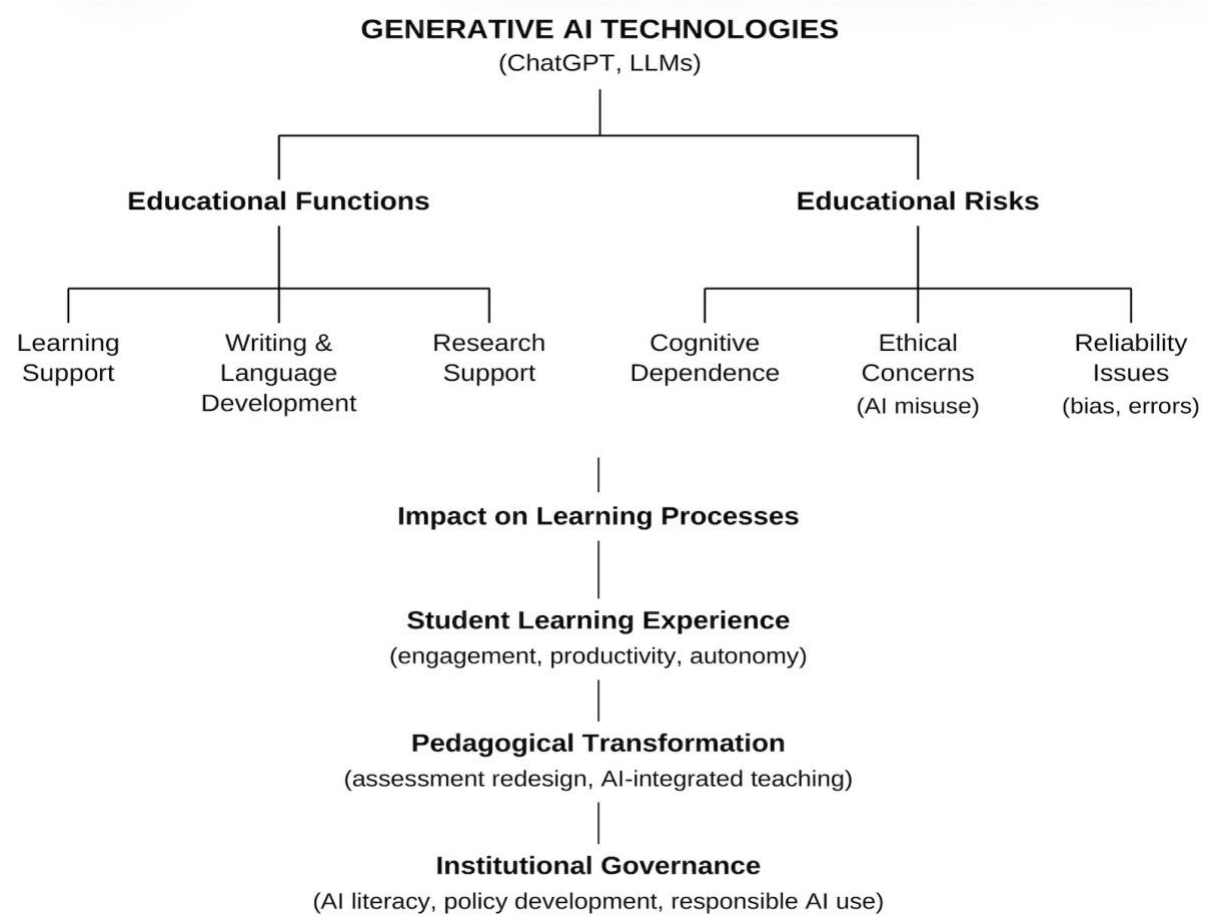


Fig. 1. Unified Conceptual Framework of GenAI Integration in Higher Education: Integrated Socio-Technical Model of Gen AI in Higher Education created by ChatGPT (ChatGPT-generated).

Conceptual Framework: GenAI in Higher Education

The conceptual framework presented in Fig. 1 was produced by ChatGPT after uploading the ChatGPT Excel workbook. It synthesizes the mixed role of generative artificial intelligence (AI) in higher education, highlighting its pedagogical values and associated risks, and tracing its implications for learning processes, student experience, instructional design, and institutional governance. The framework reflects a structured abstraction of recurrent empirical patterns and theoretical insights across the studies under review, examining tools such as ChatGPT and other large language models (LLMs).

Mixed-Role: Educational Functions and Educational Risks

At the center of the framework is the divergence between educational functions and educational risks, reflecting the two dominant narratives in AI in higher education. This dual structure reflects a growing consensus that GenAI cannot be understood exclusively as a pedagogical modernization, but rather as a sociotechnical phenomenon characterized by both empowering and limiting effects.

Educational Functions

The first dimension, educational functions, examines how GenAI enriches academic tasks and supports learning. Three subdomains emerged consistently across studies.

First, learning support reflects AI's role as an on-demand cognitive platform. GenAI tools provide explanations, clarify complex concepts, and simulate tutoring interactions, thereby supporting individualized and self-paced learning.

Second, writing and language development are among the most prominent applications of GenAI. The literature repeatedly highlights AI's capacity to assist with drafting, paraphrasing, grammar correction, and language refinement, benefiting predominantly multilingual learners and those developing academic writing proficiency.

Third, research support encompasses the use of AI for idea generation, summarization of scholarly materials, and preliminary literature exploration. Unlike writing support, this function contributes to early-stage knowledge production processes, facilitating the organization, development and synthesis of information.

Educational Risks

In contrast, the second dimension, educational risks, raises doubts about the unintentional effects of AI integration. Three primary categories were detected.

First, cognitive dependence refers to the tendency to over-reliance on AI systems, leading to reduced critical thinking, diminished engagement in problem-solving, and superficial learning. This risk is extremely significant because it directly affects the intellectual abilities underlying learning rather than only external behaviours.

Second, ethical concerns such as academic integrity, plagiarism, authorship ambiguity, and inappropriate reliance on AI-generated content emerged. These concerns are a constant focus in institutional and policy-oriented literature, highlighting tensions between technological capability and academic norms.

Third, reliability issues involve the technical limitations of GenAI, including hallucinations (i.e., the creation of illusory information), factual inaccuracies, and biases. Unlike ethical concerns, which arise from the user's behaviour, reliability issues are intrinsic to the technology itself and raise questions about trustworthiness and validity.

Collectively, these risks embrace cognitive, ethical, and technical dimensions, illustrating the multidimensional challenges of AI adoption in educational contexts.

From Technology to Learning: Impact on Learning Processes

The framework influences learning processes as a first-order outcome of the interaction between educational functions and risks. This level represents a critical analytical connection, highlighting how AI redesigns students' engagement with content, knowledge construction, and learning strategies.

The model underlines that not all risks operate equally at this level. Cognitive dependence is explicitly linked to learning processes because of its direct influence on students' reasoning ability, whereas ethical and reliability concerns have indirect effects.

Student Learning Experience

The framework identifies the student learning experience as an essential domain, encompassing engagement, productivity, and autonomy. These three dimensions represent the lived, experiential consequences of AI integration from the learner's perspective. Engagement refers to students' motivation and active participation in their learning journey. This may be enhanced by using GenAI as interactive support and reducing passive reliance. Productivity emphasizes improvements in efficiency, task completion, and academic output, which are frequently cited as primary benefits of AI tools. Autonomy considers students' ability for self-directed learning, which may be concurrently reinforced (through scaffolding) and weakened (through overdependence) by AI use.

Pedagogical Transformation

Beyond individual learning experiences, the framework illustrates how GenAI accelerates pedagogical transformation, reflecting a shift from progressive tool adoption toward more fundamental changes in teaching practices. First, assessment redesign emerges as a necessary response to the challenges posed by AI-generated content, urging educators to return to traditional evaluation methods. Second, AI-integrated teaching involves purposefully incorporating AI tools into instructional strategies, moving toward more adaptive, technology-enhanced learning environments. This transformation signifies a reconfiguration of the pedagogical model, in which AI is not only a supplementary tool but an active component of instructional design.

Institutional Governance

The framework situates these changes within the institutional governance perspective, emphasizing the general responses required to manage AI integration. Three key integration priorities are identified: 1) AI literacy requires both students and educators to develop the competencies needed to use AI technologies effectively and responsibly. 2) Policy development determines the necessary parameters for academic integrity, acceptable use, and assessment practices. 3) Adopting ethical frameworks and institutional standards demonstrates responsible AI use and confirms that it aligns with educational values, underscoring that the repercussions of GenAI extend beyond classrooms and require coordinated institutional strategies.

Integrative Perspective

This framework presents a multi-level, process-oriented model that details the influence of GenAI from technological capabilities and risks through to individual, pedagogical, and institutional consequences. It conceptualizes AI in higher education as a dynamic sociotechnical system in which benefits and risks coexist and interact, shaping both immediate learning processes and broader educational transformations. By integrating micro (student experience), meso (pedagogy), and macro (governance) perspectives, the framework proposes an inclusive understanding of the evolving role of GenAI in higher education. It also provides the foundation for future empirical research and policy development, as it is grounded in a meta-synthesis. The aim will be to leverage the identified points to maximize educational benefits while mitigating associated risks.

Human Validation of AI-Assisted Coding

While GenAI tools (ChatGPT and Copilot) were used to generate initial codes, thematic categories, and conceptual frameworks from the corpus of 154 studies, this process was conducted under continuous human supervision rather than as an autonomous analysis by AI tools. This judgment procedure was

applied independently and in parallel to each of the two workbooks; for every source article in both the Coding-ChatGPT and Coding-Copilot datasets, the text was independently read and coded according to the same thematic synthesis procedure [24] before comparing this independent coding against the corresponding AI-generated codes produced by that tool. Where human coding and a given tool's AI-generated coding converged, the code was retained. Where they diverged, the disagreement was adjudicated by returning to the original source excerpt to determine which coding, human, AI, or a revised synthesis of both, most faithfully represented the author's or participants' language and meaning. This process was carried out separately for the ChatGPT and Copilot outputs to preserve the integrity of each tool's independent dataset and to allow the resulting differences between the two conceptual frameworks (Figs. 1 and 2) to reflect genuine variation between the tools rather than cross-contamination introduced during validation. All AI-generated outputs, from the article collection, including evaluating them from inclusion/ exclusion, excerpts retrieval, individual codes, axial groupings, and the final thematic categories and conceptual frameworks for both tools, were reviewed against this standard; outputs found to be inaccurate, unsupported by the source text, or inconsistent were corrected or deleted prior to inclusion in the final synthesis. This dual, tool-specific adjudication process functioned as an additional layer of credibility and confirmability checking, consistent with Lincoln and Guba's [27] criteria, and ensured that the resulting themes and frameworks for each tool reflected a human-verified synthesis of the literature rather than an unexamined AI output.

Copilot:

The conceptual framework (Fig. 2) was produced by Copilot after uploading the Copilot Excel workbook. It positions generative artificial intelligence (GenAI) within four interconnected domains: Learning, Pedagogy/ Governance, Operations, and Inclusion/Industry, reflecting the main areas in which higher education systems are undergoing rapid transformation. These categories reflect the themes discovered across the meta-synthesis, based on the empirical analysis of GenAI in the higher education context using the Copilot tool.

The central node, GenAI in Higher Education, highlights GenAI's role as an adaptive support mechanism that accelerates cognitive, administrative, and creative processes. Solid connections represent primary domain linkages, while dashed cross-links depict secondary but recurrent interactions.

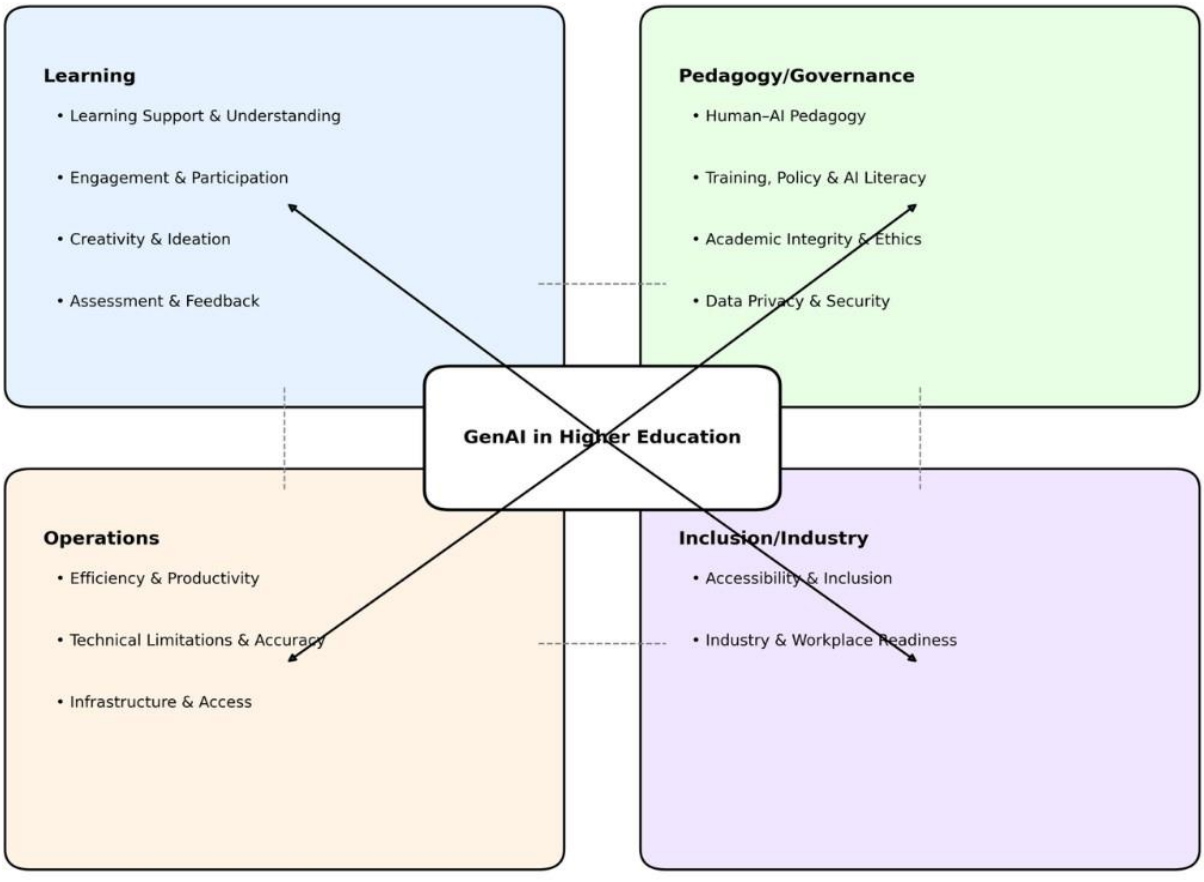


Fig. 2. System-level conceptual framework of GenAI in higher education created by Copilot

The Learning domain portrays how GenAI is reshaping student learning processes, particularly in areas such as personalization, engagement, creativity, and assessment. Multiple reviews indicate that GenAI tools can provide tailored learning opportunities, enhance academic support, and increase student involvement, thereby directly influencing their understanding and participation in meaningful ways. For example, research stresses GenAI's ability to deliver personalized learning experiences and improve student engagement through adaptive capabilities and responsive outputs. GenAI's role in shaping new forms of brainstorming and creative expression supports findings highlighting its flexibility across various educational tasks, while simultaneously raising questions about the integrity of assessments.

The Pedagogy/Governance domain acknowledges that instructional innovation must be anchored in responsible and ethical structures. Governance concerns, including privacy, data protection, fairness, and transparency for the responsible use of GenAI in education are highlighted. Analyses of GenAI governance models in educational research underline the importance of establishing adaptable, multi-level governance frameworks capable of addressing emerging pedagogical and regulatory considerations at macro-, meso-, and micro-levels. These outcomes reinforce the need for well-defined policies, faculty training, and AI literacy in ethically grounded and pedagogically rigorous structures.

The Operations domain exposes the organizational and infrastructural capacity required for effective GenAI integration. It has significant potential to improve efficiency and productivity in higher education by reorganizing administrative processes and supporting decision-making. For instance, GenAI is frequently discussed for its potential to reduce costs, enhance administrative workflows, and optimize service delivery when properly implemented. However, there is a tendency to understate its technical challenges, including limitations in accuracy, system reliability, data security, and the need for rigorous technological infrastructure to support safe use.

Finally, the Inclusion/Industry domain positions GenAI within the broader social and economic landscape, recognizing its implications for equity, accessibility, and workforce preparation. GenAI's predicted societal impact accentuates its potential to support more inclusive educational environments, but it can also aggravate existing inequalities if not implemented equitably. The rapid transformation of workforce expectations requires higher education institutions to prepare students for emerging skills involving AI ethics, AI literacy, and human-AI partnership.

These four domains explain the complex, interdependent ways GenAI is reshaping higher education. The framework highlights how changes in governance organizations shape pedagogical approaches; how operational infrastructures influence possibilities for equitable learning; how industry demands affect curriculum and workforce readiness; and how student learning needs force institutional adaptation. It offers a unified, multidimensional perception for understanding and guiding the responsible, transformative integration of GenAI within higher education.

Divergence Between the ChatGPT and Copilot Frameworks

Although both frameworks were derived from the same 154 article corpus and validated using an identical human-adjudication procedure, the ChatGPT and Copilot syntheses converged on different numbers of themes (12 and 13, respectively) and organized them according to different structural logics. This divergence is itself an informative finding rather than a limitation to be minimized. The ChatGPT framework is organized causally, tracing a pathway from educational functions and risks, through learning processes and student experience, to pedagogical and institutional consequences. The Copilot framework, by contrast, is organized categorically, grouping themes into four equal, interconnected domains (Learning, Pedagogy/Governance, Operations, and Inclusion/Industry) without positioning a directional sequence between them. This suggests that the two tools did not simply disagree about which themes were present in the excerpts but applied different underlying organizing principles to conceptually adjacent codes: one prioritizing process and causality, the other prioritizing domain classification. For example, codes related to academic integrity and data privacy were grouped by ChatGPT under a higher-ranking "Educational Risks" category linked causally to learning outcomes, whereas Copilot situated the same underlying codes within a "Pedagogy/Governance" domain alongside training and policy codes, emphasizing institutional responsibility rather than potential risk. This pattern indicates that GenAI-assisted thematic synthesis is sensitive not only to the evidence provided but to each tool's default synthesis exploration, underscoring the value of the human validation and adjudication procedure (Section 10.b) after independently verifying against source excerpts. Thus, relying on a single GenAI tool's output could mistake one tool's organizing logic for an objective feature of the underlying evidence. Future replications using additional GenAI or more advanced or paid tools would help clarify whether this divergence reflects a unique behaviour of these two specific systems or a more general property of GenAI-assisted qualitative synthesis.

Conclusions

After applying various criteria to strengthen conceptions, 154 articles were identified for detailed conceptual analysis. The figures in this paper are tentative and emergent, subject to further refinement through additional studies. It is, however, important to note the high frequency of occurrence of related themes that have thus far emerged from this study. Taken together, the five domains identified across both frameworks: learning, pedagogy, governance, operations, and inclusion/industry, offer higher education instructors and institutions a starting point for structuring their own AI adoption efforts, rather than approaching GenAI integration as a single, undifferentiated decision. For instructors, the findings suggest that instructional value is concentrated in well-scaffolded uses of GenAI (e.g., explanation, drafting support, research assistance) paired explicitly with strategies

that counter cognitive dependence, such as requiring students to critique or verify AI-generated content rather than accept it uncritically. For administrators and institutional leaders, the consistent appearance of governance-related themes (policy development, data privacy, academic integrity) across both frameworks indicates that institutional policy cannot be an afterthought to classroom-level adoption; clear, proactively communicated norms for acceptable AI use appear to be a precondition for, rather than a response to, effective pedagogical integration. For those responsible for operations and infrastructure, the recurrence of themes related to efficiency, access, and technical limitations suggests that equitable release depends as much on infrastructure investment and reliability testing as on pedagogical design. Finally, the persistence of inclusion and industry-readiness themes across both syntheses points to a need for institutions to evaluate GenAI adoption not only by classroom outcomes but by its effects on equitable access and workforce preparedness. These implications should be read as directional, given the conceptual and evolving nature of the AI in higher education; they are intended to guide further empirical validation and institutional discussion rather than to serve as a finalized implementation checklist.

Conflict-of-interest

The authors declare no financial conflicts of interest related to the research, authorship, or publication of this paper.

References

[1] H. Turillo. (Sep, 19). The Future of AI and Metaverse in Education. Available: https://businessabc.net/the-future-of-ai-and-metaverse-ineducation.
[2] S. Elayyan, "The future of education according to the fourth industrial revolution," Journal of Educational Technology and Online Learning, vol. 4, (1), 2021. . DOI: https://doi.org/10.31681/jetol.737193.
[3] H. Abbass, "Editorial: What is Artificial Intelligence?" TAI, vol. 2, (2), pp. 94–95, 2021. . DOI: https://doi.og/10.1109/TAI.2021.3096243.
[4] L. Chen, P. Chen and Z. Lin, "Artificial intelligence in education: A review," Ieee Access, vol. 8, pp. 75264–75278, 2020. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9069875. DOI: https://doi.org/10.1109/ACCESS.2020.2988510.
[5] C. DiCarlo, Building a God: The Ethics of Artificial Intelligence and the Race to Control It. Essex, CT: Prometheus Books, 2025.
[6] B. Mahesh, "Machine learning algorithms: A review," International Journal of Science and Research (IJSR).[Internet], vol. 9, (1), pp. 381–386, 2020. . DOI: https://doi.org/10.21275/ART20203995.
[7] W. Dong, D. Pan and S. Kim, "Exploring the integration of IoT and Generative AI in English language education: Smart tools for personalized learning experiences," Journal of Computational Science, vol. 82, (1), pp. 102397, 2024. DOI: https://doi.org/10.1016/j.jocs.2024.102397.
[8] A. Kumar et al, "Transforming Education Through Iot And AI Opportunities And Challenges," Educational Administration: Theory and Practice, vol. 30, (5), pp. 11610–11622, 2024. . DOI: https://doi.org/10.53555/kuey.v30i5.4982.
[9] A. Barr, Artificial Intelligence: Cognition as Computation. Stanford University, 1982.
[10] R. Luckin and W. Holmes, "Intelligence unleashed: An argument for AI in education," UCL Knowledge Lab, 2016Available: https://discovery.ucl.ac.uk/id/eprint/1475756/.
[11] J. Zhao et al, "Cognitive psychology-based artificial intelligence review," Frontiers in Neuroscience, vol. 16, pp. 1–9, 2022. . DOI: https://doi.org/10.3389/fnins.2022.1024316.
[12] D. N. Cassenti, V. D. Veksler and F. E. Ritter, "Editor's Review and Introduction: Cognition-Inspired Artificial Intelligence," Topics in Cognitive Science, vol. 14, (4), pp. 652–664, 2022. DOI: https://doi.org/10.1111/tops.1262.
[13] D. Hassabis et al, "Neuroscience-inspired artificial intelligence," Neuron, vol. 95, (2), pp. 245–258, 2017. DOI: https://doiorg.roxy.nipissingu.ca/10.1016/j.neuron.2017.06.011.
[14] R. S. Rosenberg, "Artificial intelligence and linguistics: A brief history of a one-way relationship," in Annual Meeting of the Berkeley Linguistics Society, 1975, pp. 379–392.
[15] A. Gocen and F. Aydemir, "Artificial Intelligence in Education and Schools," REM: Research on Education and Media, vol. 12, (1), pp. 13–21, 2020. . DOI: https://doi.org/10.2478/rem-2020-0003.
[16] X. Jia, L. Ren and J. Cai, "Clinical implementation of AI technologies will require interpretable AI models," Med. Phys., vol. 47, (1), pp. 1–4, 2020. . DOI: https://doi.org/10.1002/mp.13891.
[17] W. Wang and K. Siau, "Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda," Journal of Database Management (JDM), vol. 30, (1), pp. 61–79, 2019. . DOI: https://doi.org/10.4018/JDM.2019010104.
[18] H. Snyder, "Literature review as a research methodology: An overview and guidelines," Journal of Business Research, vol. 104, pp. 333–339, 2019. . DOI: https://doi.org/10.1016/j.jbusres.2019.07.039.
[19] I. Celik et al, "The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research," TechTrends, vol. 66, (4), pp. 616–630, 2022. . DOI: https://doi.org/10.1007/s11528-022-00715-y.
[20] K. Charmaz, Constructing Grounded Theory. (2nd ed.) London: Sage Publications, 2014.
[21] E. G. Guba and Y. S. Lincoln, "Competing paradigms in qualitative research," in Handbook of Qualitative Research Anonymous Thousand Oaks: Sage: 1994, pp. 105–117.
[22] D. Walsh and S. Downe, "Meta-synthesis method for qualitative research: A literature review," J. Adv. Nurs., vol. 50, (2), pp. 204–211, 2005. DOI: https://doi.org/10.1111/j.1365-2648.2005.03380.x.
[23] M. Sandelowski, J. Barroso and C. I. Voils, "Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings," Res. Nurs. Health, vol. 30, (1), pp. 99–111, 2007. . DOI: https://doi.org/10.1002/nur.20176.
[24] J. Thomas and A. Harden, "Methods for the thematic synthesis of qualitative research in systematic reviews," BMC Med. Res. Methodol., vol. 8, (1), pp. 45, 2008. . DOI: https://doi.org/10.1186/1471-2288-8-45.
[25] B. L. Paterson et al, Meta-Study of Qualitative Health Research: A Practical Guide to Meta-Analysis and Meta-Synthesis. 2001.
[26] A. Tong et al, "Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ," BMC Med. Res. Methodol., vol. 12, (1), pp. 181, 2012. . DOI: https://doi.org/10.1186/1471-2288-12-181.
[27] Y. S. Lincoln and E. G. Guba, Naturalistic Inquiry. 1985.
[28] S. P. S. Senanayake and D. N. N. J. N. Dean, "Factors Influencing Trustworthiness of Qualitative Research: A Systematic Literature Review," International Journal of Research and Scientific Innovation, vol. 12, (12), pp. 297–319, 2025. DOI: https://doi.org/10.51244/IJRSI.2025.12120028.
[29] J. Hewitt-Taylor, "Use of constant comparative analysis in qualitative research," Nursing Standard, vol. 15, (42), pp. 39–42, 2001. . DOI: https://doi.org/10.7748/ns2001.07.15.42.39.c3052.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Transforming University Teaching through Generative AI: An Exploratory Pilot Study and a Multidisciplinary Scale-Up Design

Marc Simon-Altaba
Faculty of Law, Universitat Internacional de Catalunya (UIC Barcelona), Barcelona, Spain

Article DOI: 10.68286/samyrad2026.327

ABSTRACT
Generative artificial intelligence is transforming higher education, but its educational value depends on whether it supports or displaces the reasoning that courses are intended to develop. The objective of this completed exploratory pilot was to examine how a bounded, verification-based integration of GenAI operated in first-year Business Law teaching, focusing on students' perceptions, recurrent limitations in AI-generated outputs, and students' capacity to identify and correct those limitations. The pilot was implemented at UIC Barcelona in two first-year Business Administration groups, one in Spanish and one in English. Case-based activities combined independent work, bounded AI assistance, mandatory verification, and justification of what students accepted, corrected, or rejected. Evidence comprised 65 anonymous end-of-course questionnaires, documented instructor observations, continuous-assessment activities, and an instructor activity log. Students evaluated the experience positively and reported gains in critical AI use but did not perceive an automatic performance advantage and strongly preferred an occasional or regular but supervised role for AI (94.0% and 87.5%). Documented classroom episodes showed three recurrent limitations: generic and uniform outputs in AI-assisted trademark creation, formally plausible but insufficiently specific franchise contracts, and occasional reliance on outdated legislation. These findings support a model of AI as a contestable "third voice": a source of propositions that must be tested against disciplinary criteria rather than treated as an authority or substitute for the student's final intellectual work. The VI Exea Impact Award 2026 has enabled a subsequent multidisciplinary and international scale-up involving UIC Barcelona, Universitat de Barcelona, Universitat de Girona, and United Arab Emirates University. No outcome data from that phase was available at submission. Its purpose is to test the model's transferability through a two-cycle mixed design and a common analytical-reasoning rubric. The present paper therefore reports pilot evidence, and a candidate model whose multidisciplinary and international transferability remains to be tested.

KEYWORDS Generative AI; Teaching innovation; Higher education; Case-based learning; Analytical reasoning

How to cite: Simon-Altaba, M. (2026). Transforming university teaching through generative AI: An exploratory pilot study and a multidisciplinary scale-up design. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.327

Introduction

Generative artificial intelligence (GenAI) has moved rapidly from an optional digital aid to an ordinary component of students' academic environment. The resulting challenge is not simply whether universities should permit its use. It is how teaching and assessment should be designed when students can obtain fluent, plausible, and complete responses within seconds. International guidance has therefore emphasized human agency, institutional validation, and data protection,

while the research literature identifies both opportunities for feedback and engagement and risks involving inaccuracy, academic integrity, dependency and reduced epistemic effort [1]–[3]. This challenge also has an inclusion dimension: access to the technology is insufficient if universities do not develop students' capacity to understand and use it critically [4].

The evidence is also more conditional than early claims of transformation sometimes suggest. A recent analysis shows positive average effects on academic performance, motivation, and higher-order thinking, but also identifies methodological weaknesses and the need to distinguish the quality of an AI-assisted product from traditional learning [5]. A 2026 systematic review similarly concludes that critical and creative thinking are most likely to benefit when GenAI is embedded in inquiry-oriented tasks with reflective prompts, rubrics, and verification requirements; unstructured use is more often associated with cognitive offloading [6]. Student perception studies show broad interest in GenAI but persistent concerns about reliability, privacy, and over-reliance [7]. These findings support a pedagogical rather than technological question: what division of intellectual labour between student, teacher and AI works best for what a course is intended to achieve?

This question is particularly visible in legal and business education. Legal analysis requires students to identify legally relevant facts, determine applicable rules, distinguish authority from assertion, apply criteria to a concrete case, and justify a conclusion. GenAI may generate an answer that reads convincingly while omitting a requirement, inventing a source, overlooking an exception, or failing to adapt general information to the facts. Its fluency can therefore create an illusion of competence precisely where novice students lack the domain knowledge needed to detect the defect. The same problem extends to economics, communication and the humanities whenever learning outcomes depend on interpretation, evaluation of evidence and reasoned judgment rather than textual production alone. Spanish legal-education scholarship similarly treats GenAI as a transformation of both legal learning and professional practice, reinforcing the need to connect technological experimentation with disciplinary judgment [8].

The primary contribution of this paper is the analysis of a completed exploratory pilot conducted at UIC Barcelona in two first-year Business Administration groups taking Business Law. The empirical evidence consists of 65 valid anonymous survey responses, documented instructor observations, an instructor activity log, and evidence arising from ordinary continuous-assessment activities that compared work completed with and without AI. The pilot examines how bounded AI use operated in the classroom and what recurrent limitations emerged when students were required to verify and correct generated material.

The pilot has subsequently informed a larger multidisciplinary and international project whose development received a decisive institutional boost through the VI Beca Exea Impact 2026. The expanded phase connects Law, Economics, Communication and Humanities at UIC Barcelona, Law at Universitat de Barcelona (UB), Economics at Universitat de Girona (UdG), and Economics at United Arab Emirates University (UAEU). This scale-up is presented here as a research implication and future validation stage, not as an empirical component of the present study. No outcome data from the expanded phase was available at the conference deadline, and the paper does not claim that multidisciplinary or international transferability has already been demonstrated.

Three research questions organize the analysis: RQ1 asks how students perceived the effects of structured GenAI activities on interest and understanding; RQ2 identifies recurrent limitations in AI output during case-based work; and RQ3 examines how a supervised model encouraged verification and responsible use. These questions are addressed through complete pilot evidence. The subsequent multidisciplinary phase is presented further in the manuscript in Section ‘From Pilot To Exea-Supported Multidisciplinary Scale-Up’ as the design for testing whether the candidate model transfers across disciplines, institutions, and national contexts.

Pedagogical and governance framework

From Output Generation to Reasoning Support

The model begins from constructive alignment: AI use is justified only when the activity still requires the learner to perform the reasoning expressed in the learning outcomes. Case and problem-based approaches are suitable because they organize learning around interpretation, hypothesis formation, application of knowledge and justified decision-making [9], [10]. Formative feedback is educationally valuable when it helps students compare current performance with a standard and regulate their next action, rather than merely supplying a finished response [11]. Evidence from law-school assessment likewise illustrates that apparent task competence does not make the system a dependable source of legal judgment [12]. GenAI can contribute to this process, but only if its role is deliberately limited.

The basic activity sequence is therefore: an independent or minimally assisted first attempt; a controlled AI intervention; verification of the output against course materials, authoritative sources, or an agreed rubric; correction and counterargument; and a final position for which the student remains responsible. Within that sequence, AI may generate questions, challenge an assumption, propose an alternative scenario, audit coherence, or provide a draft for criticism. As we will see, it must not operate as the sole authority, the undisclosed author of the final work or the final assessor of student performance. This division of labour treats GenAI as a contestable cognitive artifact rather than an answer service.

The distinction matters because assistance during a task is not equivalent to learning. Experimental evidence indicates that students can rely on AI rather than internalize the supported process [13]. A large field experiment in mathematics found that unrestricted assistance improved

practice performance but could impair later unaided performance; safeguards that withheld direct solutions and provided teacher-designed hints mitigated the negative effect [14]. Although that setting differs from university legal education, the mechanism is directly relevant: when the tool removes the need to reason, the quality of the immediate product may rise while the learner's capacity remains unchanged or weakens. The present model therefore evaluates not only what students produce with AI, but also how they interrogate it and what they can justify independently.

Integrity, Human Oversight and Data Protection

A supervised model requires clear ex ante rules. Each task should specify whether AI is prohibited, optional or required; which functions are permitted; what use must be declared; and what evidence of the process must be retained. This is more informative than a binary institution-wide ban because the acceptable degree of AI involvement depends on the learning outcome. An AI assessment scale offers a related approach by linking levels of permitted use to assessment design and transparency [15]. Academic integrity then concerns truthful attribution, preservation of the student's intellectual contribution and accountability for submitted claims, not merely whether a tool was opened [16], [17]. This allocation of responsibility is also consistent with broader community-wide ethical frameworks for AI in education [18].

Human oversight is equally structural. The European Union's AI Act identifies educational uses that may affect access, evaluation or monitoring as especially sensitive and requires safeguards appropriate to the risk; its broader human-oversight logic includes awareness of automation bias [19]. The European Commission's updated 2026 guidelines likewise connect ethical AI use in education with critical AI literacy, the AI Act and the General Data Protection Regulation [20]. In the present model, no AI system makes final grading decisions, and instructors remain responsible for the pedagogical level, legal accuracy and fairness of materials and assessment.

Data protection is implemented through minimization and purpose limitation. Students are instructed not to enter identifiable personal data, confidential material, or sensitive information into external systems. Activities are designed so that fictitious or sufficiently abstracted cases can achieve the learning objective. This approach is consistent with Spanish and EU data-protection guidance emphasizing compliance from the design stage when AI is incorporated into processing operations [21]. Proportionality also has a sustainability dimension: the project favours purposeful, bounded interactions over indiscriminate generation and treats computational use as justified by a defined pedagogical function.

Completed pilot: method

Context and Design

The completed phase was an exploratory classroom intervention in an introductory Business Law subject within the first year of the Bachelor's Degree in Business Administration at UIC Barcelona. The intervention was implemented in two parallel groups, one taught in Spanish and one in English. This context was pedagogically demanding, as students were acquiring foundational legal concepts while learning to apply them to business situations, and therefore faced a heightened risk of accepting plausible but incomplete AI explanations. The study combined descriptive survey evidence with documented instructor observation, an instructor activity log, and evidence arising from ordinary continuous-assessment activities. It was not designed as a randomized or quasi-experimental test of causal impact. The groups were defined by language of instruction, not by assignment to an AI treatment, and the available performance comparisons were embedded in ordinary teaching rather than standardized as equivalent outcome measures. The completed pilot is therefore characterized cautiously; the prospective Exea-supported project, rather than this phase, adopts the mixed quasi-experimental design described later.

Intervention Activities and Safeguards

The first activity type used structured comparison in practical cases. Students alternated between solving selected problems with and without AI under specified conditions. They then compared the solutions and identified omissions, unsupported assumptions, excessive generality, and conceptual errors. The exercise did not reward the mere use of AI. It assessed whether students could determine which parts of an output were legally relevant and explain why a proposed answer should be accepted, corrected, or rejected.

A second activity addressed industrial property. Students used AI to propose elements of a trademark, such as a name or a logo, and then applied legal criteria concerning distinctiveness, descriptiveness and registrability. The educational value lay in separating generation from evaluation: an apparently creative proposal still required analysis of whether it could perform the legal function of a trademark and what obstacles might arise.

A third activity concerned commercial contracting. Students first worked on the structure of a contract, including a franchise-type example, and later compared their work with an AI-generated draft. The comparison revealed a recurrent pattern: AI often produced text with a credible formal appearance but insufficient adaptation to the transaction, missing clauses, weak treatment of risk or generic formulations. Students had to identify these deficiencies and justify revisions rather than submit the generated draft as a finished instrument.

The intervention used several AI and AI-assisted tools according to the pedagogical purpose of each activity. ChatGPT Pro and Perplexity supported the creation and variation of practical cases and assessment questions.

Kahoot Plus supported AI-assisted group assessment games, while Gamma Plus was used to review and improve classroom presentation materials. Various web-based generative-design tools were used in industrial-property activities to create logos, trademarks, and related visual elements. The study did not compare platforms as technological products; each tool was treated as a pedagogical instrument subject to the same verification requirements.

The same toolset also supported the professor in producing preliminary quiz questions, interactive revision items, and variants of case statements. Every item was reviewed for accuracy, level, ambiguity, and alignment with the syllabus. This teacher-facing use accelerated drafting but did not transfer pedagogical judgment to the system; many proposed questions were too easy, overly general, or ambiguous and required substantial calibration.

Three safeguards applied across activities. First, students had to verify all AI outputs against course materials and subjects explained in class. Second, the final answer required a reasoned account of what was retained, changed, or discarded. Third, some tasks restricted AI use so that students preserved opportunities for independent performance and could compare different working processes. No automated grading was used, and the AI output was never treated as an authoritative legal source.

Participants, Data and Analysis

An anonymous end-of-course questionnaire produced 65 valid responses: 33 from the Spanish-language group and 32 from the English-language group. The questionnaire addressed prior frequency of AI use, overall evaluation of the experience, perceived effects on interest and understanding, comparison of practical work with and without AI, development of critical-use habits, detection of errors or incomplete responses, and the preferred role of AI in teaching.

During implementation, practical activities formed part of the ordinary continuous-assessment process. The professor considered not only the substantive final answer but also students' capacity to detect gaps, omissions, excessive generality, or unsupported reasoning in AI-generated material and to justify the corrections they proposed. In parallel, the instructor maintained an activity log recording the exercises implemented, recurrent characteristics of AI-generated outputs, classroom comparisons, and the principal difficulties observed. These records were used to document recurring patterns and select illustrative classroom episodes; they were not treated as an independently coded qualitative dataset.

Analysis was descriptive. The questionnaire was administered through Moodle, whose reporting functionality automatically generated the response counts and percentages used in the analysis; no separate statistical package was used. Percentages were examined within each language group and are reported without inferential comparison. Because the questionnaire measured perceptions and self-reported learning, it cannot establish objective gains in analytical reasoning. Instructor observations, the activity log, and continuous-assessment evidence were used descriptively to identify recurring patterns and illustrative episodes, but no independent coding or inter-rater reliability procedure was applied. No directly identifying information was collected for the analysis, and students were instructed not to include personal or confidential information in external AI tools.

Results

Survey and Descriptive Findings

The first finding is that AI was already normalized in students' academic practice. In the Spanish-language group, 48.5% reported using it fairly frequently and 33.3% habitually; the corresponding figures in the English-language group were 34.4% and 43.8% (Table 1). Thus, 81.8% and 78.2% respectively fell within the two highest frequency categories. The pedagogical problem was not how to introduce an unfamiliar technology, but how to convert existing use into a more accountable and evaluative practice.

Table 1. Selected student survey findings

Indicator	Spanish Participants=33	English Participants=32
Prior use: fairly frequent or habitual	81.8%	78.2%
Experience: positive or very positive	75.8%	81.3%
Course more interesting: "quite"	60.6%	50.0%
Understanding improved: quite or much	78.8%	68.8%
AI work: results "similar"	48.5%	—
AI work: "depends on activity"	—	46.9%
Critical AI use learned: quite or much	72.7%	71.9%
Errors found: quite often or most times	81.8%	56.3%
Prefer occasional or regular supervised role	94.0%	87.5%

The overall evaluation was favourable. In the Spanish group, 69.7% described the experience as positive and 6.1% as very positive; in the English group, the proportions were 59.4% and 21.9%. Students also reported motivational and comprehension benefits. The course was considered "quite" more interesting by 60.6% of the Spanish group and 50.0% of the English group. For perceived understanding, 63.6% of the Spanish group selected "quite" and 15.2% "much"; in the English group, 43.8% selected "quite" and 25.0% "much." These are perception measures, but they answer RQ1 by indicating that the structured activities were broadly

accepted and associated with greater interest and perceived comprehension.

The data do not support an automatic performance advantage. Among the Spanish-group participants in the survey, 48.5% considered the results of AI-supported practical work similar and 24.2% considered them worse. In the English group, the most frequent answer was that the outcome depended on the activity (46.9%). Instructor observation was consistent with this qualification: AI was more useful for producing a structure or draft than for applying legal criteria accurately to a specific set of facts. Where conceptual knowledge was weak, access to a fluent draft did not resolve the underlying difficulty.

The strongest result concerned critical use. In the Spanish group, 54.5% stated that they had learned to use AI critically "quite" a lot and 18.2% "much"; in the English group, the figures were 56.3% and 15.6%. Students also encountered errors or incompleteness frequently. In the Spanish group, 63.6% reported this on "quite a few" occasions and 18.2% on most occasions. In the English group, the corresponding proportions were 37.5% and 18.8%. These responses support RQ2 and RQ3: recurrent defects were visible to students, and a substantial majority perceived development in their capacity to question the output.

Students' preferred institutional model was correspondingly cautious. In the Spanish group, 48.5% favoured occasional support and 45.5% regular but supervised use. In the English group, 37.5% chose occasional support and 50.0% regular supervised use. A central role for AI was selected by only 6.1% and 9.4%, respectively. The preference is not for exclusion, but for integration within boundaries that preserve teacher oversight and student responsibility.

Across the practical activities, the instructor activity log recorded recurrent limitations including answers that were correct in the abstract but insufficient for the case; omission of legally relevant requirements; imprecise terminology; failure to distinguish closely related concepts; inadequate adaptation of contract language; and confident statements lacking qualification or current authority. These patterns explain why the tool's most productive classroom role was not to supply the correct answer, but to create an object that students could test against disciplinary criteria and official legal sources.

Illustrative Episodes of AI Output and Student Verification

The activity log and continuous-assessment discussions provide concrete examples of how this verification process operated. These episodes are illustrative rather than an independently coded qualitative dataset, but they show the mechanism underlying the broader patterns reported above.

In the industrial-property exercise, students were asked to create commercial brands for businesses specified by the professor using AI-assisted design tools. The generated proposals were technically usable but generally modest: many were uniform, generic, and insufficiently differentiated from one another. During the subsequent class comparison, students themselves identified this pattern. The exercise therefore shifted from asking whether AI could generate a brand to asking whether the proposal displayed sufficient distinctiveness and contextual fit to function persuasively as a commercial identity.

In the franchise-contract activity, AI-assisted drafts were again formally plausible and structurally correct, but their clauses remained general and lacked depth in relation to the concrete transaction. During classroom comparison, students identified that important provisions required greater specificity, including clauses dealing with indemnification in the event of breach. The problem was not the absence of a contract-like document, but the limited capacity of a generic draft to translate formal structure into sufficiently concrete rights, obligations, and consequences. Students' ability to identify such gaps and justify corrections formed part of the continuous-assessment process.

A third episode concerned temporal legal accuracy. In some practical exercises, classroom review detected that AI-generated answers relied on superseded versions of the applicable legislation. Verification of current course materials and official legal sources was therefore necessary before the reasoning could be accepted. This problem is particularly significant in legal education because fluent explanation can conceal obsolete authority, and it directly supports the model's rule that AI output must never be treated as a legal source in itself. Taken together, the three episodes illustrate different forms of contestability: students had to differentiate a generated proposal, test the substantive adequacy of an apparently complete document, and verify the currency of the authority supporting an answer. Their value lies not in demonstrating causal improvement in critical thinking, which the design cannot establish, but in showing how the classroom tasks made verification and correction observable components of student work.

Discussion

AI as a Contestable "Third Voice"

The completed pilot supports a model in which AI functions as a "third voice" in the classroom: neither the teacher, nor the student, nor an authoritative source, but an additional position that must earn acceptance through verification. Its educational value arises from contestability. A draft that is incomplete, overconfident, generic, or outdated becomes useful when the task requires students to locate the defect, reconstruct the reasoning, and defend a correction. The same output becomes educationally weak when it is copied because it appears plausible. The trademark, franchise-contract, and outdated-legislation episodes illustrate three forms of contestability: differentiating a generated proposal, testing the substantive adequacy of an apparently complete document, and verifying the temporal validity of the authority supporting an answer.

This interpretation reconciles the favourable student experience with the absence of a perceived automatic performance advantage. Motivation and perceived understanding may improve because the tool introduces variety, immediacy, and alternative formulations. Yet better assisted production does not itself demonstrate stronger unaided reasoning. That distinction is central in recent evidence reviews [5], [6] and in methodological critiques of studies that do not define the treatment, control condition, and learning outcome precisely [22]. The present pilot therefore treats the neutral or task-dependent performance result as substantively informative, not as a failure of innovation.

Guardrails Preserve the Learning Task

The verification and justification rules appear to be the mechanism most consistent with the reported development of critical habits. They prevent the students' contribution from collapsing into prompt formulation and blind acceptance of outputs. This aligns with evidence that unrestricted assistance can become a cognitive crutch, whereas carefully designed constraints can preserve engagement [14]. It also responds to systematic concerns that over-reliance may weaken decision-making and analytical judgment [23]. The relevant unit of design is thus not "use of ChatGPT" in the abstract, but a complete instructional sequence specifying what the student must do before, during and after consultation.

For assessment, this implies that authenticity should be evidenced through process as well as product. Suitable measures include a short AI-use declaration, retained prompts or versions where proportionate, an annotated verification note, identification of rejected suggestions, oral defence of a key decision and an unaided component. These measures should not become surveillance by default; they should be selected only where they validly reveal the learning outcome. Long-term assessment design should connect present tasks with the independent judgment that graduates will need beyond their studies [24].

The Professor Remains Central

The pilot also shows that GenAI changes rather than eliminates academic work. The instructor spends less time producing a first draft of a quiz or scenario, but more time calibrating difficulty, checking accuracy, anticipating misuse, and designing the comparison that will make the output pedagogically productive. Therefore, expertise from professors remains indispensable because only a subject specialist can judge whether a coherent response captures the doctrinal requirements, evidential standards, and level appropriate to the course.

This shift has institutional implications. Developers should focus not only on operating tools, but on task architecture, verification protocols, data protection, and assessment validity. Policy should allow differentiated uses rather than assuming that one permission rule fits every subject or learning outcome [17]. Clear communication also reduces inequity: students should not have to infer from informal practice whether AI is acceptable, which functions are allowed or how use must be acknowledged.

Transferability and Disciplinary Difference

The pilot's underlying reasoning cycle is potentially transferable because it is not tied to a single legal doctrine: identify relevant information; distinguish evidence from assumption; apply a disciplinary standard; test alternatives; justify a conclusion; and state uncertainty. However, transferability does not mean uniformity. Law emphasizes authority, qualification, and application to facts; economics may emphasize model assumptions, causal interpretation, and sensitivity to data; communication may emphasize audience, framing and source credibility; and the humanities may emphasize interpretation, contextualization, and competing readings. A future common rubric must therefore capture general reasoning dimensions while allowing discipline-specific indicators and examples. This is also an employability issue: technological systems are expected to restructure professional work, making expert judgment, verification, and accountability more rather than less important [25].

On top of that, the international extension adds further variables: language, local regulation, educational conventions, and differences in students' prior exposure to GenAI. For instance, UAEU's participation can test whether the model operates across a different institutional and national setting, but it also requires equivalence procedures for instruments and careful interpretation of contextual effects. The completed pilot justifies testing the model, yet it does not establish that it will transfer unchanged.

From pilot to Exea-supported multidisciplinary scale-up

The completed pilot subsequently became the basis for a broader research and teaching-innovation program. The VI Beca Exea Impact 2026 marked an inflection point by supporting the development of common instruments and a network extending beyond the original Business Law classroom. The current scope includes Law, Economics, Communication and Humanities at UIC Barcelona; Law at UB; Economics at UdG; and Economics at UAEU. This configuration creates three forms of transfer to be tested: across disciplines within one institution, across universities within Spain, and across national and educational contexts through UAEU.

The next phase is designed to test rather than assume transferability. Over two academic years, a first formative cycle will adapt cases, prompts, declarations and rubric descriptors to selected subjects and refine implementation procedures; a second cycle will use the revised materials to generate more comparable evidence. The shared core will preserve bounded AI roles, explicit permitted uses, verification, and a final accountable student judgment. The principal planned outcome is an analytical reasoning index based on argumentative structure, integration of evidence,

treatment of counterarguments, justificatory coherence, and explicit handling of uncertainty, with discipline-specific descriptors where necessary.

The mixed evaluation will combine objective performance tasks with pre- and post-questionnaires, declared AI-use records, student focus groups, staff interviews and, where feasible and proportionate, aggregated learning-platform data. Selected unaided components are intended to distinguish assisted output from retained capability, while qualitative evidence will examine how students decide whether an AI output is trustworthy and how teachers calibrate the tool's role. Any use of learning-platform data will be governed by proportionality, privacy, transparency and ethical review [26]. No outcome data from this expanded phase is available in the present paper. The pilot has therefore generated a structured candidate model and testable hypotheses for multidisciplinary and international evaluation, not evidence that such transfer has already succeeded.

Limitations

The completed study has several limitations. The sample was limited to 65 respondents from two groups in one first-year subject at one institution. Participation reflected ordinary course conditions, and the questionnaire captured perceptions rather than a validated measure of learning. The language groups were not randomized treatment groups and were not compared inferentially. The activity log provides documented instructor observations, but these were not independently coded or subjected to inter-rater analysis. Continuous-assessment activities were not designed as standardized outcome measures and are used here to illustrate classroom mechanisms, not to demonstrate gains. The study did not include standardized pre- and post-tests, a control condition, or independently assessed performance measures. Consequently, causal claims about motivation, understanding, critical thinking, or analytical reasoning would be unwarranted.

The intervention also combined several activities and uses of AI, so the descriptive findings cannot isolate which component produced a particular perception or classroom response. Tool behaviour may change with model versions, language, and prompts, limiting technological reproducibility. The three classroom episodes reported in the Results Section are documented illustrations rather than a formally coded qualitative sample. Finally, the expanded project is prospective. Its broader institutional composition increases relevance but also introduces disciplinary, linguistic, institutional, and national variability that the common protocol must manage. These limitations explain the scale-up's emphasis on defined treatments, comparable measures, objective reasoning outcomes, and two cycles of refinement.

Conclusion

The pilot indicates that GenAI can contribute to university teaching without becoming a substitute for student reasoning, but the benefit is conditional on design. Students evaluated the experience positively and reported greater interest, perceived understanding, and critical-use habits. At the same time, they did not perceive an automatic advantage in practical results and overwhelmingly preferred an auxiliary, supervised role for AI. The documented classroom episodes add practical detail to those perceptions: AI-assisted brands tended toward uniformity and generality, franchise-contract drafts were formally plausible but insufficiently specific, and some practical answers relied on outdated legislation. The most defensible interpretation is therefore not that GenAI produces better learning by itself, but that it can create productive occasions for verification, correction, and justified decision-making.

The contestable "third voice" model captures this contribution. AI proposes; the student evaluates; pre-established disciplinary sources and criteria determine whether the proposal is adequate; and the professor designs, documents, and supervises the process. This architecture preserves human accountability and makes the imperfections of the tool part of the learning task rather than a hidden risk. The current pilot provides descriptive evidence of how that mechanism operated, while its limitations preclude claims of causal improvement in analytical reasoning.

The VI Exea Impact Award 2026 has accelerated the project's evolution from a local legal-education pilot to a multidisciplinary and international evaluation program involving UIC Barcelona, UB, UdG and UAEU. The next phase will test whether the candidate reasoning framework can transfer across Law, Economics, Communication and Humanities and across different institutional and national contexts. Its contribution will depend on maintaining the evidential discipline applied here: separating assisted performance from learning, completed results from expected outcomes, and technological novelty from the pedagogical conditions that make it educationally useful.

Acknowledgment

The author acknowledges the institutional support associated with the VI Beca Exea Impact 2026 and the contribution of the participating teaching teams at UIC Barcelona, Universitat de Barcelona, Universitat de Girona and United Arab Emirates University to the development of the expanded project.

References

[1] UNESCO, Guidance for Generative AI in Education and Research. Paris, France: UNESCO, 2023.

[2] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learn. Individ. Differ., vol. 103, Art. no. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

[3] M. Giannakos et al., "The promise and challenges of generative AI in education," Behav. Inf. Technol., vol. 44, no. 11, pp. 2518–2544, 2025, doi: 10.1080/0144929X.2024.2394886.

[4] P. Gallego Rodríguez, "El derecho constitucional a la educación y la IA generativa: propuestas para mitigar la exclusión digital educativa," Estudios de Deusto, vol. 73, no. 1, pp. 183–212, 2025, doi: 10.18543/ed.3329.

[5] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," Comput. Educ., vol. 227, Art. no. 105224, 2025, doi: 10.1016/j.compedu.2024.105224.
[6] C. Li, H. Cui, and L. S. Hagedorn, "The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes," Comput. Educ.: Artif. Intell., vol. 10, Art. no. 100571, 2026, doi: 10.1016/j.caeai.2026.100571.
[7] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," Int. J. Educ. Technol. Higher Educ., vol. 20, Art. no. 43, 2023, doi: 10.1186/s41239-023-00411-8.
[8] A. Abadías Selma and D. González Uriel, Eds., El impacto de la IA en el aprendizaje y en la práctica del Derecho. Las Rozas, Spain: Aranzadi LA LEY, 2024.
[9] C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?" Educ. Psychol. Rev., vol. 16, pp. 235–266, 2004, doi: 10.1023/B:EDPR.0000034022.16470.f3.
[10] J. Biggs and C. Tang, Teaching for Quality Learning at University, 4th ed. Maidenhead, U.K.: Open Univ. Press, 2011.
[11] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," Stud. Higher Educ., vol. 31, no. 2, pp. 199–218, 2006, doi: 10.1080/03075070600572090.
[12] J. H. Choi, K. E. Hickman, A. B. Monahan, and D. Schwarcz, "ChatGPT goes to law school," J. Legal Educ., vol. 71, no. 3, pp. 387–400, 2022.
[13] A. Darvishi, H. Khosravi, S. Sadiq, D. Gašević, and G. Siemens, "Impact of AI assistance on student agency," Comput. Educ., vol. 210, Art. no. 104967, 2024, doi: 10.1016/j.compedu.2023.104967.
[14] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without guardrails can harm learning: Evidence from high school mathematics," Proc. Natl. Acad. Sci. U.S.A., vol. 122, no. 26, Art. no. e2422633122, 2025, doi: 10.1073/pnas.2422633122.
[15] M. Perkins, L. Furze, J. Roe, and J. MacVaugh, "The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment," J. Univ. Teach. Learn. Pract., vol. 21, no. 6, 2024, doi: 10.53761/q3azde36.
[16] D. R. E. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," Innov. Educ. Teach. Int., vol. 61, no. 2, pp. 228–239, 2024, doi: 10.1080/14703297.2023.2190148.
[17] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," Int. J. Educ. Technol. Higher Educ., vol. 20, Art. no. 38, 2023, doi: 10.1186/s41239-023-00408-3.
[18] W. Holmes et al., "Ethics of AI in education: Towards a community-wide framework," Int. J. Artif. Intell. Educ., vol. 32, pp. 504–526, 2022, doi: 10.1007/s40593-021-00239-1.
[19] European Parliament and Council, "Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," Off. J. Eur. Union, L 2024/1689, Jul. 12, 2024.
[20] European Commission, Guidelines on the Ethical Use of Artificial Intelligence and Data in Teaching and Learning. Brussels, Belgium: European Commission, 2026.
[21] Agencia Española de Protección de Datos, Una aproximación para la adecuación al RGPD de tratamientos que incorporan inteligencia artificial. Madrid, Spain: AEPD, Feb. 2020.
[22] J. Weidlich, D. Gašević, H. Drachsler, and P. Kirschner, "ChatGPT in education: An effect in search of a cause," J. Comput. Assist. Learn., vol. 41, no. 5, Art. no. e70105, 2025, doi: 10.1111/jcal.70105.
[23] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review," Smart Learn. Environ., vol. 11, Art. no. 28, 2024, doi: 10.1186/s40561-024-00316-7.
[24] D. Boud and N. Falchikov, "Aligning assessment with long-term learning," Assess. Eval. Higher Educ., vol. 31, no. 4, pp. 399–413, 2006, doi: 10.1080/02602930600679050.
[25] R. Susskind and D. Susskind, The Future of the Professions: How Technology Will Transform the Work of Human Experts. Oxford, U.K.: Oxford Univ. Press, 2015, doi: 10.1093/oso/9780198713395.001.0001.
[26] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," Am. Behav. Sci., vol. 57, no. 10, pp. 1510–1529, 2013, doi: 10.1177/0002764213479366.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Incorporating Domain Knowledge into Student Risk Prediction: An Ablation Study Using Higher Education Enrollment Records

Sara A. Khan
SaraKhan@uowdubai.ac.ae

Faculty of Engineering and Information Systems, University of Wollongong in Dubai, Dubai, UAE

Article DOI: 10.68286/samyrad2026.332

ABSTRACT
This paper examines how domain knowledge can be used to predict student academic risk using higher-education administrative data. The study uses 37,583 anonymized student enrollment records from Peru and applies an ablation-style experimental architecture. The target variable is modeled using a two-stage approach: (1) risk occurrence and (2) risk severity. Four feature conditions are tested: Raw-only (student profile variables), Cluster-only (unsupervised cluster labels), and Hybrid configurations with or without domain-based weighting. The domain-based weighting scheme scales feature groups using theory-backed weights. Results show that the Raw-only model establishes a credible baseline with an error profile suited for early intervention. Standalone Cluster-only models performed poorly across both stages. Under 5-fold cross-validation, the weighted Hybrid model detected significantly more at-risk students than Raw-only (McNemar p = 0.0045), but not significantly more than the unweighted Hybrid model (p = 0.25). When concluding the severity predictions, a students' risk severity was directly correlated to their number of enrolled courses. Therefore, the severity stage largely recovered course load as the severity prediction. Once applied, weighting reduced this reliance on course load. Overall, theory-driven weighting is a transparent, testable mechanism for incorporating domain knowledge, with a modest effect on at-risk detection. Future work should evaluate severity on data where it varies independently of course load and calibrate decision thresholds against alert fatigue.

KEYWORDS domain knowledge, ablation framework, regression models, student records.

How to cite: Khan, S. A. (2026). Incorporating Domain Knowledge into Student Risk Prediction: An Ablation Study Using Higher Education Enrollment Records. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.332

Introduction

Despite constant advancements in higher education, institutions continue to face student dropout and academic failure, creating a need for early-warning methods that can identify at-risk students before this risk develops into disengagement or withdrawal [1].

Educational Data Mining (EDM) allows educators to analyze student records through statistical, machine learning, and data mining methods [2]. Prior work shows that behavioral engagement, attendance, study habits, and administrative or demographic attributes can all contribute to a student's risk profile [3][4][5]. However, predictive pipelines often treat all these variables as equally meaningful and assume that any patterns

discovered in cleaned data are useful without modifications. Such an assumption can obscure contextual factors that matter with individuals embedded in social and institutional environments such as higher education institutions [6][7].

Clustering is frequently introduced to facilitate the grouping of students with similar profiles into categories [8]. Theoretically, clustering helps capture latent structures within data. Practically, when applied in a purely data-driven manner, clustering might lead to the formation of conclusions not completely aligned with established knowledge (i.e., the field of student learning and academic success). Evaluating schedule choice based on chronotype helps illustrate this problem. Despite its relevance and importance in academic performance and students' retention levels, chronotype is rarely incorporated explicitly into predictive models of academic risk, leaving a gap in understanding students in higher education [9] [10]. This makes it possible for the same recorded behavior to have different implications for students operating under different environments.

This study evaluates domain knowledge as a modeling intervention rather than as post-hoc narrative context. It follows the theory-guided data-science principle that established knowledge can be encoded directly within a computational pipeline [11]. Bean's student-attrition framework is used to assign fixed weights to feature groups according to their theoretical proximity to attrition [12]. The contribution is tested through a controlled ablation design in which the dataset, cross-validation folds, preprocessing, estimators, and metrics remain constant while only cluster-derived features and domain-informed scaling are changed. Risk is separated into occurrence, whether at least one course is at risk, and severity, the quantity of at-risk courses while risk is present. This separation is intended to support identification and intervention prioritization.

The paper addresses three main questions:

RQ1: *To what extent can cleaned raw student profile attributes predict academic risk outcomes at the individual level?*

RQ2: *What is the effect of incorporating clustering into academic-risk prediction when compared to prediction using cleaned raw student profile data alone?*

RQ3: *What is the effect of incorporating domain knowledge into clustering (through domain-informed feature weights) on predictive outcomes, compared to purely data-driven clustering and unweighted hybrid feature sets?*

Related work

Educational Data Mining and Domain Knowledge

Student-retention research has increasingly moved from isolated academic explanations toward multi-dimensional accounts of student attrition. Tinto emphasized a focus on academic and social integration [13], whereas Bean emphasized the effect of environmental, organizational, and background determinants in shaping a student's intent to leave [12]. These theories provide a structured basis for interpreting administrative variables such as tuition behavior, study mode, schedule, enrollment, and student background. Demographic and pre-admission features have also been shown to support student-performance prediction; particularly when behavioral logs through learning platforms are unavailable [14].

These demographic and pre-admission factors can further be categorized into varying levels of priority based on the accumulated knowledge of experts (i.e., domain knowledge). The use of domain knowledge is important because high predictive performance does not itself establish explanatory meaning [15]. Theory-guided modeling makes assumptions explicit and allows them to be tested as part of a model pipeline. In the present study, the weights are not optimized to maximize a metric; their use aims to incorporate a literature-grounded hierarchy whose behavioral effect can be compared under controlled conditions. Educational prediction commonly draws from either behavioral activity in higher education learning-platforms or administrative and demographic records. While platform behavior can form a detailed activity log, the findings become limited to what students do inside the observed system [16]. Administrative records provide less behavioral detail yet remain available before and outside a student's platform engagement. This makes them particularly relevant to early screening, provided that their interpretation remains sensitive to institutional conditions and does not reduce demographic associations to causal explanations [11].

Clustering and Ablation-Style Evaluation

Supervised classifiers remain common in EDM, while clustering can often be used to discover student profiles before classification [17]. Hybrid architectures have reported improvements in a model's predictive accuracy when clustering is able to add information which is not already represented within a cleaned dataset [18] [19]. For example, a cluster label derived from the same raw attributes may function mainly as a significantly compressed summary of existing information. Its descriptive usefulness cannot guarantee an improved individual-level prediction.

Breaking down a system into its component parts to test each component in isolation (i.e., ablation) provides a direct way to evaluate the influence of incorporating this distinction. Component-wise ablation has been used in knowledge-tracing models to isolate the effect of embeddings or modules [20], while feature-level ablation quantifies whether raw or derived variables contribute independent information to a prediction model [21]. This is particularly relevant when clustering might result in cluster groups that do not align with the domain knowledge developed in a scientific field [22]. The present design therefore treats raw attributes, cluster labels, their combination, and theory-driven weighting as separate, testable conditions to identify the effect of each condition on predictive outcomes.

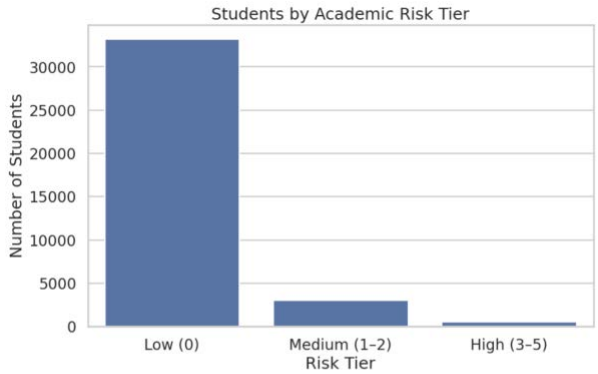
A central focus of this study is in distinguishing between a variable's descriptive and predictive utility. A cluster can be internally coherent and still be weak as a predictor. Conversely, a supervised model can discriminate risk without producing easily isolated student archetypes. Hybrid EDM studies often combine the two functions, but the resulting gain must be attributed

carefully: improvements may arise for various reasons including the utilization of complementary data, different inductive biases, or additional model complexity [17]–[19]. Keeping the preprocessing and evaluation conditions fixed allows this study to focus on the effect of the testable conditions alone. The desire to preserve interpretability motivated the use of linear occurrence and count models. While more complex architectures would have allowed this study to capture nonlinear interactions, they would have limited the ability to determine an exact source of the deviations in model performance [23]. Logistic and Poisson regression provide probability and count outputs that can be inspected directly, while the ablation conditions expose how the same evaluation metrics respond to models with raw, clustered, and theory-weighted datasets.

Methodology

Dataset and Preprocessing

The source dataset contains 37,583 anonymized enrollment records from Peru for 2023 [24]; as a publicly accessible, anonymized, secondary dataset, a formal ethics approval was not required. Each record includes mixed categorical and numeric attributes. The outcome (i.e., the number of at-risk courses) is a discrete count from 0 to 5. Two targets were derived: risk occurrence (count ≥ 1) and risk severity (the count, modeled only for records with observed risk). A reporting variable grouped counts into low (0), medium (1–2), and high (3–5) risk tiers. For all 3,597 at-risk records, the severity variable count equaled the number of enrolled courses, so severity was determined by course load in this dataset. Columns with more than 50% missingness (i.e., type of educational institution, institution status, and educational institution name) were removed. Listwise deletion was used to remove 736 records. The resulting dataset contained 36,847 records and 17 columns. Gender labels were normalized from a mix of numeric and categorical coding. The final class distribution was strongly imbalanced with 33,187 records in the low tier, 3,056 in the medium tier, and 541 in the high tier, as shown in Fig. 1.



risk_tier	count
Low (0)	33187
Medium (1–2)	3056
High (3–5)	541

Fig. 1. Distribution of students across academic risk tiers after cleaning.

Performance metrics represent 5-fold stratified cross-validation means (stratified by risk occurrence, random seed 42), applied through shared indices so that every ablation condition used identical records. To ensure differences in at-risk detection between models exceeded what would be expected by chance, McNemar's exact test was applied to the out-of-fold predictions. A stratified 70/30 split was kept for Figs. 2–4. Numeric features were passed through, while categorical features were one-hot encoded with unknown categories ignored. Preprocessing was implemented within scikit-learn Pipeline and Column Transformer objects fitted on each training fold. To maintain identical transformations and prevent leakage in encoding, the cluster label was assigned from predictors before splitting.

Clustering and Ablation Conditions

Clustering was used only as a feature-construction step; the outcome was excluded. Hierarchical clustering required a full pairwise Gower distance matrix, whose memory grows with the square of the number of records. This would have required about 5.4 GB for all 36,847 records. Running it on the full dataset exceeded the memory available in a standard Google Colab session. Therefore, it was run on a random sample of 15,000 records (0.9 GB) to select k (k = 3 to 8). Silhouette scores were low (0.294 at k = 3, 0.281 at k = 4, 0.226 at k = 5, 0.1866 at k = 6), indicating weak but non-random structure relative to a random-label baseline, which explains the limited predictive value of the cluster label; k = 4 was retained for granularity. K-prototypes was then used to assign the final four-cluster categorical label to all records. It agreed poorly with the hierarchical partition (adjusted Rand index = 0.235) but was stable between sample and full-data fits (0.700). Thus, the label could act as a plausible partition rather than for the purpose of forming distinct segments.

Four conditions were evaluated: (1) Raw-only, using enrollment, gender, program major, shift schedule, study mode, department, number of enrolled courses, tuition-payment indicators for March 2022 and March 2023, and benefit discounts; (2) Cluster-only, using the cluster label alone; (3) Hybrid unweighted, combining raw attributes with the cluster label; and (4) Hybrid weighted, applying theory-informed scaling to the encoded Hybrid matrix before fitting.

Bean-Based Feature Weighting

The weighting intervention grouped variables into four pillars derived from Bean's attrition model [12]. Environmental variables (i.e., tuition-payment indicators and benefit discounts) and Academic variables (i.e., enrollment and number of enrolled courses) received a weight of 0.50. Organizational variables—shift schedule and study mode—received a 0.30. Background variables—gender, department, and program major—received a 0.20. One-hot expanded columns inherited the weight of their source feature. A sparse-safe Function Transformer applied the scale factors after encoding and before model fitting. The 'sensitivity variant' values represent a theory-driven design choice rather than a parameter set optimized for the best metrics. Two alternatives breaking the Environmental/Academic tie (Table 1) tested sensitivity to this choice.

Table 1. Domain-informed feature weights

Pillar (weight)	Mapped variables	Sensitivity variant
Environmental (0.50)	Tuition payment 2022/2023; benefit discounts	0.60 / 0.40
Academic (0.50)	Enrollment; number of enrolled courses	0.40 / 0.60
Organizational (0.30)	Shift schedule; study mode	0.30
Background (0.20)	Gender; department; program major	0.20

Predictive Models and Evaluation

Stage 1 modeled risk occurrence with logistic regression. The estimator used a maximum of 2,000 iterations and balanced class weights so errors on the minority at-risk class contributed more strongly to training. Models were evaluated using accuracy, at-risk precision, at-risk recall, F1, confusion matrices, and receiver operating characteristic—area under curve (ROC-AUC). Because the use case prioritizes early intervention, recall and false negatives were prioritized over accuracy alone.

Stage 2 modeled risk severity with Poisson regression (2,000 iterations, L2 alpha = 0.1) among students with observed risk; since severity equals course load here, it tests how closely each configuration recovers this relationship. MAE and RMSE quantified continuous error, the Pearson χ^2 statistic divided by residual degrees of freedom (χ^2/df) checked the Poisson mean–variance assumption (above 1 indicates overdispersion, below 1 under dispersion), and rounded predictions were checked for exact and ±1/±2 agreement. The stages were combined as expected at-risk courses = P(risk) × predicted severity, with severity predicted for every student.

As nonlinear reference points outside the ablation, Random Forest, XGBoost, and CatBoost were trained on the Raw-only features with the same folds, using class-balanced weighting and, for XGBoost and CatBoost severity models, a Poisson objective.

All analysis was executed in Python within Google Colab using pandas and NumPy for data handling, matplotlib and seaborn for visualization, scikit-learn for preprocessing and modeling, XGBoost and CatBoost for gradient-boosted benchmarks, stats models for McNemar's test, and SciPy for sparse matrix operations. The same fitted preprocessing objects were reused within each pipeline during testing. This was important for the weighted condition because one-hot expansion changes the number and order of columns. Therefore, weights were mapped to encoded columns only after the training preprocessor had established the feature space.

To provide a holistic view of results, the evaluation incorporates multiple metrics into predicting both the likeliness (i.e., risk classification) and severity of risk. For classification, ROC-AUC checks the accuracy of the model in differentiating between safe and at-risk students, while recall and precision show the quality and probability of identifying at-risk students. For severity, MAE and RMSE look at continuous instead of rounded values, while agreement metrics help confirm the practical reliability of predicted scores for advisors in higher education.

Results

Risk Occurrence

Table 2 summarizes the occurrence results as means across the five cross-validation folds. The Raw-only model established a credible baseline with ROC-AUC = 0.7631 and recall = 0.7067. Given the severe class imbalance in the dataset (approximately 9 safe students for every at-risk student), the observed precision of 0.2010 represents an enrichment over random chance. While false positives are elevated due to the low base rate (10%), the model successfully captures 70.7% of true at-risk students. This trade-off may be appropriate for a screening layer in which a supportive review follows the automated flag.

Table 2. Risk OCCURRENCE RESULTS (5-FOLD MEANS; SD ≤ 0.018)

Model	Acc.	Recall	Prec.	AUC
Raw	0.6966	0.7067	0.2010	0.7631
Cluster	0.8122	0.3228	0.2061	0.6298
Hybrid-U	0.6961	0.7125	0.2017	0.7631
Hybrid-W	0.6948	0.7162	0.2016	0.7635

Cluster-only achieved the highest accuracy (0.8122) but the weakest intervention profile. Recall fell to 0.3228 and ROC-AUC to 0.6298, showing that a broad cluster label could not preserve enough individual-level information to distinguish which students inside a cluster were at risk. The accuracy increase is therefore attributable largely to the majority no-risk class rather than to improved risk detection.

The confusion matrices from the illustrative 70/30 split make the difference concrete. Raw-only identified 756 at-risk students and missed 323, while generating 3,074 false positives and 6,883 true negatives. Cluster-only identified only 346 at-risk students and missed 733, although its majority-class predictions produced 8,594 true negatives and only 1,363 false positives. Fig. 2 serves as a visual representation of the aforementioned values. The accuracy advantage appears to come from a more conservative alert policy which substantially increased the number of struggling students left undetected.

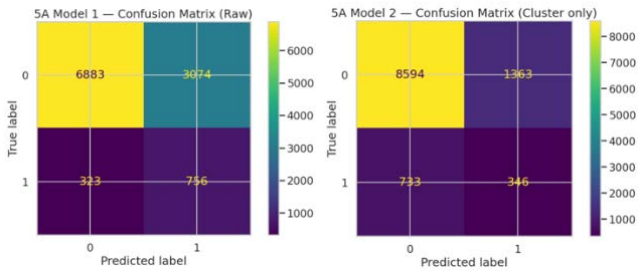


Fig. 2. Confusion matrices for the Raw and Cluster-only occurrence model.

Adding the cluster label to the raw attributes produced almost no change: the unweighted Hybrid model reached ROC-AUC =

0.7631 and recall = 0.7125. The weighted Hybrid model, by contrast, shifted the decision profile slightly further toward higher sensitivity. Recall increased to 0.7162 while ROC-AUC remained essentially unchanged at 0.7635. Accuracy decreased slightly to 0.6948, consistent with the model accepting more false positives to reduce false negatives. Across the out-of-fold predictions, the weighted Hybrid model detected 34 more at-risk students than Raw-only (McNemar p = 0.0045) but only 13 more than the unweighted Hybrid model, which was not statistically significant (p = 0.2503). On the illustrative 70/30 split, Fig. 3 shows that weighting changed the Hybrid confusion matrix from 326 false negatives and 753 true positives to 313 false negatives and 766 true positives.

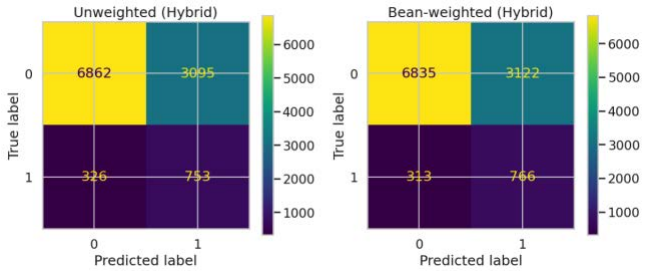


Fig. 3. Confusion matrices for the unweighted and Bean-weighted Hybrid occurrence models.

Risk Severity

Table 3 reports severity error (5-fold means) among at-risk students, measured as MAE and RMSE between the predicted and actual number of at-risk courses. Because this number equals each student's course load, lower error indicates how closely a configuration reproduces the number of enrolled courses rather than an independent measure of risk. Raw-only and Hybrid unweighted were almost identical (MAE 0.1234 vs. 0.1232; RMSE 0.1583 vs. 0.1585), so the cluster label added little beyond the profile variables from which it was derived.

Table 3. Risk severity results (5-fold means; sd ≤ 0.035)

Model	MAE	RMSE
Raw	0.1234	0.1583
Cluster	0.6051	0.7275
Hybrid-U	0.1232	0.1585
Hybrid-W	0.2081	0.2304

Cluster-only performed worst, with MAE = 0.6051 and RMSE = 0.7275. Its predictions effectively approximated an average severity for each cluster and could not represent variations within clusters, because the cluster label does not contain course load. The weighted Hybrid model showed a more complex pattern. Continuous error increased to MAE = 0.2081 and RMSE = 0.2304, consistent with the weighting reducing the influence of the course-load variable, yet every rounded prediction fell within ±1 course in all five folds. On the 70/30 split, approximately 97.3% of rounded predictions exactly matched the observed integer count, and train and test MAE were nearly identical (0.208 and 0.209), which does not indicate obvious overfitting.

Weighting also redistributed severity estimates rather than uniformly increasing them. On the 70/30 split, median predicted severity rose from 1.0938 to 1.2204, while the mean decreased slightly from 1.5604 to 1.5477. The intervention therefore moved central predictions upward while reducing some higher-end estimates. This explains why integer-level agreement remained high even though the continuous residuals increased.

The end-to-end expected-risk construction preserves the distinction between risk classification and intensity, offering a record-level summary that institutions can inspect. Because severity equals course load in this dataset, the expected count approximately equals the probability of risk multiplied by the number of enrolled courses, so a student with a heavier course load may warrant earlier review even if both students have the same predicted probability of risk. Retaining the component outputs avoids the loss of information that would occur if the system exposed only a single rounded value of risk. The expected count is therefore best viewed as an ordering aid rather than a replacement for probability, severity, and contextual review.

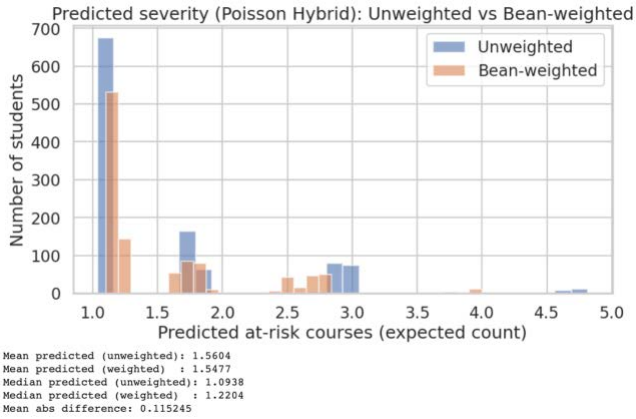


Fig. 4. Distribution of predicted severity for the unweighted and Bean-weighted Hybrid Poisson models.

Weighting Sensitivity and Nonlinear Benchmarks

Breaking the Environmental/Academic tie in either direction left occurrence results unchanged (recall 0.7162–0.7175; ROC-AUC 0.7635 throughout). Severity MAE depended on it (0.1752 with Academic higher, 0.2081 with equal weights, 0.2611 with Environmental higher) since the Academic variables include course load.

The nonlinear benchmarks (Table 4) achieved higher occurrence discrimination than logistic regression (ROC-AUC 0.8096-0.8158 vs 0.7631) and reproduced severity almost exactly (MAE ≤ 0.0009), the latter because tree-based models map course load directly onto the count.

Table 4. Nonlinear benchmarks on raw-only features (5-fold means)

Model	AUC	Severity MAE
Logistic / Poisson	0.7631	0.1234
Random Forest	0.8096	0.0005
XGBoost	0.8149	0.0003
CatBoost	0.8158	0.0009

Discussion

What the Ablation Reveals

Across both stages, raw student attributes carried most of the predictive signal, consistent with Bean's model, whose environmental, organizational, and background determinants map directly onto these attributes [12]. Occurrence discrimination was moderate, and the low severity error reflects severity equaling course load. Cluster labels alone were insufficient for individual prediction, consistent with their weak structure, and added little to the raw dataset. This does not imply that clustering lacks value: the four clusters can still support profiling, communication, or exploratory analysis. A single label is simply too coarse to replace the underlying attributes when deciding which individual student should be flagged.

The result also demonstrates why accuracy is an inadequate headline metric for imbalanced early-warning systems. A model can achieve high accuracy by predicting the majority class and still miss the minority class (i.e., students who are at risk). The Cluster-only occurrence model is the clearest example: its 0.8122 accuracy appears favorable until its 0.3228 recall and 0.6298 ROC-AUC are considered. Ablation makes this failure mode visible because the folds, preprocessing, and evaluation metrics are kept constant.

The near-equivalence of Raw-only and Hybrid unweighted clarifies the role of derived features: a cluster label generated from the same profile matrix summarizes patterns the linear models can already access. Larger gains would require a derived variable capturing external, longitudinal, or theoretically constrained information unavailable in the raw data, so cluster labels are best used as descriptive context unless ablation shows independent predictive value.

Domain Knowledge as a Modeling Intervention

The principal effect of Bean-based weighting was behavioral rather than a uniform gain. In occurrence prediction, the weighted Hybrid model had the highest recall without changing ranking ability and detected significantly more at-risk students than Raw-only ($p = 0.0045$), but its gain over Hybrid unweighted was not significant ($p = 0.2503$), so the effect of weighting alone could not be distinguished from chance. Domain knowledge can nonetheless be encoded as an explicit, reproducible pipeline component [25] [26]. In Bean's terms, the weights emphasize environmental pulls such as tuition payment and benefit discounts over more distal background attributes [12]. The intervention changed the cost profile in a direction aligned with early intervention: more struggling students were detected than by the Raw-only model, at the expense of a small increase in overall at-risk detections.

The severity result requires cautious interpretation. The weighted Poisson model had higher MAE and RMSE than Raw-only and Hybrid unweighted. Since severity equals course load here, this difference reflects how much weight each configuration gives to the enrolled courses rather than predictive insight. Error was lowest when the Academic pillar, which contains course load, was weighted higher. This dataset therefore cannot show whether weighting improves severity estimates; the stage's practical value is combining risk probability with course load to prioritize cases.

Practical Implications

The results support an early-warning architecture in which administrative records provide the core individual-level signal, clustering supports descriptive segmentation, and domain weighting is treated as a policy-sensitive calibration step. The occurrence model should operate to support decision making rather than facilitating automated judgment. Its precision of about 0.20 (four in five flags are false alarms) makes this stage the main source of alert fatigue. This is acceptable because a flag triggers a low-cost supportive review. Thresholds can then be calibrated to the number of advisors or interventions available; allowing the institution to function based on capacity and prevent alert fatigue [27]. A practical deployment could therefore use a tiered workflow: a low threshold for an initial supportive check, a higher threshold for resource-intensive intervention, and the expected count (risk probability × course load) to order cases by priority. Such a workflow would preserve human review and allow thresholds to update when institutional capacity (e.g., available staff, venue, or educational resources) changes.

The role of schedule-related variables also remains contextual. Chronotype is not directly measured in the dataset, and shift schedule remains a proxy for the institutional timing conditions under which students study. Prior work on social jet lag and performance shows why such variables may deserve interpretation beyond their raw statistical coefficient [16]. The present weighting scheme is one transparent way to express that interpretation, but it does not establish causal effects. Model predictions should therefore augment, not replace, human advisor review.

Limitations and future work

The study uses one anonymized 2023 cohort from a single repository, so the results may not generalize across institutions, policies, economic conditions, or academic years. The at-risk course count is an operational proxy for academic struggle and may reflect course difficulty, grading practice, or assessment design. Additionally, it also translates to course load for at-risk students, so partial risk is not captured. The models estimate associations rather than causal drivers of dropout. The approximately 9:1 class imbalance also means that accuracy must be interpreted alongside recall, precision, and ROC-AUC.

Logistic and Poisson regression were chosen for transparency and controlled comparison, not maximum performance. Nonlinear benchmarks reached higher occurrence discrimination (ROC-AUC up to 0.8158), indicating unmodeled interactions. Across all students, the at-risk course count is overdispersed (variance-to-mean ratio ≈ 1.77), driven by the large share of students with no at-risk courses; a single count model fitted to every student would therefore require negative binomial regression [28] or a hurdle or zero-inflated model [29]. The present two-stage design follows the hurdle approach by modeling occurrence separately and severity only among at-risk students. Within the severity stage, however, the divergence between MAE and rounded agreement is not due to overdispersion: Pearson χ^2/df was 0.013-0.034, indicating strong underdispersion, as expected when severity equals course load. Evaluating severity therefore requires data in which it varies independently of course load.

The pillar weights were literature-grounded but not optimized; occurrence results were robust to the Environmental/Academic ratio, whereas severity error was not. Broader alternatives, such as 0.6/0.3/0.1 and 0.5/0.3/0.2, should be tested across multiple cohorts. Decision thresholds should be selected against institutional capacity, and fairness analysis could compare false-negative rates across relevant subgroups. Because demographic variables can improve prediction while also creating ethical and deployment concerns, protected attributes require explicit governance and robustness checks [30]. Additional longitudinal features, including payment stability and schedule consistency, may improve interpretability and generalization.

Conclusion

This study used a two-stage ablation framework to separate risk occurrence from risk severity and to test raw attributes, cluster labels, their combination, and Bean-based feature weighting under controlled conditions. Raw student-profile variables provided a strong baseline. Cluster labels alone performed poorly and added little incremental predictive value when combined with the same attributes from which they were derived. Domain-informed weighting did not produce a universal accuracy gain, but it shifted occurrence prediction toward higher recall while maintaining ROC-AUC, detecting significantly more at-risk students than the Raw-only model, although its gain over the unweighted Hybrid model was not statistically significant. In severity prediction, weighting increased continuous error but retained high rounded agreement; because severity equaled course load for every at-risk student, this stage mainly recovered course load, and the higher error reflects the weighting reducing the influence of that variable. The findings position domain knowledge as a transparent modeling intervention whose effect should be evaluated against the institutional decision objective, rather than assumed to improve every metric. Future work should evaluate severity on data in which it varies independently of course load, compare count models suited to bounded, under dispersed counts, and calibrate support thresholds to balance institutional capacity (e.g., available staff, venue, or educational resources) against alert fatigue.

References

[1] M. O. González-Morales, D. López-Aguilar, P. R. Álvarez-Pérez, and P. A. Toledo-Delgado, "Dropping out of higher education: Analysis of variables that characterise students who interrupt their studies," Acta Psychologica, vol. 252, Art. no. 104669, 2025, doi: 10.1016/j.actpsy.2024.104669.

[2] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans. Syst., Man, Cybern. C, vol. 40, pp. 601–618, 2010, doi: 10.1109/TSMCC.2010.2053532.

[3] A. Almalawi, B. Soh, A. Li, and H. Samra, "Predictive models for educational purposes: A systematic review," Big Data Cogn. Comput., vol. 8, no. 12, 2024, doi: 10.3390/bdcc8120187.

[4] A. Hysaj, G. Farouqa, SA Khan, and L. Hiasat, "Reducing Stress Through Formative Assessments: A Case of the Digital Platform," *Lecture Notes in Computer Science*, pp. 486–500, Jan. 2023, doi: 10.1007/978-3-031-35927-9_33.

[5] G. Farouqa *et al.*, "Undergraduate Students' Journey with AI in the United Arab Emirates," *Lecture Notes in Computer Science*, pp. 155–171, 2025, doi: 10.1007/978-3-031-93539-8_11.

[6] I. Schoon and J. Heckhausen, "Conceptualizing individual agency in the transition from school to work: A social-ecological developmental perspective," Adolescent Res. Rev., vol. 4, 2019, doi: 10.1007/s40894-019-00111-3.

[7] J.-A. Botha and I. Potgieter, "Socio-demographical variables as predictors of academic self-directedness," Int. Rev. Res. Open Distrib. Learn., vol. 25, no. 4, pp. 38–51, 2024, doi: 10.19173/irrodl.v25i4.7516.

[8] Y. Lu, S. Yeom, J. Maktoubian, M. M. Rahman, and S.-H. Kim, "Improve student risk prediction with clustering techniques: A systematic review in education data mining," Educ. Sci., vol. 15, no. 12, 2025, doi: 10.3390/educsci15121695.

[9] T. Enright and R. Refinetti, "Chronotype, class times, and academic achievement of university students," Chronobiol. Int., vol. 34, pp. 1–6, 2017, doi: 10.1080/07420528.2017.1281287.

[10] L. S. Kalbacher and C. Randler, "Exploring the nexus between sleep, chronotype, and non-cognitive predictors in university students: Implications for academic success," Chronobiol. Int., vol. 41, no. 8, pp. 1165–1175, 2024, doi: 10.1080/07420528.2024.2383396.

[11] A. Karpatne et al., "Theory-guided data science: A new paradigm for scientific discovery from data," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2318–2331, 2017, doi: 10.1109/TKDE.2017.2720168.

[12] J. P. Bean, "The synthesis of a theoretical model of student attrition," 1981.

[13] V. Tinto, "Stages of student departure: Reflections on the longitudinal character of student leaving," J. Higher Educ., vol. 59, no. 4, pp. 438–455, 1988, doi: 10.2307/1981920.

[14] M. Bilal, M. Omar, W. Anwar, R. H. Bokhari, and G. S. Choi, "The role of demographic and academic features in a student performance prediction," Sci. Rep., vol. 12, Art. no. 12508, 2022, doi: 10.1038/s41598-022-15880-6.

[15] López-Meneses, E., Mellado-Moreno, P. C., Herrerías, C. G., & Pelícano-Piris, N. (2025). Educational Data Mining and Predictive Modeling in the Age of Artificial Intelligence: An In-Depth Analysis of Research Dynamics. *Computers*, *14*(2). https://doi.org/10.3390/computers14020068.

[16] B. Smarr and A. Schirmer, "3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance," Sci. Rep., vol. 8, 2018, doi: 10.1038/s41598-018-23044-8.

[17] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," IEEE Access, vol. 5, pp. 15991–16005, 2017, doi: 10.1109/ACCESS.2017.2654247.

[18] D. Hooshyar, Y.-M. Huang, and Y. Yang, "A three-layered student learning model for prediction of failure risk in online learning," Human-Centric Comput. Inf. Sci., vol. 12, pp. 357–372, 2022, doi: 10.22967/HCIS.2022.12.028.

[19] S. Fida, N. Masood, N. Tariq, and F. Qayyum, “A novel hybrid ensemble clustering technique for student performance prediction,” J. Universal Comput. Sci., vol. 28, no. 8, pp. 777–798, 2022, doi: 10.3897/jucs.73427.
[20] S. Pandey and G. Karypis, “A self-attentive model for knowledge tracing,” 2019.
[21] I. Lazic, C. Barà, M. Iovino, S. Stramaglia, N. Jakovljevic, and L. Faes, “Information-theoretic quantification of high-order feature effects in classification problems,” arXiv:2507.04362, 2025, doi: 10.48550/arXiv.2507.04362.
[22] A. Jasinska-Piadlo et al., “Data-driven versus a domain-led approach to k-means clustering on an open heart failure dataset,” Int. J. Data Sci. Anal., vol. 15, no. 1, pp. 49–66, 2023, doi: 10.1007/s41060-022-00346-9.
[23] Kim, S., Choi, E., Jun, Y.-K., & Lee, S. (2023). Student Dropout Prediction for University with High Precision and Recall. *Applied Sciences*, *13*(10). https://doi.org/10.3390/app13106275
[24] M. M. Diaz, “Peru student enrollment data (2023),” Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/miguelmallqui17/peru-student-enrollment-data-2023
[25] M. Melchior, “Incorporating domain knowledge for learning interpretable features,” Archives of Data Science, Series A, vol. 8, no. 2, 2022, doi: 10.5445/IR/1000150142.
[26] T. Miller, I. Durlik, A. Łobodzińska, L. Dorobczyński, and R. Jasionowski, “AI in context: Harnessing domain knowledge for smarter machine learning,” Appl. Sci., vol. 14, no. 24, 2024, doi: 10.3390/app142411612.
[27] F. Voutsas, J. Violos, and A. Leivadeas, “Mitigating alert fatigue in cloud monitoring systems: A machine learning perspective,” Comput. Netw., vol. 250, Art. no. 110543, 2024, doi: 10.1016/j.comnet.2024.110543.
[28] Y. Tiara, M. Aidi, E. Erfiani, and R. Rachmawati, “Overdispersion handling in Poisson regression model by applying negative binomial regression,” BAREKENG, vol. 17, pp. 417–426, 2023, doi: 10.30598/barekengvol17iss1pp0417-0426.
[29] M. Flynn and L. A. F. Fcas, “More flexible GLMs: Zero-inflated models and hybrid models,” Casualty Actuarial Society E-Forum, 2009.
[30] R. Yu, H. Lee, and R. F. Kizilcec, “Should college dropout prediction models include protected attributes?” arXiv:2103.15237, 2021, doi: 10.48550/arXiv.2103.15237.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Enhancing geometry exemplification through integration of digital technology tools in South African higher education mathematics classrooms

Annie M. Kgosi and Anthony A. Essien
University of the Witwatersrand, South Africa

Article DOI: 10.68286/samyrad2026.339

ABSTRACT
Geometry teaching in South African higher education continues to face challenges related to student engagement, conceptual understanding, and visualising abstract spatial relationships. Traditional exemplification, reliant on static diagrams and chalkboard illustrations, limits opportunities for pre-service teachers to explore geometric concepts dynamically. This paper reports an emerging, work -in-progress study investigating how digital tools, including dynamic geometry software, document-camera visualisers, and 3D visualisation platforms, enhance geometric exemplification in undergraduate mathematics classrooms. Framed by variation theory and a qualitative, interpretivist case study design, the study examines how seven mathematics teacher educators across five South African institutions choose and use examples during technology-mediated instruction. Data are collected through pre- and post-observation interviews and video-recorded classroom observations; the pre-observation phase was completed in semester one, with the classroom observations continuing and post-observation interviews in semester two of 2026. The findings reported here are preliminary and drawn from one of the seven participating teacher educators. Preliminary analysis of classroom video data indicates that this teacher educator paired document-camera visualisers with physical manipulatives to model worked examples, such as calculating the surface area of cylinders and triangular prisms, moving systematically from concrete objects to generalised formulae. This pairing appears in this single case, to support contrast and generalisation, key variation-theory patterns, while raising early considerations around instructor readiness and equitable access that remain to be tested against the full participant sample.

KEYWORDS Geometry; digital technology integration; exemplification; variation theory; South African higher education; conceptual understanding.

How to cite: Kgosi, A. M., & Essien, A. A. (2026). Enhancing geometry exemplification through integration of digital technology tools in South African higher education mathematics classrooms. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.339

Introduction

Mathematics education stands at a critical juncture where traditional pedagogical approaches must evolve to meet the demands of an increasingly digital society. Higher education institutions increasingly recognise technology as fundamental to preparing pre-service teachers (PSTs) for twenty-first-century classrooms [1]. The National Council of Teachers of Mathematics (NCTM, 2023) describes technology as a catalyst for change and innovation and advocates its strategic integration to enrich mathematical learning [2]. Technology use can shift classroom focus from routine calculation toward strategic thinking, fostering deeper conceptual understanding [1], and recent evidence links effective integration to more positive student attitudes and engagement in STEM [3,4].

However, research says little about how exemplification is supported by technology integration. Exemplification, understood as the discursive act of illustrating a mathematical generality through a particular example, is a commonplace of mathematics teaching that offers insight into teachers' instructional decisions and underlying pedagogical reasoning [5,6,7]. Structured exemplification is critical for helping PSTs grasp the object of learning that underlies broader generalisation. Despite the recognised potential of digital tools, teacher educators (TEs) often lack the pedagogical and technological knowledge to integrate examples purposefully when teaching geometry [8], leaving a gap this study addresses by examining how examples are chosen and used when technology is integrated into geometry instruction. This paper reports the study's design and preliminary observations. As an emerging, work-in-progress account, the findings below are illustrative of a single case within the planned sample of seven teacher educators and are not yet generalisable across it.

Theoretical Framework: Variation Theory

This study is framed by variation theory, rooted in phenomenography, which examines how learners discern the critical aspects of an object of learning [9]. A critical aspect is an essential feature of a concept that learners must experience as variation, against a background of invariance in other features, in order to discern it [10,11]. Effective teaching therefore requires purposeful variation of critical features through patterns such as contrast, separation, generalisation, and fusion. Variation theory also distinguishes between the intended, enacted, and lived object of learning: what a teacher plans to teach, what is actually made available in the lesson, and what learners come to discern [7,11]. Combined with the Mathematics Discourse in Instruction framework's account of exemplification [6] variation theory allows this study to analyse not only which examples teacher educators select when integrating digital tools, but how the pattern of variation across those examples shapes what is made available for PSTs to learn about three-dimensional geometry.

Literature Synthesis

Global adoption of digital technologies in mathematics education continues to grow, with evidence pointing to gains in conceptual understanding and engagement (NCTM, 2023). Participatory digital tools have been shown to stimulate mathematical discussion and improve classroom communication [12], while dynamic and interactive tools support hands-on visualisation of abstract concepts [13]. In geometry specifically, tools such as GeoGebra, augmented reality, graphing calculators, and interactive whiteboards have demonstrated effectiveness across educational levels [8]. However, integration remains uneven: teacher readiness, curriculum alignment, and equitable access continue to constrain the extent to which digital tools translate into enriched exemplification [1,3]. Few studies, however, examine the specific examples teacher educators choose and how these are made available through technology that is the focus taken up by the research questions below.

Research Questions

Main Research Question

What examples are chosen and used when integrating digital technology tools in mathematics instruction to promote deep learning in geometry?

Sub-Research Questions

1a. Which exemplification affordances are provided during the integration of technologies in supporting specific geometric concepts?

1b. How does exemplification support the intended, enacted and lived object of learning during the teaching of three-dimensional geometry when integrating technology-enhanced instruction?

1c. What do teacher educators perceive to be the impact of exemplification on students' conceptual understanding and problem-solving skills.

The RQ1b addresses directly how exemplification links intended, enactment and PST learning, the analytical core of variation theory as applied in this study which receives the extensive treatment in the findings below. Whereas RQ 1c is deliberately framed around teacher educators' perceptions of impact, since the study's current data sources are TEs facing rather than PST-facing (see. Research Methodology and Design and Discussion below).

Research Methodology and Design

The study adopts a qualitative, interpretivist case study design [14] to obtain detailed, contextualised accounts of how mathematics teacher educators use exemplification when integrating technology into geometry instruction. Qualitative case study is well suited to examining the complexity and dynamic character of teaching practice within its natural setting [15].

The research is being conducted at five public universities and universities of technology in South Africa that prepare pre-service mathematics teachers. Institutions and participants were purposively selected based on established technology-teaching infrastructure, accessibility, and location convenience. Seven mathematics teacher educators who use digital technology in their practice form the study sample; access is negotiated through faculty heads of department during the ethical clearance process. This study notes that all seven participants use different digital technology tools as resources like dynamic software (e.g., GeoGebra and document-camera visualisers and others). The case reported below (TE1) uses a document -camera visualiser for a 3D visualisation platforms and PowerPoint rather than a dynamic geometry software.

Three data sources are used: semi-structured pre-observation interviews, video-recorded classroom observations, and semi-structured post-observation interviews. The pre-observation interviews, completed in

semester one of 2026, explored participants planned use of examples and technology. Video-recorded classroom observations capture actual classroom exemplification practice, including teacher educators' talk, gestures, use of physical manipulatives, and use of digital tools such as document-camera visualisers and dynamic geometry software. Post-observation interviews, scheduled for semester two of 2026, will probe teacher educators' reflections on the examples used and their intended versus enacted object of learning during the video-recorded classroom observation lessons. Together these methods allow triangulation between planned, enacted, and reflected-upon exemplification practice.

Data Analysis

Data analysis follows a hybrid deductive-inductive thematic analysis [16]. Deductive codes were derived a priori from variation theory and the Mathematics Discourse in Instruction framework: example type (worked/numerical, physical/manipulative, symbolic), pattern of variation (contrast, generalisation, separation, fusion), and object of learning (intended, enacted, lived). Inductive codes were allowed to emerge from repeated viewing of the video data to capture practices not anticipated by the framework, such as the specific affordances of the classroom document-camera visualiser.

The single-case analysis reported below was triangulated across two data sources for the same teacher educator (TE1): video-recorded classroom observation and a semi-structured post-observation interview, itself video-recorded and transcribed. Both were coded using the same intended/enacted/lived object-of-learning categories, together with codes for exemplification affordances and in-the-moment adjustments, allowing the researcher to move beyond what was visible on screen to the teacher educator's own account of her planning, in-the-moment decisions, and reading of PSTs' understanding.

Coding of this material, alongside the pre-observation interview transcripts, is ongoing across the full data set of seven teacher educators. The findings below represent a first cycle of analysis for a single participant, organised by research question rather than by named theme; a fuller thematic synthesis across all seven cases will follow once classroom observations and post-observation interviews are completed in semester two of 2026.

Ethical Considerations

Video recording of classroom interactions was conducted in accordance with relevant institutional research-ethics requirements. Participants were informed about the purpose of the study and the use of recordings and transcripts. Identifying information was removed from reported excerpts, and pseudonyms or participant labels were used where appropriate.

Findings

This preliminary analysis draws on video-recorded observation footage of one teacher educator's lesson on the surface area of three-dimensional solids, captured via a ceiling-mounted projector and a document-camera visualiser. Across six short video excerpts, TE1 worked through two examples in sequence: the total surface area of a cylinder and the total surface area of a triangular prism with the dimensions originally given as base 5 cm, slant sides 3 cm and 4 cm, height 20 cm. TE1 wrote a general formula before substituting the given values step by step for the cylinder (e.g., SA = $2\pi r^2 + 2\pi rh$) and the triangular prism SA = 2(½bh) + H (l + b +Hyp), considering the hypotenuse side of the right-angled triangle, so that the same underlying structure surface area equals the sum of the areas of all faces, was held invariant while the solid itself, and the particular face-area formulae required, varied. For each solid, she wrote a general formula before substituting the each given value step by step. TE1 defines SA = ½bh as half times base times h which refers only to one part of a right -angle triangle only, then TE further explained the addition of 2 times ($½bh$) which is the area of the triangular area with base length of a triangular face (b) and height altitude of a triangular face. Therefore, TE1 continued to explain the perimeter of a triangular base ($l + b + hyp$), multiplied by the height of the prism (H) or Length of the prism (L). These finding provide the exact formula as written on the visualiser. This matches the variation-theory pattern of contrast, varying the solid, held against generalisation, the constant underlying structure.

The document-camera visualiser was the central technological affordance observed across these excerpts, it allowed the teacher educator's real-time handwriting to be magnified and projected for the whole class, functioning as a dynamic substitute for the chalkboard while preserving worked steps for later review, its specific affordances are detailed under RQ1a, below. In several excerpts, this projected worked example was paired with a physical manipulative - a small cylindrical object held up by the TE1, and in an earlier segment, a fold-and-learn 3D geometric-shapes kit visible on the lecture table linking the concrete solid to its symbolic and numerical representation.

RQ1: What examples are chosen and used? Preliminary evidence drawn from a single case and not yet generalised across the full sample - indicates that teacher educators favour worked numerical examples of real-world solids (cylinders, triangular prisms) rather than purely abstract exemplars, and that these examples are built up systematically from a general formula to a numerical solution. Physical manipulatives using everyday objects and commercial fold-and-learn geometric kits were chosen alongside the technology-mediated example, suggesting a deliberate move from concrete to symbolic representation rather than reliance on digital tools alone. TE1 confirmed in the post observation interview that this combination was planned in advance rather than incidental: she reported using PowerPoint slides, the document camera visualiser, and physical solids together in the same lesson "because I already knew that there are my students who learn better when they see solids or when they touch them", alongside others who "learn from diagrams".

RQ1a: Exemplification affordances of the technologies used. The document-camera visualiser afforded three

things a traditional chalkboard could not as easily provide: (i) whole-class visibility of fine handwritten detail through magnification and projection, (ii) preservation of the full worked sequence on screen for learners to revisit, and (iii) flexibility to combine the projected symbolic work with a physical object held in view of the camera. The post-observation interview corroborates this: TE1 described calling a PST forward mid-lesson to draw and project a supporting sketch of the triangular prism once she noticed some PSTs "could not follow" the worked example and separately described using the camera visualiser to let PSTs project and present their own worked solutions for whole-class critique. This suggests the visualiser's affordance extends beyond teacher-led modelling to peer-mediated exemplification and real-time formative feedback ("we critique… you missed this, this is where you went wrong"). These affordances appear to support the coherence of the example space; the visible link between the concrete object, the general formula, and its numerical instantiation; though this requires confirmation against the fuller dataset, beyond this single teacher educator's self-report and one set of video excerpts.

RQ1b: Intended, enacted, and lived object of learning. This question is addressed most directly by triangulating the video data with TE1's post-observation account, which separates the three constructs clearly. The intended and enacted descriptions below draw on both sources; the lived description draws solely on TE1's self-report, since no direct PST-side data is intended to be collected, the study focus is only on TEs practice.

Intended: TE1 reported planning a multi-representational route to the object of learning, the general procedure for finding the total surface area of a 3-D solid, before the lesson began, deliberately combining slides, the visualiser, and physical solids "because I was aware that there would be students who would learn better from the drawings" and others who would not understand "just from the use of solids or by reading the slides".

Enacted: In the video data, the enacted lesson followed this plan through two contrasting worked examples (cylinder, then triangular prism) that varied the solid while keeping the additive face-area structure invariant. TE1's interview reveals a further, unplanned adjustment: when the originally chosen triangular prism was not right-angled and several PSTs "could not follow" the calculation, she substituted a simplified right-angled example ("let's assume the base is a right-angled triangle of dimensions 3, 4 and 5") and brought in a student-drawn, visualiser-projected diagram. This is a clear instance of the enacted object of learning being actively renegotiated in response to visible non-discernment, with the same technological resources used to close the gap.

Lived: TE1's account of what PSTs appeared to take away goes beyond procedural recall to structural understanding: she reported that PSTs came to understand "why one has to calculate areas of the faces of a 3-D solid and add them up", and could explain, for a cylinder, that "the length of a rectangle becomes a circumference and the width of the rectangle becomes a height, that's why we have 2πr [multiplied by] height" linking the net of the solid to a previously learned area formula. She further reported that PSTs generalised the procedure beyond the taught example ("for any triangular prism to calculate surface area… it depends on the dimensions of the triangle and the height of the prism") and, when subsequently given a task without accompanying drawings, were able to visualise the solid mentally evidence, in TE1's account, of a shift from concrete to abstract representation.

Relationship among the three: Asked directly about the relationship between what she intended, enacted, and what PSTs appeared to experience, TE1 stated that "what happened is everything that I planned to achieve", attributing the alignment to a multi-representational design that had anticipated diverse PST learning preferences in advance. In this case, then, enacted and (by the teacher educator's own report) lived objects of learning tracked closely with what was intended, in part because the one significant contingency that arose mid-lesson was resolved using the same representational and technological resources that had been planned for other purposes. This is a single, teacher-reported case, and PST-side data (from PST interviews or work samples) will be needed to corroborate the lived object of learning independently of the teacher educator's perspective.

RQ1c: Perceived impact on conceptual understanding and problem-solving: This study's data sources are entirely teacher-educator-facing at this stage; this sub-question is addressed here as TE1's *perceived* impact rather than as independently measured PST outcomes; direct evidence of PST problem-solving will require PST-facing data collection in later phases. Full impact cannot yet be claimed given the study's ongoing status and a single case scope. However, converging evidence from the video and interview data points to a consistent early pattern in TE 1's account: the visualiser-supported, step-by-step reconstruction of formulae appeared to make the logical structure of each calculation explicit, and TE1 reported that PSTs performed well on subsequent tasks, with one PST presenting a worked solution on the visualiser for the class to check against their own answers and, where errors occurred, to jointly identify "where you went wrong". This peer-checking routine, combined with TE1's account of PSTs solving a later task without diagrams, suggests, in this teacher educator's perception, that real-time, technology-mediated modelling may support both error-noticing and a degree of conceptual transfer. Confirmatory evidence, particularly direct evidence of PSTs' problem-solving rather than the teacher educator's report of it, will depend on the completed interview and observation dataset across all seven participants and ideally on the addition of a PST-facing data source (see discussion below).

Discussion

Based on variation theory, TE1's case offers a concrete illustration of how contrast and generalisation can be engineered through technology-mediated exemplification: holding the additive face-area structure invariant while varying the solid (cylinder, then triangular prism) is a clear instance of the contrast/generalisation pairing described by variation theory [9,10], and the document-camera visualiser appears, in this case, to

have been the technological resource that made that pairing visible and reviewable for the whole class in close to real time - consistent with the broader literature on participatory and interactive digital tools supporting mathematical communication and hands-on visualisation [12,13]. The mid-lesson substitution of the triangular-prism example is, on this reading, a particularly clear instance of the enacted object of learning being renegotiated in response to visible non-discernment, and it is the strongest evidence in this case for a claim variation theory makes but that is not often documented directly in classroom data: that teachers can and do adjust their pattern of variation in situ when a first attempt does not secure discernment.

At the same time, the case is a single one, and its findings should be read as illustrative rather than representative. Three limits are worth stating explicitly, since they bear directly on how the emerging implications below should be read. First, the Lived object of learning and the perceived impact reported under RQ1c rest entirely on TE1's self-report; no PST interviews, work samples, or assessment data have yet been collected, so what PSTs actually discerned remains inferred rather than observed. Second, TE1's technology use in the analysed excerpts - a document-camera visualiser and PowerPoint is narrower than the dynamic geometry software and 3D visualisation platforms foregrounded in the Introduction and Literature Synthesis; whether or how the six other participants use such tools is not yet established in this report. Third, the equitable-access and infrastructure concerns raised in the literature [1,3] are not directly evidenced in this single case, since TE1's institution evidently had working visualiser and projection infrastructure throughout; any claim about infrastructure access at this stage draws on the wider literature rather than on this study's data.

Emerging Implications

This single case suggested for further inquiry and practice, rather than as generalisable recommendations; they will be revisited once data collection across all seven participating teacher educators is complete.

First, TE1's practice suggests that mathematics teacher education programmes could consider pairing digital visualisation tools (document-camera visualisers, and for participants who use dynamic geometry software) with physical manipulatives to support movement between concrete, pictorial, and symbolic representations of geometric concepts; whether this pattern holds across the wider sample remains to be seen.

Second, the mid-lesson substitution described under RQ1b suggests that professional development for teacher educators could usefully address the design of contrasting example sequences, informed by variation theory, and the in-the-moment adjustment of those sequences when a first example does not secure discernment, rather than treating technology adoption as sufficient in itself - though this, too, is based on one teacher educator's practice and account.

The equitable-access question raised in the wider literature [1,3] whether institutions can reliably provide the projection and visualiser infrastructure this kind of teaching depends on, is an important one, but this study's data do not yet speak to it directly; we flag it here as a question for the completed dataset, rather than as a recommendation this case supports.

Conclusion

This paper has reported the design and preliminary, single - case, observations of an ongoing, emerging study into how South African mathematics teacher educators exemplify geometric concepts through digital technology integration. Early evidence from TE1's classroom video and post-observation interview data suggests that a document-camera visualisers, used alongside physical manipulatives can support a structured, generalisable approach to teaching three-dimensional geometry consistent with variation theory including an instance of the enacted object of learning being renegotiated mid-lesson in response to PST's visible non-discernment. As data collection continues into semester two of 2026 across all seven participating teacher educators, and as PST-facing data are considered for later phases, subsequent analysis will more fully address the affordances, enacted objects of learning, and perceived impact on PSTs' conceptual understanding across the participating institutions, with implications for mathematics teacher preparation and curriculum design in South African higher education to be developed once that fuller picture is available.

References

[1] Atteh, E., Assan-Donkoh, I., Ayiku, F., Nkansah, E., & Adams, A. K. (2020). The Use of Technology among School Mathematics Teachers and Students: The New Wave of Recommended Instructions. Asian Research Journal of Mathematics, 16(5), 18–29. https://doi.org/10.9734/arjom/2020/v16i530189

[2] National Council of Teachers of Mathematics. (2023). Equitable integration of technology for mathematics learning [Position statement]. https://www.nctm.org/standards-and-positions/equitable-integration-of-technology

[3] Hidayat, A., & Firmanti, P. (2024). Navigating the tech frontier: a systematic review of technology integration in mathematics education. Cogent Education, 11(1). https://doi.org/10.1080/2331186X.2024.2373559

[4] Pitchford, N.J., Gulliford, A., Outhwaite, L.A., Davitt, L.J., Katabua, E., Essien, A.A. (2021). Using Interactive Apps to Support Learning of Elementary Maths in Multilingual Contexts: Implications for Practice and Policy Development in a Digital Age. In: Essien, A.A., Msimanga, A. (eds) Multilingual Education Yearbook 2021. Multilingual Education Yearbook. Springer, Cham. https://doi.org/10.1007/978-3-030-72009-4_8

[5] Adler, J., & Davis, Z. (2006). Opening another black box: Researching mathematics for teaching in mathematics teacher education. Journal for Research in Mathematics Education, 37(4), 270–296.

[6] Adler, J., & Ronda, E. (2015). A framework for describing mathematics discourse in instruction and interpreting differences in teaching. African Journal of Research in Mathematics, Science

and Technology Education, 19(3), 237–254. https://doi.org/10.1080/10288457.2015.1089677
[7] Essien, A.A. (2021). Rethinking exemplification in mathematics teacher education multilingual classrooms. In: Kollosche D (ed.), Exploring New Ways to Connect: Proceedings of the eleventh international mathematics education and society conference. Tredition, Austria, 24–29 September, vol. 1. pp. 39–55. https://doi.org/10.5281/zenodo.5457116
[8] Sunzuma, G. (2023). Technology integration in geometry teaching and learning: A systematic review (2010–2022). LUMAT: International Journal on Math, Science and Technology Education, 11(3), 1–18.
[9] Marton, F., & Booth, S. (1997). Learning and awareness. Lawrence Erlbaum Associates.
[10] Marton, F., & Pang, M. F. (2006). On Some Necessary Conditions of Learning. Journal of the Learning Sciences, 15(2), 193–220. https://doi.org/10.1207/s15327809jls1502_2.
[11] Pang, M. F., & Marton, F. (2005). Learning theory as teaching resource: Enhancing students' understanding of economic concepts. Instructional Science, 33(2), 159–191. https://doi.org/10.1007/s11251- 005-2811-0
[12] Calor, S. M., Dekker, R., van Drie, J. P., & Volman, M. L. (2022). Scaffolding small groups at the group level: Improving the scaffolding behaviour of mathematics teachers during mathematical discussions. Journal of the Learning Sciences, 31(3), 369–407.
[13] Attard, C., & Holmes, K. (2020). "It gives you that sense of hope": An exploration of technology use to mediate student engagement with mathematics. Heliyon, 6(1), e02945. https://doi.org/10.1016/j.heliyon.2019.e02945
[14] Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE.
[15] Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE.

[16] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Beyond Prompting: Embedding Ethical and Equity-Oriented Artificial Intelligence Literacy within the Canadian Dental Public Health Competency Framework

Anil Menon[1], Abdulrahman Ghoneim[2], Keith Da Silva[3], Vinay Pilly[4], and Kalyani Baldota[5]
(corresponding author e-mail: anil.menon@umanitoba.ca)

[1]Dr. Gerald Niznick College of Dentistry, University of Manitoba, Winnipeg, MB, Canada
[2]Faculty of Dentistry, Dalhousie University, Halifax, NS, Canada
[3]Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada
[4]Private dental practitioner in Texas, USA
[5]Royal College of Dental Surgeons of Ontario, Toronto, ON, Canada

Article DOI: 10.68286/samyrad2026.342

ABSTRACT
Artificial intelligence (AI) is rapidly reshaping health professions education. However, structured AI literacy has not been consistently incorporated into dental public health (DPH) curricula. The Discipline Competencies for Dental Public Health developed by the Canadian Association of Public Health Dentistry (CAPHD) remain a key framework for articulating the knowledge, skills, and attitudes of DPH in Canada; published in 2008, they predate the contemporary generative AI era. Risks posed by unstructured and unsupervised AI use in DPH education extend beyond academic integrity and include inaccurate interpretation of population oral health data, fabricated or unverifiable evidence, algorithmic bias affecting underserved populations, privacy risks involving patient and community information, insufficient attention to Indigenous data governance and cultural safety, and weakened accountability in policy and advocacy work. This conceptual paper proposes an author-developed, five-level AI literacy model mapped onto the eight CAPHD competency domains and four competency layers (Basic Shared, Foundation, Advanced, Expert). The five levels are foundational AI literacy, responsible student and practitioner use, dental public health applications, equity, ethics, and population health judgment, and professional accountability. The model has not yet undergone external validation. It is intended to support curriculum planning, graduate and residency training, community oral health programs, continuing professional development, and future consensus-building around competency renewal.

KEYWORDS AI literacy, dental public health, discipline competencies, health professions education, equity, cultural safety, Indigenous data governance, professional accountability

How to cite: Menon, A., Ghoneim, A., Da Silva, K., Pilly, V., & Baldota, K. (2026). Beyond prompting: Embedding ethical and equity-oriented artificial intelligence literacy within the Canadian dental public health competency framework. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.342

Introduction

Generative artificial intelligence (AI) tools have entered routine use among health professions students and practitioners with a speed that has outpaced curricular response. In dental public health (DPH), residents increasingly use large language models for literature search, evidence synthesis, surveillance interpretation, program planning, and academic writing, often without formal preparation in their accuracy, limitations, or population-level implications. As AI tools become increasingly available to students and practitioners, DPH residents may develop AI-use habits without explicit competencies, formal supervision, or shared standards.

DPH education and workforce development in Canada are anchored by the Discipline Competencies for Dental Public Health (2008), developed by the Canadian Association of Public Health Dentistry (CAPHD) with support from the Public Health Agency of Canada [1]. The framework articulates eight competency domains across four layers and informs curriculum planning, graduate and residency training, and continuing professional development. It does not, however, address AI because it predates the contemporary generative AI era.

This paper proposes a five-level AI literacy model mapped to the CAPHD Discipline Competencies to guide curriculum development, residency training, continuing professional development, and future competency renewal in Canadian dental public health. The model draws on emerging international and national AI literacy, digital health, health professions education, and public health workforce models and extends them to the distinctive responsibilities of DPH. The model proposed in this manuscript is intended as a scholarly contribution and does not represent an official CAPHD position.

Background

The CAPHD Discipline Competencies

The CAPHD Discipline Competencies (Release 4.0, 2008) describe the essential knowledge, skills, and attitudes that are required for the practice of dental public health in Canada [1]. The framework is organized into eight domains: Oral Public Health Sciences; Oral Health Assessment and Analysis; Oral Health Program Planning, Implementation and Evaluation; Oral Health Policy Planning, Implementation and Evaluation; Partnerships, Collaboration and Advocacy; Diversity and Inclusiveness; Communication; and Leadership. Within each of these domains, competencies are articulated across four progressively complex layers: Basic Shared, Foundation, Advanced, and Expert. The DPH framework is explicitly grounded in core public health values of equity, social justice, sustainable development, community participation, and respect for diversity.

The Educational Gap

Three gaps motivate this work. First, the CAPHD discipline competencies, while comprehensive, contain no AI-specific competency statements. Second, DPH curricula across Canadian universities have not yet adopted a consistent structure for introducing AI literacy at the undergraduate, graduate, or continuing education level. Third, existing AI literacy frameworks developed for general education or clinical disciplines have not appropriately addressed the equity, cultural safety, and Indigenous data governance dimensions of AI use within oral health, especially for First Nations, Inuit, and Métis communities and other equity-deserving populations.

Educational and Public Health Frameworks Informing the Proposed Model

A widely cited formulation of AI literacy was provided by Long and Magerko (2020), who define AI literacy as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" [2]. Through an interdisciplinary review of literature on AI education, perceptions, and practice, they organized seventeen AI literacy competencies under five thematic questions: what AI is, what it can do, how it works, how it should be used, and how people perceive it, and articulated fifteen design considerations to guide learner-centred AI education. This foundational framework informs much of the subsequent literature on AI literacy in education and the workplace, and is adopted in the present paper as the working definition of AI literacy. The model proposed below extends this general definition for the specific responsibilities of dental public health practice.

The proposed model is informed by emerging AI and digital health education frameworks from Europe and North America. A dental-specific precedent is provided by Schwendicke and colleagues, who developed a core curriculum for artificial intelligence in oral and dental healthcare through expert input and Delphi consensus involving the IADR e-oral health network, the ITU/WHO Focus Group AI for Health, and the Association for Dental Education in Europe [3]. Their curriculum identifies foundational AI knowledge, oral and dental AI use cases, evaluation of AI applications, and broader issues including generalizability, representativeness, explainability, autonomy, accountability, and governance. This work provides an important foundation for AI education in dentistry. However, it is not specifically mapped to dental public health competencies and does not operationalize AI literacy through population oral health surveillance, public health policy, program evaluation, community engagement, Indigenous data governance, or equity-oriented professional accountability. The proposed model builds on this dental AI curriculum precedent by translating AI literacy into the CAPHD dental public health competency architecture. UNESCO's AI competency frameworks for students and teachers emphasize human-centred AI, ethics, foundational knowledge, application, and progressive development [4,5]. In health professions education, the Association of American Medical Colleges (AAMC) is developing AI competencies across the learning continuum, while the DECODE framework illustrates how digital health competencies can be organized into domains and learning outcomes [6,7]. In public health, the Association of Schools and Programs of Public Health (ASPPH) final AI framework report and Canada's AI for Public Health (AI4PH) initiative highlight the need for responsible, ethical, and equitable AI capacity-building in public health education, research, policy, and practice [8,9]. Canada's updated public health core competencies also emphasize data, equity, ethical practice, and technologies, although AI literacy is not named as a discrete competency area [10]. The NHS/Health Education England DART-Ed framework further supports the

idea of AI literacy as a capability layer that extends rather than replaces existing professional competencies [11].

Building on these frameworks, the proposed model maps AI literacy onto the CAPHD Discipline Competencies to address the distinctive responsibilities of dental public health, which include oral health surveillance, program planning and evaluation, policy advocacy, community engagement, cultural safety, Indigenous data governance, and professional accountability.

Distinctive AI Competency Needs in Dental Public Health

Currently, existing AI competency frameworks address valuable but incomplete dimensions of the responsibilities held by DPH practitioners. General AI literacy frameworks, including Long and Magerko's foundational definition [2] and UNESCO's competency frameworks for students and teachers [4,5], emphasize universal foundational concepts, critical evaluation, and ethical use but are oriented to the general public and non-technical learners rather than to health system contexts. Medical AI competency models such as the AAMC's address clinical and educational applications but are oriented toward individual patient encounters and physician practice. Public health AI initiatives (ASPPH and AI4PH) address population-level applications but do not address the distinct content of oral health surveillance, prevention programming, and oral health policy. Digital health frameworks (DECODE and the NHS DART-Ed capability framework) address health workforce digital readiness broadly but are not adapted to the specific knowledge base and professional accountabilities of dental public health specialists.

DPH introduces several context-specific AI competency requirements. First, oral health surveillance relies on a small number of population data sources (such as the Canadian Health Measures Survey oral component and the Canadian Community Health Survey oral health module), each with specific sampling, weighting, and methodological constraints that AI tools are unlikely to handle correctly without expert oversight. Second, oral health policy areas such as community water fluoridation, public dental insurance, and access for equity-deserving populations are politically contested, evidence-sensitive, and especially vulnerable to misinformation, which can be amplified by generative AI. Third, dental public health practice in Canada is tied closely to Indigenous oral health as well as community-based prevention programs, where Indigenous data governance principles, cultural safety, and community participation are constitutive of competent practice rather than optional considerations. Fourth, dental public health specialists are increasingly being called upon to advise federal, provincial, territorial, regional, and municipal governments on oral health programs and policy, creating a need for explicit competencies in AI procurement, governance, and accountability that are not addressed by general public health AI frameworks. These context-specific requirements support the case for a dedicated DPH AI literacy model rather than relying on adjacent frameworks alone.

Conceptual Development Approach

This conceptual model was developed through a structured mapping exercise rather than a systematic review. The approach is best characterized as a directed competency mapping informed by framework analysis, in which existing AI and digital health competency frameworks were used as conceptual anchors and CAPHD competencies were used as the operational target architecture.

Mapping decisions followed three decision rules. First, CAPHD competencies were selected when the underlying public health function was directly affected by the use of AI, including assessment, surveillance, communication, program evaluation, policy development, advocacy, leadership, and community engagement. Second, competencies were assigned to an AI literacy level according to the primary educational task required: understanding AI concepts, using AI responsibly, applying AI to DPH functions, judging equity and ethical consequences, or exercising professional accountability. Third, each mapping was classified as one of three forms of alignment: an existing CAPHD competency that already implies a relevant capability, an existing competency that requires reinterpretation in an AI-enabled environment, or a new AI-specific capability that extends the 2008 framework. This distinction was used by the authors to avoid simply re-labelling existing competencies and to identify where new AI-related educational content is required.

Ambiguous mappings were handled through internal co-author discussion. When an AI-related capability could plausibly fit more than one CAPHD domain, the authors prioritized the domain in which the intended learner performance would most likely be taught or assessed. For example, AI-generated surveillance interpretation was mapped to Oral Health Assessment and Analysis, while AI procurement, monitoring, and organizational accountability were mapped to Leadership and Partnerships, Collaboration and Advocacy.

The selection of source frameworks was purposive. Frameworks were considered if they (i) explicitly addressed AI literacy or digital health competency within an educational or workforce context, (ii) were issued or endorsed by a national, international, or sector-level organization (rather than by a single research group), and (iii) were publicly available in English at the time of writing. The frameworks selected include UNESCO's AI competency frameworks for students and teachers, the AAMC AI competencies for medical education, the DECODE digital health competency framework, the ASPPH AI for Public Health initiative, the Canadian AI4PH initiative, Canada's updated public health core competencies, and the NHS/Health Education England DART-Ed capability framework. Together, they represent perspectives of education, health professions education, and public health workforce across multiple jurisdictions. Other frameworks may also be relevant, and the absence of a formal scoping review is acknowledged as a limitation.

Source influence was overlapping rather than exclusive. Foundational AI literacy (L1) drew chiefly from Long and Magerko, Schwendicke et al., and UNESCO; responsible use (L2) from UNESCO, AAMC, and the dental AI curriculum; applied DPH use (L3) from Schwendicke et al., DECODE, ASPPH, and AI4PH; equity and population health judgment (L4) from UNESCO, ASPPH, AI4PH, and WHO guidance; and professional accountability (L5) from AAMC, NHS DART-Ed, ASPPH, and WHO guidance. These sources were used as conceptual anchors rather than as one-to-one precedents for the CAPHD mappings.

The CAPHD Discipline Competencies were reviewed against each source framework to identify AI-related learning needs that map naturally to existing CAPHD domains and competency statements. Mapping decisions were guided by their alignment with the foundational CAPHD values, relevance to core DPH functions (surveillance, prevention, program planning and evaluation, policy, advocacy, and community engagement), and suitability across undergraduate, graduate, residency, and continuing professional development contexts. The proposed five-level structure emerged from a synthesis of the source frameworks; in particular, the layered progression draws on UNESCO's progressive development model and the NHS DART-Ed capability-layer construct, while the integration with CAPHD domains preserves the eight-domain, four-layer architecture of the existing Canadian DPH competency framework.

Initial mapping was developed by the lead author and progressively refined through internal review and discussion with co-authors. The mapping should be understood as a structured author-developed proposal intended to support discussion and consultation, and not as a validated competency framework. Subsequent steps should include external consultation, Delphi consensus development, and pilot implementation, as described in Section XI.

Proposed Model: Five-Level AI Literacy for Dental Public Health

The proposed model articulates five progressive levels of AI literacy, each integrated with specific CAPHD domains and layers. The levels are not strictly sequential within a single course but are intended to scaffold AI competency development across the trajectory of a DPH learner, from undergraduate dental student to expert specialist. Fig. 1 presents the overall conceptual structure.

The five AI literacy levels should not be interpreted as a one-to-one replacement for the four CAPHD competency layers. The CAPHD layers describe increasing professional scope and complexity, from Basic Shared to Expert, whereas the five AI literacy levels describe domains of AI-related capability. Foundational AI literacy and responsible use may begin early in training, while governance, procurement, monitoring, and organizational accountability are most relevant at advanced and expert levels. The model is therefore a competency-extension scaffold mapped onto, rather than substituted for, the CAPHD architecture.

Level 1: Foundational AI Literacy

This level would establish a working understanding of AI concepts, capabilities, applications, and limitations that are relevant to public health practice. Learners would distinguish between rule-based, machine learning, and generative AI systems; understand the role of training data and recognize common failure modes such as hallucination, bias, and miscalibrated confidence. This level integrates with CAPHD Domain 1 (Oral Public Health Sciences), specifically competencies 1.1, 1.2, 1.3, and 1.6, framing AI as a contemporary public health science. It also engages Domain 7 (Communication), competency 7.3, situating AI as a communication technology that requires deliberate, informed use.

Level 2: Responsible Student and Practitioner Use

This level would inform learners about the appropriate use of AI tools in academic and professional work, including validation, disclosure, attribution, citation verification, privacy protection, and the avoidance of plagiarism or fabrication. Level 2 corresponds to CAPHD Domain 8 (Leadership), particularly competency 8.3 (to act ethically with clients, information, and resources) and 8.7 (manage self, information, and resources in a way that honours public health ethics), and Domain 7 (Communication), competencies 7.4 and 7.6.

Level 3: Dental Public Health Applications

At level 3, learners would develop applied competence in using AI tools to support core DPH functions: population health assessment, surveillance interpretation, program planning, evidence synthesis, and policy analysis. Learners would be taught to critically appraise AI-generated outputs relative to established epidemiologic and biostatistical standards, recognizing where AI tools can misrepresent sampling design, confounding, or causal reasoning errors. This level engages CAPHD Domain 2 (competencies 2.10, 2.13, 2.16), Domain 3 (competencies 3.7, 3.8, 3.14), and Domain 4 (competencies 4.7, 4.9).

Level 4: Equity, Ethics, and Population Health Judgment

This level represents the conceptual core of the five-level model. Learners would develop judgment to identify, critique, and mitigate algorithmic bias against equity-deserving populations; interrogate the representativeness of training data; recognize that the First Nations principles of Ownership, Control, Access, and Possession (OCAP®) are specific to First Nations data governance rather than a pan-Indigenous framework [12]; attend to First Nations, Inuit, and Métis-specific governance principles, community authority, consent processes, and culturally safe approaches to data collection, interpretation, and use; safeguard cultural safety in AI-mediated practice; and weigh population-level benefits against potential surveillance harms. This level integrates with CAPHD Domain 6 (Diversity and Inclusiveness), competencies 6.4, 6.7, 6.8, and 6.9, and Domain 5 (Partnerships, Collaboration and Advocacy), competencies 5.6 and 5.9.

Level 5: Professional Accountability

At the highest level, learners would situate AI use within the regulatory, institutional, and political contexts of public health practice. Competencies here include developing policies on AI procurement and deployment for organizations, contributing to professional standards, advising public health organizations on responsible AI use in publicly funded oral health programs, and modelling ethical AI practice as a leader. This level engages CAPHD Domain 8 (Leadership), competencies 8.13, 8.14, and 8.16, and Domain 5 (Partnerships, Collaboration and Advocacy), competencies 5.12 and 5.15.

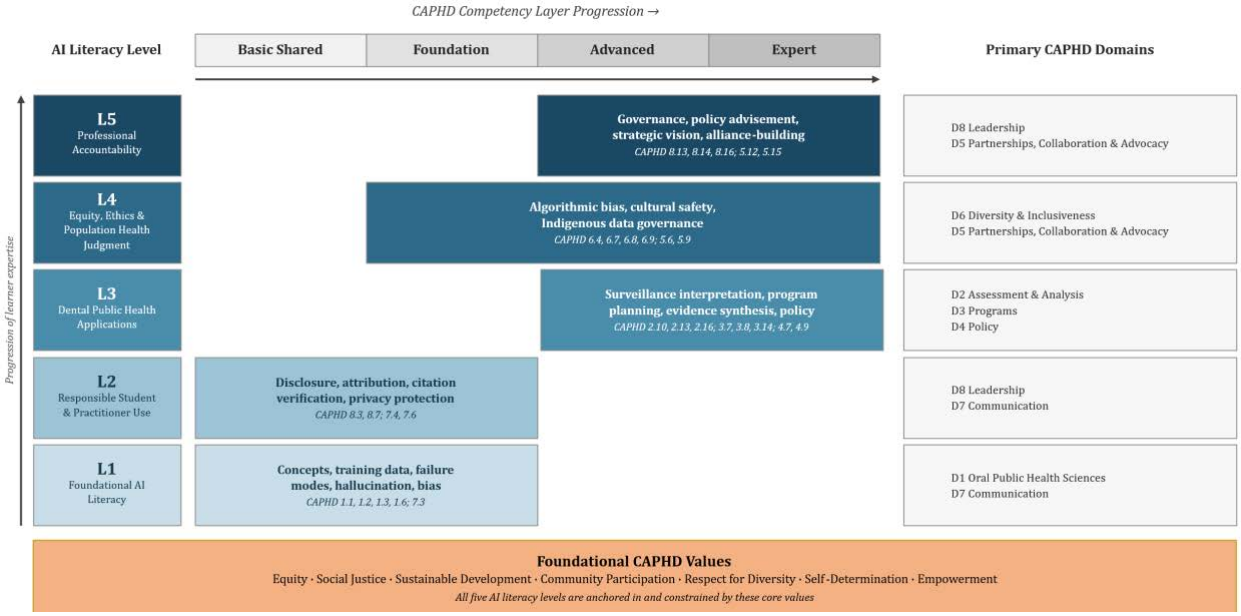


Fig. 1. Conceptual representation of the five-level AI literacy model, showing each level's primary location within the CAPHD competency layer progression, the principal CAPHD domains engaged, and the foundational CAPHD values that anchor all levels.

Table 1. Alignment of Proposed AI Literacy Levels with CAPHD Discipline Competency Domains, Statements, and Layers.

AI Literacy Level	CAPHD Domain(s)	Mapped Competencies†	CAPHD Layer
L1: Foundational AI Literacy	D1; D7	1.1, 1.2, 1.3, 1.6; 7.3	Basic Shared → Foundation
L2: Responsible Student and Practitioner Use	D8; D7	8.3, 8.7; 7.4, 7.6	Basic Shared → Foundation
L3: Dental Public Health Applications	D2; D3; D4	2.10, 2.13, 2.16; 3.7, 3.8, 3.14; 4.7, 4.9	Advanced → Expert
L4: Equity, Ethics, and Population Health Judgment	D6; D5	6.4, 6.7, 6.8, 6.9; 5.6, 5.9	Foundation → Expert
L5: Professional Accountability	D8; D5	8.13, 8.14, 8.16; 5.12, 5.15	Advanced → Expert

Note. CAPHD domains abbreviated: D1 Oral Public Health Sciences; D2 Oral Health Assessment and Analysis; D3 Oral Health Program Planning, Implementation and Evaluation; D4 Oral Health Policy Planning, Implementation and Evaluation; D5 Partnerships, Collaboration and Advocacy; D6 Diversity and Inclusiveness; D7 Communication; D8 Leadership. Competency numbers refer to statements within each domain as published in CAPHD (2008) [1].

Table 2. Alignment Between CAPHD Competencies and Proposed AI-Specific Extensions.

AI Literacy Level	Existing CAPHD Basis†	AI-Specific Extension††
L1 Foundational AI Literacy	Public health sciences; communication technology	Generative AI concepts, training data, hallucinations, model limitations, miscalibrated confidence
L2 Responsible Use	Ethical use of information; professional responsibility; communication	AI-use disclosure, citation verification, privacy protection with third-party tools, avoidance of fabricated evidence
L3 DPH Applications	Assessment, epidemiology, program planning, policy analysis	Critique of AI-supported interpretations of surveillance data, sampling weights, confounding, causal inference, and policy evidence
L4 Equity, Ethics, Population Health Judgment	Diversity, inclusiveness, cultural safety, advocacy, community development	Algorithmic bias assessment, representativeness of training data, Indigenous data governance, monitoring for disparate impacts
L5 Professional Accountability	Leadership, partnerships, strategic vision, policy engagement	AI procurement governance, transparency standards, accountability mechanisms, monitoring AI-supported public health decisions

Note. † Existing CAPHD basis refers to competencies or values already present in the 2008 framework. †† AI-specific extensions refer to proposed additions or reinterpretations needed to operationalize those competencies in an AI-enabled DPH context. These extensions are author-proposed and do not represent official CAPHD competency revisions.

Table 3. Worked Examples and Suggested Educational Settings for the Proposed AI Literacy Framework.

Level	Educational Setting	Worked Example
L1	Early residency	A first-year DPH resident will be asked to differentiate between machine learning, large language models, and generative AI; describe the concept of training data and its limitations; and identify why an LLM produces a confidently stated but incorrect answer about community water fluoridation guidance. This activity scaffolds CAPHD competencies 1.1, 1.3, and 7.3 by introducing AI as a contemporary public health science and a communication technology.
L2	Residency; CPD	During a literature review, residents will be required to provide rationale for their selected AI tool(s), maintain a structured AI disclosure log, attribute generated text correctly, verify all citations against primary sources, and reflect on the implications for patient privacy when entering clinical scenarios into a third-party model. The exercise enacts CAPHD competencies 8.3 (act ethically with clients, information and resources) and 7.4 (interpret information for non-professionals and community audiences related to oral health issues) under contemporary conditions.
L3	Residency; graduate training	Residents will use an AI-assisted analytic workflow to interrogate publicly available oral health surveillance data (for example, the Canadian Health Measures Survey oral component or the Canadian Community Health Survey oral health module), critically appraise the model's interpretation against published epidemiologic methods, and identify where AI outputs misrepresent sampling weights, confounding, or causal inference. This corresponds to CAPHD 2.10 (analyze oral health information to determine implications, uses, gaps and limitations) and 2.13 (apply research methods used in dental public health including biostatistics and epidemiology).
L4	Residency; CPD; faculty development	Residents will critique an AI risk-stratification model for caries prevention applied to an Indigenous community, examining whether training data adequately represent the relevant community; whether outputs respect appropriate community data governance, including OCAP® principles for First Nations contexts when applicable; and whether deployment is consistent with cultural safety. This operationalizes CAPHD 6.7 (apply knowledge of oral health inequities and inequalities in designing oral health programs and policies), 6.9 (create culturally safe and supportive environments within public health programs and organizations), and 5.6 (build capacity for oral health and general wellbeing through community development strategies).
L5	Senior residency; specialist CPD; leadership	Senior residents and practicing specialists will develop a position statement for their regional public health authority on the responsible procurement and deployment of AI tools in publicly funded oral health programs, including governance, accountability, monitoring for disparate impact, and stakeholder consultation. This aligns with CAPHD 8.14 (develop a strategic vision for dental health within the organization), 8.16 (build alliances and partnerships within changing political environments), and 5.12 (facilitate dialogue among governments, community partners and other stakeholders to support healthier communities).

Cross-Walk: AI Literacy by CAPHD Competency Architecture

Table 1 summarizes the alignment between the proposed AI literacy levels, the CAPHD competency domains, the specific competency statements engaged, and the CAPHD layer at which the literacy level is most relevant. Table 2 further characterizes the nature of each alignment. Together, these tables are intended as curricular planning tools for DPH educators and program directors.

Illustrative Worked Examples

Table 3 presents one illustrative worked example for each level, drawn from typical DPH educational activities, with the suggested educational setting indicated. These examples are intended to help educators in operationalizing the five-level model and are not exhaustive.

Assessment Strategies and Learning Outcomes

Effective adoption depends on the availability of defensible learning outcomes and assessment strategies. Table 4 maps each level of the model to a sample of measurable learning outcomes and to suggested assessment modalities consistent with established health professions education practice. The proposed modalities include short-answer assessments and concept-mapping at the foundational level; structured AI-use disclosure logs, citation-verification audits, and reflective writing at the level of responsible practice; rubric-based appraisal of AI-supported analytic outputs and portfolio entries at the applied level; case-based equity-and-bias critiques and Objective Structured Clinical Examination (OSCE)-style stations at the equity and ethics level; and policy memoranda, position statements, and oral defences at the professional accountability level. Across all levels, longitudinal portfolio assessment is recommended to track competency progression and consolidation across an educational program.

Educational outcomes can further be evaluated both cross-sectionally and longitudinally. Cross-sectional evaluation might include pre- and post-instruction AI literacy self-assessment, structured reflective writing, and rubric-based grading of AI-supported assignments. Longitudinal evaluation might include portfolio assessment across the course of a DPH residency, faculty review of integrated AI-related performance over time, as well as tracking of graduate practice patterns through alumni surveys. Given the rapid pace of change in AI tools, learning outcomes and assessment instruments will require periodic refresh.

Table 4. Sample Learning Outcomes and Suggested Assessment Modalities for Each AI Literacy Level.

Level	Sample Measurable Learning Outcomes	Suggested Assessment Modalities
L1	Learners can describe how generative AI models are trained and identify at least three common failure modes (hallucination, bias, miscalibrated confidence) using DPH-relevant examples.	Short-answer quiz; concept-map exercise; structured class discussion with rubric.
L2	Learners can produce a written work product accompanied by a rationale for using the selected AI tool(s); complete AI-use disclosure log; verify all AI-generated citations against primary sources; and articulate privacy implications of entering identifiable data into AI tools.	Structured AI-use disclosure log; citation-verification audit of submitted assignments; reflective writing rubric.
L3	Learners can apply AI tools to a defined DPH task (e.g., surveillance interpretation, program planning) and critically appraise AI-generated outputs against epidemiologic and biostatistical standards.	Rubric-based assessment of AI-supported analytic outputs; structured critique exercise comparing AI output to published epidemiologic evidence; portfolio entry.
L4	Learners can evaluate an AI tool for algorithmic bias affecting equity-deserving populations, assess alignment with Indigenous data governance principles, and propose mitigation strategies that uphold cultural safety.	Case-based equity-and-bias critique with rubric; OSCE-style station evaluating AI appraisal in a population-health scenario; reflective portfolio integrating cultural-safety considerations.
L5	Learners can draft an organizational or jurisdictional policy on responsible AI use in public oral health programs and defend governance, accountability, and equity-monitoring components.	Capstone policy memorandum or position statement with structured rubric; oral defence; longitudinal portfolio assessment across residency.

Risks and Limitations of Generative AI in Dental Public Health Education

Although the five-level model supports thoughtful use of AI in dental public health education, several characteristics of contemporary generative AI tools warrant explicit critical engagement. These risks shape the content of the framework and underline the importance of structured AI literacy.

Generative AI models are well known to fabricate citations, misattribute findings, and generate authoritative-sounding statements that are not supported by the underlying evidence base [13]. In a discipline where policy and program decisions are tied to surveillance data, evidence synthesis, and contested issues such as community water fluoridation, this raises major risks of misinformation entering educational and professional work without detection. Too much reliance on AI in evidence synthesis can also reduce learners' engagement with primary sources, weakening the methodological skills that the CAPHD explicitly identifies as core competencies.

AI models developed on internet-scale corpora reflect the biases and inequities of their training data. Norori and colleagues have described how data-driven, algorithmic, and human sources of bias in health AI can propagate and amplify societal inequities, with especially adverse consequences for populations historically underrepresented in biomedical datasets [14]. Within oral health, this can produce outputs that under-represent or misrepresent equity-deserving populations, including First Nations, Inuit, and Métis communities, newcomer and refugee populations, low-income groups, people with disabilities, and rural and remote communities. Without specific instruction on algorithmic bias, learners may apply AI outputs in ways that reproduce or amplify structural inequities, contrary to the foundational values of the CAPHD discipline competencies.

Most consumer-grade generative AI models are proprietary, with limited clarity about training data, model updates, content moderation policies, and downstream uses of submitted prompts. Opacity of proprietary models raises concerns about confidentiality of patient and community information, accountability for model behaviour, and the reproducibility of analyses that rely on rapidly changing models. The environmental and resource footprint of large-scale AI deployment, including energy and water use associated with model training and inference, is also a relevant population-health consideration that aligns with the CAPHD value of sustainable development and with the WHO principle of sustainable AI [15]. Finally, unequal access to advanced AI tools, including paid tiers, further widens educational and professional inequities within DPH training programs, between institutions, and across jurisdictions. A core purpose of structured AI literacy is to ensure that future DPH practitioners can recognize and respond to these risks rather than be passively exposed to them.

Discussion and Implementation Considerations

The proposed five-level model is designed to be implementable across multiple educational contexts: undergraduate dental curricula, DPH residency and graduate training, community-based oral health programs, and continuing professional development. Because it has been mapped onto the existing CAPHD architecture, it does not require institutions to adopt a parallel competency framework but instead provides curricular content that can be integrated into existing competency-based structures.

Three implementation aspects would warrant emphasis. First, the five-level model prioritizes equity, cultural safety, and Indigenous data governance as constitutive of competent AI use, not as optional add-ons. This reflects the explicit

foundational values of the CAPHD competencies and is consistent with global commitments to health equity articulated in the United Nations Sustainable Development Goals 3, 4, and 10. Second, consistent with the NHS DART-Ed view of AI literacy as a capability layer that extends rather than replaces existing professional competencies, the model intentionally distinguishes between the use of AI as a learner support tool and the use of AI as an object of professional critique; both are necessary, but they require distinct pedagogical approaches. Finally, the five-level model is intended to support future discussion on competency renewal, recognizing that the 2008 release predates the contemporary AI era.

Several implementation barriers should be anticipated. Faculty AI literacy is currently uneven, and faculty development will be needed before the five-level model can be delivered consistently. Schools can begin with low-risk introductory modules, shared AI-use disclosure templates, common rubrics, and case libraries that can be adapted locally. Faculty development should precede high-stakes learner assessment and should address AI limitations, privacy, citation verification, bias, and the use of institutional rather than personal or identifiable data in demonstrations. Institutional resources, including access to AI tools, faculty time, and ethics review capacity, vary substantially across Canadian universities. The swift pace of change in AI tools means that specific tool-based content will date quickly, requiring ongoing curriculum maintenance. Regulatory and accreditation guidance on AI use in health professions education is still emerging at federal, provincial, and territorial levels, creating uncertainty about how AI literacy should be assessed in licensure and specialty certification. These barriers support positioning the model as an implementable starting point rather than a finished prescription, allowing adaptation to local contexts.

Future Implementation and Evaluation

Future work should focus on both implementation and evaluation. Implementation can include introductory AI literacy modules, structured AI-use disclosure logs, reflective exercises, case-based critiques of equity and algorithmic bias, and rubric-based assessment of AI-supported assignments. Evaluation can include pre- and post-instruction AI literacy self-assessment, student reflections, faculty review, and review of AI-supported outputs. Pilot implementation could begin within DPH graduate or residency training and be refined through consultation with DPH educators, residents, practitioners, First Nations, Inuit, and Métis partners, community representatives, and professional organizations.

Several next steps would strengthen the scientific and practice basis of the five-level model. A modified Delphi consensus exercise involving Canadian DPH educators, residency program directors, public health dental specialists, learner representatives, First Nations, Inuit, and Métis oral health partners, and experts in health equity, data governance, and AI ethics could refine the model's content, learning outcomes, and assessment strategies. Participants would review each AI literacy level, the proposed CAPHD mappings, Table 2's extension categories, and the learning outcomes in Table 4; rate their clarity, relevance, feasibility, and completeness; and identify missing or ambiguous capabilities. National stakeholder consultation through CAPHD, the Royal College of Dentists of Canada, the Commission on Dental Accreditation of Canada, and provincial and territorial regulatory bodies can situate the five-level model within the broader Canadian DPH education and regulatory ecosystem. Pilot curriculum implementation within one or more Canadian DPH residency or graduate programs, with pre- and post-evaluation and qualitative learner feedback, can generate evidence on feasibility and educational impact. Formal validation studies, including psychometric evaluation of AI literacy self-assessment instruments and rubric-based assessment tools, would support broader adoption. Findings from these steps can inform future discussions on renewing Canadian DPH competencies for an AI-enabled public health environment.

Limitations

This paper proposes an author-developed conceptual five-level model and does not present empirical evaluation data. The mapping reflects the authors' interpretation of the CAPHD Discipline Competencies and should be refined through consultation with DPH educators, residents, practitioners, First Nations, Inuit, and Métis partners, community representatives, and professional organizations. This paper did not include a formal scoping review, and additional AI literacy and digital health competency frameworks may be relevant. This proposal is not intended to replace formal competency-renewal processes or to serve as an official CAPHD position.

Conclusion

The use of generative artificial intelligence is now an unavoidable feature of the educational and professional environment of dental public health. By building on international AI literacy and digital health frameworks, Canada's updated public health core competencies, and on the values-based architecture of the Canadian CAPHD Discipline Competencies, the proposed five-level model offers a contemporary, equity-oriented structure for embedding AI literacy across DPH education and practice. It also provides a foundation for future discussions on updating Canadian DPH competencies for an AI-enabled public health environment.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgment

The views expressed in this paper are those of the authors and do not represent an official position or approved revision of the Canadian Association of Public Health Dentistry. The

authors used Grammarly for limited language editing, structural refinement, and revision support. The conceptual content, framework design, competency mapping, analysis, and interpretation are the authors' own, and the authors take responsibility for the final manuscript.

References

[1] Canadian Association of Public Health Dentistry, "Discipline Competencies for Dental Public Health in Canada, Release 4.0," Canadian Association of Public Health Dentistry, Mar. 2008. Available: https://caphd.ca/wp-content/uploads/2022/06/DisciplineCompetenciesVersion4_March31.pdf

[2] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. 2020 CHI Conf. Human Factors in Computing Systems (CHI '20), New York, NY: Association for Computing Machinery, 2020, pp. 1–16, doi: 10.1145/3313831.3376727.

[3] F. Schwendicke, A. Chaurasia, T. Wiegand, S. E. Uribe, M. Fontana, I. Akota, O. Tryfonos, and J. Krois, on behalf of the IADR e-oral health network and ITU/WHO Focus Group AI for Health, "Artificial intelligence for oral and dental healthcare: Core education curriculum," J. Dent., vol. 128, 104363, 2023, doi: 10.1016/j.jdent.2022.104363.

[4] UNESCO, AI Competency Framework for Students. Paris: United Nations Educational, Scientific and Cultural Organization, 2024. Available: https://www.unesco.org/en/articles/ai-competency-framework-students

[5] UNESCO, AI Competency Framework for Teachers. Paris: United Nations Educational, Scientific and Cultural Organization, 2024. Available: https://www.unesco.org/en/articles/ai-competency-framework-teachers

[6] Association of American Medical Colleges, "Artificial intelligence competencies across the learning continuum," AAMC, Washington, DC. Available: https://www.aamc.org/about-us/medical-education/ai-competencies. Accessed Sep. 10, 2026.

[7] J. Car, Q. C. Ong, T. Erlikh Fox et al., "The Digital Health Competencies in Medical Education Framework: an international consensus statement based on a Delphi study," JAMA Netw. Open, vol. 8, no. 1, e2453131, 2025, doi: 10.1001/jamanetworkopen.2024.53131.

[8] Association of Schools and Programs of Public Health, Harnessing Innovation: Artificial Intelligence for Academic Public Health: A Framework for the Responsible and Ethical Use of Artificial Intelligence in Public Health Education, Research, and Practice. Washington, DC: ASPPH, 2026. Available: https://aspph.org/initiatives/ai-for-public-health/ai-framework-report/. Accessed Sep. 10, 2026.

[9] C. Yip, L. C. Rosella, S. Nayani, C. Sanmartin, D. Hennessy, L. Faye, and A. Lvin, Laying the Groundwork for Artificial Intelligence to Advance Public Health in Canada. AI4PH, Statistics Canada, and Public Health Agency of Canada, 2024. Available: https://ai4ph-hrtp.ca/wp-content/uploads/2024/06/AI-to-advance-public-health-in-Canada-2024-FINAL.pdf

[10] National Collaborating Centres for Public Health, "Core competencies for public health in Canada: Release 2.0," NCCPH, 2025. Available: https://nccph.s3.amazonaws.com/uploads/2025/09/Core-Competencies-2.0-EN-Aug29.pdf

[11] Health Education England and University of Manchester, "Artificial Intelligence and Digital Healthcare Technologies Capability Framework," HEE and University of Manchester, London/Manchester, 2023. Available: https://digital-transformation.hee.nhs.uk/support-for-organisations/research-and-publications/dart-ed/horizon-scanning/ai-and-digital-healthcare-technologies

[12] First Nations Information Governance Centre, "The First Nations Principles of OCAP®," FNIGC, Ottawa, ON. Available: https://fnigc.ca/ocap-training/. Accessed Aug. 11, 2026.

[13] W. H. Walters and E. I. Wilder, "Fabrication and errors in the bibliographic citations generated by ChatGPT," Sci. Rep., vol. 13, art. 14045, Sep. 2023, doi: 10.1038/s41598-023-41032-5.

[14] N. Norori, Q. Hu, F. M. Aellen, F. D. Faraci, and A. Tzovara, "Addressing bias in big data and AI for health care: A call for open science," Patterns, vol. 2, no. 10, 100347, Oct. 2021, doi: 10.1016/j.patter.2021.100347.

[15] World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva, Switzerland: World Health Organization, 2021, ISBN 9789240029200. Available: https://www.who.int/publications/i/item/9789240029200

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Human-Curated AI-Assisted Scenario Design for Game-Based Sustainability Learning: A Design-Oriented Case Study

Teodor Daskalov, Neviana Krasteva

Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

Article DOI: 10.68286/samyrad2026.345

ABSTRACT

This paper presents a design-oriented case study of a human-centered model integrating artificial intelligence (AI) into game-based sustainability education. Unlike dominant AI-in-education approaches focused on automation, adaptive tutoring, or learner modeling, this study positions AI exclusively as a preparatory tool for scenario generation, while preserving full pedagogical authority for human facilitators. The model was implemented in a non-formal scout-based learning environment with 24 adolescents working in cooperative teams. Participants engaged in a tabletop simulation addressing water stress and interdependencies among forest ecosystems, water bodies, and urban systems. Descriptive pre–post differences were observed in ecological reasoning, collaborative engagement, and related learner-response indicators. Cognitive scores increased across all measured indicators, while behavioral data showed higher participation, initiative, and decision justification over time. Findings suggest that structured scenario variability - generated with AI during design and curated by educators before implementation - can be integrated into socio-ecological learning while preserving human pedagogical control. The study introduces a conceptual framework for "AI-as-design-augmentation," emphasizing human-centered control in educational design. Despite its exploratory nature, results support the feasibility of AI-assisted scenario design for sustainability education.

KEYWORDS artificial intelligence in education, game-based learning, sustainability education, water stress, serious games, systems thinking, non-formal learning

How to cite: Daskalov, T., & Krasteva, N. (2026). Human-curated AI-assisted scenario design for game-based sustainability learning: A design-oriented case study. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.345

Introduction

Water stress is an increasingly complex socio-ecological challenge shaped by climate change, urbanization, ecosystem degradation, and uneven resource distribution. It should be understood not as a simple resource shortage but as a systemic imbalance involving ecological, infrastructural, and governance interactions [1]. This complexity requires educational approaches that extend beyond factual knowledge toward systems thinking and causal reasoning. Non-formal education environments, such as scouting programs, provide suitable conditions for experiential and collaborative learning due to their emphasis on teamwork, responsibility, and action-based engagement [2]. Game-based learning further strengthens these conditions by embedding learning in structured decision-making systems that simulate real-world complexity [3]. At the same time, AI is increasingly used in education for adaptive instruction, analytics, and content generation [4], [5]. Much of this work treats AI as an instructional, adaptive, or automated component. The present study examines a narrower design-stage role: AI generates candidate scenario variability, while

human facilitators retain responsibility for selection, revision, pedagogical alignment, and final authorization of all learner-facing content.

The intervention is based on a cooperative simulation titled "Let Us Save the Forest", modeling interactions among forest, water, and urban systems under stress conditions. Rather than testing the causal effect of AI on learning, the study examines a documented human-curated design process in which AI expands the space of possible scenarios without receiving pedagogical authority and then describes implementation patterns observed in the resulting simulation.

The study makes three contributions. First, it proposes a human-curated AI-assisted scenario design framework that explicitly separates AI-based variability generation from pedagogical authorization. Second, it documents a transparent curation workflow, including generation, screening, adaptation, and retention of scenario materials. Third, it provides descriptive evidence from a non-formal sustainability simulation illustrating how such a workflow can be implemented without attributing learner outcomes causally to AI.

Research questions

RQ1. How can a human-curated AI-assisted workflow be structured to generate scenario variability while preserving full human pedagogical authority?

RQ2. How are AI-generated scenario elements filtered, adapted, and retained through human pedagogical curation?

RQ3. What descriptive patterns of systems-oriented reasoning, collaboration, and learner engagement are observed during implementation of the resulting simulation?

Background and related work

Water stress is widely conceptualized as a socio-ecological system outcome rather than a purely hydrological phenomenon. It emerges from interactions among natural processes, land use, governance structures, and consumption patterns. United Nations [1] emphasizes the importance of capturing feedback loops and multi-level dependencies in education. Non-formal education supports experiential learning and learner agency, defined as the ability to actively shape learning processes [6]. Scout-based environments are particularly suitable for sustainability education due to their emphasis on cooperation and real-world problem-solving [2].

Game-based learning integrates educational content into structured systems of rules and feedback, enabling learners to engage with dynamic problem spaces [3]. Gamification and simulation-based learning should be distinguished analytically: gamification introduces game-design elements into non-game contexts [7], whereas the present intervention uses a rule-based simulation in which learner decisions modify an interconnected system state. AI in education is primarily associated with adaptive systems, automation, analytics, and increasingly content generation [4], [5]. These applications raise questions about transparency and the location of pedagogical decision-making. For the present study, the relevant distinction is therefore not simply between AI use and non-use, but between AI as an instructional agent and AI as a bounded design-support mechanism operating before learner interaction. Recent work has also examined the intersection of artificial intelligence and gamification, including adaptive, personalized, and feedback-oriented uses of AI in gamified educational environments [8]. This literature demonstrates the growing integration of AI into gamification, but its primary focus differs from the present study. Rather than using AI to adapt the learner experience or make learner-facing decisions, the present study restricts AI to the preparatory generation of candidate scenario materials, with final pedagogical selection and authorization remaining exclusively with human facilitators.

These studies establish an important broader foundation for human–AI collaboration, instructional-design support, and human oversight in educational contexts [9]–[11]. Recent work also demonstrates the growing integration of AI and gamification, particularly through adaptive and learner-facing applications [8]. The contribution of the present study is narrower and procedural rather than a claim of novelty for human-in-the-loop AI itself. It operationalizes human oversight as an explicit separation between AI-based candidate generation and human pedagogical authorization within a game-based sustainability-learning workflow. AI is used to expand the pool of possible scenarios and decisions, whereas facilitators independently screen, adapt, and authorize the materials before they become part of the learner-facing simulation. The study therefore contributes a documented curation procedure and a pilot illustration of how bounded, design-stage AI assistance can be integrated into scenario-based sustainability learning while preserving human pedagogical authority.

This distinction defines the research gap addressed here. The study focuses on a workflow in which AI-generated outputs remain candidate design material until they are reviewed and authorized by human facilitators. The contribution is therefore centered on the separation of scenario generation from pedagogical authority and on the documentation of human curation as an explicit stage of educational design.

Human-curated AI-assisted design framework

The proposed framework integrates three components: systems thinking, simulation-based learning, and AI-assisted scenario variability. Learning takes place within a coupled forest–water–urban system where each subsystem interacts through feedback loops and threshold effects. AI-assisted variability is treated as an expansion of the design space rather than as instructional agency (Table 1).

AI operates strictly at the design level, generating candidate scenarios and decision options. Human facilitators select, revise, reject, and structure these outputs according to pedagogical goals before any content enters the learner-facing simulation. Learners then engage only with human-authorized scenarios through iterative cycles of decision-making, feedback interpretation, and reflection. The proposed framework is illustrated in Fig. 1. This separation ensures that AI contributes scenario diversity without receiving instructional authority. In the proposed framework, AI output is treated as candidate design material rather than educational content until it has been reviewed and authorized by a human facilitator.

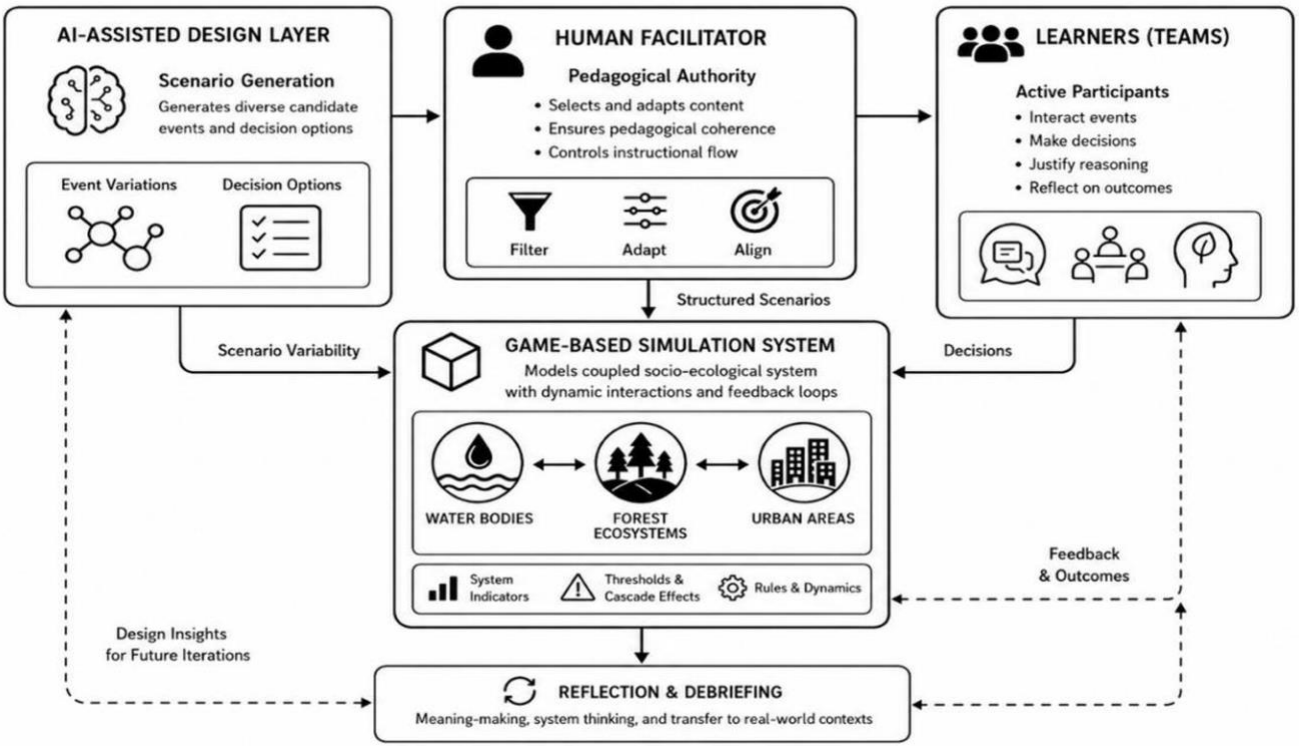


Fig. 1. Human-Curated AI-Assisted Scenario Design Framework for Game-Based Sustainability Learning.

Table 1. Role Distribution in the Human-Curated Design Framework.

Component	Function	Control
AI System	Generates candidate scenario variations	No pedagogical authority
Facilitator	Selects, revises, and authorizes content	Full pedagogical authority
Game system	Simulates ecosystem dynamics	Rule-based structure
Learners	Make decisions and reflect	Active participants

Methods or analytical approach

An exploratory design-oriented pilot case study with descriptive pre–post and structured observational components was used to examine feasibility and the design process rather than establish causality. The study involved 24 adolescents (ages 12–16) organized into eight teams of three. All participants were members of a scout organization with prior experience ranging from 5 months to 2 years. The study was not designed to estimate an independent causal effect of AI on learner outcomes.

Implementation was organized in two consecutive sessions involving the same eight teams of three participants. Session 1 used the basic scenario set (events E01–E12; decisions D01-D10) to familiarize participants with the rule system and the core forest–water–urban interdependencies. Session 2 used an extended scenario set (events E01–E24; decisions D01–D16), which incorporated the Session 1 materials together with additional scenarios and decisions to increase scenario variety and complexity. Thus, some scenario and decision content was repeated across sessions, while the second session introduced additional material. Because the same teams participated in both sessions, increasing familiarity with the game rules and task structure may have contributed to observed differences between earlier and later assessments. Decisions and assessments were organized collectively, so the primary analytical unit was the small team (n = 8), rather than the individual participant.

Simulation design

The simulation “Let Us Save the Forest” models interactions among three ecological subsystems using indicators on a 1–10 scale: Water Bodies, Forest Ecosystems, and Urban Areas. Each of eight gameplay rounds includes scenario interpretation, decision-making, and system updates.

A Sustainability Index was calculated as the mean of the three subsystem values, providing an internal summary of the simulated system state rather than a validated ecological measurement. Water stress was triggered when Water Bodies reached 3 or below. Once activated, at the end of each subsequent round Forest Ecosystems and Urban Areas each decreased by 1 point unless the selected decision included restoring or retaining water. The same decision could not be used in two consecutive rounds. These rules produced cascading system effects, delayed consequences, and constrained repetition of strategies. The basic configuration began with Water Bodies, Forest Ecosystems, and Urban Areas each set to 5, yielding an initial Sustainability Index of 5.00. The second session used a more difficult predefined starting configuration; because the preserved summary does not retain the complete

numerical starting configuration for that session, its starting Sustainability Index is not reconstructed here. The second session incorporated the materials from Session 1 together with additional events and decisions, thereby increasing scenario variety and complexity rather than constituting an entirely separate scenario system. Within each round, one event card was drawn without replacement from the active session set, so the same event could not recur within that session; its indicator effects were then applied; the team then discussed the resulting system state, selected one decision from the available decision set, applied its effects, and checked the water-stress threshold before advancing to the next round. The three system indicators are summarized in Table 2.

Table 2. System Indicators

Indicator	Description	Scale
Water Bodies	Water availability and quality	1–10
Forest Ecosystems	Ecological resilience and stability	1–10
Urban Areas	Urban system stability	1–10

AI-assisted scenario generation and human curation

AI was used only in the preparatory phase to generate candidate ecological events and decision options. Preliminary scenario materials were generated in ChatGPT using the GPT-5.4 Thinking model (OpenAI) via the standard web interface, accessed on 5 March 2026. The same model configuration was used across six prompting iterations. Across these iterations, 32 event drafts and 24 decision drafts were produced; 24 events and 16 decisions were retained after human review, corresponding to retention rates of 75.0% and 66.7%, respectively.

Prompting and curation procedure

The prompting procedure followed two recurring patterns: (a) event-generation prompts requesting short, age-appropriate situations representing natural, anthropogenic, or mixed pressures on the forest–water–urban system; and (b) decision generation prompts requesting ecologically meaningful response options compatible with the three-indicator rule structure. Outputs were screened for thematic relevance, internal logical consistency, age appropriateness, comprehensibility, sufficient variability, strategic balance, and compatibility with the rule system. Materials were excluded when they were overly generic, duplicated already retained scenario logic, were strategically unbalanced or pedagogically unclear, or produced effects too extreme to be meaningfully represented within the 1–10 indicator structure. Retained outputs were pedagogically edited by standardizing wording, calibrating difficulty, aligning effect structures across cards, and removing ambiguities in causal relations.

Representative prompt templates

To illustrate the structure of the prompting approach, two representative prompt templates are provided below. These templates reflect the two recurring prompt functions used in the scenario-generation process; they are presented as representative templates rather than as verbatim records of individual prompts. Event-generation template: “Generate a short, age-appropriate ecological scenario involving interactions among water bodies, forest ecosystems, and urban areas. The scenario should represent a natural, anthropogenic, or mixed environmental pressure and should be compatible with a 1–10 indicator system. Specify the affected indicator(s) and indicate the direction and magnitude of the expected effect.”

Decision-generation template: “Generate an ecologically meaningful decision option for a team-based simulation involving water bodies, forest ecosystems, and urban areas. The decision should provide a strategically relevant response to an environmental scenario, have identifiable consequences for the three indicators, and remain compatible with the existing water-stress and system-update rules.”

Accordingly, AI functioned as a creativity-support and variability-generation mechanism, while instructional design and final content authorization remained human-controlled. No AI-generated scenario or decision option was treated as learner-facing educational content solely on the basis of generation.

For procedural transparency, the scenario-generation record is specified as follows: AI system/model: ChatGPT, GPT-5.4 Thinking (OpenAI); interface: standard web interface; access date: 5 March 2026; number of generation iterations (n = 6); number of facilitators conducting curation (n = 3); review mode: independent. Each facilitator screened candidate items independently against the stated criteria. The archived record does not preserve item-level curator votes or a formal majority-vote or unanimity threshold; therefore, the manuscript reports the three independent reviews and the final retained set without claiming a voting rule. Because the archived curation record does not preserve the item-level ratings assigned independently by each facilitator, an inter rater agreement coefficient cannot be calculated retrospectively. Retained materials were implemented through two scenario sets rather than assigned as fixed card sequences to individual rounds: Session 1 used E01–E12 and D01–D10, while Session 2 used E01–E24 and D01–D16; in each round one event was drawn without replacement, preventing event repetition within the same session, and one decision was selected by the team.

Data collection and descriptive analysis

Data sources included system performance indicators, the Results Assessment Sheet, the Observation Assessment Form, and structured facilitator observations. Participant discourse was used as contextual observational evidence for reasoning patterns rather than as a separately reported qualitative dataset. The Results Assessment Sheet recorded the start, end, and change in the three system indicators and the overall Sustainability Index, together with three cognitive outcomes: understanding of water stress, understanding of the role of forests and ecosystem services, and sustainable strategy choice. Cognitive outcomes used a 1–3 rubric, where 1 indicated limited or

inaccurate understanding, 2 partial understanding, and 3 a clear, accurate, and convincing explanation supported by the post-game discussion.

The Observation Assessment Form was completed in real time at intervals of approximately 10 minutes and used a 1–5 scale for participation, initiative, collaboration, ecological reasoning, and decision justification. The supporting rubric distinguished between almost absent participation or collaboration at the lower end of the scale and active, consistent, and well-argued participation, initiative, collaboration, and systems-oriented reasoning at the upper end. Behavioral dimensions included participation, initiative, collaboration, ecological reasoning, and decision justification. Two facilitators independently rated observations using predefined criteria after prior rubric calibration. The reported >85% refers to the aggregate percentage agreement recorded for paired facilitator ratings. Because the archived summary does not preserve the raw paired ratings or the exact denominator, the calculation cannot be reconstructed more precisely; accordingly, agreement is reported descriptively and no chance-corrected inter-rater reliability coefficient is claimed. Given the ordinal 1–3 cognitive scale and the exploratory case-study design, pre–post results are reported descriptively as initial, final, and change scores. No normalized-gain or causal-effect estimate is used in the revised analysis.

All descriptive values were aggregated at the team level (n = 8), consistent with the collaborative organization of the intervention. Each cognitive construct was represented by one rubric-scored indicator. For each cognitive indicator, the reported Initial and Final values are the arithmetic means of the eight team-level rubric scores obtained from the pre-game and post-game discussions, respectively. The reported Initial and Final behavioral values are the arithmetic means of the eight team-level scores recorded at the first and final structured observation points, respectively. For each team, the Sustainability Index was calculated as (Water Bodies + Forest Ecosystems + Urban Areas) / 3. For Session 1, the predefined starting value was therefore 5.00, based on the three initial indicators set at 5. The reported overall final value of 6.13 is the arithmetic mean of the eight teams' final Sustainability Index values. Because Session 2 used a different predefined starting configuration, its numerical baseline is not reported here. The available summary data did not include dispersion statistics; accordingly, descriptive means are reported without SD or IQR estimates.

The final methodological sequence was: AI-assisted scenario design/generation; facilitator screening, adaptation, and authorization; initial team assessment before gameplay; Session 1 implementation with structured observations; post-game reflection and assessment; Session 2 implementation with the extended scenario set and structured observations; and final team assessment after the second session. Accordingly, the reported cognitive pre–post comparison refers to the initial assessment conducted before gameplay and the final assessment conducted after completion of the second session. The behavioral comparison refers to the first and final structured observation points across the two-session implementation. Because the same teams participated throughout both sessions, these comparisons represent descriptive changes over the overall learning sequence and cannot isolate the effects of the extended scenario set or AI assisted scenario design.

Results and findings

Human-curated scenario design

The curation process resulted in the retention of 24 of the 32 candidate event drafts (75.0%) and 16 of the 24 candidate decision drafts (66.7%). Thus, 8 event drafts and 8 decision drafts were not retained for implementation. These retention patterns provide descriptive evidence of the role of human curation: AI expanded the pool of candidate content, whereas human facilitators determined which materials entered the educational simulation. The retained materials were subsequently standardized and pedagogically adapted by the facilitators, including adjustments to wording, difficulty, card effect structures, and causal clarity. Because item-level records do not preserve the frequency of individual rejection reasons or distinguish systematically between unchanged and adapted retained items, the results are reported as aggregate retention outcomes rather than as item-level modification statistics.

Simulation implementation

Across the two consecutive sessions, each structured around eight gameplay rounds, system performance was examined descriptively. The arithmetic mean of the eight teams' final Sustainability Index values was 6.13. For reference, the basic configuration used in Session 1 started at an Index of 5.00. Because Session 2 used a different predefined starting configuration, the overall final value of 6.13 is reported as a descriptive final system-state value and is not interpreted as a cross-session change from the initial configuration. Facilitator observations suggested more coordinated strategies and more stable system regulation in later gameplay than in earlier gameplay. Because round-level quantitative values are not available here, this pattern is treated as contextual observational evidence rather than as a quantified temporal trend or an independent effect of AI-assisted scenario generation.

Descriptive learner outcomes

Table 3 presents the descriptive cognitive outcomes. Facilitator observations suggested that participants increasingly connected forest conditions with water regulation and urban system stability. This is treated as contextual evidence compatible with systems-oriented reasoning, not as a direct measurement of a validated systems-thinking construct.

Table 3. Descriptive Cognitive Outcomes (1-3 Scale)

Indicator	Initial Score	Final Score	Change
Water stress understanding	1.0	2.5	+1.5
Forest ecosystem role	1.0	2.625	+1.625
Sustainable decision-making	1.0	2.0	+1.0

Because the measures are descriptive and the design lacks a comparison condition, the observed pre–post differences are not attributed independently to AI. At the pre-game assessment, all eight teams received the minimum score of 1 on each of the three cognitive indicators. This pattern reflects the scoring rule applied to the initial team discussion: responses that showed limited or inaccurate understanding were assigned 1, whereas scores of 2 and 3 required progressively more explicit and accurate explanations. The uniform baseline should therefore be interpreted descriptively rather than as evidence that the rubric was highly discriminative at the lower end of the scale. Because all teams started at the same minimum score, the observed pre–post differences should likewise be interpreted cautiously and descriptively rather than as evidence of a quantified treatment effect.

All recorded behavioral indicators were higher at the final observation than at the initial observation. These descriptive differences are presented in Table 4 and are interpreted as learner-response patterns observed during implementation of the human-curated simulation.

Table 4. Descriptive Behavioral Outcomes (1-5 Scale)

Indicator	Initial Score	Final Score	Change
Participation	2.0	4.0	+2.0
Initiative	3.0	4.5	+1.5
Collaboration	3.0	4.5	+1.5
Ecological reasoning	2.0	4.0	+2.0
Decision justification	2.0	4.0	+2.0

Facilitator observations suggested more explicit use of system relations and decision justification in later gameplay. This pattern is reported descriptively and is not interpreted as evidence of a measured developmental transition.

Discussion

The central contribution of this case study is the separation between AI-based generation and human pedagogical authorization. AI expanded the range of candidate scenarios and decision options, but it did not select instructional content, interact with learners, assess performance, or determine pedagogical goals. This design makes human curation a constitutive stage of the educational process rather than a post hoc correction.

The simulation component remains consistent with prior work emphasizing experiential feedback, structured decision-making, and learner agency in game-based and collaborative environments [3], [6]. In the present framework, however, these learning processes are analytically separated from the AI contribution: the AI contributes candidate variability at the design stage, while the educational mechanism is implemented through the human-authorized simulation. Unlike AI-in-education applications centered on automation or adaptive instructional functions [4], [5], the present framework constrains AI to design augmentation. AI-generated outputs are therefore provisional rather than pedagogically authoritative. The observed retention rates further illustrate this principle: a meaningful share of generated material was not carried forward, demonstrating that generation and educational authorization were separate decisions. Human curation should not be treated as evidence of validity. It functions as a governance safeguard, but facilitator review can still reproduce ecological oversimplifications, inconsistent assumptions, or normative biases. Future implementations should therefore document validation criteria and review decisions and, where feasible, include independent expert checking of ecological content. The descriptive learner patterns suggest that the resulting simulation was workable as a collaborative learning environment. Repeated decision cycles were accompanied by higher final descriptive scores for participation, collaboration, ecological reasoning, and decision justification, while cognitive scores were also higher at final than at initial measurement. These observations support the feasibility of the overall human-curated design, but they cannot isolate the contribution of AI-generated variability from the effects of simulation, facilitation, repeated practice, teamwork, or novelty.

The study is limited by its small sample, absence of a control or comparison condition, single non-formal educational setting, potential novelty effects, and reliance on facilitator-rated measures. The cognitive and behavioral rubrics used in the present pilot were study-specific descriptive instruments rather than validated standardized measures. The available summary data did not preserve the raw paired ratings or provide dispersion statistics; consequently, formal inter-rater reliability beyond descriptive percentage agreement could not be estimated, and variability around the reported team-level means could not be assessed using SD or IQR. The reported values should therefore be interpreted as aggregate descriptive evidence rather than as individual-level statistical estimates. The same facilitators were involved in implementation and observation, which may have introduced observer or expectancy effects, and no long-term retention measure was included. In addition, because the same eight teams participated in both consecutive sessions and the second session used the more complex extended scenario set, part of the observed improvement may reflect practice effects and increasing familiarity with the game rules and task structure. Future work should compare human-only scenario design with human-curated AI-assisted scenario design in order to preserve raw paired ratings and dispersion statistics, estimate inter-rater reliability from the underlying observations, and assess whether any observed differences persist over time.

Conclusion

This study introduced a human-curated AI-assisted scenario design framework for sustainability education. Its defining feature is a three-stage separation between AI generated variability, human pedagogical curation and authorization, and simulation-based learning execution. AI was used to expand the pool of possible scenarios, while all pedagogical decisions remained under human control. Descriptive cognitive and behavioral changes observed during implementation support the practical feasibility of the overall design, but they do not establish an independent causal effect of AI. The stronger contribution of the case study is therefore procedural and conceptual: it shows how AI can be bounded to a design-support role without acquiring pedagogical authority. Future research should

validate the framework across larger and more diverse educational contexts and directly compare human-only scenario design with human curated AI-assisted design. Such studies should also test long term learning retention and examine whether structured human curation can preserve transparency and pedagogical accountability while benefiting from AI-generated scenario diversity.

Declarations

Ethics approval and consent: Under the applicable local institutional context, the activity was treated as a low-risk non-formal educational practice and did not require separate institutional ethics committee approval. Participation was voluntary; participants and their guardians were informed of the educational and research character of the activity, and prior parental/guardian consent was obtained. Data were minimized and anonymized and were not used for formal assessment. The AI component operated only in preparatory scenario design: it did not process participant profiles, create individual profiles, classify learners, or make automated pedagogical, assessment-related, or organizational decisions concerning participants.

AI use disclosure: The authors used ChatGPT and NotebookLM during manuscript preparation for language editing, grammar correction, text polishing, and clarity. Perplexity was used as a search aid to identify candidate publications; every cited source and bibliographic entry was independently checked by the authors against the original publication. The design-stage use of ChatGPT (GPT-5.4 Thinking) to generate candidate scenario items formed part of the research procedure and is reported transparently in the Methods. The authors take full responsibility for the manuscript, data, interpretations, and conclusions.

Acknowledgment

This research was funded by the Ministry of Education and Science (MES) under the National Science Fund, Contract No. КП-06-Н85/3 (05.12.2024).

References

[1] United Nations, The United Nations World Water Development Report 2024: Water for Prosperity and Peace. Paris, France: UNESCO, 2024. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000388948
[2] World Organization of the Scout Movement, “Introduction to the Scout Method,” World of Scouting Learning Zone, Mar. 15, 2023. [Online]. Available: https://learn.scout.org/resource/introduction-scout-method
[3] J. L. Plass, B. D. Homer, and C. K. Kinzer, “Foundations of game-based learning,” Educational Psychologist, vol. 50, no. 4, pp. 258–283, 2015, doi: 10.1080/00461520.2015.1122533.
[4] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education—Where are the educators?,” International Journal of Educational Technology in Higher Education, vol. 16, Art. no. 39, 2019, doi: 10.1186/s41239-019-0171-0.
[5] H. Crompton and D. Burke, “Artificial intelligence in higher education: The state of the field,” International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 22, 2023, doi: 10.1186/s41239023-00392-8.
[6] OECD, “Conceptual learning framework: Student agency for 2030,” 2019. [Online]. Available: https://www.oecd.org/education/2030project/teaching-andlearning/learning/studentagency/Student_Agency_for_2030_concept_note.pdf
[7] S. Deterding, M. Sicart, L. Nacke, K. O’Hara, and D. Dixon, “Gamification: Using game-design elements in non-gaming contexts,” in CHI ’11 Extended Abstracts on Human Factors in Computing Systems. New York, NY, USA: ACM, 2011, pp. 2425–2428, doi: 10.1145/1979742.1979575.
[8] P. N. Adi, T. Köhler, M. B. Triyono, Priyanto, S. A. Handayani, and R. F. Maruanaya, “AI-enhanced gamification in education: An integrative review of trends, impacts, and corrective role potential,” Journal of Computers in Education, advance online publication, 2026, doi: 10.1007/s40692-026-00387-0.
[9] D. Ifenthaler, R. Majumdar, P. Gorissen, M. Judge, S. Mishra, J. Raffaghelli, and A. Shimada, “Artificial intelligence in education: Implications for policymakers, researchers, and practitioners,” Technology, Knowledge and Learning, vol. 29, pp. 1693–1710, 2024, doi: 10.1007/s10758-024-09747-0.
[10] P. N. Kibar and H. Ilgaz, “The intersection of artificial intelligence and instructional design practice: A systematic review,” Educational Technology Research and Development, advance online publication, 2026, doi: 10.1007/s11423-026-10624-z.
[11] T. Li, J. S. Krajcik, and R. Spiro, “The transformative collaboration of human intelligence and artificial intelligence in designing knowledge-in-use science assessment for learning,” Journal of Science Education and Technology, vol. 35, pp. 688–714, 2026, doi: 10.1007/s10956-025-10275-4.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

From Static LMS to Autonomous Agents: Reimagining Moodle-Based Open and Distance e-Learning Delivery for the Agentic AI Era in Kenya

Sadati Okeyo[1], Joel Nakitare[2]

(email. sokeyo@umma.ac.ke), (email. jnakitare@rongovarsity.ac.ke)

[1]Systems Librarian Umma University Kajiado, Kenya

[2]University Librarian Rongo University Rongo, Kenya

Article DOI: . https://doi.org/10.68286/samyrad2026.348

ABSTRACT

Artificial intelligence (AI) in education has progressed from generative content tools toward agentic systems capable of autonomous planning, decision-making and multi-step task execution. Learning Management Systems (LMS) such as Moodle have underpinned Open and Distance e-Learning (ODeL) delivery across Kenyan higher education institutions, yet their architecture remains largely static, offering limited capacity for autonomous, adaptive learner support. This study addressed a critical gap: the absence of empirical evidence on institutional readiness for agentic AI integration within Moodle-based ODeL ecosystems in Kenya. A descriptive cross-sectional survey was administered online to respondents drawn from institutional leadership, ICT/eLearning directors, ODeL coordinators and registered ODeL students at Kenyan higher education institutions offering ODeL. Seventeen (N = 17) respondents took part, with question-level completion ranging from 8 to 17 owing to conditional branching and voluntary attrition. Findings show that 88.9% of respondents whose institutions had customised Moodle reported moderate-to-full customisation, and 87.5% reported that their institution had piloted or implemented at least one adaptive or automated function, most commonly automated plagiarism detection (100%) and adaptive learning paths (71.4%); because current implementations cluster around integrity and pathway automation rather than autonomous, goal-directed agents, these findings are best read as evidence of preparedness for a prospective transition to agentic AI rather than of its current widespread deployment. A Fisher's exact test found no statistically significant association between perceived ICT budget adequacy and adoption of adaptive functionality ($p = 1.00$), and a Welch's independent samples-t-test found no statistically significant difference in perceived institutional readiness between staff ($M = 3.25$, $SD = 1.71$) and students ($M = 4.50$, $SD = 0.58$), $t(3.68) = -1.39$, $p = .244$, although the small sample severely limits statistical power. Infrastructural constraints, weak institutional AI policy, and equity and data-governance concerns emerged as the principal qualitative barriers. The study proposes a four-domain conceptual readiness framework, specified a priori from the technology-adoption and governance literature (infrastructure, policy and governance, human capability, and equity), to guide Kenyan ODeL institutions in transitioning from static LMS infrastructure toward agentic-AI-driven delivery, in alignment with Sustainable Development Goal 4. Given the modest and private-institution-skewed sample, the findings are best read as an exploratory baseline rather than a population-representative estimate, and the paper closes with a call for a larger, multi-institution replication.

KEYWORDS agentic AI, Moodle , ODeL, learning management systems, Kenya, higher education, institutional readiness

How to cite: Okeyo, S., & Nakitare, J. (2026). From static LMS to autonomous agents: Reimagining Moodle-based open and distance e-learning delivery for the agentic AI era in Kenya. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.348

Introduction

The global higher-education technology landscape has shifted decisively from generative AI tools that respond to prompts toward agentic AI systems that plan, act and adapt with minimal human supervision [1], [4]. Agentic AI is distinguished from earlier intelligent tutoring systems and conversational assistants by its capacity to pursue multi-step, long-horizon goals, coordinate tools and agents, and adjust its own plans as circumstances change [2], [4], [5]. In higher education, this is beginning to manifest as autonomous learning companions, adaptive assessment agents and proactive student-support agents that act on institutional data rather than merely displaying it [6].

Kenyan higher education institutions have widely adopted Moodle as the backbone of Open and Distance e-Learning (ODeL) delivery, a trend accelerated by the COVID-19 pandemic and by continuing demand that outstrips the capacity of conventional face-to-face provision [7], [8], [9]. However, Moodle deployments in the region have historically served as repositories for content delivery and administrative workflow rather than as platforms for autonomous, adaptive learner support [7]. Whether Kenyan ODeL institutions currently possess the technical, human, policy and equity foundations required to responsibly adopt agentic AI within Moodle remains an open empirical question.

This study addressed that gap directly. It asked: (i) to what extent have Kenyan ODeL institutions customised Moodle and piloted adaptive or automated functionality; (ii) how adequate do stakeholders judge the ICT budget, infrastructure, policy and governance foundations for agentic AI; (iii) do perceptions of readiness differ between institutional staff and students; and (iv) what do stakeholders identify as the principal barriers to, and enablers of, a transition toward agentic-AI-driven Moodle delivery. The paper contributes an empirically grounded, exploratory baseline and a four-domain conceptual readiness framework intended to guide systems librarians, ICT directors and ODeL administrators as they plan this transition, in a manner consistent with Sustainable Development Goal 4 (quality education).

The contribution of this paper is threefold. It provides, to the authors' knowledge, an early empirically grounded readiness assessment of agentic AI within Moodle-based ODeL delivery specific to the Kenyan context, at a moment when agentic functionality is moving rapidly from pilot to mainstream deployment internationally [5], [6]. It demonstrates, transparently and with appropriate statistical caution, what a small-sample institutional survey can and cannot support, a methodological contribution relevant to systems librarians and ODeL administrators who often work with similarly constrained data-collection capacity. Finally, it offers a four-domain conceptual framework, specified a priori from the Technology, Organisation and Environment (TOE) tradition [14], [15] rather than induced from the survey data, that later, larger-scale studies can operationalise and test.

Related work

Agentic AI in Education

Agentic AI is consistently defined in the recent literature as AI that pursues complex, often long-horizon goals with minimal human intervention, planning, using tools, maintaining memory and self-adjusting to changing environments [1], [2]. This marks a qualitative shift from earlier intelligent tutoring systems, which operated within a fixed curriculum, and from generative AI assistants, which remain fundamentally reactive to user prompts [4]. Surveys of the field classify emerging applications into intelligent tutoring, adaptive assessment, autonomous learning companions and multi-agent orchestration layers that sit across an institution's existing systems [2], [3]. Commentators note that the shift from copilot to autopilot modes of operation raises new questions of trust, transparency and governance that are only beginning to be addressed in policy [1], [3], [5], and that adoption in the sector is accelerating rapidly even as organisational trust in autonomous systems lags behind investment [5], [6].

Moodle and ODeL in Kenya

Kenyan and regional studies confirm that Moodle is now a common LMS across the higher-education sector, frequently supplemented by institution-specific customisation [7]. Institutional-factors research on Kenyan e-learning adoption identifies organisational leadership support, staff readiness and resourcing as significant predictors of successful technology transitions [8], while national e-readiness surveys commissioned through the Kenya Education Network document persistent gaps in device-to-student ratios and ICT human capacity across the university sector [9]. Implementation-focused work further reports that device and connectivity access remain uneven, particularly at public universities, and recommends participatory planning involving leadership, staff and students, together with explicit institutional ODeL policy, as preconditions for successful technology transitions [10], a recommendation this study revisits in the specific context of agentic AI.

Data Governance and Equity

Agentic AI systems that act on learner data raise data protection obligations that did not attach to static LMS use in the same way. Kenya's Data Protection Act, 2019 requires lawful, transparent and accountable processing of personal data and establishes the Office of the Data Protection Commissioner (ODPC) as the enforcing authority [11]; the ODPC's sector guidance note for education explicitly extends these obligations to institutions offering remote e-learning [12]. Separately, equity scholarship on African distance education consistently documents an urban and rural digital divide in device ownership, bandwidth and consequently in the ability to benefit from technology-enhanced learning, with rural and historically disadvantaged students most exposed [13]. Any readiness assessment for agentic AI in an African ODeL context must therefore treat data governance and equity as first-order dimensions rather than afterthoughts.

Theoretical Grounding: Institutional Readiness Models

Institutional technology adoption is commonly theorised through the Technology, Organisation and Environment (TOE) framework, which holds that adoption outcomes are jointly shaped by technology characteristics (complexity, compatibility, relative advantage), organisational characteristics (leadership support, financial resources, internal readiness) and environmental characteristics (policy pressure, regulatory context, competitor behaviour) [14]. Applying TOE specifically to generative AI adoption in higher education, Kalmus and Nikiforova found that environmental and organisational constraints, including regulatory ambiguity and unclear institutional policy, shaped educator resistance even where individual attitudes toward the technology were otherwise favourable [15], echoing this study's own finding that policy and governance items lagged infrastructure items despite generally favourable individual attitudes. This study's four thematic blocks (infrastructure, budget/organisation, policy/environment, and equity) were specified a priori from this literature, before data collection, and are used to structure the proposed conceptual framework as a literature-derived contribution rather than a pattern induced from the survey data.

AI Governance and Accountability in Higher Education

Agentic systems complicate conventional accountability structures because a single student-facing outcome may result from a multi-step, tool-using pipeline rather than a single identifiable human decision [3], [16]. Emerging governance scholarship addresses this by proposing that a named human accountability owner be assigned to each deployed agentic workflow and that agents be granted scoped, time-limited permissions rather than static, standing access to personal data, extending established privacy-engineering principles to autonomous, tool-using systems [16]. These developments reinforce why this study treats a formal AI-in-teaching policy and an explicit accountability framework as first-order readiness indicators rather than administrative afterthoughts.

Methodology

Research Design

The study adopted a descriptive, cross-sectional survey design, appropriate for establishing a baseline profile of institutional readiness across multiple stakeholder perspectives at a single point in time. The design combined closed-ended items (single-choice, multiple-choice and five-point Likert matrices) with open-ended items to allow both quantifiable comparison and qualitative elaboration of barriers and support needs.

Population, Sampling and Participants

The target population comprised stakeholders at Kenyan higher education institutions that had formally embraced ODeL delivery, spanning four strata: institutional leadership (Deans/Directors of ODeL), ICT/eLearning directors, ODeL coordinators, and registered ODeL students. Purposive and snowball sampling within these strata were used to reach respondents through institutional ODeL and ICT departments. The instrument was administered online via SurveyMonkey between 20 June and 7 July 2026 and yielded seventeen respondents (N = 17): 11.8% from public institutions and 88.2% from private institutions, distributed across the four strata as registered ODeL students (53.3%), ICT/eLearning directors (20.0%), ODeL coordinators (20.0%) and institutional leadership (6.7%) among the fifteen respondents who disclosed a category. The pronounced private-institution skew is a material feature of the sample, not merely a demographic footnote, and is used in the Results Section (Budget, Infrastructure, Policy and Equity Perceptions) to explain an otherwise counter-intuitive result. The survey did not capture a unique institutional identifier, so N = 17 is reported as individual respondents rather than as a count of distinct institutions, and demographic detail beyond institution type and respondent category (e.g., years of experience, age) was not collected; both points are addressed in the Limitation Section.

Instrument

The questionnaire comprised eight thematic blocks: respondent and institution profile; Moodle customisation; adaptive/automated functionality already piloted; ICT budget and its adequacy; four five-point Likert agreement matrices covering (i) overall readiness, (ii) infrastructural barriers, (iii) policy and data-governance maturity, and (iv) equity; and three open-ended questions on barriers, support needed, and additional comments. The four Likert blocks correspond directly to the four framework domains specified a priori in the Related Work Section (Theoretical Grounding: Institutional Readiness Models) and the framework is later discussed as a literature-derived structure that the instrument operationalises, not as a finding the analysis selected among alternatives. Likert items were anchored 1 = Strongly disagree to 5 = Strongly agree. Skip logic meant later blocks were shown only to respondents who indicated relevant prior experience (e.g., the functionality-detail block was shown only to those who had piloted adaptive features), which explains the declining n across blocks (17 down to as few as 6 or 8) reported in the Results Section.

Data Analysis

Closed-ended responses were analysed descriptively (frequencies, percentages, means and standard deviations) using the response summaries generated in SurveyMonkey. Two inferential tests were run on the fully-completed Likert and categorical items (n = 8): a Welch's independent-samples t-test comparing perceived overall institutional readiness between staff (institutional leadership, ICT directors and ODeL coordinators pooled) and students, and a Fisher's exact test examining the association between perceived ICT budget adequacy (collapsed into adequate/neutral versus inadequate/very inadequate) and whether the institution had piloted adaptive functionality. Welch's t-test and Fisher's

exact test were selected in preference to the pooled-variance t-test and the standard Pearson chi-square test because both are more robust to the unequal variances and small expected cell counts (< 5) that characterise a sample of this size; a Mann-Whitney U test and an uncorrected chi-square are reported alongside as robustness checks. Degrees of freedom for the Welch's t-test were computed using the Welch-Satterthwaite equation rather than assumed from sample size, since the two are not interchangeable under unequal variances. All descriptive standard deviations are reported using the sample (n-1) formula for consistency with the inferential statistics. Open ended responses were analysed thematically by inductive coding of recurring terms. All quantitative analysis was performed in Python 3 using SciPy 1.x.

Ethical Considerations

Content validity was addressed by deriving each thematic block directly from constructs established a priori in the adoption and equity literature reviewed in the Related Work Section, namely TOE-aligned infrastructure, organisational and environmental items [14], [15], data-governance items aligned with the Kenya Data Protection Act, 2019 [11], [12], and equity items drawn from the African digital divide literature [13]. Given the exploratory nature and small sample of this study, formal internal-consistency statistics (e.g., Cronbach's alpha) were not computed for the Likert matrices, since such coefficients are unstable and can be misleading at n = 8; this is flagged explicitly as a target for the larger replication study recommended in the Future Research Directions Section. Face validity was supported by piloting the questionnaire informally with two ODeL practitioners prior to distribution, who confirmed that item wording was unambiguous in the Kenyan institutional context.

Validity and Reliability

Participation was voluntary and anonymous. The instrument's introductory page informed respondents that the survey was part of an academic research study, that responses were confidential, and that completion was expected to take twelve to fifteen minutes; no personally identifying information (name, email address, or institution name) was collected. Because the Content validity was addressed by deriving each thematic block directly from constructs established a priori in the adoption and equity literature reviewed in the Related Work, namely TOE-aligned infrastructure, organisational and environmental items [14], [15], data-governance items aligned with the Kenya Data Protection Act, 2019 [11], [12], and equity items drawn from the African digital divide literature [13]. Given the exploratory nature and small sample of this study, formal internal-consistency statistics (e.g., Cronbach's alpha) were not computed for the Likert matrices, since such coefficients are unstable and can be misleading at n = 8; this is flagged explicitly as a target for the larger replication study as recommended in 'Proposed conceptual readiness framework' section. Face validity was supported by piloting the questionnaire informally with two ODeL practitioners prior to distribution, who confirmed that item wording was unambiguous in the Kenyan institutional context did not record institutional identity, individual institutions cannot be singled out from the published results, which further protects respondent confidentiality but also constrains the analysis, as discussed in the Limitation Section. Respondents could exit the survey at any point, which accounts for the partial-completion pattern evident in the declining question-level n reported in the Results Section.

Results and findings

Respondent and Institutional Profile

Table 1 summarises the respondent and institutional profile; while the report of the completion count (n) for every item discussed below is presented later so that percentages computed over different respondent subsets remain traceable. The sample was dominated by private-institution respondents (88.2%) and by registered ODeL students (53.3%), a skew used to interpret an otherwise counter-intuitive result and flagged as a limitation of the study.

Table 1. Respondent and Institutional Profile

Element	Example value	n	%
Institution type (n=17)	Public	2	11.8
	Private	15	88.2
Respondent category (n=15) Leadership 1 6.7	Leadership	1	6.7
	ICT/eLearn. Dir.	3	20.0
	ODeL Coord.	3	20.0
	ODeL Student	8	53.3

traceable. The sample was dominated by private-institution respondents (88.2%) and by registered ODeL students (53.3%), a skew used in the Results Section (Budget, Infrastructure, Policy and Equity Perceptions) to interpret an otherwise counter-intuitive result and flagged as a limitation.

Moodle Use and Customisation

Of the ten respondents who answered, 90.0% confirmed Moodle as their institution's primary LMS for ODeL delivery. Among the nine who detailed the extent of customisation, 55.6% reported extensive customisation integrated with other systems, 22.2% moderate customisation (plugins added), 11.1% full customisation with AI-enhanced plugins, and 11.1% no customisation beyond default installation. Among the ten who itemised specific customisations (multiple selection permitted), additional plugins (80.0%), custom theming (70.0%) and API integrations (70.0%) were the most common, followed by analytics dashboards (60.0%), third-party integrations such as payment or SMS gateways (50.0%) and mobile app integration (40.0%); only 10.0% reported no customisation at all.

Adaptive and Automated Functionality

Among the eight respondents who answered, 87.5% reported that their institution had piloted or implemented at least one adaptive or automated learning function within Moodle, against 12.5% who had not. Table 2 shows the specific functionalities reported by the seven respondents who answered "yes" and went on to itemise which functionalities were in use; the deployment-stage question that follows was instead shown to all eight respondents who answered the piloting question, including the one respondent who had not piloted any functionality and who selected "discontinued use", which accounts for the differing n between these two items. Automated plagiarism detection was universal among adopters (100%), followed by adaptive learning paths (71.4%) and automated grading (57.1%); chatbot/virtual-assistant support and AI-generated content were each reported by only 14.3%, suggesting that current "adaptive" practice in Kenyan Moodle deployments still clusters around integrity and pathway tools rather than conversational or generative agentic functions, and that the readiness this paper assesses is preparedness for a prospective transition rather than evidence of agentic AI already being widely deployed. Of the eight respondents who indicated a deployment stage, 50.0% reported full institutional deployment, 25.0% partial rollout, 12.5% pilot/testing and 12.5% discontinued use (Table 2).

Table 2. Adaptive/Automated Functionalities Piloted (n=7, multi-select)

Functionality	%
Automated plagiarism detection	100
Adaptive learning paths	71.4
Automated grading	57.1
Predictive analytics for at-risk students	42.9
Chatbot / virtual assistant support	14.3
AI-generated content / quizzes	14.3

Budget, Infrastructure, Policy and Equity Perceptions

Budget visibility was low: of eight respondents, 62.5% (five) did not know what share of the institutional budget was allocated to ICT, and 75.0% (six) did not know whether a dedicated LMS/AI innovation budget line existed; the remaining 25.0% (two respondents) confirmed that a dedicated line did exist, and none reported affirmatively that no such line existed. Perceived budget adequacy for agentic AI specifically was mixed: 37.5% rated it neutral, 25.0% adequate, 25.0% very inadequate, and 12.5% inadequate, with none rating it very adequate.

Table 3 reports mean agreement (on a 1 to 5 scale, sample standard deviation) across the four Likert matrices (n = 8 throughout). Readiness-related items clustered in the neutral-to-agree range (M = 3.50 to 3.88), as did policy and governance items, with self-reported compliance with the Data Protection Act, 2019 rated highest overall (M = 4.13). Notably, the infrastructural-barrier items were rated closer to the disagree end of the scale (M = 2.25 to 3.00), indicating that respondents in this sample did not, on balance, regard connectivity, device shortages, hosting capacity, power outages or technical-support gaps as acute; this is a more optimistic picture than the barriers literature on Kenyan e-learning generally reports [8], [9], [10], and is most plausibly explained by the sample's 88.2% private institution skew (Table 1), since private institutions in Kenya are generally better resourced than the public universities on which that literature disproportionately reports, rather than by sample size alone. Equity items were more concerning: 75.0% agreed or strongly agreed that rural or remote students face disadvantages in accessing agentic AI features (M = 3.75), the highest mean of any equity item.

Table 3. Mean agreement by domain (1=strongly disagree, 5=strongly agree, n=8, sample SD)

Item	M	SD
Institution has technical infrastructure for agentic AI	3.63	0.75
Staff have adequate AI-management skills	3.75	0.89
Students prepared for autonomous AI tools	3.75	0.89
Institution has clear AI strategic plan	3.5	0.93
Leadership visibly committed to AI adoption	3.63	1.19
Confident institution can manage data-privacy risk	3.50	0.76
Overall: institution ready for agentic-AI Moodle	3.88	1.36
Unreliable internet hampers LMS use	2.75	1.17
Insufficient computing devices	2.50	1.20
Server/hosting capacity inadequate	2.25	1.28
Frequent power outages disrupt learning	3.00	0.93
Lack of dedicated technical support staff	2.25	1.04
Institution has formal AI-in-teaching policy	3.75	1.28
Clear data-governance guidelines for AI data	3.25	1.04
Complies with Kenya Data Protection Act (2019)	4.13	0.83
Clear accountability framework for AI decisions	3.63	0.92
Equal device/internet access for AI-enhanced learning	3.25	1.04
Students with disabilities adequately supported	3.00	1.30
Rural/remote students disadvantaged in AI access	3.75	0.46

Staff vs. Student Perceived Readiness

Restricting to the eight fully-completed responses, staff (institutional leadership, ICT directors and ODeL coordinators, n = 4) reported an overall-readiness mean of 3.25 (SD = 1.71; individual scores 1, 3, 4, 5), while students (n = 4) reported a higher mean of 4.50 (SD = 0.58; scores 4, 4, 5, 5). A Welch's independent-samples t-test, with degrees of freedom computed via the Welch-Satterthwaite equation rather than assumed from group size, found this difference

was not statistically significant, t(3.68) = -1.39, p = .244; a Mann-Whitney U test corroborated this, U = 4.0, p = .288. The mean difference was 1.25 points on the five-point scale, 95% CI [-1.34, 3.84], with a large point-estimate effect size (Hedges' g = 0.85), but the interval spans zero and covers both a negligible and a very large true effect, the expected signature of a real difference tested with too few observations to pin down, not evidence that no difference exists. The direction of the gap (students were somewhat more optimistic than staff) is the reverse of what is often assumed in the technology-adoption literature and of what the original study proposal anticipated, but with n = 4 per group and fewer than four whole degrees of freedom the test is severely underpowered and this result should be treated as inconclusive rather than as evidence of no true difference.

Budget Adequacy and Adoption of Adaptive Functionality

A 2×2 contingency table (n = 8) cross-tabulating collapsed budget adequacy (adequate/neutral vs. inadequate/very inadequate) against whether the institution had piloted adaptive functionality (yes/no) is shown in Table 4. A Fisher's exact test (preferred here because two of the four expected cell counts fell below 5) found no statistically significant association, p = 1.00, odds ratio = 0.00. An uncorrected Pearson chi-square test is reported for reference only, χ^2 (1, n = 8) = 0.69, p = .408; given the small expected counts this chi-square result should not be relied upon. In short, this sample provides no evidence that perceived budget adequacy is associated with whether adaptive functionality has been piloted; this null result is plausibly attributable to the very small sample rather than to a genuine absence of relationship, and it reverses the significant association hypothesised in the original study proposal.

Table 4. Budget Adequacy × Adaptive-Functionality Adoption (n=8)

Budget adequacy (collapsed)	Piloted: Yes	Piloted: No
Adequate / Neutral	4	1
Inadequate / Very inadequate	3	0

Qualitative Barriers and Support Needs

Six respondents identified the single greatest barrier to agentic AI adoption. Data privacy, security and regulatory-compliance concerns were the most frequently cited theme, followed by inadequate staff skills and technology apprehension, gaps in leadership attention and resourcing, general lack of awareness, and technical/systems failures. On support needed (n = 6), financial investment was named in almost every response, typically paired with technical capacity-building or training, and with calls for clearer AI governance policy. Additional open comments (n = 5) emphasised that adoption should be approached strategically and responsibly, with attention to staff readiness and data security, and that sensitisation ahead of adoption was necessary; one respondent offered no further comment and one described the topic positively without elaboration (Table 5).

Table 5. Sample Size (n) by Survey Item

Survey block	Piloted: No
Moodle as primary LMS	10
Extent of customisation	9
Specific customisations (multi-select)	10
Piloted adaptive functionality (yes/no)	8
Specific functionalities piloted (multi-select)	7
Deployment stage	8
ICT budget share / dedicated line / adequacy	8 each
Four Likert matrices (19 items)	8 each
Open-ended barriers / support / comments	6 / 6 / 5

Discussion

Three findings stand out. First, Kenyan ODeL institutions in this sample have already moved well beyond a default installation of Moodle: extensive customisation, API integrations and analytics dashboards are common, and a majority report having piloted at least one adaptive or automated function, most often plagiarism detection and adaptive pathways rather than conversational or generative agents. This suggests the technical substrate for agentic AI (APIs, integrations, data pipelines) is more developed in this sample than the static-LMS framing in prior literature might imply, though functionality remains concentrated in integrity and routing tools rather than autonomous, goal-directed agents as defined in the agentic AI literature [1], [2], [4]; the study therefore measures readiness for a prospective transition, not the extent of current agentic deployment.

Second, neither of the two inferential hypotheses tested here (a staff-student readiness gap and a budget-adoption association) was supported at conventional significance levels. This null result is itself informative: it cautions against assuming, without institution-level data, that budget alone drives adoption, or that frontline staff are systematically more sceptical than students. It also underscores a structural limitation of this study, discussed in the Limitation Section, rather than a substantive conclusion that can be generalised beyond this sample.

Third, the qualitative and equity data together paint a more cautious picture than the readiness means alone. Data governance and compliance concerns dominate the open-ended barrier responses even though self-reported Data Protection Act compliance was the single highest-rated item on the closed-ended matrices (M = 4.13); this is a tension that suggests respondents may distinguish formal legal compliance from confidence in operational data governance for autonomous systems specifically.

Similarly, 75% agreement that rural and remote students face disadvantages in accessing agentic AI features echoes the wider African ODeL equity literature [13] and signals that

agentic AI, if adopted without deliberate equity safeguards, risks widening rather than narrowing existing access gaps.

Read through a TOE lens, the findings suggest an uneven profile across the three dimensions [14], [15]. The technology dimension appears comparatively mature, evidenced by widespread API integration, analytics dashboards and third-party integrations. The organisational dimension is mixed: leadership commitment and staff skill were both rated moderately, yet budget visibility was strikingly low, with roughly two-thirds of respondents unable to say what share of the institutional budget supports ICT. The environmental dimension, comprising policy and regulatory context, was the most consistently flagged as unresolved in the open-ended responses, even though formal legal compliance with the Data Protection Act was rated highly on the closed-ended items. This pattern is consistent with Kalmus and Nikiforova's TOE-informed finding that environmental and governance constraints, rather than technology or organisational readiness alone, most often shape resistance to AI adoption in higher education [15].

For practice, the implication is that institutions preparing for agentic AI in Moodle should not treat budget allocation as a proxy for readiness, given the null association observed in the Results Section (Budget Adequacy and Adoption of Adaptive Functionality) and the low-budget visibility reported by respondents; a dedicated, transparent LMS/AI budget line, tracked and communicated to staff, may do more to signal institutional readiness than the raw size of the ICT budget. The emerging practice of naming a human accountability owner for each deployed agentic workflow, and granting agents scoped rather than static permissions to student data [16], offers a concrete governance mechanism that Kenyan ODeL institutions could adapt directly, and would address the accountability gap that respondents flagged even where formal legal compliance was rated highly.

Limitations

Sample size and power: N = 17 overall, with usable n for inferential tests as low as 4 per group, severely limits statistical power; the null results reported in the Results Sections (Staff vs. Student Perceived Readiness and Budget Adequacy and Adoption of Adaptive Functionality) should be read as inconclusive, not as evidence of absence. Institutional identification: the instrument did not capture a unique institution identifier, so the sample cannot be characterised with confidence as spanning a known number of distinct institutions, and clustering of respondents within the same institution cannot be ruled out. Sample composition: the sample is skewed toward private institutions (88.2%) and registered students (53.3%), while institutional leadership was represented by a single respondent; the Result Section (Budget, Infrastructure, Policy and Equity Perceptions) shows this skew, rather than sample size alone, most plausibly explains the sample's unexpectedly benign infrastructural-barrier ratings, and it limits how far any finding can be generalised to the public universities that dominate national ODeL provision. Attrition and skip logic: question-level n fell from 17 to as low as 6 or 8 for later blocks introducing potential non-response bias toward more engaged or more technically confident respondents. Self-report: all measures, including infrastructural barrier items, rely on respondent perception rather than independently verified institutional data (e.g., audited ICT budgets, measured connectivity). Respondent detail: beyond institution type and respondent category, the instrument did not collect further demographic detail (e.g., years of experience, age, gender), limiting the granularity with which sub-groups can be described or compared.

Proposed Conceptual Readiness Framework

Status of the Framework and Its Relation to Prior Models

The four domains presented in this section were specified a priori from the technology-adoption and governance literature reviewed (principally TOE [14], [15] and the AI-governance literature [3], [16]), and structured both the survey instrument and the discussion. They are a literature-derived conceptual contribution, not an inductive finding the survey data selected over alternative groupings; the close correspondence between the domains and the instrument's Likert blocks reflects that the instrument operationalises the framework, not the reverse. One reasonable objection is that TOE's environmental dimension could already accommodate regulatory and equity pressures, making a separate equity domain redundant with TOE. We retain equity as an explicit, cross-cutting domain because, in the pre-scaling audit use this framework targets equity failures such as unequal device access can persist even where an institution's regulatory environment is otherwise favourable; folding equity into environment risks the downstream, secondary treatment both this paper's respondents and the equity literature warn against [13].

Table 6 positions the proposed framework against the three adoption models most used in the Kenyan and regional e learning literature reviewed. TAM-based studies centre on individual perceived usefulness and ease of use; UTAUT2-based studies add habit, price value and facilitating conditions but remain individual-level; TOE based studies operate at institutional level across technology, organisation and environment but were not developed for autonomous, decision-making systems and do not, without extension, treat data-governance accountability or learner equity as separate, named constructs [14], [15]. The proposed framework is deliberately institutional rather than individual (like TOE, unlike TAM/UTAUT2), and names two constructs, explicit AI accountability and cross-cutting equity, separately from TOE's environmental dimension, on the grounds that these become first-order, trackable concerns only once a system acts autonomously on learner data rather than merely being used by a learner.

Table 6. Prior Models Versus the Proposed Framework

Model	Level	Names AI accountability?	Names equity?
TAM	Individual	No	No
UTAUT2	Individual	No	No
TOE [14]	Institutional	Not separately	Not separately
Proposed (this study)	Institutional	Yes	Yes (cross-cutting)

Four-Domain Structure

The framework treats infrastructure, policy and governance, human capability, and equity as interdependent rather than sequential domains, summarised in Table 7: strength in one (e.g., extensive Moodle customisation) does not substitute for weakness in another (e.g., unclear data-governance accountability for autonomous decisions), and the equity domain is deliberately positioned as a cross-cutting check on the other three rather than a final add-on.

Table 7. Four-Domain Agentic-AI Readiness Framework

Domain	Core question	Illustrative indicator
Infrastructure	Can the technical base sustain autonomous, always-on agents?	Hosting/API capacity, connectivity, device access
Policy & Governance	Are decisions and data flows accountable?	Formal AI-in-teaching policy, DPA-aligned data governance, accountability framework
Human Capability	Can staff and students operate, oversee and trust agentic tools?	Staff AI-management skills, student digital readiness, leadership commitment
Equity	Does the transition avoid widening access gaps?	Rural/remote parity, disability accommodation, affordability

Operationally, the framework suggests a staged pathway: (1) baseline audit of the four domains using instruments similar to the one applied here, ideally at institutional rather than individual level; (2) closing governance gaps (a formal AI-in-teaching policy and an explicit accountability framework for AI-driven decisions) before scaling any pilot beyond integrity or routing tools; (3) targeted capacity-building for both staff and students, since this sample suggests neither group can be assumed to lag the other; and (4) an equity gate at each stage, requiring evidence that rural, remote and disability-related access gaps are being actively monitored and mitigated before a pilot is declared ready for full institutional deployment.

Practical Readiness-Audit Checklist

Table 8 translates the framework into a short checklist that a systems librarian or ICT/eLearning director can apply directly during an institutional audit, prior to approving any expansion of adaptive or agentic functionality beyond a contained pilot.

Table 8. Pre-Scaling Readiness Checklist

Element	Example value
Infrastructure	Hosting/API capacity load-tested; connectivity mapped by campus/region
Policy & Governance	AI-in-teaching policy adopted; named accountability owner per workflow
Human Capability	Staff AI-management training completed; student orientation delivered
Equity	Rural/remote and disability access gaps assessed and mitigation funded

Future Research Directions

Building on the shortcomings identified in the Limitation Section, four directions for future research follow. First, a larger, adequately powered replication should recruit a stratified, institution-identified sample balanced between public and private provision, so the individual- and institution-level analyses originally envisaged for this study can be run with appropriate power. Second, self-reported infrastructure and budget items should be paired with independently verified institutional data (audited ICT expenditure, measured connectivity) to test whether the benign infrastructural-barrier ratings observed here hold once triangulated and once the public-university share is increased. Third, since current adoption in this sample clusters around integrity and routing tools rather than autonomous agents, a follow-up study once genuinely agentic tools are deployed at scale would test whether these readiness patterns persist further along the agentic spectrum [1], [2], [4]. Fourth, the accountability-owner and scoped-permission governance mechanisms noted in the Discussion Section [3], [16] merit a dedicated case study within a Kenyan ODeL institution to assess their practicality under local resource constraints.

Conclusion and recommendation

This study set out to establish an empirical baseline for agentic AI readiness within Moodle-based ODeL delivery in Kenya. Within a modest, exploratory and private-institution-skewed sample of seventeen respondents, the picture that emerges is one of technically active but governance-cautious institutions: extensive Moodle customisation and a majority uptake of automated integrity and pathway tools, set against low visibility of ICT/AI budgets, self-identified gaps in formal AI policy and accountability, and persistent concern about rural and remote student disadvantage. Neither a staff-student readiness gap nor a budget-adoption association reached statistical significance in this sample, a finding that should motivate a larger, adequately powered, multi-

institution study rather than be read as a substantive conclusion.

For systems librarians, ICT directors and ODeL administrators, four recommendations follow directly from the data. First, formalise AI-in-teaching and data-governance policy ahead of, not after, scaling adaptive functionality, given that policy items scored no higher on average than infrastructure items despite Data Protection Act compliance being rated highly. Second, invest deliberately in equity safeguards for rural and remote learners before or alongside any agentic AI rollout, rather than treating equity as a downstream concern, since 75% of respondents in this sample agreed that such learners already face disadvantages in accessing AI-enhanced features. Third, commission institution-level, not only individual-level, readiness audits with verified rather than self-reported budget and infrastructure data, deliberately including public universities, to enable the kind of adequately powered inferential analysis this exploratory study could not support. Fourth, adapt the accountability-owner and scoped-permission governance practices summarised in the Discussion Section to the Kenyan ODeL context, so that responsibility for any agentic decision affecting a student remains traceable to a named individual even as functionality becomes more autonomous.

Systems librarians occupy a distinctive position from which to lead this transition: they already sit at the intersection of information governance, digital infrastructure and user support, and their existing responsibility for metadata, access control and information ethics maps closely onto the policy, human-capability and equity domains of the framework proposed in the proposed conceptual readiness framework. Positioning the library and ICT/eLearning functions jointly as institutional stewards of the readiness audit recommended above, rather than leaving agentic AI adoption to be driven by vendor offers or ad hoc departmental pilots, would give Kenyan ODeL institutions a coherent, accountable route from static LMS infrastructure toward agentic-AI-driven delivery.

Acknowledgment

The authors thank the institutional leaders, ICT/eLearning directors, ODeL coordinators and students who took part in the survey. The authors used Claude (Anthropic) to organise this manuscript into the required IEEE conference structure.

Declarations

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. A. Raidas and R. Bhandari, "Agentic AI in Education: Redefining Learning for the Digital Era," in Artificial Intelligence in Education: Transforming Learning for the Future, London, U.K.: Scholars' Press, 2025, p. 89.

[2] G. Kostopoulos, V. Gkamas, M. Rigou, and S. Kotsiantis, "Agentic AI in Education: State of the Art and Future Directions," IEEE Access, vol. 13, pp. 177467–177491, 2025.

[3] I. Adabara, B. O. Sadiq, A. N. Shuaibu, Y. I. Danjuma, and V. Maninti, "Trustworthy Agentic AI Systems: A Cross-Layer Review of Architectures, Threat Models, and Governance Strategies for Real-World Deployment," F1000Research, vol. 14, p. 905, 2025.

[4] L. Yan, "From Passive Tool to Socio-cognitive Teammate: A Conceptual Framework for Agentic AI in Human-AI Collaborative Learning," arXiv:2508.14825, 2025.

[5] S. Murugesan, "The Rise of Agentic AI: Implications, Concerns, and the Path Forward," IEEE Intelligent Systems, vol. 40, no. 2, pp. 8–14, Mar./Apr. 2025.

[6] N. Kshetri, "Revolutionizing Higher Education: The Impact of Artificial Intelligence Agents and Agentic Artificial Intelligence on Teaching and Operations," IT Professional, vol. 27, 2025.

[7] R. N. Kibuku, D. O. Ochieng, and A. N. Wausi, "E-Learning Challenges Faced by Universities in Kenya: A Literature Review," Electronic Journal of e-Learning, vol. 18, no. 2, pp. 150–161, 2020.

[8] A. S. Mulwa and D. N. Kyalo, "The Influence of Institutional Factors on the Readiness to Adopt E-Learning in Kenyan Universities," Journal of Education and Practice, vol. 2, no. 11, pp. 1–9, 2011.

[9] M. Kashorda and T. Waema, E-Readiness Survey of Kenyan Universities Report, Nairobi, Kenya: Kenya Education Network (KENET), 2014.

[10] J. W. Mwangi, "Challenges Facing the Implementation of E-Learning in Kenyan Universities," International Journal of Academic Research in Progressive Education and Development, vol. 5, no. 3, pp. 1–12, 2016.

[11] Republic of Kenya, The Data Protection Act, No. 24 of 2019, Kenya Gazette, vol. CXXI, no. 156, 15 Nov. 2019.

[12] Office of the Data Protection Commissioner, Guidance Note for the Education Sector, Nairobi, Kenya: ODPC Kenya, 2024.

[13] R. Lembani, A. Gunter, M. Breines, and M. T. B. Dalu, "The Same Course, Different Access: The Digital Divide Between Urban and Rural Distance Education Students in South Africa," Journal of Geography in Higher Education, vol. 44, no. 1, pp. 70–84, 2020.

[14] L. G. Tornatzky and M. Fleischer, The Processes of Technological Innovation. Lexington, MA: Lexington Books, 1990.

[15] J.-E. Kalmus and A. Nikiforova, "To Accept or Not to Accept? An IRT TOE Framework to Understand Educators' Resistance to Generative AI in Higher Education," in Proc. Int. Workshop on AI in Education and Educational Research (AIEER), co-located with ECAI 2024, arXiv:2407.20130, 2024.

[16] K. Navaie, "From Rights to Runtime: Privacy Engineering for Agentic AI," AI Magazine, vol. 46, no. 4, e70036, 2025.

[17] A. Azevedo et al., "Institutional Policies on Artificial Intelligence in Higher Education: Frameworks and Best Practices for Faculty," New Directions for Adult and Continuing Education, Hoboken, NJ: Wiley, 2025.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

AI Literacy, Digital Capabilities and Intellectual Property Awareness: Evidence for Preparing Future Professionals in the Digital Age

Dimitri Chapkanov, Martin Najdov

(phone: +38976420540; e-mail: d.chapkanov@is.edu.mk), (e-mail: m.najdov@is.edu.mk)

Institute for Intellectual Property – Skopje, 1000 Skopje, North Macedonia

Article DOI: https://doi.org/10.68286/samyrad2026.351

ABSTRACT

The rapid expansion of artificial intelligence (AI), generative technologies, and digital platforms is transforming the ways knowledge is created, shared, and protected across educational and professional environments. As AI-driven tools become increasingly integrated into learning, research, and creative activities, intellectual property (IP) issues, including authorship, ownership, plagiarism, copyright protection, and the ethical use of digital content, have become more prominent. Despite growing attention to AI literacy, digital competencies, and IP governance, empirical evidence examining the relationship between digital capabilities and IP awareness in AI-driven environments is still limited. This study addresses this gap by examining the relationships between digital capabilities, AI-related environments, and IP awareness among individuals preparing to participate in increasingly digital and AI-enabled societies. It also investigates whether demographic characteristics are associated with perceptions of AI-related opportunities and risks associated with IP. By exploring these relationships, the study contributes to a better understanding of the factors that support responsible AI use, strengthen IP awareness, and promote the development of digital competencies. The findings provide practical implications for higher education institutions and policymakers seeking to foster ethical AI adoption and effective IP management in the digital era.

KEYWORDS AI literacy, digital capabilities, intellectual property, higher education, professional preparation

How to cite Chapkanov, D., & Najdov, M. (2026). AI Literacy, Digital Capabilities and Intellectual Property Awareness: Evidence for Preparing Future Professionals in the Digital Age. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.351

Introduction

Public and private sector organizations and higher education institutions play a vital role in knowledge creation and development. They face challenges in managing and leveraging intellectual property (IP) in the context of rapid digital transformation (DT) [1], [2]. The utilization of digital technologies, such as digital platforms, e-commerce, social media and digital marketing, affects the valuation and protection of IP, including trademarks, patents, and

copyrights, demonstrated in several analyses and empirical studies [3], [4], [5], [6]. This research paper was developed within the project "Factors of Intellectual Property Management and Digital Transformation in Macedonia" (2025–2026), funded by the Ministry of Education and Science of the Republic of N. Macedonia, building on previous empirical research [7], [8], [9], and this study examines the relationship between DT, IP management, and AI literacy while taking gender-related differences into account. It is conceptually based on the resource-based view and dynamic capabilities theory, according to which AI-related environments and digital capabilities are regarded as strategic resources that enable organizations to create, manage, and protect IP in rapidly evolving environments.

Technological and Human Perspectives on Digital Transformation

DT encompasses more than the adoption of digital technologies; it involves the ability of organizations, and individuals to adapt to an increasingly digital environment. Previous research [10], [11] demonstrates that digital technologies contribute to higher productivity, greater competitiveness, and enhanced innovation, while also underscoring the critical role of human resources, AI literacy and digital competencies in achieving successful transformation.

Effective DT requires a strategic approach that aligns organizational goals with digital capabilities. Through process optimization, continuous improvement, and effective governance, organizations can enhance adaptability, sustain innovation, and achieve long-term resilience in a rapidly changing digital environment [12]. Okene et al. [13] highlights the challenges of balancing IP protection with public access in the digital environment, particularly regarding digital piracy, AI-generated content, and globally distributed works. Their study emphasizes differences among international legal frameworks and the need for harmonized approaches through legal adaptation, international cooperation, and public awareness to support innovation while ensuring fairness and accessibility.

The successful implementation of digital technologies depends largely on developing individuals' digital capabilities [14]. AI literacy, digital skills, and adaptability are essential for supporting DT, encouraging innovation, and enabling effective IP creation and protection. Similarly, Gyamerah et al. [15] highlights that AI literacy, and digital competencies are often overlooked in strategic planning, despite their significant impact on innovation and IP development. Their findings underline the importance of investing in formal and non-formal education to strengthen AI literacy, digital capabilities and improve the effective use of digital technologies.

As DT progresses, the development of digital capabilities becomes essential for effective technology adoption and process improvement. AI literacy and the ability to use digital tools, such as mobile applications and social media platforms, support organizational agility, innovation, and competitive advantage [16]. Strengthening these capabilities enables organizations to maximize the value of technological investments and achieve improved operational and strategic outcomes.

Digital Transformation and Its Implications for Intellectual Property

DT is transforming the IP landscape by enabling new forms of innovation, knowledge sharing, and value creation, while simultaneously increasing challenges related to piracy, misuse, and protection [17]. Emerging technologies, including AI, blockchain, and digital platforms, require adapted legal, institutional, and organizational frameworks to ensure effective IP governance and maintain a balance between innovation and protection [18].

DT has transformed the IP landscape by enabling innovation, value creation, and wider dissemination of intellectual assets, while also increasing risks related to digital reproduction, AI-generated content, and global distribution [19]. Zinych et al. [20] highlight persistent challenges in digital IP protection, including weak enforcement, inconsistent legal practices, and limited effectiveness of existing frameworks. They emphasize the need for stronger legislation, improved enforcement mechanisms, and institutional reforms to support innovation and digital development.

Sheikh [21] highlights the importance of IP rights in the digital economy, emphasizing challenges related to digital reproduction, ownership, licensing, and enforcement. The study suggests that stronger IP protection mechanisms are essential for supporting innovation and sustainable competitiveness. Hantoush [22] examines the integration of open access resources in academic libraries, emphasizing the need for effective licensing and management strategies to balance knowledge accessibility with IP governance. Similarly, Samoilenko, Bieliaieva, and Borodii [23] explore digital approaches for protecting authorship rights, highlighting the role of digital tools in verifying ownership, preventing plagiarism in AI-driven creative environments.

The Role of Artificial Intelligence and Digital Technologies in Intellectual Property Management

In the context of rapid technological development, advanced digital technologies, especially AI and digital tools, are becoming increasingly significant [24]. These technologies reshape IP practices by enabling more efficient processes for creation, management, monitoring, and protection of IP assets [25].

Kilimnik [26] examines the challenges of IP protection in the digital economy, highlighting the impact of AI, blockchain, and digital platforms on IP rights, governance, and regulatory frameworks. The study proposes an integrated legal and technological approach to strengthen IP protection in digital markets. Sanek et al. [27] examine the impact of digital assets, including cryptocurrencies, NFTs, and AI-generated works, on existing IP frameworks within the European Union context. The study identifies regulatory challenges related to ownership, attribution, and enforcement, emphasizing the need for harmonized legal frameworks and effective mechanisms for IP protection in the digital era. Finally, Rubio-Alfageme [28] argues that traditional copyright frameworks are insufficient for addressing digital piracy and rapidly evolving digital markets. The study emphasizes the

importance of adaptive business models, accessible digital services, and balanced AI-supported enforcement approaches that protect creators' rights while maintaining public access to digital content.

A conceptual distinction is made in this study between AI literacy and broader digital capabilities. AI literacy refers specifically to the ability to understand, critically evaluate, use, and interact responsibly with AI systems and AI generated outputs, including awareness of their limitations, risks, ethical implications, and potential consequences for intellectual property. Digital capabilities, in contrast, refer to broader competencies required to operate effectively in digitally mediated environments, including information search and evaluation, use of digital platforms and services, digital communication, security practices, and confidence in using digital technologies. Although these constructs are related and may mutually reinforce one another, they are not conceptually identical. This distinction is important because the present study examines their relationship with IP awareness while also recognizing that the available questionnaire data combines several AI-related and broader digital competency items into a composite measure. Accordingly, the empirical results should not be interpreted as demonstrating that AI literacy alone explains differences in IP awareness.

Background and related work

Research Question

Drawing on the theoretical foundations and empirical gaps identified in previous research and the existing literature, this study addresses the following research questions:

1. To what extent are AI literacy and digital capabilities associated with awareness of IP?

2. How do respondents perceive the positive and negative implications of digital transformation for intellectual property?

3. To what extent do AI literacy, digital capabilities, and awareness of IP differ across gender groups?

Research Hypotheses

Based on established theories and empirical gaps, this study tests hypotheses regarding AI literacy, digital transformation, and intellectual property:

H1 (Main hypothesis): AI literacy and digital capabilities are positively and significantly related to individuals' knowledge and awareness of IP.

H1a (Sub-hypothesis): There is a statistically significant difference in AI literacy and digital capabilities between respondents who correctly identify the types of IP and those who do not.

H1b (Sub-hypothesis): AI literacy and digital capabilities have a positive and statistically significant effect on awareness of IP.

H2 (Main hypothesis): Respondents perceive both positive and negative implications of DT for IP.

H2a (Sub-hypothesis): There is a statistically significant positive association between the positive and negative implications of DT on IP.

H3 (Main hypothesis): Statistically significant differences exist in AI literacy, digital capabilities, and awareness of IP between male and female respondents.

H3a (Sub-hypothesis): Gender significantly predicts AI literacy and digital capabilities.

H3b (Sub-hypothesis): Gender significantly predicts awareness of IP.

Variables

The AI Literacy and Digital Capabilities Index (AILI) was constructed from individual's overall score on the questionnaire items covering two related domains: (1) AI related competencies, including understanding and use of AI tools, evaluation of AI-generated outputs, recognition of AI limitations and risks, and responsible use of AI; and (2) broader digital capabilities, including information search and evaluation, use of digital services and platforms, digital security, and trust in digital systems. Items were measured on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The composite score was calculated as arithmetic means of valid item responses.

The Intellectual Property Awareness (IPA) refers to the level of an individual's knowledge and understanding of the concept of IP and their various key forms. The measure included multiple choice items and five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Where applicable, correct identification of IP types was analyzed separately as a knowledge indicator rather than being treated as equivalent to the broader IPA construct. Digital Transformation Impact (DTI) refers to respondents' perceptions of how AI and digital technologies influence the creation, dissemination, protection, and potential misuse of IP. The construct is conceptualized as comprising two core perceptual dimensions:

1. The positive dimension – Intellectual Property Positive Index (IPPI) reflects perceived enabling effects of DT on IP and constructed from Likert-scale items assessing perceived enabling effects of digital transformation on IP, including innovation, dissemination, accessibility, authorship verification, and modern licensing practices.

2. The negative dimension – Intellectual Property Negative Index (IPNI) captures perceived risks associated with DT, constructed from items assessing perceived risks, including piracy, unauthorized sharing, misuse of digital content, difficulties in enforcement, and trademark violations in online environments.

Gender is measured as a nominal variable with two categories (male, female).

Methods or analytical approach

Participants

The study sample consists of 102 respondents from the public sector in N. Macedonia. A convenience sampling

method was applied. The sample includes both male and female respondents, enabling comparative, correlational and regression analyses of AI literacy, digital capabilities and awareness of IP. Participation was voluntary and anonymous.

Questionnaire

The questionnaire included three main sections: demographic information, AI literacy and digital skills, and IP management. Responses were collected using multiple choice and yes/no questions, as well as a five-point Likert scale (1 = strongly disagree, 5 = strongly agree), with an additional option of "Don't know / No answer" for participants who were unsure or preferred not to answer.

Data Collection and Analysis Procedure

The study was conducted using an online questionnaire. Prior to participation, all respondents were informed about the purpose and duration of the study and were assured that their responses would be used exclusively for scientific purposes. Participants received clear instructions for completing the questionnaire. This method allows standardized, systematic data collection. The analysis includes descriptive statistics, correlation, t-tests for group comparisons, and regression analysis to examine relationships between DT, AI literacy and digital capabilities and IP variables. Responses coded as "Don't know / No answer" were treated as missing values. Therefore, the effective sample size varies across statistical tests depending on the availability of valid responses for the variables included in each analysis. Before conducting the inferential analyses, the relevant statistical assumptions were assessed, including the distributional characteristics of the variables and the assumptions relevant to correlation, group comparison, and regression analyses. Where the assumption of equal variances was not met, Welch's t-test was used as an alternative to the conventional independent samples T test. The results were interpreted with consideration of the relatively small effective sample sizes in some analyses.

Results and findings

Regarding the H1 (Main hypothesis): AI literacy and digital capabilities are positively and significantly related to individuals' knowledge and awareness of IP, the results supported the hypothesis, as the findings indicated a statistically significant positive correlation between the AILI and IPA, r (80) = 0.294, p = 0.007, indicating a weak-to moderate positive association between the two variables.

Regarding the H1a (Sub-hypothesis): There is a statistically significant difference in AI literacy and digital capabilities between respondents who correctly identify the types of IP and those who do not, the results supported the hypothesis, as the findings indicated a statistically significant difference in AILI scores between respondents with correct knowledge of types of IP and those with incorrect knowledge of types of IP, t (40.15) = 2.20, p = 0.034. The group with correct knowledge of types of IP had a higher AILI score (M = 4.37, SD = 0.42) than the group with incorrect knowledge of types of IP (M = 4.10, SD = 0.57).

Regarding the H1b (Sub-hypothesis): AI literacy and digital capabilities have a positive and statistically significant effect on awareness of IP, the results supported the hypothesis, as the findings indicated that the regression model was statistically significant, F (1,80) = 9.79, p = 0.002, explaining 10.9% of the variance in IPA (R^2 = 0.109). Furthermore, AILI emerged as a significant positive predictor of IPA (β = 0.33, p = 0.002). These findings indicate that respondents with higher levels of AILI tend to report higher levels of IPA. However, because the data are cross-sectional, the regression result should be interpreted as a statistical predictive association rather than evidence that AILI causes changes in IPA.

Regarding the H2 (Main hypothesis): Respondents perceive both positive and negative implications of DT for IP, the results supported the hypothesis, as the findings indicate a strong positive perceived impact of DT on IP, reflected in the Intellectual Property Positive Index (IPPI) (M = 4.21, SD = 0.61), as well as a notable negative perceived impact, reflected in the Intellectual Property Negative Index (IPNI) (M = 3.80, SD = 0.65), confirming the coexistence of both enabling and risk-related effects of DT on IP management. These findings indicate that respondents simultaneously recognized potential enabling effects and potential risks of DT for IP.

Regarding the H2a (Sub-hypothesis): There is a statistically significant positive association between the positive and negative implications of DT on IP, the results supported the sub-hypothesis, as the findings indicated a statistically significant moderate positive correlation was found between Intellectual Property Positive Index (IPPI) and Intellectual Property Negative Index (IPNI), r (33) = 0.412, p = 0.014. This finding indicates that respondents who reported stronger perceived benefits of DT also tended to report greater awareness of its potential risks. The result suggests that positive and negative perceptions of DT are not mutually exclusive.

Regarding the H3 (Main hypothesis): Statistically significant differences exist in AI literacy, digital capabilities, and awareness of IP between male and female respondents, the findings provided partial support for the hypothesis. First, as the findings indicated a statistically significant difference in AILI scores between male and female respondents, t (80) = 1.989, p = 0.050. Male respondents reported a higher mean AILI score (M = 4.40, SD = 0.58) compared with female respondents (M = 4.16, SD = 0.54), indicating higher levels of AI literacy among male participants. Second, no statistically significant difference was identified in IPA scores between male and female respondents, t (80) = 0.031, p = 0.975. The mean scores were identical for male (M = 3.83, SD = 1.21) and female respondents (M = 3.83, SD = 1.25), suggesting that gender does not significantly influence IPA.

Regarding the H3a (sub-hypothesis): Gender significantly predicts AI literacy and digital capabilities, the findings supported the hypothesis, as the regression model was marginally statistically significant, F (1,80) = 3.96, p = 0.050, explaining 4.7% of the variance in AILI scores (R^2 = 0.047). Gender was identified as a positive predictor of AILI (B = 0.245, β = 0.217, p = 0.050), indicating that male respondents reported slightly higher levels of AILI compared with female respondents. These findings suggest that gender contributes to differences in AILI, although the overall explanatory power of the model is limited.

Regarding the H3b (Sub-hypothesis): Gender significantly predicts awareness of IP, the results did not support the hypothesis, as the regression model was not statistically significant, F (1,80) = 0.001, p = 0.975, explaining virtually

none of the variance in IPA scores ($R^2 < 0.001$). Gender was not found to be a significant predictor of IPA (B = 0.008, β = 0.003, p = 0.975). The results show that male and female respondents reported almost identical levels of IPA.

Discussion

The discussion is structured around the three research questions. The first research question concerns the association between AI-related competencies, digital capabilities, and IPA; the second addresses respondents' perceptions of the positive and negative implications of DT for IP; and the third examines gender differences in these constructs. The interpretation below follows this structure to distinguish the empirical findings from their broader conceptual and educational implications.

The descriptive statistics provided an overview of respondents' self-reported proficiency in AI literacy and digital competence, their use of digital technologies, and their awareness of IP rights.

The findings reveal generally high levels of AI literacy and digital competence among the respondents, with all assessed dimensions recording mean values above the midpoint of 3. In contrast, comparatively lower scores were observed in relation to trust in digital systems and AI, as well as awareness of IP rights. This pattern indicates potential areas for further development, particularly in strengthening digital trust and improving understanding of IP protection within the increasingly digital environment.

The correlation analysis conducted to test Hypothesis H1 indicates a positive association between AI literacy, digital capabilities, and IPA. This relationship is particularly relevant in AI-driven environments, where a higher level of AI literacy may contribute to individuals' ability to recognize and address key IP-related issues, including authorship, ownership of AI-generated content, plagiarism, the ethical use of generated outputs, and the protection of IP. These findings further emphasize the growing importance of AI and digital competencies in supporting responsible and informed engagement with AI-generated content.

The significant association between AILI and IPA indicates that individuals with higher levels of AI literacy and digital competence tend to exhibit greater awareness of IP rights and the responsibilities associated with their use in digital environments. This finding contributes to existing literature and previous research by providing empirical evidence for the relationship between AI-related competencies and IP awareness in an increasingly AI driven context [14], [16]. However, the relatively modest strength of the correlation indicates that AI literacy and digital capability alone do not fully explain variations in IPA, suggesting that other individual, organizational, or contextual factors may also influence this relationship.

The findings for the H1a sub-hypothesis demonstrate a statistically significant difference in AILI scores between respondents who correctly and incorrectly identified the different types of IP. This result supports H1a and suggests that respondents with stronger knowledge of IP concepts tend to have higher levels of AILI. One possible explanation is that individuals with greater AI literacy and digital competence may be better positioned to acquire, understand, and retain knowledge related to IP. Conversely, greater awareness of IP rights may also promote more responsible, informed, and systematic engagement with digital technologies, particularly in AI-supported creative and content-generation processes.

The findings related to the H1b sub-hypothesis indicate that AI literacy and digital capability significantly contribute to the prediction of IPA. The positive association between AILI and IPA suggests that individuals with higher levels of AI literacy and stronger digital competencies are more likely to recognize and understand IP rights. Although the model explains a relatively modest proportion of the variance in IPA ($R^2 = 0.109$), the results nevertheless highlight the relevance of AI literacy and digital competence in shaping individuals' understanding of IPA. From a practical perspective, incorporating IP-related topics into digital literacy and competency development initiatives could help strengthen IPA and encourage more informed and responsible behavior in digital environments.

The findings of this study contribute to the growing body of research on the role of AI literacy in IP management by demonstrating that AI-related competencies and digital capabilities are positively associated with higher levels of IPA. Whereas previous research has predominantly viewed digital literacy as a fundamental prerequisite for technology adoption and innovation, the present study broadens this perspective by emphasizing the relevance of AI literacy for understanding and addressing IP-related challenges in AI enabled environments. As AI becomes increasingly embedded in content creation, knowledge production, and information processing, individuals need to develop not only technical competencies but also the capacity to recognize and appropriately address issues concerning authorship, ownership, originality, ethical use, and the protection of AI-assisted outputs. In this context, AI literacy can be regarded as an important complementary competence that supports responsible digital practices and more informed decision-making concerning IP rights. The findings therefore underscore the value of integrating AI literacy into IP education and broader digital competency development, particularly within higher education, where students are being prepared to operate in increasingly AI driven professional environments.

The findings related to the H2 hypothesis provide empirical evidence of the dual role of DT in the context of IP. Consistent with previous research, the results suggest that DT simultaneously creates opportunities for innovation and knowledge dissemination while also generating challenges related to accessibility, misuse, and IP protection [19], [20]. The relatively high IPPI score (M = 4.21, SD = 0.61) indicates that respondents perceive DT as an important enabler of innovation, knowledge sharing, and broader access to digital content. At the same time, the IPNI score (M = 3.80, SD = 0.65) reflects a considerable level of awareness regarding the risks associated with misuse, piracy, and the protection of IP in digital environments. Taken together, these findings highlight the complex and multidimensional role of DT in shaping contemporary IP practices.

Overall, the relatively high ratings across both the positive and negative dimensions suggest that respondents hold a balanced view of DT. While they recognize the potential of DT to facilitate innovation, knowledge sharing, and broader

access to information, they also acknowledge the challenges these technologies create for effective IP protection and governance. This balanced perception highlights the dual nature of DT and the need to maximize its benefits while addressing the associated risks to IP.

The findings for the H2a sub-hypothesis reveal a statistically significant, moderate positive association between respondents' positive and negative perceived impacts of DT on IP ($r = 0.412$, $p = 0.014$). This result suggests that respondents who perceive greater benefits of DT, including enhanced knowledge dissemination, innovation, and accessibility, also tend to demonstrate greater awareness of the associated risks, such as misuse and IP infringement. Rather than viewing the benefits and risks of DT as mutually exclusive, respondents appear to recognize both dimensions simultaneously. This pattern points to a relatively balanced perception of DT, in which greater recognition of its potential advantages is accompanied by increased awareness of its implications for IP protection.

The findings for H3, including the H3a and H3b sub hypotheses, suggest that gender has a relatively limited role in explaining differences in AI literacy and digital capabilities and does not have a significant effect on IPA. Although male respondents reported slightly higher AILI scores than female respondents, the observed difference appears to be modest, indicating that gender may account for only a small proportion of the variation in AI literacy and digital competence. This interpretation is further supported by the regression results, which show that gender explains only a small proportion of the variance in AILI. Consequently, other factors, including educational background, professional experience, access to digital training, and organizational context, may be more important in shaping individuals' AI literacy and digital capabilities.

In contrast, gender was not found to have a statistically significant effect on IPA, with male and female respondents reporting broadly comparable levels of IPA. This finding suggests that IPA may be more closely related to factors such as knowledge, prior experience, and exposure to digital environments than to demographic characteristics. Overall, the results indicate that gender represents a relatively weak predictor of both AILI and IPA when compared with broader individual, educational, and contextual factors.

These findings also carry important implications for higher education institutions. Integrating AI literacy, digital competencies, and IP education into academic curricula could help equip future professionals with the knowledge and skills necessary to navigate AI-enabled environments responsibly. Such an approach would support not only technological competence, but also greater awareness of ethical, legal, and IP-related considerations associated with the use of AI and digital technologies.

Conclusion

The empirical findings provide several important insights into the links between AI literacy, digital capabilities, awareness of IP, and the role of DT in IP management. However, limited empirical evidence exists on whether AI literacy and digital capabilities are associated with awareness of IP at the individual level. This study addresses this gap by providing empirical evidence on how AI literacy and digital capabilities relate to individuals' awareness of IP.

First, the findings show that AI literacy and digital capabilities are closely related to IPA. Respondents with higher AILI scores reported a better understanding of IP concepts, rights, and responsibilities. The regression results also show a positive relationship between AILI and IPA, indicating that individuals with stronger digital skills are more likely to recognize the importance of IP, understand its implications, and appreciate its value in digital environments.

The findings also point to a clear relationship between IP knowledge and AILI. Respondents who were able to correctly identify different types of IP tended to have higher levels of AILI, suggesting that AI literacy and knowledge of IP are closely connected. This finding underlines the need for educational approaches that develop technological skills alongside a basic understanding of IP principles and their application in digital environments.

Second, the findings show that DT has both positive and negative implications for IP management. Respondents recognize that digital technologies can create important opportunities by supporting innovation, improving access to information, facilitating the distribution of IP assets, and introducing new ways of verifying authorship, including blockchain-based solutions. At the same time, they are aware of the risks associated with piracy, unauthorized sharing, trademark misuse, and limitations in existing IP protection mechanisms. The recognition of both these opportunities and risks suggests that respondents have a relatively balanced understanding of the challenges and benefits that DT brings to IP management.

The results also indicate that individuals who perceive stronger positive effects of DT tend to be more aware of its potential risks. This suggests that responsible participation in digital environments depends not only on technological skills but also on an understanding of the legal and ethical issues associated with digital technologies.

Third, the findings related to demographic characteristics show only limited gender differences in AILI. Male respondents reported slightly higher levels of AILI, while gender did not have a significant effect on IPA, with male and female respondents showing nearly identical levels of IPA. These findings suggest that factors such as education, professional experience, exposure to digital technologies, organizational support, and access to AI and digital training may play a more important role in shaping AILI and IPA than gender alone.

A key contribution of this study is the identification of AI literacy as an emerging competence that can strengthen awareness of IP rights in digitally transformed environments. Compared with traditional digital skills, AI literacy also involves the ability to critically assess AI generated content, understand issues related to authorship, and recognize the ethical and IP implications of using AI technologies. The findings highlight the need to integrate AI literacy, digital capabilities, and IP education into higher education curricula and professional development programs. As future professionals are increasingly expected to work in AI-driven environments, developing AI literacy can contribute to a better understanding of IP rights, more responsible use of

technology, ethical content creation, and the protection of innovation. In this way, education can help students develop not only the technical skills needed to work with digital technologies, but also the critical, ethical, legal, and professional competencies required for responsible participation in AI-enabled workplaces and society. Although the findings may have implications for higher education, these implications should be understood as conceptual and practical proposals rather than direct empirical conclusions about students or academic institutions. The present sample consists of Macedonian public-sector respondents and does not include a dedicated student or higher-education population. Accordingly, the proposed integration of AI literacy, digital capabilities, and IP education into higher education curricula requires empirical validation in student and academic staff populations before broader conclusions can be drawn.

From this perspective, the digitalization of higher education can be understood through three closely related dimensions: digital infrastructure, AI and digital competence, and IP, ethics, and responsible AI use. The first dimension refers to the technological and organizational foundation needed to support digital education, including digital platforms and e-services, cloud-based and collaborative technologies, digital security, data protection, and access to emerging technologies. The second dimension focuses on the development of AI and digital skills, such as searching for and evaluating information, digital communication and collaboration, effective use of AI tools, critical assessment of AI-generated content, responsible use of generative AI, and the ability to continuously adapt to technological change. The third dimension addresses IP and ethical competencies, including knowledge of authorship, copyright, plagiarism, ownership, licensing, privacy, transparency, accountability, bias, and the responsible use of AI. Together, these dimensions provide a broader framework for supporting the responsible and effective DT of higher education.

Based on the empirical findings and theoretical discussion, the study proposes a conceptual educational framework that can be represented as: Student → Digitally Competent → AI-Literate → IP-Aware → Ethically Responsible → AI-Ready Professional.

This sequence should not be interpreted as an empirically established developmental pathway or as evidence that individuals necessarily progress through these stages in a fixed order. Rather, it represents a proposed conceptual framework illustrating how digital capabilities, AI literacy, IPA, and ethical responsibility may be integrated within broader professional preparation. The empirical results of the present study establish associations among selected constructs, but they do not demonstrate that these competencies develop sequentially or causally. Future longitudinal and multi-stage research is required to determine whether the proposed framework represents a developmental pathway, a set of mutually reinforcing competencies, or a broader model of professional readiness.

From a practical and institutional perspective, higher education institutions can apply this framework to assess their level of digital maturity, identify gaps in students' and staff competencies, redesign curricula, and establish measurable learning outcomes related to AI literacy. It can also support the integration of IP and AI ethics into teaching and learning, as well as the assessment of students' digital and AI readiness. In this context, AI literacy should be viewed as a transversal competence that is developed across study programs and linked to disciplinary knowledge, professional practice, ethical decision-making, and IPA.

Such integrated educational approaches can help students participate more responsibly in digital environments and become better prepared to respond to emerging technological challenges. In this regard, higher education institutions have an important role in developing not only students' technical AI skills, but also their ethical, legal, and IP knowledge. This broader approach can better prepare graduates to use AI responsibly and to respond effectively to the demands of AI-enabled workplaces and society.

From a policy and organizational perspective, the findings emphasize the need for evidence-based approaches to digital IP management. Policymakers, higher education institutions, and organizations should support training and development initiatives that bring together emerging technologies, AI-driven environments, digital competencies, and IP governance. In this context, strengthening AI literacy should be regarded as an important priority for higher education. Developing these competencies can encourage more responsible use of AI, increase awareness of IP rights, support ethical decision making, and better prepare graduates for the growing demands of an AI-enabled labor market.

The broader contribution of this approach lies in shifting the focus of DT from technology adoption towards competence development. The success of DT should therefore not be assessed only by the extent to which institutions adopt new technologies, but also by how effectively these technologies contribute to developing professionals who are digitally competent, able to think critically, aware of ethical responsibilities, and knowledgeable about IP. This perspective places human capabilities and responsible technology use at the center of DT in higher education.

This study has few limitations. The use of a convenience sample and a cross-sectional research design may limit the generalizability of the findings and does not allow for clear causal conclusions. Future research should therefore examine and further validate the proposed framework using larger and more diverse samples, including students and academic staff from different higher education institutions, sectors, and countries. Further studies could also consider additional individual and organizational factors that may influence AI literacy, digital capabilities, and the awareness of IP rights, such as organizational support, previous digital experience, and the intensity and frequency of AI use.

An additional consideration concerns the self-reported nature of AI literacy and digital capability measures. The study assesses respondents perceived competencies through questionnaire responses rather than through objective performance-based tasks. Consequently, the reported AILI scores may reflect respondents' perceptions of their abilities rather than their objectively demonstrated AI and digital skills. Future research could strengthen the measurement approach by combining self-report scales with performance-based assessments, scenario-based tasks, or validated AI literacy instruments.

In conclusion, the findings indicate that AI literacy, digital capabilities, and IPA are closely interconnected and should be viewed as important components of responsible DT. Effective DT should not be understood simply as the adoption of AI technologies. Rather, it is achieved when the use of these technologies contributes to the development of professionals who can apply AI critically, safely, ethically, and responsibly, while also understanding its implications for IP.

Declarations

Funding: This research was funded by the Ministry of Education and Science under the project "Factors of Intellectual Property Management and Digital Transformation in Macedonia (2025-2026)", implemented by the Public Scientific Institution Institute for Intellectual Property – Skopje.

References

[1] V. Buchkovski, D. Chapkanov, and G. Naumovski, “Digital management of copyright and related rights from the point of view the Macedonian legislation,” in Knowledge Society and 21st Century Humanism, Sofia: UniBIT, 2020.

[2] D. Chapkanov, G. Naumovski, G. Manchevski, and A. Klimovski, “Analysis of the System for Electronic Evidence for Broadcasted Copyrighted Works and Objects of Related Rights in Macedonia,” Socioeconomica, The Scientific Journal for Theory and Practice of Socio-economic Development, vol. 5, no. 9, pp. 13–22, 2016.

[3] D. Chapkanov, “Вреднување на добро познатите трговски марки на телекомуникацискиот пазар од студенти на Економскиот факултет и Правниот факултет „Јустинијан Први“,” Зборник на текстови од меѓународна конференција на Филозофскиот факултет во Скопје, pp. 727–754, 2016.

[4] D. Chapkanov, “Предизвиците на дигиталното управување со правата на уметниците-изведувачи во контекст на македонското законодавство,” ArsAcademica, vol. 7, no. 1, pp. 173–183, 2019.

[5] G. Naumovski and D. Chapkanov, “Convergence of Trademark Law and E-Commerce: Overview of US, EU and China Regulations on Trademarks and Domain Names,” Mizan Law Review, vol. 8, no. 4, pp. 424–438, 2014.

[6] G. Naumovski and D. Chapkanov, “Empirical analysis of trademark perception of students enrolled at legal and economic students in The Republic of Macedonia,” in 9th International Conference of Education, Research and Innovation, Seville, Spain, pp. 7634–7639, 2016.

[7] V. Buchkovski, G. Naumovski, D. Chapkanov, N. Nikoloska, F. Vuchen, and M. Najdov, “Анализа на односот меѓу интернет домен имињата и трговските марки,” Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2024.

[8] N. Nikoloska, D. Chapkanov, F. Vuchen, and M. Najdov, “Преглед на решенија од компаративното право во однос на трговските марки на интернет,” in Односот меѓу интелектуалната сопственост и информатичката технологија, Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2025.

[9] D. Chapkanov and N. Nikoloska, “Valuation of trademarks for sport equipment,” Research in Kinesiology, vol. 53, no. 1, 2025.

[10] G. Naumovski, Љ. Stefanoski, M. Bogeva Micovska, and M. Najdov, “Теоретски основи на правото на информатичката технологија. Односот меѓу интелектуалната сопственост и информатичката технологија,” Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2025.

[11] D. Chapkanov, M. Najdov, F. Vuchen, and N. Nikoloska, “Основни социолошки теории за информатичката технологија. Односот меѓу интелектуалната сопственост и информатичката технологија,” Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2025.

[12] E. Jager, “Digital transformation,” in Mastering the TOGAF® Standard, pp. 757–779, 2025. doi: 10.1007/979 8-8688-1814-1_19.
[13] N. Okene, N. W. Atejimah, and B. B. Timothy, Intellectual Property in the Digital Age, Rivers State University, January 2025.

[14] C. Blanka, B. Krumay, and D. Rueckel, “The interplay of digital transformation and employee competency: A design science approach,” Technological Forecasting and Social Change, vol. 178, 121575, 2022.

[15] S. Gyamerah, L. Afshari, and D. Asante, “Digital transformation in the SME context: The nexus between leadership, digital capabilities, and digital strategy,” International Small Business Journal, vol. 43, no. 3, pp. 303–328, 2025.

[16] D. Cetindamar, B. Abedin, and K. Shirahada, “The role of employees in digital transformation: A preliminary study on how employees’ digital literacy impacts the use of digital technologies,” IEEE Transactions on Engineering Management, vol. 71, pp. 7837–7848, 2021.

[17] F. Vuchen, D. Chapkanov, M. Najdov, and N. Nikoloska, “Осврт на користењето на трговските марки на интернет,” in Односот меѓу интелектуалната сопственост и информатичката технологија, Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2025.

[18] Љ. Stefanoski, M. Najdov, N. Nikoloska, and F. Vuchen, “Авторското право и сродните права на интернет,” in Односот меѓу интелектуалната сопственост и информатичката технологија, Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2025.

[19] T. Hongyun, J. M. Sohu, A. U. Khan, I. Junejo, S. N. Shaikh, S. Akhtar, and M. Bilal, “Navigating the digital landscape: Examining the interdependencies of digital transformation and big data in driving SMEs’ innovation performance,” Kybernetes, vol. 54, no. 3, pp. 1797–1825, 2025.

[20] L. Zinych, I. Petrovska, O. Burlak, and N. Polishko, “Problems of intellectual property rights protection in the digital environment,” Law Human Environment, vol. 16, no. 4, pp. 135–157, 2025. doi: 10.31548/law/4.2025.135.

[21] Z. Sheikh, “From code to copyright: A comprehensive study of intellectual property rights in the evolving digital economy,” Intellectual Property Rights: Issues and Challenges (Conference), 10.51767/c250617. July 2025. doi:

[22] A. K. Hantoush, “Developing digital collections in an open access environment with intellectual property challenges: Mengembangkan koleksi digital dalam lingkungan akses terbuka dengan tantangan hak kekayaan intelektual,” Academia Open, vol. 11, no. 1, 2026. doi: 10.21070/acopen.11.2026.13877.

[23] O. Samoilenko, A. Bieliaieva, and Y. Borodii, “Using digital technologies to certify intellectual property rights,” Technical Sciences and Technologies, May 2025. doi: 10.25140/2411-5363-2025-1(39)-125-133.

[24] N. Nikoloska and M. Najdov, “Промоција на географските називи за македонските производи во дигиталното опкружување како алатка за унапредување на туризмот. Некои аспекти на заштитата на географските називи,” Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2024.

[25] Љ. Stefanoski, V. Buchkovski, F. Vuchen, and N. Nikoloska, “Имиња на домени и правото на индустриска сопственост,” in Односот меѓу интелектуалната сопственост и информатичката технологија, Скопје: ЈНУ Институт за интелектуална сопственост – Скопје, 2025.

[26] I. Kilimnik, “Contemporary challenges in intellectual property protection: Legal transformations and technological disruptions,” in Intellectual Property: Protection in Modern Conditions, pp. 45–67, 2025. doi: 10.36690/IPP-45-67.

[27] K. A. Sanek, I. Kamarudin, and A. Kamarudin, “Redefining Ownership: Digital Assets, Intellectual Property, and Emerging

Technologies," International 41, Journal of Law Government and Communication, vol. 10, no. pp. 385–396, Sep. 2025. doi: 10.35631/IJLGC.1041025.

[28] B. M. Rubio-Alfageme, "Intellectual property challenges in the 21st century: A study of digital piracy in the European Union," Intangible Capital, vol. 21, no. 2, pp. 335–358, 2025. doi: 10.3926/ic.3040.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Cultivating Faculty's AI Literacy through the Use of Open-Access Platform

Alexa Alice Joubin

(e-mail: ajoubin@gwu.edu)

George Washington University, DC 20052, United States of America

Article DOI: https://doi.org/10.68286/samyrad2026.353

ABSTRACT

This design-based research investigates the development of faculty's self-efficacy in understanding AI's pedagogical and technological implications. Generative AI (gen AI) is a pluralistic boundary object in higher education. Faculty disagree on how to approach it, but they tend to agree that its presence cannot be ignored. However, wide-spread discussion of the abstract concept does not correlate to actionable knowledge. Studies of literacy indicate that high fluency often masks low adequacy. This project's purpose is to give faculty first-hand experiences of designing AI in order to enhance literacy. We do so by building the open-access Teach Anything platform as a trust inducing environment. It provides backend access and testing of several open-source large language models (LLM), and not just one LLM. Faculty use their own files to design their own applications in the context of experimentation or in their courses. We use Teach Anything to test the hypothesis that hands-on knowledge empowers faculty. Our key finding is that educators are empowered by this community-oriented, participatory approach, because they are proactive designers of AI, rather than passive consumers. They curate the retrieval corpus that grounds their open-source AI applications. They incorporate their new knowledge into adaptive teaching and students' learning.

KEYWORDS open source, open access, generative AI, educator literacy, participatory design

How to cite: Joubin, A. A. (2026). Cultivating faculty's AI literacy through the use of open-access platform. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.353

Introduction

In higher education, gen AI is happening to faculty and students rather than with them or for them, leading to social and financial inequalities, among other spillovers. Multiple AI tools are imposed on campuses by administration and large corporations. AI has diluted and diminished the value of learning experiences, with its misinformation affecting the trustworthiness in human behaviors and AI-infused society. Academics can be fluent in talking about AI, but their deeper competence and pedagogical integration remain limited [1]. In cross-disciplinary surveys, educators describe repositioning themselves as curators and verifiers of AI output rather than confident designers of it [2], and another study records high faculty awareness of generative AI alongside markedly lower applied competence [3]. We need an inclusive environment and a safe testing ground for literacy-driven trustworthy behaviors that push beyond stock arguments for or against gen AI. Trust emerges from gaining access to processes rather than merely accepting the verifiability of explicit answers.

The National Institute of Standards and Technology defines trust as a property of a system through metrology [4], defining trustworthiness through metrics of transparency, explainability, and interpretability, assessing an AI system's adherence to precursors of trust. This in part has led to the techno-solutionism in educational AI design that assumes trust will emerge from systems that produce predictable performance. I hypothesize that trust in AI-infused environment is not just a technical issue, but also a relational question [5]. Trust is not only about explainability or interpretability of a model. It concerns fostering deep social connections between educators and students. In this dynamic environment with many stakeholders, trust is not a property of an object but rather a type of relation and action [6]. Shifting our perspective on generative artificial intelligence (gen AI) in this way helps reduce misplaced trust in AI's outputs while enabling students to engage with AI as a tool for interpretation and exploration rather than a product for passive consumption.

My primary research question is: How do faculty, as designers of AI, develop AI literacy? My hypothesis is that faculty's self-efficacy in understanding AI's pedagogical and technological implications will be enhanced by hands-on experiences, unavailable through commercial AIs. No commercial AI platforms allow full backend access for faculty. Customization of AI through API or fine-tuning LLMs has a high barrier of technical knowledge that most faculty do not possess. To test this hypothesis, Alexa Alice Joubin and Akhilesh Rangani built the open-access Teach Anything, https://www.teachanything.ai, as a "sandbox" for experimentation. Teach Anything gives educators access to the configuration and operation of open models to enhance meta-cognition skills. Faculty do not need any prior technical knowledge to use Teach Anything to design their own AI. Once faculty's AI literacy develops in a controlled environment, they will be in a better position to foster trustworthy behaviors in their classrooms. They can then incorporate their new knowledge into their design of activities and assignments.

Background and related work

Faculty Report Low Self-efficacy

While a great deal of research in education focuses on AI's techno-solutionist effects [7], it often fails to identify how it can do harm and how educators might counter it.

Faculty predominantly lack AI literacy and opportunities to customize existing AI tools, with the majorities of "college and university leaders" indicating that "their institutions are not very or not at all ready to use GenAI for such things as: preparing students for the future, 56% say their schools are not prepared for this" [8, p. 3]. A multi-university study identified low self-reported AI literacy among faculty [9]. AI competencies are central determinants of educator AI literacy [1], and AI literacy is the strongest predictor of pedagogically relevant use or rejection of generative AI [10]. A systematic review finds learner autonomy and literacy emerging as central pedagogical concerns, yet documents little attention to faculty as designers of AI rather than users of it [11].

Commercial AI Systems Inadequate

While commercial AI chatbots provide immediate responses to students' questions, their design often causes the loss of human nuance and discourages critical thinking, a problem that is corroborated by controlled evidence [12]. A study of a freely available tutor built on metacognitive scaffolding found that pedagogical design, not model fine-tuning, determined whether students learned or merely offloaded the task [13]. An evaluation of ChatGPT as a teaching assistant in introductory programming found it fluent yet inclined to supply finished answers where guided questioning was pedagogically required [14]. Effective communication today requires critical thinking to navigate diverse perspectives [15], foster relational understanding [16], and reflect on personal and others' worldviews [17]. These are essential skills for academic and career success [18] [19]. Research confirms that gen AI frequently acts as a "power persuader," shaping professional judgment and disputing human's validation [20].

Critical Thinking in Higher Education

This highlights the urgent need for an open-access and customizable AI system to enhance critical thinking. Drawing on research on the positive effect of AI-framed questioning [21], we built Teach Anything to offer adaptive responses and ask students questions in order to scaffold their critical thinking process. Rather than functioning merely as an information retrieval tool, the faculty-designed AI applications on our platform can stimulate critical thinking.

Why We Built Teach Anything

Teach Anything addresses several of these problems. Some faculty regard AI as an employable skill for students, while others reject AI because it disrupts academic integrity. Faculty cannot agree on what the problem is, and each solution has a trade-off [22], [23]. Higher education does not intrinsically need AI, but faculty need space for experimentation. We need an inclusive environment to share untold stories and to tell new stories about society and civic technology.

Built by educators for educators, Teach Anything is a permanently free platform for professors to design their own open-access AI. No technical knowledge is required for faculty to use it.

Teach Anything, https://www.teachanything.ai, is open source and open access. It models ethical practices in human-AI collaboration and uses inclusive design principles. Teach Anything is a civic science project in which faculty and students collaborate on a solution that is public interest technology. Civic science empowers citizens to actively shape technologies that impact their life.

Teach Anything's goal is to strengthen faculty's ability to independently evaluate AI systems and to make decisions about pedagogy with or against gen AI in our times. Teach Anything's humanistic methods and community-oriented principles move away from a binaristic model of humans versus machines (or consumers versus corporations) to hypothesize an ecology of ideas suitable for higher education contexts. Teach Anything prioritizes positive outcome for educators and students.

Moreover, all AI applications currently in use in education are commercial and were not designed specifically for the purposes of teaching or learning. Commercial LLMs generally do not allow changes to their parameters. Free tier AI

applications are not open-access resources, because they set usage limits, run advertisements, and turn the users' data into products. Teach Anything is free and open access. It does not use faculty data or students' chats to train any LLMs.

Since faculty build their custom AI applications on Teach Anything with their own data and system prompts, they have more control over its behavior than a commercial product allows. Control here operates at the application layer: over the system prompt, the curated corpus, and the choice of base model. It does not extend to the pre-training of the open-weight model itself, a distinction that matters when open and proprietary systems are compared directly [24]. The retrieval corpus for their applications is transparent and intentioned, not indiscriminately scraped from unvetted sources.

We are committed to open culture. Faculty users are not locked into any one company's ecosystem. They can freely experiment with more than 7 medium-weight, opensource LLMs (such as Mistral and Llama, between 10-100 billion training parameters) and switch them at will. The pre-trained parameters of these open-weight models are publicly available for developers to fine tune, though Teach Anything itself does not require or perform fine-tuning.

Unlike for-profit, proprietary systems, Teach Anything provides transparency and inclusivity by allowing professors to curate, inspect and modify the retrieval corpus, the system prompt, and the generation settings, including the temperature (or creativity) and length of AI's outputs. This approach not only democratizes advanced technologies but also fosters a collaborative community of educators and researchers committed to ethical AI implementation. Transparency in AI design ensures that the tools developed remain accountable. While commercial culture is prone to simplification and sloganization [25], open culture encourages nuanced deliberation [26].

Key Benefits

Faculty as AI designers can act as "translator-mediators" [27] rather than victimized consumers. This is the case because faculty have considerably more control over the retrieval corpus and their applications' behaviors than any commercial platform permits. It is an important first step toward building literacy to learn to curate datasets, configure a tool, and customize system prompts. Open-weight models are not themselves exempt from biases. A prospective controlled evaluation of an open model against a proprietary one found differences in factual reliability running in both directions [24]. However, instructors do not need full control of the underlying foundation model to develop literacy. Through their pre-defined corpus for retrieval and system prompts, instructors could minimize biases and misinformation.

Through Teach Anything, what faculty deploy is not a base model but an application layer over open-source LLMs. Through a curated corpus and comprehensive, domain-specific system prompts, faculty govern the model's behavior which the instructors can inspect and improve upon. The open-access Teach Anything helps faculty develop AI literacy in order to guide students effectively in responsible AI use and rejection [28].

In building their own "narrow" AI applications, faculty create and define their corpus for retrieval. In contrast, commercial AI scrapes data from the web in a nontransparent manner [29], raising privacy, copyright and trust issues. Those LLMs are trained on datasets with inherent biases of white, masculine, and Western perspectives baked into them [30], [31]. Further, university faculty and students are marginalized in deployment decisions of AI [32], [33]. This problem emerged from depriving faculty of their agency and from training data that has been indiscriminately sourced from the web regardless of quality [34]. Research characterizes these harms as ongoing asymmetries rather than incidental errors [35].

Our argument, however, does not rest on the base model. Using our system, instructors choose from available open-source LLMs to build a customized application. Comparative research supports our claim: educational value is determined by pedagogical design rather than by model access [13]; retrieval errors concentrate where the supplied context is inappropriate or insufficient [36]; and the trade-offs between open and proprietary systems rest on how each is configured for its domain [37]. Biases within open-source LLMs are minimized by the instructors' dataset, system prompts, and retrieval instructions. For instance, instructors could program Teach Anything to respond to particular types of questions by drawing on only designated files.

Key steps for faculty to build their own AI

A Gold open access, SOC2-compliant platform, Teach Anything supports the creation of custom-configured, multilingual AI chatbots by faculty partners. The architecture is modular, and instructors, rather than vendors, control every layer of it. Files the faculty member uploads are extracted, indexed, and embedded into a retrieval corpus, and a retrieval-augmented generation (RAG) pipeline grounds each response in that corpus rather than in the open web. A discipline-specific system prompt governs how the application behaves—whether it answers, withholds, or asks a question in return. Faculty select from more than seven open-weight models and can switch between them at any point, including after deployment, so that no application is bound to one vendor's model. A temperature control from 0 to 100 calibrates how conservative or exploratory the output is, and a token setting determines how long each response runs. None of these operations trains or fine-tunes a model; they are acts of curation and configuration, and it is at this layer, rather than in the weights of the base model, that pedagogy is decided [13], [36]. Without technical knowledge, faculty in any discipline could follow these steps to create and experiment with their custom-configured, domain-specific applications:

(1) Faculty would begin by selecting one of the medium-weight, open-source LLMs. They can swap the LLM at any time even after their application is deployed.

(2) Faculty would upload their files as the foundational data set. When interacting with students, the AI provides citations down to the page. It refers to specific sections of specific files.

(3) Faculty can also use Teach Anything's web crawler to index web resources, such as open education resources (OER) or open-source materials.

(4) Faculty would then write, and refine, their custom system prompts in any of the supported natural languages. They could design their AI to be a provocateur or a Socratic partner that asks students questions rather than write essays for students.

(5) Faculty can adjust their AI's "temperature" of creativity. They can set the number of tokens to control how long their AI's output would be in each instance.

(6) When completed, faculty can easily share their AI via permalinks that do not require students to login. The permalink can, for example, be added to existing learning management systems (LMS).

(7) Faculty also has the option of using the automatically generated HTML or Next.js code to embed their chatbot as a floating button on any website.

(8) After implementation, faculty can check, on their dashboard, anonymized student chat records. There are also analytics available, including commonly asked questions by students, outputs using source content (RAG hit rate), average duration of chat sessions, and low-confidence queries. Faculty use this insight to adjust their pedagogy and refine their AI.

The faculty-designed chatbots will only reference the uploaded files and the faculty designated web sources. Keeping the inference window small significantly reduces the chances for factual errors and fabricated information. The baseline risk is substantial. Measured hallucination rates in systematic-review tasks reach 39.6% for GPT-3.5, 28.6% for GPT-4, and 91.4% for Bard. The primary problem is fabricated information [38]. Invented sources and quotations are especially corrosive in humanities teaching, where novices are least equipped to detect them [39]. Constraining the model through system prompts and a bounded corpus reduces this risk significantly even if it does not completely remove it [36]. The corpus includes faculty's files and web sources designated by faculty. Our web crawler is designed with the ethical principle to respect robots.txt, a file at each website's root folder. This file directs web crawlers on the paths they may access and paths that are off-limits. Honoring robots.txt and the site owner's wishes is standard for ethical crawling.

These chatbots, operating in 58 languages with textual and voice input, can offer adaptive responses to suit the learner's needs, taking as a cue the ways in which the student frames a question. Students' full privacy is guaranteed. Faculty can upload material and write system prompts in different languages. Their AI applications will respond in the language which students choose to use in their queries.

Teach Anything uses only open-source, mediumweight LLMs. When processing student queries and generating outputs, Teach Anything uses both the embedding and retrieval augmented generation (RAG) methods to ground responses in pre-defined, curated content. This smaller inference window enables more accurate operations, because the faculty-designed applications reference small, pre-defined datasets and follow specific system-level instructions tuned for pedagogical benefits. A small corpus with clear boundaries is more effective for RAG than a large dataset. Research on retrieval systems indicates that errors occur when the retrieved context is unclear to answer the query [36]. Context-aware retrieval architectures are shown to be more effective than generalists tools, such as ChatGPT, in specialist domains [40]. After generating each output, the chatbot reveals its reasoning process for transparency and pedagogical benefits. Students enjoy complete privacy because Teach Anything does not store their personal information or preferences. Their interactions are not sent off to train the LLM. Commercial LLMs are trained on stolen texts that are not disclosed or authorized. Further, users' chats are used by AI vendors to train their LLM. Our system is transparent and ensures data privacy and security.

Commercial gen AI products store usage history that harms the users [41]. In fact, prolonged interaction with commercial AI systems that store user information increases the models' digital sycophancy [42]. These models mirror users' views regardless of whether those perspectives are beneficial to users' education. Teach Anything is designed to be productively "amnestic" in that it does not store student profiles or personal preferences. This feature prevents the issue of sycophancy common in commercial AI products. Interacting with digital platforms without leaving traces is a rare respite in the AI-dominated world. Every time a student interacts with a faculty-designed AI on Teach Anything, it is a fresh start.

Results and findings

More than 400 faculty members participate in Teach Anything. They are based in East Asia, India, Australia, the Middle East, Turkey, Egypt, Western Europe, Norway, the UK, North Africa, Canada, the US, Brazil, Argentina, and elsewhere. A small number of them took the extra step of sharing reflections on their experience.

As a result of using Teach Anything, these participants (n=18) reported improvement in self-efficacy in understanding AI (Table 1). 70% gained understanding of AI's pedagogical implications, and 30% developed knowledge of AI's technological implications through their control of its temperature, system prompts, and inference window (files they uploaded). In terms of the next steps, 90% indicated that they were better positioned to make responsible, effective, and context-appropriate decisions about how to teach with or against AI systems. 10% indicates that they had not yet decided on AI's role in their class but had stronger understanding of AI's operation.

Developing self-efficacy is the first step toward pedagogy that addresses challenges posed by gen AI. We thus contribute to a shared public knowledge base about AI systems and their societal impacts, particularly in the domain of higher education. We enhance literacy and self-efficacy by hands-on practice.

Faculty participants also suggested and beta-tested new features in collaboration with our team, such as allowing students to generate practice quizzes. Aligning with participatory design principles, Teach Anything strengthens a broad range of stakeholders' ability to understand, evaluate, and responsibly use AI systems. Through using our open-access platform, faculty stakeholders gain first-hand experience in order to better understand how AI systems work and where they can fail. Collectively, we evaluate the quality and limitations of AI-generated outputs.

Teach Anything is an open-source, open-access infrastructure that enables faculty across disciplines to

experiment with and design custom AI applications. As part of our efforts to promote open culture, Teach Anything offers faculty access to the proverbial engine room, with opportunities to adjust the retrieval corpus, system prompt, and generation settings for trustworthy outputs. By identifying ethical considerations and trust issues related to the use of AI in learning and development, this design-based research contributes to the ongoing dialogue about responsible AI implementation. These insights will be crucial for organizations, educational institutions, and policymakers as they navigate the integration of AI into learning and development processes.

Teach Anything also promotes equity by making advanced tools more accessible to a wider range of faculty, regardless of financial resources and technical knowledge. This research created the following digital deliverables:

1. A permanently free, open-access platform for higher education faculty to create open-source chatbots that support student learning in their specific courses.

2. A series of open-access chatbots specializing in the disciplines of faculty participants.

3. Monthly newsletters to faculty members and platform upgrades including new features.

Faculty and students evolve their modes of interaction with AI created via Teach Anything in comparison to general-purpose, commercial models.

Table 1. The impact of Teach Anything on the users across diverse demographics

Objective	Activity	Outcomes
Increase faculty's AI literacy (how AI systems work and where they can fail) through professional development opportunities offered by Teach Anything	More than 400 faculty members utilize the open-access Teach Anything to design their own domain-specific AI applications, following our illustrated, step-by-step guide.	• Faculty use the open-access infrastructure to create custom-configured AI chatbots • Faculty develop learning activities using their custom AI • Faculty test new assignments with students
Evaluate the limitations of AI-generated outputs	Faculty assess anonymized student chat records to gauge the pedagogical effectiveness of their instruction and of their custom AI. Faculty form a community of practice to share insights and strategies.	Faculty evaluate the quality and limitations of AI generated outputs and adjust • their AI applications • their pedagogical approaches accordingly
Make responsible and context appropriate decisions on AI and education	18 faculty refine their system prompts and custom-configured AI based on evaluation of classroom activities and anonymized student chat histories	These 18 faculty members discuss their development of self-efficacy with project leader in a preliminary survey

Comparison with Other Platforms

Similar projects exist, but they are by invitation only, commercial (BoodleBox), operating as fee-for-service and for K-12 teachers (Playlab), not sharable (NotebookLM), or designed for a single course. NotebookLM locks users into Google's ecosystem. It is a commercial platform with a free tier; it is not open source and open access. Users can only upload files and do not have backend access. Public sharing is difficult. Playlab and BoodleBox only allow users to choose from among commercial LLMs.

UCLA and Arizona State University, respectively, engage third-party AI vendors to build textbooks and assignments that students have to pay for. UCLA offers a comparative literature course with a controversial AIgenerated textbook [43], [44] and class assignments provided by Kudu, a third-party platform [45]. Without faculty consent or knowledge [46], Arizona State University quietly used online instructional material from faculty to create AI-generated learning modules [47]. These partnerships with for-profit entities are often ineffective and untrustworthy. ASU Atomic generated faulty and decontextualized information [46].

Little Language Models by MIT Media Lab [48] is by invitation only and aims at elementary education. CS50 Duck, a virtual rubber duck chatbot, allows students in Harvard's Introduction to Computer Science to "interact directly with GPT-4 in a controlled manner" [49]. These models do not provide backend access, are not open access or open source.

Teach Anything is a grass-root, faculty-student collaborative learning project that serves as public interest technology. Our open-access platform allows educators everywhere to create custom AI for free. It does not write essays or code for students, and it gives educators and backend access to participate in civic science.

Conclusion: GenAI is not a Black Box

Teach Anything demonstrates that gen AI need not be a black box. In fact, it is AI companies that create and propagate the myth of a black box technology. Engineers know how its parts work by simply generating texts or images "appropriate to the scenario described by the prompts" that evoke "certain conceptual spaces" [50]. For its own interest, the AI industry has woven the so-called black boxes "into the stories they tell." If it appears to be a black box it is only because of the companies' choice of using "an impenetrable cluster of numerical values, linking vectors clumped together incomprehensibly in the training phase" [50].

The open-access Teach Anything changes these stories by empowering faculty members as designers. Large corporations offer AI models that cannot be edited or finetuned by consumers. Faculty do not have control over these models that lack interpretability. When addressing literacy challenges, bigger is not always better, as "large foundation models are giving way to smaller, more task specific models" [51], [52]. Teach Anything supports the creation of custom and "narrow" AI applications for specific

subjects or pedagogical methods. Teach Anything gives faculty more control.

Acknowledgment

Led by Alexa Alice Joubin, Teach Anything is co-built by Akhilesh Rangani at George Washington University. In collaboration with Joubin and Rangani, Ryan Kosloff joined later and contributed several new features.

Declarations

Funding: Research supported by George Mason University Institute for Humane Studies and by George Washington University's Public Interest Technology Scholar Program, GW Coders, Taiwan Studies Open Access Award, and Columbian College of Arts and Sciences Nick-of-Time Impact grant.

References

[1] A. M. Ayyoub et al., "Advancing higher education with GenAI: Factors influencing educator AI literacy," Front. Educ., vol. 10, 2025, doi: 10.3389/feduc.2025.1530721.
[2] J. Du and L. Kohnke, "Repositioning teachers in the age of AI: A cross-disciplinary study of perceptions and practices in China," Teach. Educ., vol. 37, no. 3, pp. 281–305, 2026, doi: 10.1080/10476210.2025.2590694.
[3] N. Najjar, M. Rouphael, M. El Hajj, T. Bitar, P. Damien, and W. Hleihel, "Faculty perceptions and adoption of AI in higher education: Insights from two Lebanese universities," Educ. Sci., vol. 16, no. 1, art. 55, 2026, doi: 10.3390/educsci16010055.
[4] National Institute of Standards and Technology, "Trustworthy and responsible AI — Artificial intelligence risk management framework: Generative artificial intelligence profile," NIST, Gaithersburg, MD, USA, 2025. [Online]. Available: https://www.nist.gov/trustworthy-and-responsible-ai
[5] D. Broniatowski, A. A. Joubin, A. A. Susan et al., "The epistemologies of trust: Conflicting worldviews in the trustworthy AI discourse," AI Ethics, vol. 6, art. 260, 2026.
[6] X. Tan, G. Cheng, and M. H. Ling, "Artificial intelligence in teaching and teacher professional development: A systematic review," Comput. Educ.: Artif. Intell., vol. 8, art. 100355, 2025.
[7] N. B. Milman, "Exploring the crossroads of technology, society, and the human condition: GW education and generative artificial intelligence (AI)," presented at the Society of the Emeriti September Lecture, Sep. 26, 2025.
[8] C. E. Watson and L. Rainie, "Leading through disruption: Higher education executives assess AI's impacts on teaching and learning," AAC&U, Washington, DC, USA, Jan. 2025. [Online]. Available: https://dgmg81phhvh63.cloudfront.net/content/user photos/AACU_AI_Report_2025.pdf
[9] D.-K. Mah and N. Groß, "Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs," Int. J. Educ. Technol. High. Educ., vol. 21, art. 58, 2024, doi: 10.1186/s41239-024-00490-1.
[10] N. Jogezai, F. A. Baloch, M. Jaffar, and G. Khljl, "From technology to pedagogy: Determinants of university faculty's pedagogically relevant use of generative AI," Qual. Assur. Educ., vol. 33, no. 4, pp. 618–634, 2025.
[11] Y. Qian, "Pedagogical applications of generative AI in higher education: A systematic review of the field," TechTrends, vol. 69, no. 5, pp. 1105–1120, 2025, doi: 10.1007/s11528-025-01100-1.
[12] J. A. Bowen and C. E. Watson, Teaching with AI: A Practical Guide to a New Era of Human Learning. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2024.
[13] L. P. Cordova, T. Mendoza, S. Holmes, K. Megginson, B. Webster, and D. Gregory, "Not all chatbots teach: Evidence for pedagogical design in AI-assisted technical education," in Proc. 26th ACM Annu. Conf. Cybersecurity Inf. Technol. Educ. (SIGITE '25), Nov. 2025, doi: 10.1145/3769694.3771107.
[14] Z. Ahmed, S. S. Shanto, and A. I. Jony, "Potentiality of generative AI tools in higher education: Evaluating ChatGPT's viability as a teaching assistant for introductory programming courses," STEM Educ., vol. 4, no. 3, pp. 165–182, 2024, doi: 10.3934/steme.2024011.
[15] Y. T. Nakamura and J. Milner, "Inclusive leadership via empathic communication," Organ. Dyn., vol. 52, no. 1, art. 100957, 2023, doi: 10.1016/j.orgdyn.2023.100957.
[16] Y. T. Nakamura, J. Hinshaw, and R. Burns, "Developing women's authenticity in leadership," Merits, vol. 2, no. 4, pp. 408–426, 2022, doi: 10.3390/merits2040029.
[17] Y. T. Nakamura, J. Milner, and T. Milner, "Inclusive-empathy in leadership," J. Appl. Behav. Sci., vol. 58, no. 1, pp. 161–163, 2022, doi: 10.1177/0021886320982022.
[18] Y. T. Nakamura, J. Hinshaw, and D. Yu, "The role of empathy in developing ethical leadership: Neurobiology and video-based approaches," in Ethics and Human Resource Development, A. Alizadeh and D. F. Russ-Eft, Eds. Cham, Switzerland: Palgrave Macmillan, 2024, pp. 449–468.
[19] Y. T. Nakamura, J. Milner, D. Yu, and J. Hinshaw, "The role of empathy in leadership ethics: Examining relational leadership practice through video-based methods," Bus. Ethics Q., 2025, in press.
[20] S. Randazzo, H. Lifshitz, A. Joshi, F. Dell'Acqua, K. C. Kellogg, and K. R. Lakhani, "GenAI as a power persuader: How professionals get persuasion bombed when they attempt to validate LLMs," SSRN, Jul. 1, 2025, doi: 10.2139/ssrn.5678644.
[21] V. Danry et al., "Don't just tell me, ask me: AI systems that intelligently frame explanations as questions improve human logical discernment accuracy over causal AI explanations," in Proc. CHI Conf. Human Factors Comput. Syst., Apr. 2023.
[22] T. Corbin, M. Bearman, D. Boud, and P. Dawson, "The wicked problem of AI and assessment," Assess. Eval. Higher Educ., Sep. 3, 2025, doi: 10.1080/02602938.2025.2553340.
[23] S. Otterman, "A.I. is coming to class. These professors want to ease your worries," The New York Times, Jan. 17, 2026. [Online]. Available: https://www.nytimes.com/2026/01/17/nyregion/ai college-classes.html
[24] S. Eminovic, B. Levita, R. Schmidt, M. Lindholz, D. Desser, T. Penzkofer, M. P. Wattjes, and J. Nawabi, "Open-source versus proprietary large language models for academic writing in medicine: A controlled evaluation of DeepSeek-R1 and ChatGPT-o1 in neuroradiology," Digit. Health, vol. 12, Feb. 2026, doi: 10.1177/20552076261464745.
[25] A. Sayer, "(De)commodification, consumer culture, and moral economy," Environ. Plan. D: Soc. Space, vol. 21, pp. 341–357, 2003.
[26] J. Riskin, "AAUP: Academic freedom on the line. Viewpoint diversity isn't about viewpoint diversity," The Stanford Daily, Feb. 4, 2026. [Online]. Available: https://stanforddaily.com/2026/02/04/aaup-viewpoint-diversity/
[27] B. Czerkawski, "AI and the learning experience design: From divergent creativity to convergent precision," TechTrends, vol. 69, pp. 260–270, 2025.
[28] M. Kassorla, M. Georgieva, and A. Papini, "AI literacy in teaching and learning: A durable framework for higher education," EDUCAUSE Working Group, 2024. [Online]. Available: https://www.educause.edu/content/2024/ai-literacy-in-teaching-and learning/executive-summary
[29] Y. Guo, M. Guo, J. Su et al., "Bias in large language models: Origin, evaluation, and mitigation," arXiv:2411.10915, 2024.
[30] A. Birhane, "Algorithmic injustice: A relational ethics approach," Patterns, vol. 2, no. 2, art. 100205, 2021, doi: 10.1016/j.patter.2021.100205.
[31] S. Cave and K. Dihal, "The whiteness of AI," Philos. Technol., vol. 33, no. 4, pp. 685–703, 2020, doi: 10.1007/s13347-020-00415-6.
[32] T. L. Howard and G. W. Ulferts, "Artificial intelligence and the redefinition of higher education," Higher Educ. J., vol. 46, pp. 1–21, 2025.
[33] J. Batista, A. Mesquita, and G. Carnaz, "Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review," Information, vol. 15, no. 11, art. 676, 2024, doi: 10.3390/info15110676.
[34] I. O. Gallegos, R. A. Rossi, J. Barrow et al., "Bias and fairness in large language models," Comput. Linguist., vol. 50, no. 3, pp. 1097-1179, 2024.
[35] G. Waters and P. Honenberger, "AI biases as asymmetries: A review to guide practice," Front. Big Data, vol. 8, art. 1532397, 2025, doi: 10.3389/fdata.2025.1532397.

[36] H. Joren, J. Zhang, C.-S. Ferng, D.-C. Juan, A. Taly, and C. Rashtchian, "Sufficient context: A new lens on retrieval augmented generation systems," in Proc. Int. Conf. Learn. Representations (ICLR), 2025.
[37] Y. Sugihara, A. Milosavljevic, S. Jankovskaja, and M. Falk, "A review for navigating the trade-offs: Evaluating open-source and proprietary large language models for clinical and biomedical information extraction," Front. Digit. Health, vol. 8, art. 1778786, Apr. 2026, doi: 10.3389/fdgth.2026.1778786.
[38] M. Chelli, J. Descamps, V. Lavoué, C. Trojani, M. Azar, M. Deckert, J.-L. Raynier, G. Clowez, P. Boileau, and C. Ruetsch Chelli, "Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis," J. Med. Internet Res., vol. 26, art. e53164, 2024, doi: 10.2196/53164.
[39] Z. Ahmad, W. Kaiser, and S. Rahim, "Hallucinations in ChatGPT: An unreliable tool for learning," Rupkatha J. Interdiscip. Stud. Humanit., vol. 15, no. 4, 2023, doi: 10.21659/rupkatha.v15n4.17.
[40] A. Sriram, N. Maheswaran, B. Sundan, and S. Krishnamoorthy, "Context-aware retrieval-augmented generation for artificial intelligence in urology," Cureus, vol. 17, no. 7, art. e88167, Jul. 2025, doi: 10.7759/cureus.88167.
[41] A. Mahdawi, "Beware what you tell your AI chatbot. It's not a shrink — it's a snitch," The Guardian, May 13, 2026. [Online]. Available: https://www.theguardian.com/commentisfree/2026/may/13/beware what-you-tell-your-ai-chatbot-its-not-a-shrink-its-a-snitch
[42] S. Jain, C. Park, M. Viana, A. Wilson, and D. Calacci, "Interaction context often increases sycophancy in LLMs," in Proc. 2026 CHI Conf. Human Factors Comput. Syst., Apr. 2026, doi: 10.1145/3772318.3791915.
[43] C. Dutton, "The course is about literature. Its textbook was generated by AI," The Chronicle of Higher Education, Dec. 12, 2024. [Online]. Available: https://www.chronicle.com/article/the course-is-about-literature-its-textbook-was-generated-by-ai
[44] K. Palmer, "The AI-generated textbook that's making academics nervous," Inside Higher Ed, Dec. 13, 2024. [Online]. Available: https://www.insidehighered.com/news/faculty/learning assessment/2024/12/13/ai-assisted-textbook-ucla-has-some academics-worried
[45] UCLA Newsroom, "Comparative lit class will be first in Humanities Division to use UCLA-developed AI system," Regents of the Univ. of California, 2025. [Online]. Available: https://newsroom.ucla.edu/stories/comparative-literature-zrinka stahuljak-artificial-intelligence
[46] E. Whitford, "Faculty concerned about ASU's 'Frankensteinian' AI course builder," Inside Higher Ed, Apr. 29, 2026. [Online]. Available: https://www.insidehighered.com/news/tech innovation/artificial-intelligence/2026/04/29/faculty-concerned about-asus-new-ai-course
[47] ASU Atomic, "Level up your professional potential," Arizona Board of Regents, 2026. [Online]. Available: https://atomic.asu.edu
[48] M. Dhariwal and S. Dhariwal, "Little language models: Empowering children to be future AI modelers," LinkedIn, Aug. 12, 2024. [Online]. Available: https://www.linkedin.com/pulse/little language-models-empowering-children-future-ai-modelers-e5cpc/
[49] R. Liu, C. Zenke, C. Liu, A. Holmes, P. Thornton, and D. J. Malan, "Teaching CS50 with AI: Leveraging generative artificial intelligence in computer science education," in Proc. 55th ACM Tech. Symp. Comput. Sci. Educ. V. 1, 2024, pp. 750–756, doi: 10.1145/3626252.3630938.
[50] E. Salvaggio, "The black box myth: What the industry pretends not to know about AI," Tech Policy Press, Jun. 17, 2025. [Online]. Available: https://www.techpolicy.press/the-black-box-myth-what the-industry-pretends-not-to-know-about-ai/
[51] L. Hinkel, "Creating AI that matters: How the MIT-IBM Watson AI Lab is shaping AI-sociotechnical systems for the future," MIT News, Oct. 21, 2025. [Online]. Available: https://news.mit.edu/2025/creating-ai-that-matters-1021
[52] J. Lin, J. Tang, H. Tang, S. Yang, W.-M. Chen, W.-C. Wang, G. Xiao, X. Dang, C. Gan, and S. Han, "AWQ: Activation-aware weight quantization for on-device LLM compression and acceleration," in Proc. 5th MLSys Conf., Santa Clara, CA, USA, 2024.

S A M Y R A D 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Skill (Mis)Alignments: A case study of a recent educational program against industry needs

Juan A. Talamás-Carvajal (e-mail: juan.talamas@tec.mx), Gustavo De Los Rios (e-mail: A01410922@tec.mx), and Patricia Caratozzolo (e-mail: pcaratozzolo@tec.mx)

[1] Institute for the Future of Education, Tecnologico de Monterrey, Mexico

[2] School of Engineering and Sciences, Tecnologico de Monterrey, Mexico

Article DOI: 10.68286/samyrad2026.356

ABSTRACT

The persistent misalignment between Higher Education curricula and rapidly evolving industry requirements challenges institutions seeking to prepare graduates for dynamic labor markets. This study presents a data-driven approach for assessing curriculum–industry alignment through a case study of a recently developed Information Technology engineering program and occupational demand in the INFOCOMM (Informatics and Communication) industry. We extracted Knowledge, Skills, and Abilities (KSAs) from educational program content and job postings from Mexico and the United States using Natural Language Processing and Large Language Models, then classified them using a domain-specific Infocomm taxonomy. We then used weighted Jaccard similarity to quantify alignment between the educational program and selected occupations over seven months. Results reveal substantial variation across occupations and KSA dimensions. Skills showed the highest overall alignment, followed by Abilities, whereas Knowledge exhibited the greatest misalignment. Occupations closely associated with software engineering generally showed stronger correspondence with the curriculum, while more specialized or adjacent occupations showed substantially lower alignment. The findings show that curriculum–industry alignment is multidimensional and occupation-dependent, not adequately captured by a single aggregate skill-gap measure. The proposed approach provides universities with an evidence-based mechanism for identifying curricular strengths, gaps, and potential areas for targeted intervention as industry requirements evolve.

KEYWORDS Higher Education, Informatics and Communication, Skills, and Abilities, Natural Language Processing, Large Language Models

How to cite: Talamas-Carvajal, J. A., De Los Rios, G., & Caratozzolo, P. (2026). Skill (mis)alignments: A case study of a recent educational program against industry needs. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.356

Introduction

One of the longest-standing gaps that educational institutions face is the disconnect between what Higher Education Institutions (HEIs) teach and what the targeted industries need. Recent graduates are often told they will “learn the job when they get the job” and that much of what they were taught is either obsolete or not appropriate for certain roles. This misalignment stems from 2 main factors: industries that constantly push for innovation and optimization, and the slow response time of Educational Institutions to changes outside them [1]. Paradoxically, while many state-of-the-art discoveries occur within these institutions, the adoption of novel technologies and strategies is much slower [2]. Recent studies have further highlighted that digital transformation, automation, and Industry 4.0 paradigms continue to accelerate changes in workforce requirements, intensifying the need for agile educational adaptation [3]. While this gap has generated tension among industry, universities, and students, it also creates an opportunity for educational

innovation. Although eliminating the gap may not be feasible because of the need for technological maturation, proprietary industrial knowledge, accreditation constraints, and institutional inertia, it can certainly be reduced. One frequently proposed strategy is to develop stronger collaborations between HEIs and industry stakeholders, enabling curricula to evolve in response to emerging professional demands [3]. Such collaborations can also improve graduate employability, facilitate workforce transition into emerging sectors, and support regional economic competitiveness. A promising approach to reducing this gap is to systematically analyze competencies and workforce skills. The European Commission Joint Research Center defines competencies as the "*demonstrated ability to use knowledge, know-how, experience and – job-related, personal, social or methodological – skills, in work or learning situations and in professional and personal development*" [4]. Competencies are commonly operationalized through classifications such as Knowledge, Skills, and Abilities (KSAs), which provide a structured framework for analyzing workforce preparedness and occupational requirements [5].

In this study, we use the term "Skills" as a general reference to the components that comprise competencies within the KSA framework. The term "Knowledge" is self-explanatory, but we differentiate Skills and Abilities as follows: a Skill is a capacity necessary to complete a particular task, and an Ability relates to having the sensory, physical, psychomotor, or cognitive means to perform an action. Any competent professional carries a specific set of KSAs necessary to perform effectively in their occupational domain. Consequently, individuals possessing those same KSAs are theoretically more likely to succeed in equivalent professional roles. Educational institutions can therefore mitigate curriculum obsolescence by continuously monitoring evolving industry-relevant KSAs and incorporating them into academic programs through targeted learning experiences, competency-based curricula, and updated course content. This competency-centered perspective has become increasingly important as labor markets evolve more rapidly and interdisciplinary technical profiles become more common. One way this alignment has strengthened in recent years is through automated extraction and analysis of skills from labor market data. Advances in Natural Language Processing (NLP) and Machine Learning (ML) have enabled large-scale analysis of job postings, allowing researchers and policymakers to identify emerging competencies and changing occupational demands with greater accuracy and speed. Taxonomies such as ESCO in Europe and SINCO in Mexico, along with recent industrial skill studies, have demonstrated the usefulness of computational approaches for mapping workforce requirements and supporting educational planning [6], [7], [8], [9], [10]. Most existing studies focus on labor market analysis or curriculum design independently, rather than systematically comparing the two ecosystems. Despite these advances, little research directly compares educational program content with large-scale industry skill taxonomies to identify concrete curricular gaps and intervention opportunities. Methodologies are still needed to systematically evaluate how well modern educational programs align with rapidly evolving industry ecosystems, such as the Infocomm (Informatics and Communication) industry, which encompasses information technologies, software services, digital communication systems, telecommunications, and networking.

In this paper, we present a case study demonstrating how a taxonomy developed for the Infocomm industry can be used to evaluate educational programs and identify gaps, obsolete topics, and core KSAs relative to industry needs. We also show how the results can inform interventions that improve curriculum alignment and strengthen graduate preparedness. To guide this study, we propose the following research questions:

RQ1. *To what amount do modern educational programs match the needs of their target industry in terms of developing the necessary skills for jobs in that industry?*

RQ2. *Can the differences found be leveraged towards improving educational programs, student preparedness, and/or collaborations between industry and HEIs?*

Background

The Skill Gap

As mentioned before, a persistent gap exists between what students learn in school and what they need to succeed in their occupations after graduation. This gap reflects differences in KSAs between what students graduate with and what industry requires. In short, it is a difference in competence. This phenomenon has been widely discussed in educational policy, workforce development, and labor economics literature, particularly in the context of digital transformation and rapidly evolving technological ecosystems [4]. This gap goes beyond what might traditionally be called "corporate culture," meaning the rules, attitudes, and processes specific to individual workplaces. Instead, it reflects structural differences between educational preparation and occupational expectations. The European Union defines this phenomenon as the "Skill Gap" or "Skill Needs," which remains the subject of extensive research and policy development initiatives [4]. Similar concerns have also been identified in studies focused on Industry 4.0, digital economies, and emerging technical professions, where workforce demands evolve faster than institutional curricula [3].

This issue is problematic because no single solution exists, and many factors widen the gap. The time, investment, and effort required to develop the necessary skills are among the biggest barriers to entry for newly graduated students entering the workforce. Furthermore, the increasingly interdisciplinary nature of technical occupations means graduates are expected to have not only technical knowledge but also communication, adaptability, teamwork, and problem-solving skills. Therefore, both industry and HEIs have an interest in ensuring students are as prepared as possible before entering the job market. A seemingly straightforward way to address this issue is to continuously develop specific KSAs as needed. For example, continuous professional training could keep the workforce up to date on technological breakthroughs, practical courses could strengthen operational competencies, and holistic educational approaches could support the development of broader abilities. However, the challenge lies not only in developing KSAs but also in identifying which KSAs are most relevant, emerging, or strategically important for a given occupational ecosystem. This challenge has motivated the development of large-scale labor market analysis methodologies that systematically

identify and monitor workforce competencies using computational approaches and data-driven models.

Automated Skill Extraction and Labor Market Intelligence

One current solution is to automate skill extraction from job markets. Traditional analyses of occupational skills were often manual, time-intensive, and hard to scale because job postings are unstructured. While basic information such as job titles and company names tends to follow relatively standardized formats, the most relevant information for workforce analysis, namely the competencies being requested, is commonly embedded within *free-form* job descriptions. Recent advances in Natural Language Processing (NLP), Machine Learning (ML), and Large Language Models (LLMs) have enabled the automated extraction, classification, and analysis of competencies from large collections of online job postings. These approaches allow researchers and policymakers to process massive volumes of labor market information from platforms such as *LinkedIn*, *Indeed*, and *JORA* far more efficiently and consistently than manual methods. Several initiatives have demonstrated the value of computational skill extraction methodologies. Examples include the ESCO taxonomy developed by the European Commission, the SINCO occupational classification system developed by INEGI in Mexico, and recent studies focused on the Automotive and Infocomm industries [6-10]. These initiatives support workforce planning, educational policy development, occupational mobility analysis, and strategic upskilling efforts. One key advantage of automated skill extraction approaches is their ability to capture much larger and more representative samples of industry needs than traditional manual analyses. As a result, the insights generated from computational approaches can more accurately reflect the dynamic evolution of labor market demands. Additionally, continuous monitoring of online job postings enables near-real-time identification of emerging technologies, evolving professional profiles, and evolving competency requirements. Despite these advances, extracting skills from labor market data alone does not fully address the educational alignment problem. Most existing approaches focus primarily on industry demand, while comparatively few studies compare extracted competencies with educational program content and graduate profiles. Consequently, methodologies that bridge labor market intelligence, curriculum analysis, and educational evaluation remain needed.

Competency Taxonomies and Educational Alignment

Competency taxonomies play a critical role in connecting labor market analysis with educational program evaluation. By organizing competencies into structured categories such as KSAs, these frameworks provide a common language for systematically comparing workforce requirements and educational outcomes. Taxonomies such as ESCO and SINCO have demonstrated that standardized competency representations can support multiple objectives, including occupational mobility, reskilling, upskilling, workforce forecasting, and educational planning [9], [10]. In educational contexts, competency frameworks also facilitate curriculum mapping, learning outcome assessment, and identifying gaps between intended graduate profiles and actual industry requirements.

The growing availability of computational methods for competency extraction has enabled more sophisticated alignment methodologies between industry and academia. In particular, similarity-based approaches and semantic matching techniques enable researchers to compare educational content with labor market requirements at scale, revealing both strengths and weaknesses in academic programs. However, research still has limited focus on applying these methodologies directly to detailed educational program structures, including course objectives, content, and graduate profiles. Most existing studies focus on labor market analysis or curriculum design independently, rather than on systematic comparisons between the two ecosystems. This limitation becomes especially relevant in rapidly evolving sectors such as the Infocomm industry, where educational programs must continuously adapt to technological transformation and shifting occupational demands.

This study addresses this gap by using an Infocomm competency taxonomy to compare workforce requirements with the content and graduate profiles of a modern engineering program. Through this approach, we aim to identify curricular gaps, core competencies, and opportunities for improving alignment between HEIs and industry needs.

Materials and Methods

The present study comprises three main stages: skill extraction from job postings obtained from online sources relevant to the targeted industry, Skill extraction from the educational program(s) being evaluated, and, finally, a comparison/matching methodology to measure the amount of shared and missing Skills between them. The Infocomm taxonomy has already been developed, and its results are shown in [6]. The process described below has been validated in [11,12] through multiple review rounds and cross-validation across annotators.

Stage 1: Labeling and processing pipeline for job postings

The process of building the job postings dataset is composed of several steps: data sweeping or scanning (scraping) for relevant job postings on several job search websites (LinkedIn, JORA, Indeed), atomic idea extraction related to the KSA-O framework from the retrieved job postings, and classification of such ideas into a Skill, Knowledge, Ability, or Other. In the case of “Other”, this refers to requirements for the job that do not fall into the regular KSA terminology, but are nevertheless required. For example, specific credentials, availability, or certifications. This final step also groups variants of a similar Skill into one using a representative term.

We scrape the data through the APIFY platform and retrieve it as JSON or CSV files containing the information summarized in Table 1. The extracted features provide the necessary information for Skill extraction and classification and record the industry's evolution over time.

After obtaining enough data, we used a local version of ChatGPT called TecGPT, based on GPT-4o, to process the information, extract the essential KSA-Os, and place them into a structured Python list. We used a set of specific prompts developed after several iterations. These prompts did the following: create a structured list for readability with

requirements, certifications, essential skills, and other specific knowledge or expertise as separate items. The following is the base prompt used for skill extraction from job descriptions using GPT-4:

System: You are an information-extraction system specialized in identifying skills in job descriptions. Your task is to recognize both technical skills and interpersonal skills explicitly mentioned in the text.

User: *The following is the job description. Extract all explicitly mentioned skills and return them as a Python list.*

The model then returns the results via a .txt file, which can then be added to the dataset. The next step in the pipeline is classifying the recovered atomic ideas into specific KSAs under the taxonomy. This is done by running the output from the local GPT system through a fine-tuned (one-shot prompted) GPT-4o model via the OpenAI API. In this process, we use the Infocomm taxonomy as a reference through a process called Retrieval-Augmented Generation [13], producing a set of Skills, Knowledge, and Abilities aligned with what we found while developing the original taxonomy. To avoid instances where a Skill is expressed in similar terms ("teamwork" and "works in teams", for example), we created hypernym- and hyponym-based groups [14], which were then used in the RAG process. This generalization aimed to convert industry Skills into a form that could be more easily compared with Higher Education Skills.

Table 1. Extracted features from job posting websites

Feature name	Description
salary	Yearly salary range being offered
postedAt	Time since the job was posted
externalApplyLink	Link to direct application page
positionName	Job position being offered
jobType	Full-time or part-time
company	Company offering the job
location	Address of the job being offered
rating	Rating of the job posting
reviewsCount	Number of reviews of the job posting
urlInput	URL used to search for the job
url	URL of the job position
id	Unique ID of the job posting
scrapedAt	Date of scraping
postingDateParsed	Date of publishing
description	Overall description of the job posting
descriptionHTML	HTML version of the description
searchInput	Keyword specified to the scraper if not using a URL
isExpired	Indicates whether the job is still available at the time of scraping
q_extracted	Text query used to scan for a particular position
q_extracted_norm	Normalized version of the search query

To bridge the language differences between HEIs and industry, we took two additional generalization steps: we obtained canonical (more general) hypernyms [15], [16] for the obtained KSAs, and we separated them into parent categories. This way, we can properly identify specific KSAs that might initially register as "not matching" or "missing" due to granularity. One common example is specific programming languages or structures. Cases like SQL, DHCP, and Python can be properly categorized into general groups, and then into parent domains like "Databases", "Networks", and "Programming Language", respectively. This allows for a much clearer and fairer comparison between the two sets of KSAs. Using the previous process, we scraped seven months' worth of job postings.

This resulted in a total of 4032 job postings analyzed over the seven months, as shown in Table 2. The overall distribution of KSAs in the dataset was 66% Knowledge, 25% Skills, and 9% Abilities.

Table 2. Job postings and KSAs per Month

Month	Job Postings	K	S	A
Jul-25	540	11420	8969	5174
Aug-25	519	11209	9061	4971
Sep-25	560	12062	9550	5427
Oct-25	624	14569	10616	5715
Nov-25	547	12036	9235	5213
Dec-25	520	11523	8663	4959
Jan-26	722	15604	11694	6811

Stage 2: Skill extraction from educational program

Building the educational program dataset followed the same process as the job postings dataset. Because the educational models came from an official source, we did not need the initial data-scraping step; the remaining steps matched those described above: atomic idea extraction and classification of those atomic ideas into KSAs via RAG. As with the job postings dataset, a local version of ChatGPT performed atomic idea extraction to process the information, extract the essential KSA-Os, and place them into a structured Python list for later classification using the same taxonomy used for the job posting Skills.

One important detail is that the educational programs used are the "detailed" versions. This means we have not only the names and order of all the courses that comprise the majors, but also the general objective of those courses, the course content, and the specific topics covered in each course. We also recovered the graduate profile for each major, which provides a general description of the capabilities of students who graduate from the major, as well as a summary of the most relevant KSAs they will obtain through the program. The program selected for this comparison is Engineering in Information Technology (ITC), which prepares students for jobs in the Infocomm industry.

Stage 3: Matching methodology

Having obtained both the industry and Educational Institution KSAs, we proceeded to the comparison/matching stage. One aspect of the comparison is that a full comparison of the KSA portfolios of both industry and HEIs is not valuable, as the complete set of KSAs in the industry is much larger than what any institution can teach. Instead, we compare the top Skills, Abilities, and Knowledge needed by industry occupations against the top ones developed by the educational programs. We use the Weighted Jaccard Similarity for this comparison. Jaccard Similarity (or Naïve Jaccard) is a statistical measure that quantifies how similar two sets are. It equals the ratio of the intersection size to the union size. Jaccard similarity ranges from zero to one: zero indicates the sets share no elements, and one indicates they are identical. We use the similarity index to calculate document similarity. In this case, a document is the set of KSAs from both industry and HEIs [17]. Jaccard Similarity is defined as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Weighted Jaccard extends the Jaccard index to numerical vectors as well, if these vectors are non-negative. Weighted Jaccard measures similarity by "calculating the sum of minimum weights divided by the sum of maximum weights for each dimension, resulting in a value between 0 and 1 that represents the overlap magnitude". It is commonly used to compare documents in data mining. Weighted Jaccard is defined as follows:

$$J_w(x,y) = \frac{\sum_i min(x_i, y_i)}{\sum_i max(x_i, y_i)}$$

We obtained different similarity scores for the program against selected occupations for each monthly extraction. In cases where weight scores were obtained, the weights were defined as *0.5 Knowledge + 0.3 Skills + 0.2 Abilities*, based on the observed KSA frequency, as well as the impact of the different categories in determining job matches. It is important to note that the "Others" category was discarded because it doesn't represent skills, but traits like experience, certifications, or place of residence, for example.

Results

Similarity matching between the ITC program curriculum and job postings yielded the following: we observed high match scores (~50% knowledge) for Software and Artificial Intelligence Engineer occupations, and low scores (~5 or less) for ICT security and marketing manager occupations. These results cover 7 months of job postings, during which the match percentage changes each month. Fig. 1 shows these results as a heat map summarizing the results by occupation and month. One important detail is that some months had no job postings for one or more occupations; in those cases, the space was left blank to indicate the absence of a match score. Additionally, most occupations averaged 20% to 30% similarity, while only the first-ranked ones averaged above 40%.

Occupations are listed by the average rank they achieved across all months; for example, Software Engineer averaged 42% similarity throughout the experiment. Some cases show similarity varying from a high score of 43% one month to a low score of 12% another, as in the Data Scientist occupation. The occupations with the fewest registered appearances were Computer System Analyst and Data Architect, with each case missing 5 and 4 months, respectively.

To avoid cases where missing months might disproportionately affect rankings, we excluded missing similarity scores when computing average scores. Additionally, because job postings follow industry needs rather than a structured calendar, we assess whether occupations released less often have their knowledge needs met.

For Knowledge, we observed that the occupations with the best overall matches were those directly related to software (software engineer, software architect, software developer, etc.) or those with "engineering" in their title. These two groups consistently ranked in the top 50% of all occupations, while occupations further from classical engineering had the lowest similarity scores. Marketing manager, product designer, and database administrator, for example, are occupations adjacent to classic engineering jobs but still require related knowledge.

On Skills, we observed higher overall matches, with scores around 55% and a higher bottom of 10%. Aside from a few isolated cases, almost all occupations had their base Skill needs covered to an average of around 40%. Fig. 2 makes it clear that Skills are more commonly matched between industry needs and the educational program being compared. Aside from a few cases, the general tendency seen with Knowledge matches is maintained: engineering occupations were more commonly seen toward the top occupations, while more adjacent cases were seen toward the bottom.

Finally, for Abilities (Fig. 3), we observed a higher ceiling than for the other two components of the KSA framework, with some cases achieving a 60% similarity rate, and a similar floor to the Knowledge case, with some cases hovering around the 5% mark. Overall, Abilities fell between the previous two components, with better similarity than the Knowledge case but lower than Skills. In this case, there appears to be no difference made towards specific occupations that mention "engineering" as part of their description, unlike the previous two cases.

Overall matching, considering all 3 aspects of the KSA framework, showed similarity scores averaging 30% to 35%, mostly due to the influence of the Knowledge aspect. This is partly because, in the blended case, we prioritized Knowledge due to its importance.

Discussion

As mentioned before, the very nature of the Skill Gap makes it a complex and persistent problem. This is reflected in our matching algorithm results, as the highest match we observed was around 62% for the Ability comparison. The score alone might raise concerns; the issue is exacerbated by the fact that this specific program was scheduled for a 2026 implementation date. It is important to note that the program was accredited by CACEI (Council of Accreditation of Engineering Teaching for its initials in Spanish), ABET (formerly Accreditation Board for Engineering and Technology, now working under the acronym exclusively), FIMPES (Federation of Mexican Private Higher Education Institutions), along with experienced teachers and members of the industry, and the national educational organism (SEP). This speaks to the complexity of building an educational program that meets requirements from different government agencies and accreditation institutions, provides the foundational knowledge needed to train future professionals to understand the underlying principles of their work rather than being purely technically trained, and delivers the state-of-the-art skills industries require.

Another possible explanation is the fast-changing nature of the INFOCOMM industry. New programming languages, hardware advances, the rapid evolution of AI tools, and global industry interest make it highly volatile. The speed at which this industry changes has far outpaced the traditional structure of a university degree, further widening the Skill Gap between newly graduated workers and industry needs. While the case of INFOCOMM-related programs is clearer and the surrounding circumstances make it more pronounced, this phenomenon has been happening across almost all University programs, especially engineering-based ones [19].

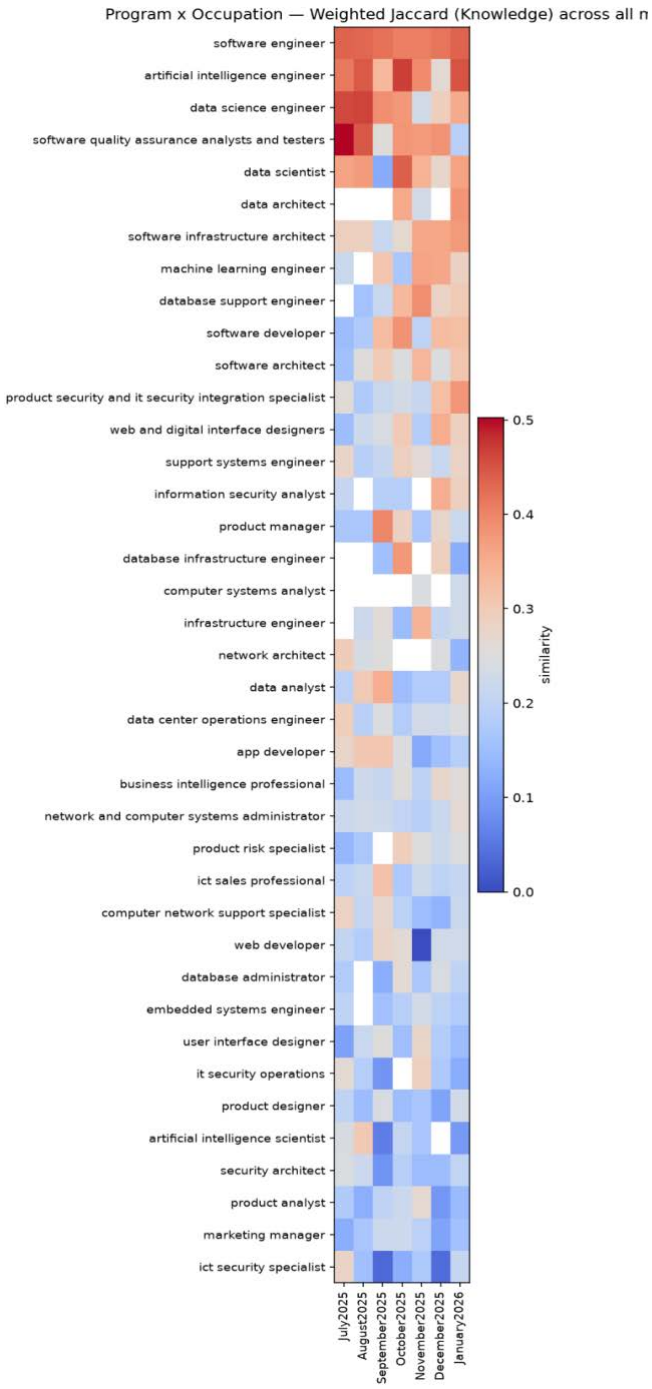


Fig. 1. Weighted Jaccard similarity scores between Job Postings and the ITC program per month for Knowledge.

Similarly, but in the opposite direction, the skill gap (specifically in the INFOCOMM sector) might be reversed: academia is ahead of industry (at least in the INFOCOMM sector). This is the case specifically for legacy systems, systems that have aged past their time and are now considered outdated, but still perform mission-critical processes: "*large software systems that we don't know how to cope with but that are vital to our organization*" [20]. In this case, job postings show skills that are "outdated" relative to the industry's current state. Examples include the need for platform-specific software like .NET, hardware-specific needs like a PLC running exclusively on Windows XP or requiring Visual Basic, or network architectures that don't support modern industry practices. This generates an interesting dilemma both for universities and the industry: the industry benefits from following the state-of-the-art and best practices, as it assures security, reliability, and compatibility with most modern needs, but a large percentage of their systems are outdated and would require extensive work and investment to modernize, leading to legacy systems that require maintenance and personnel. Meanwhile, universities aspire to prepare students for their future careers and often look to current best practices and developments for guidance, leaving legacy systems behind, as their name implies.

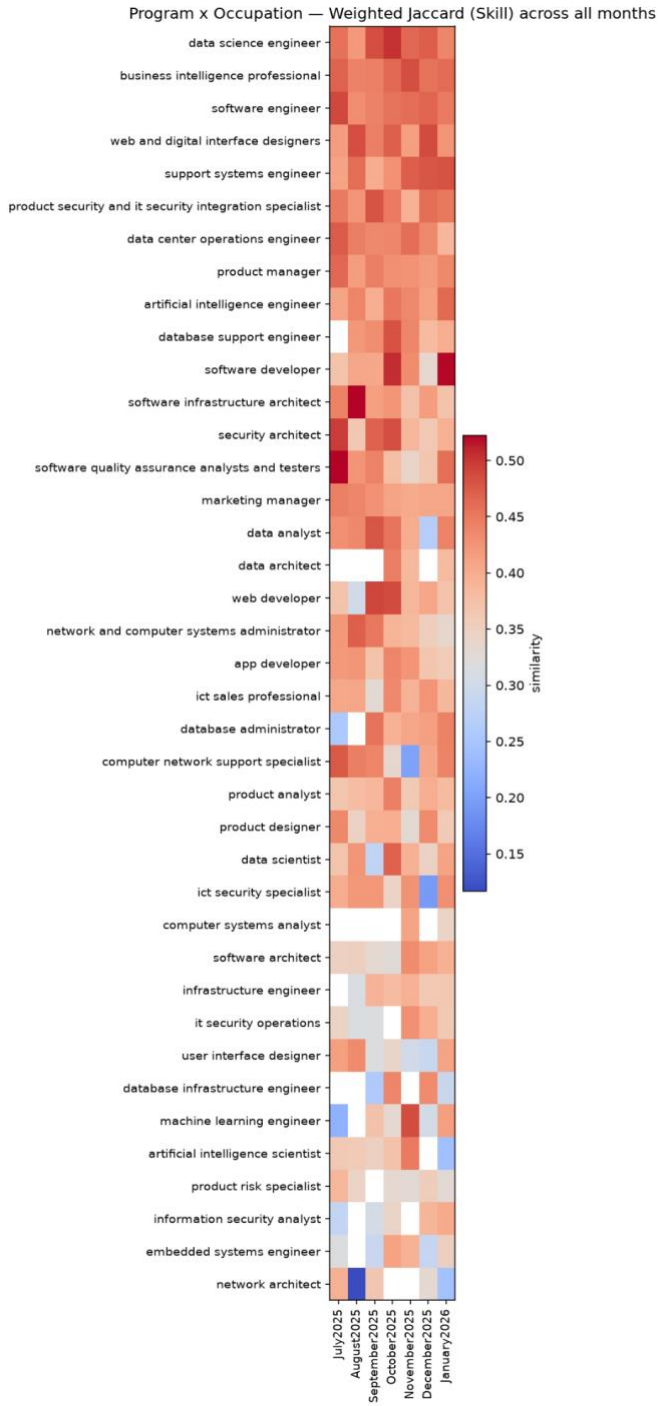


Fig. 2. Weighted Jaccard similarity scores between Job Postings and the ITC program per month for Skills.

Skills and Abilities represent a different situation altogether, as they reflect broader aspects of professional life that do not change as easily. Teamwork, Time management, attention to detail, and many other Skills and Abilities are, to some degree, evergreen and, as such, not a problem for the Skill Gap. Low skill-matching scores can occur because skills

are not always stated explicitly in a course's description or objectives. This omission does not mean the skill is not developed; it is developed indirectly, as there is no such thing as a “teamwork learning course” in most University courses.

It is also important to consider the nature of each component of the KSA framework and industry needs: while Knowledge is desired, to what degree is it necessary? In some cases, Knowledge is required only to the level of a technician, and not an engineer. Similarly, Skills might be more favored, as they represent the specific capabilities needed. Finally, Abilities are broader capabilities that are generally desired and commonly developed in many educational programs. Our similarity scores show a clear pattern: Skills are most similar between educational programs and job postings, while Knowledge is the least similar.

One final aspect worth discussing is the presence of occupations with very diverse similarity scores across months. For example, the Data Scientist occupation showed similarity scores ranging from ~40% to ~10% in consecutive months for Knowledge. We believe this month-to-month variability reflects the industry's diverse needs, even within the same occupation, and the industry's constantly changing nature.

Regarding RQ1: We found that, even in the best-case scenario, no cases existed where the program covered every single industry requirement. This could be due to the specialization of different businesses versus the more general knowledge base most universities develop in their students, the fast-changing nature of the industry (where the state-of-the-art shifts month to month), or the considerable need for legacy system support in the industry.

In terms of RQ2: The reality of the Skill Gap between recent graduates and Industry needs is clear, and it is unlikely to be solved by following the traditional structure of an educational program. Instead, it may be worth moving toward a specialization period in the later stages of degrees: cover the basics of the Skills required for employment in the initial and middle stages, then prepare students through elective courses, preferably developed in collaboration with industry. In this regard, it might benefit students and their future employers if these courses also functioned as micro-credentials, so they can demonstrate their experience and Skills.

Regarding limitations: the job postings were obtained from websites mainly serving the North American and Mexican markets and, as such, are tailored to those regions. Other markets may prioritize different Skills, so the results might differ. The educational program, while accredited by several organizations, does not represent all Information Technology majors. However, it is the program offered by one of the top universities in its region.

Conclusion

The results of this study suggest that curriculum–industry alignment should not be understood as a binary condition in which an educational program either does or does not satisfy industry needs. Instead, alignment varies substantially across occupations and across the Knowledge, Skills, and Abilities that constitute professional competence.

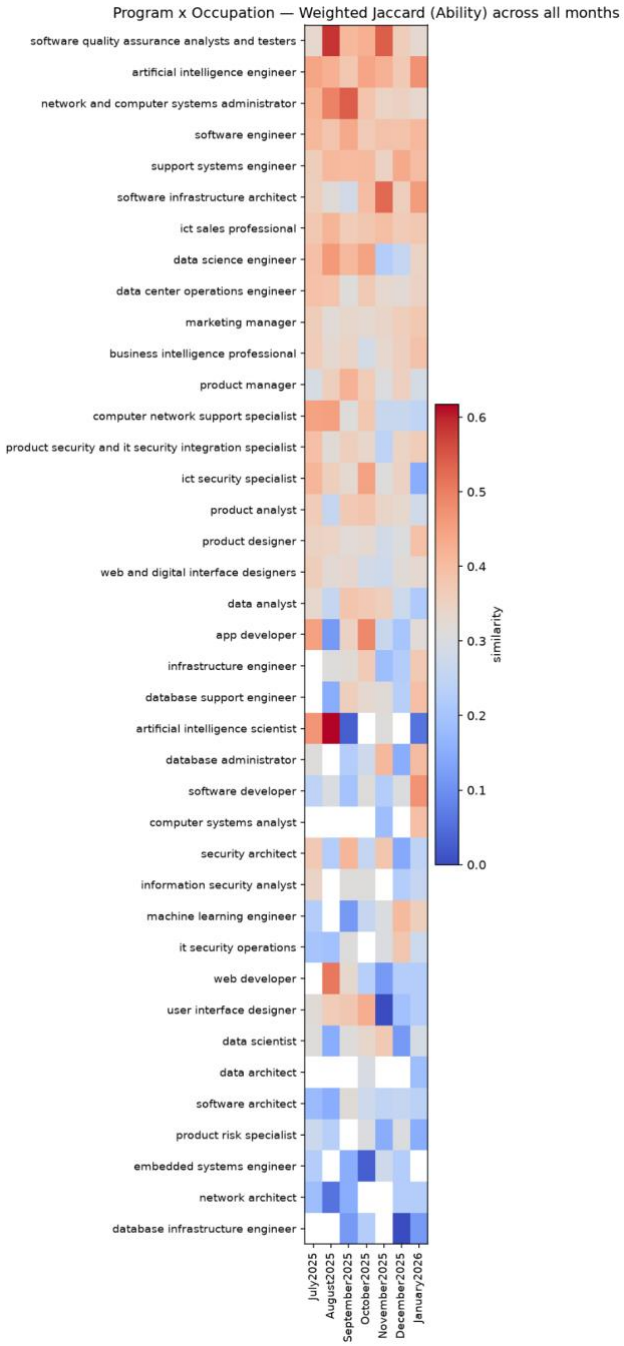


Fig. 3. Weighted Jaccard similarity scores between Job Postings and the ITC program per month for Abilities.

This distinction matters in rapidly evolving sectors such as INFOCOMM, where universities must balance foundational education with increasingly specialized and dynamic occupational requirements. From an educational perspective, the objective should therefore not necessarily be to maximize curriculum similarity with current job postings. Incorporating every industry requirement into a degree program would be neither feasible nor desirable, particularly when some requirements are highly specialized, rapidly changing, or tied to legacy technologies; systematic evidence of alignment and misalignment can help institutions distinguish competencies that belong in the core curriculum from those better addressed through electives, industry collaborations, professionalizing courses, certifications,

micro-credentials, or workplace learning. Beyond the specific Information Technology program examined here, the methodology provides a mechanism to periodically assess how educational offerings align with evolving labor-market demand. Applying the approach longitudinally and across programs, institutions, industries, and geographic contexts could support more evidence-informed curriculum review and reveal whether gaps are persistent, emerging, or temporary. Future work could also incorporate educational evidence beyond formal course descriptions, including learning activities and assessment outcomes, to better capture Skills and Abilities that may develop implicitly. Ultimately, the value of curriculum–industry alignment analysis lies not in eliminating the skill gap, but in making it visible and actionable. Combining labor market intelligence with curriculum evidence can provide universities with a more dynamic basis for deciding what should remain foundational, what should evolve, and where complementary learning pathways are needed to prepare graduates for changing professional environments.

Acknowledgment

This research was funded by the Challenge-Based Research Funding Program, Grant No. FP_IFE_EDU_S_2, of the Institute for the Future of Education, Tecnologico de Monterrey, Mexico.

References

[1] UNESCO, *Transforming higher education: global collaboration on visioning and action*. UNESCO, 2026. doi: 10.54675/SNJW1822.

[2] M. Schmidt, J. K. McDonald, and S. Moore, “The research we don’t need will persist until we dismantle the systems that sustain it,” *J Comput High Educ*, vol. 37, no. 2, pp. 507–542, Jun. 2025, doi: 10.1007/s12528-025-09446-4.

[3] D. Horváth and R. Zs. Szabó, “Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?,” *Technological Forecasting and Social Change*, vol. 146, pp. 119–132, Sep. 2019, doi: 10.1016/j.techfore.2019.05.021.

[4] Cedefop, *Terminology of European education and training policy*. Publications Office of the European Union, 2024. Accessed: May 27, 2026. [Online]. Available: https://data.europa.eu/doi/10.2801/991753

[5] P. O. Caratozzolo Martelliti, Y. V. Rueda Castro, M. Gutierrez Aguilar, J. D. Azofeifa Ugalde, L. J. González Gómez, "Discover tomorrow, shape today : shaping skills report", Reporte/report, ITESM, 2024. [Online]. Available: https://hdl.handle.net/11285/703615

[6] Azofeifa, Jose Daniel, Valentina Rueda-Castro, Luis Jose Gonzalez-Gomez, Sonia M. Gómez-Puente, Julieta Noguez, and Patricia Caratozzolo. "Unlocking the future of infocomm workforce: a visual KSA matrix taxonomy approach to education and occupational profiles." In *2024 IEEE Global Engineering Education Conference (EDUCON)*, pp. 01-09. IEEE, 2024. https://doi.org/10.1109/EDUCON60312.2024.10578632

[7] H. G. Ceballos Cancino, S. Butt, G. d. l. Ríos Alatorre, D. P. Madera Espíndola, "Skills report on automotive industry in Mexico : shaping skills framework", Reporte/report, ITESM, 2025. [Online]. Available: https://hdl.handle.net/11285/703348

[8] S. Butt, H. G. Ceballos Cancino, M. Fung, "Shaping skills for the Mexican workforce : Data-based insights into the Mexican automotive industry.", Reporte/report, ITESM, 2025. [Online]. Available: https://hdl.handle.net/11285/704782

[9] European Commission, “European Skills, Competences, Qualifications and Occupations (ESCO),” 2025. Accessed: May 27, 2026. [Online]. Available: https://esco.ec.europa.eu/en/classification/skill_main

[10] INEGI (Mexico), “National Occupational Classification System, SINCO,” 2020. Accessed: May 27, 2026. [Online]. Available: https://en.www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825198411

[11] L. J. Gonzalez-Gomez et al., “Dynamic taxonomy generation for future skills identification using a named entity recognition and relation extraction pipeline,” Frontiers in Artificial Intelligence, vol. 8, July 2025, doi: 10.3389/frai.2025.1579998.

[12] Butt, S., Alatorre, G. D. L. R., González Gómez, L. J., & Ceballos, H. G. (2025). Semi-supervised taxonomy expansion and completion in dynamic taxonomies. Applied Sciences, 15(12), 6517.

[13] M. Klesel and H. F. Wittmann, “Retrieval-Augmented Generation (RAG),” *Bus Inf Syst Eng*, vol. 67, no. 4, pp. 551–561, Aug. 2025, doi: 10.1007/s12599-025-00945-3.

[14] C. S. G. Khoo and J.-C. Na, “Semantic relations in information science,” in *Annual Review of Information Science and Technology*, Information Today, Inc., 2005, pp. 157–228.

[15] P. Langlais and T. L. Gao, “RaTE: a Reproducible automatic Taxonomy Evaluation by Filling the Gap,” *Proceedings of the 15th International Conference on Computational Semantics*, pp. 173–182, 2023.

[16] S. Meisenbacher and P. Norlander, “Building a Custom Taxonomy of AI Skills and Tasks from the Ground Up with Job Postings,” *Proceedings of the Second Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)*, pp. 117–130, 2026, doi: 10.18653/v1/2026.customnlp4u-1.11.

[17] J. Noble, “What is Jaccard Similarity? | IBM.” Accessed: Aug. 19, 2026. [Online]. Available: https://www.ibm.com/think/topics/jaccard-similarity

[18] V. Garousi, G. Giray, E. Tuzun, C. Catal, and M. Felderer, “Closing the Gap Between Software Engineering Education and Industrial Needs,” *IEEE Software*, vol. 37, no. 2, pp. 68–77, Mar. 2020, doi: 10.1109/MS.2018.2880823.

[19] P. Mhaske, B. Bhattacharjee, N. Haldar, P. Upadhyay, and A. Mandal, “Bridging digital skill gaps in the global workforce: A synthesis and conceptual framework building,” *Research in Globalization*, vol. 11, p. 100311, Dec. 2025, doi: 10.1016/j.resglo.2025.100311.

[20] S. M. Hussain, S. N. Bhatti, and M. F. Ur Rasool, “Legacy system and ways of its evolution,” in *2017 International Conference on Communication Technologies (ComTech)*, Apr. 2017, pp. 56–59. doi: 10.1109/COMTECH.2017.8065750.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

The GenAI Roundabout: From Enforcing Control to Building Capability

John Palfreyman[1], Chrissi Nerantzi[2], John Hammersley[3], Gerasimos Chatzidamianos[4], Vasiliki Kioupi[5]

(email: j.palfreyman@leeds.ac.uk), (email: c.nerantzi@leeds.ac.uk), (email: j.hammersley1@leeds.ac.uk), (email: GChatzidamianos@acg.edu), (email: v.kioupi@leeds.ac.uk)

[1] Leeds University Business School, University of Leeds, Leeds, United Kingdom

[2] School of Education, University of Leeds, Leeds, United Kingdom

[3] School of Design, University of Leeds, Leeds, United Kingdom

[4] Department of Psychology, American College of Greece, Athens, Greece

[5] School of Earth, Environment and Sustainability University of Leeds, Leeds, United Kingdom

Article DOI: 10.68286/samyrad2026.357

ABSTRACT

Institutional responses to generative AI (GenAI) in higher education have been largely rules-based and assessment focussed. Whilst clear, these are poorly suited to a fast-changing, complex challenge. The Roundabout [1] is a framework that shifts from enforcing compliance to building capability. Trust sits at its centre, understood as psychological safety, with four principles built on it: transparency, criticality, creativity and responsibility. The wider Roundabout project includes co-creation with fellow educators; ethics approval for that fieldwork is just confirmed. As pre-work we report a collaborative autoethnography among the five co-authors, who are also the Roundabout's developers, each writing a first-person account of one moment from their teaching where the shift from enforcing control to building capability was tested. Turning that lens on ourselves first fits the trust and transparency the Roundabout asks of everyone else, and lets us refine the framework before testing it more widely. Trust and Responsibility recurred across most accounts, while Criticality appeared only where someone went looking for it, and one account questioned whether the framework accounts for confidence in oneself, not just trust in others. We close by setting out the next stage: testing the Roundabout with the people it's for.

KEYWORDS generative AI, Roundabout framework, collaborative autoethnography, curriculum re-imagining, assessment

How to cite: Palfreyman, J., Nerantzi, C., Hammersley, J., Chatzidamianos, G., & Kioupi, V. (2026). The GenAI roundabout: From enforcing control to building capability. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.357

Introduction

Most universities have responded to generative AI (GenAI) by setting rules rather than by building capability. Traffic-light policies (red, amber, green) became common, striving for clarity at a time of considerable uncertainty [1], [2], [3]. But GenAI is now part of everyday teaching and learning, and these rules show their limits. Rules tell students what they can do in an assignment. They do not help students or staff learn how to use GenAI responsibly as part of professional and learning practice [1], [4].

The Roundabout [1] is a framework built on principles, not rules, that shifts the focus from enforcing compliance to building capability. The Roundabout project forms part of the activities of the Imaginative Curriculum Centre in the School of Education at the University of Leeds.

We plan wider co-creation with fellow educators; ethics approval for this was confirmed whilst we were finalising this paper. In the spirit of trust and transparency central to the Roundabout we turned the lens on ourselves.

This paper reports the results of a collaborative autoethnography (CAE) among the five authors, built around one question:

What does our own experience of building and using the Roundabout principles tell us about moving from control to capability in GenAI-enabled teaching?

The Roundabout moves in the same direction as two established frameworks. UNESCO's AI Competency Framework for Teachers [16] sets out five competency areas, from ethics to pedagogy, as a global standard for training and progression. The Open University's Critical AI Literacy Framework [17] takes a principles-based approach centred on equality, diversity and epistemic justice. Both frameworks share the shift from compliance to capability that runs through the Roundabout. Neither, though, puts a single relational principle at its centre. We place Trust, understood as psychological safety, at the Roundabout's centre, with transparency, criticality, creativity and responsibility built on it.

The Roundabout Framework

The Roundabout [1] builds on Aaron Dignan's [5] assertion that organisations are either legacy or evolutionary. Legacy organisations run on strict rules, enforced hierarchy and total compliance. In contrast, evolutionary organisations are guided by principles, light-touch constraints, and encourage constant experimentation. Traffic-light GenAI policies are a legacy organisation-style response. They work well when a problem is complicated: many parts, but still predictable. GenAI in higher education is complex, not complicated: outcomes emerge from evolving, people-centric interactions [1], [2], and can't be reliably predicted.

Trust sits at the centre of the Roundabout framework. Trust here embodies psychological safety [6] – people feel able to try new things, get them wrong, learn and improve within an encouraging, supportive environment. Around that centre sit four principles, in Table I illustrated through the voice of the students and staff who use them. This is a working model, not a fixed hierarchy, and open to revision as the framework is tested further. Rules do not disappear. They still matter for risk management, assessment integrity, and other high-stakes situations where clear boundaries are needed. Principles complement rather than replace them, supporting the judgement calls no rulebook can fully anticipate.

Table 1. The Roundabout Principles

Trust (foundation)	'I feel safe to question and try new things, even getting it wrong, because I know my tutor will guide me and not judge me.'
Transparency	'I tell my students exactly how and why I've used GenAI in this task, so they can learn from me.'
Criticality	'I check what GenAI gives me against what I know and against trusted sources, rather than taking it as the finished answer.'
Creativity	'I add GenAI's ideas into the class brainstorm alongside our own, then let students later decide what's worth keeping.'
Responsibility	'It's my choice how I use GenAI here, and I can explain and stand behind that choice if asked.'

Note: Statements in Table 1 are illustrative, composed to reflect common experience of each principle; they are not verbatim participant data.

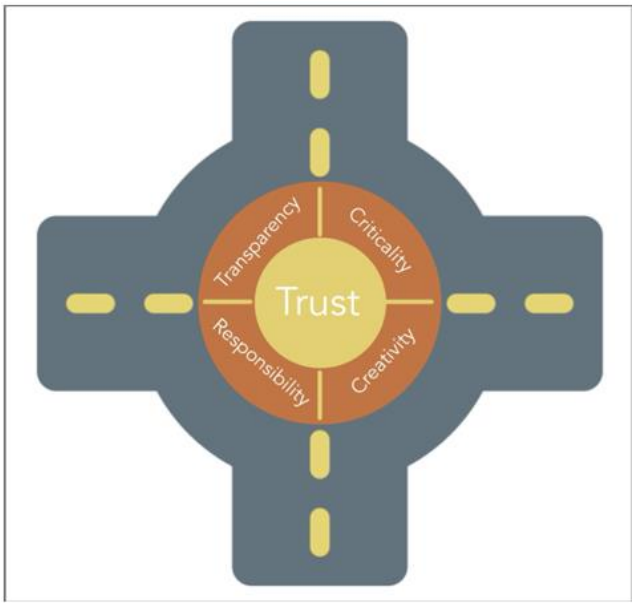


Fig. 1. The Roundabout framework. Based on [1].

Methodology: Collaborative Autoethnography

Collaborative autoethnography (CAE) asks two or more researchers to each write about their own personal experience of the same question, then read and analyse each other's accounts, looking for what matches and what does not [7], [8]. The five co-authors are also the Roundabout's developers, its participants, and its analysts in this study. This makes it a reflexive exploration and refinement of the framework, not an independent test of it, and we recognise the risk of bias that comes with developing and assessing our own framework. We chose CAE for two reasons. First, it follows BERA's ethical guidelines [13]: the authors provide the material themselves, not students or colleagues, so no separate ethics board review was required. References to students, colleagues, and institutions in the narratives are kept general, with no names or identifying details included. Second, CAE asks researchers to be open and to hold each

other to account [8], exactly what the Roundabout asks of students and educators.

Table 2 summarises the five co-authors' backgrounds: different disciplines, different lengths of teaching experience, different levels of experience with GenAI.

Table 2. Co-Author Backgrounds

Author	Years in HE	Years using GenAI	Teaching focus
John Palfreyman	9	3+	Strategy and Digital Transformation, including responsible GenAI usage.
Chrissi Nerantzi	20	3+	Creativity and play for learning, Digital education, GenAI in education, Research methods
John Hammersley	20	1+	Design Studies, Art and Design Studio
Gerasimos Chatzidamianos	20	1+	Research methods, statistics for psychology, psychopharmacology and biopsychology.
Vasiliki Kioupi	3	3	Sustainable Development

Note: Figures are self-reported by each author.

Each of the five authors wrote a short, first-person account of one specific moment from their own teaching, addressing the question above. We deliberately asked for real doubt or disagreement, not a tidy success story. CAE teams can settle too quickly on one shared, comfortable version of events; asking for doubt guards against that [8]. We then read each other's accounts and challenged them as critical peers, before analysing them together for what matched and what did not. We analysed the accounts using Framework Analysis [14], a deductive approach that codes data against categories set in advance rather than letting themes emerge inductively.

Claude produced an initial Framework Analysis matrix, coding all five narratives against the five Roundabout principles. John P then coded the same five narratives independently, by hand, highlighting each principle directly in the text. We reconciled the two matrices cell by cell: where they agreed, where the hand-coding caught something Claude's had missed, and where Claude's flagged something the hand-coding had not.

John P reviewed the reconciled matrix alone before it went to the remaining co-authors as a single combined round, asking each to check their own row and comment on the whole. Where two of us read the data differently, we report that as a finding rather than resolving it by agreement. All synthesis and interpretation in Sections V and VI remained with the human authors.

Section IV sets out the stories; Section V reports what we found.

Our Narratives

Chrissi

Without trust we are nothing.

I don't like policing my students. Yet, shortly after GenAI entered our practice, I found myself increasingly playing the gatekeeper. Until 2025/26, the institution used a traffic-light system for GenAI use in assessment, and most of my modules landed on amber. Amber generated constant questions: 'Can I do this?' 'Is this allowed?' Each time, I felt pushed towards a role I didn't want.

As an experimenter who operates in the 'uncomfort zone' [10], I question whether hard rules help in contexts this uncertain, or whether assessment should be our exclusive focus. I am in favour of exploring possibilities, trusting the process, and learning along the way. I do this with transparency and invite my students to do the same.

One example: redesigning a Problem-Based Learning module using GenAI [11], [12]. We worked with students to rethink scenarios and learning activities with Secure Copilot and explored possibilities for a GenAI-enabled curriculum that could maximise engagement and learning. The conversations required trust, transparency and a willingness to work with uncertainty.

The traffic-light system focused on assessment. My bigger question is how GenAI can deepen and diversify learning, which doesn't happen just by reading about it. Getting our hands dirty, making mistakes, reflecting and learning from them, matters more. What I've come to value is helping students develop the capability, confidence and judgement to navigate uncertainty for themselves.

Vasiliki

During one session of the Introduction to Creating Sustainable Futures module, I asked students to use GenAI to propose solutions to a sustainability business challenge involving a mining company seeking to minimise its impact on a local community. Students were then asked to critically evaluate the AI-generated responses. One group's answer confidently recommended a range of 'green' interventions that initially appeared convincing. However, as the discussion unfolded, students identified examples of greenwashing, oversimplified assumptions and tokenistic solutions. Rather than addressing the underlying causes of poor health and water contamination, the response focused on treating symptoms while overlooking the lived experiences, priorities and voices of the affected community. What began as an exercise in using GenAI quickly became an exercise in questioning it.

This session, together with my involvement in developing the Roundabout principles, particularly Criticality and Creativity, prompted me to reconsider what it means to move from control to capability in GenAI-enabled teaching.

Sustainability students often bring strong ethical commitments to their studies. While some value GenAI for the efficiency and support it offers, others remain hesitant because of concerns about its environmental footprint, labour practices and wider social implications. Watching students dismantle an apparently persuasive AI response made me realise that capability may not simply be about learning to use GenAI effectively. It may also involve developing the confidence to interrogate its outputs, recognise its limitations and make informed decisions about when its use is appropriate.

The discussion also revealed the tensions surrounding Creativity. After critiquing the initial response, students prompted GenAI to generate alternative approaches from different stakeholder perspectives. Although this broadened the conversation, several commented that the suggestions remained familiar and generic, reproducing dominant sustainability narratives rather than offering genuinely new ways of thinking.

This single teaching episode has shaped my contribution to the development of the Roundabout principles. Rather than seeing the move from control to capability as a straightforward progression towards greater use of GenAI, I now see it as creating space for informed judgement. For me, the unresolved question is whether capability lies in using GenAI effectively, in knowing when not to use it, or in navigating thoughtfully between those positions.

John H

A recent talk at an all-school meeting on AI literacy for teaching and research has made me pause and rethink how AI might be disrupting the trust relationships central to teaching. This feels especially keen as I am returning to teaching after a long period of illness and preparing to develop the inclusion of AI for the core year one programme module on a creative visual arts course. This period of illness brings with it a sense of being out of the loop more than perhaps usual.

What struck me first about the presentation on AI literacy is a sense that I felt I was being encouraged to just skip ethical concerns and move on as I/we/students/UK HE are falling behind, and this is likely to only get worse as the pace of technological change increases. My reaction since has been to sign up for more talks and start to experiment more to consider how AI can be integrated into the creative processes of the module.

Perhaps the presentation was designed to create this sense of anxiety and responsibility, but it has also made me ask to what extent I feel able to respond to the challenge of AI. In turn this has made me reconsider the issue of self-trust, as messages of falling or being left behind erode my sense of self-confidence and self-trust. This is an issue for me, especially as building relationships of trust with new students can have a significant impact on cohort cohesion and student motivation and engagement.

So I now wonder if the Roundabout over-emphasises a sense of trust in others, the other of AI, the othering relationship of learner-educator, and whether it takes significant enough account of the impact of AI on self-trust and confidence of all involved in learning and teaching.

Gerasimos

I moved to Athens, Greece, five years ago. London, where I only stayed for a year, is a city I know, while Athens is not. What makes London easier to navigate and Athens a labyrinth?

In London, I immersed myself in the unknown; I got lost and got found. In Athens, I skipped all of this and simply put my trust in Google Maps. I have never been late for an appointment or got lost, but my knowledge of the city itself is limited.

Applying this to teaching and learning in HE: how do we navigate knowledge acquisition in an AI era?

The University I work for instructs that unauthorised use of AI is prohibited. No traffic-light systems, no roundabouts, just orders expecting compliance. During a graduate research methods class, an elaborate-worded prompt appeared on one of my presentation slides, which students had to use in any GenAI platform. The aim was to critically review possible responses. No one responded. None. Awestruck by the lack of engagement, I explored the issue further. Most claimed they do not, or have never, used AI. The most confident ones reminded me of the policy. After I confirmed I had authorised its use, they admitted that most, if not all, had an upgraded paid subscription. 'I trust Google Maps more than my students trust me,' I thought. What has happened to our relationship? How can learning happen without trust among those involved?

Technology evolves faster than university policy. Policies that force students into compliance blunt the exact skills education aims to develop. We punish students for trusting AI more than they trust us for knowledge generation.

Trusting the technology to navigate Athens was efficient. If I were tested on the city itself, I would fail. True knowledge and efficiency together are only possible by learning Athens with AI, not instead of it. Isn't it time for meaningful AI policies in universities, rooted in shared principles such as trust, creativity, criticality and shared responsibility?

John P

At the start of the semester, we needed to assign a GenAI category to our team-taught undergraduate module with around 340 students spread across several tutor-led parallel workshops. In line with the University's ambition to move more modules towards Green, I argued that this module should be Green too. I was unable to convince my colleagues, so we settled on Amber.

At one level, that reflects understandable caution. A large, shared module carries more perceived risk than one taught by a single academic. But I don't think it was only about risk. Moving to Green would also have been counter cultural. It required a level of confidence in both the students and the institution that not everyone yet shared. I use Green across the modules that I lead myself, but I couldn't persuade colleagues to take the same step where the decision had to be made collectively.

Reflecting on this, I realised some of the gap was mine. If trusting students more had gone badly in one of my own

modules the consequences for me would have been manageable. For colleagues earlier in their career, teaching a 'flagship module' with visibility at the highest levels of the University, the professional risks are different. And I don't think I gave that enough weight when making my case.

That experience reinforced something important for me. I still believe trust (as a short form for Psychological Safety) is the right foundation. The challenge is not whether trust matters, but whether the conditions exist for everyone to act on it in a climate where they feel everyone 'has their backs'. Until they do, moving from control to capability will remain a challenge.

Analysis

Table 3 summarises how the five author accounts map onto the Roundabout's five principles, using Framework Analysis. This fixes the principles as categories in advance rather than letting themes emerge from the data.

Table 3. Framework Analysis Matrix

Author	Trust	Transparency	Criticality	Creativity	Responsibility
Chrissi	Trusting an experimental process with students	Openness with students, stated as practice	—	Co-designing a GenAI-enabled curriculum	Building students' capability and judgement
Vasiliki	—	—	Students identifying greenwashing in a GenAI response	GenAI's suggestions seen as generic, not novel	Informed judgement on when GenAI use fits
John H	Self-trust rather than trust in others	—	—	—	Resisting pressure to skip ethical concerns
Gerasimos	Trust between tutor and student	A prompt shared openly with the class	Students asked to critically review AI-generated responses	—	Guarding education against AI risk
John P	Confidence in students and the institution	—	—	—	Self-accountability for underweighting a colleague's career risk

Note: Cells condense narrative evidence from the project's Framework Analysis matrix; an absent entry reflects the coding definition or narrative scope, not importance (see below).

Trust appears in four accounts, taking a different form in each: confidence in students and the institution (John P), trust in an experimental process (Chrissi), trust between tutor and student (Gerasimos), and self-trust as well as trust in others (John H).

Responsibility appears in all five accounts, in two forms: guiding judgement, whether students' (Chrissi, John P) or their own (Vasiliki), and guarding education itself against risk (Gerasimos, John H).

Transparency surfaces twice, in Chrissi's and Gerasimos's accounts, but for different reasons. Chrissi names it as a stated practice, openness with her students, while Gerasimos's is incidental, a prompt shared openly without ever being called transparency.

Creativity also surfaces twice, in Chrissi's and Vasiliki's accounts, and in opposite directions: Chrissi's shows GenAI expanding students' creative thinking, Vasiliki's shows it narrowing it into generic, unoriginal suggestions.

Criticality appears twice. Vasiliki's account meets the definition directly, students interrogating a GenAI response for greenwashing, as does Gerasimos's, students asked to critically review AI-generated responses. Chrissi's questioning of her own practice strictly falls outside the Roundabout's definition, which is about questioning GenAI's output.

Two things are also worth mentioning. Each author was asked to focus on one or two principles, leading to a sparse matrix. So the number of times that a principle is mentioned cannot be correlated with its importance. Notably, three of the five accounts challenge how a principle is defined: John H on self-trust versus trust in others, Gerasimos on what university GenAI policy assumes about trust, and the coding process itself on where Criticality's boundary should sit.

Findings and Discussion

We set out to explore what our experiences of building and using the Roundabout principles tell us about moving from control to capability in GenAI-enabled teaching. Our reflections show that we don't see the move as one shift. It's more likely several uneven changes running at different speeds and driven by different people across different organisations. We see this as consistent with our collaborative approach to future Roundabout development.

Trust runs through every account, though it takes a different form each time. Chrissi says capability building only started once someone trusted the process. John P says trust in the institution has to come first, before staff can use GenAI responsibly. Gerasimos frames trust two ways in the same story: trusting Google Maps to navigate a new city, and the erosion of trust between him and students who hid their own AI use. John H raises a different form again: self-trust, his own confidence rather than his trust in anyone else. These are related ideas, not the same one, and the Roundabout's current definition, built around psychological safety between people, doesn't fully capture self-trust or trust in a technology.

Sahlberg [15] observes that trust works between two people, within a team, and across a whole institution, and

three of our accounts map onto that: Chrissi's and Gerasimos's interpersonal trust with students, and John P's institutional trust. John H's self-trust sits outside all three levels, which suggests Sahlberg's framing, like the Roundabout's, may need extending to cover trust in oneself rather than trust in others. Gerasimos's trust in Google Maps points to a further, technological form the framework doesn't yet name. Naming these distinctions matters more than deciding which is most important; we treat this as an open question for the wider study, not a gap to resolve here.

Our reflections on responsibility show the tension between capability growth and the need to control. Gerasimos and John H (without comparing notes) both land on the same dual role for educators: we take responsibility for guiding students to use GenAI responsibly, and at the same time we take responsibility for guarding our institutions' academic integrity. Chrissi calls the second of these 'policing'. This clearly weighs heavily on the minds of many academics and could explain a reluctance to move away from control-based approaches. It's always harder to build capability than to set rules.

As a team we examine criticality through different lenses. Vasiliki sees a gap: too little critical engagement with GenAI shows up across the matrix, which concerns her in a field that claims to value critical thinking. Chrissi takes the opposite view: criticality, interrogating outputs, questioning assumptions, spotting generalisations, was central to her experience, providing valuable learning. Same principle, same project, two different readings.

On the topic of creativity John P reflects on Chrissi's use of GenAI to help re-imagine her module delivery, effectively challenging and extending student thinking. He also acknowledged Vasiliki's view of the opposite effect where students found GenAI's generic response as narrowing the thought processes. The Roundabout principle would expand the opportunities GenAI creates for broadening thinking, while staying alert to where it constrains it.

Chrissi 'nails' transparency, stating that educators need to demonstrate this, not just expect it from students. Gerasimos showed transparency in action by sharing his GenAI prompts with the students, and the discussions that followed were insightful.

What does this tell us about moving from exercising control to building capability? It won't be one coherent shift. Trust enables the rest. Responsibility splits into guiding and guarding. Creativity cuts both ways, expanding thinking or flattening it into sameness. Transparency boils down to educators eating their own cooking. In our accounts, criticality appeared only when someone went looking for it, an early result the wider study will test and may revise. If there's guidance here for GenAI-enabled teaching and associated educational practices, it's that these move separately, at different speeds, not together.

Conclusion

This collaborative autoethnography set out to explore what our own experience of the Roundabout tells us about moving from control to capability in GenAI-enabled teaching. Five educators, five incidents, five different classrooms; what emerged was an early picture of how the Roundabout's principles interact in practice: unevenly, and sometimes in tension with each other.

That unevenness is the finding, not a limitation to explain away. Trust surfaced in most accounts; Responsibility surfaced in all of them; Criticality surfaced only where the task specifically called for it, not by default. Two of us split Responsibility along the same line without comparing notes. One of us found a question Trust (as the Roundabout defines it) was never built to answer.

This was always meant as a first step. The wider co-creation study, which the CAE prepares us for, is where these open questions belong: do students show Criticality the same way we did, only when someone asks for it? Does Responsibility still split into guiding and guarding when the focus extends to understanding students' expectations? And where does confidence in one's own ability, rather than trust in others, actually get built?

As a pilot, this study has clear limits: five self-selected accounts, each anchored to one incident, can't show how the principles play out at scale or for the students the Roundabout is ultimately for. For educators facing their own GenAI policy decisions, our suggestion isn't to adopt all five principles as a checklist, but to notice which ones are constraining their own practice and start there.

The CAE study gives us momentum and direction to build the Roundabout collaboratively and test it with the people it was built for.

Declarations

Conflicts of interest: The authors declare no conflicts of interest.

AI use disclosure: We acknowledge the use of Claude (Anthropic) in preparing this paper, including support with the Framework Analysis coding and with drafting. This was used openly, consistent with the transparency the Roundabout itself asks of anyone using GenAI. All generated outputs were critically reviewed and edited by the co-authors, who take full responsibility for what is included; any errors of interpretation remain our own.

Acknowledgment

We thank the students in our own classes whose questions and challenges shaped the incidents described in Section IV.

References

[1] C. Nerantzi and J. Palfreyman, "From traffic lights to roundabouts: Towards an AI-enabled curriculum built on trust," position paper, University of Leeds, 2025.
[2] University of Leeds, "Generative AI guidance for taught students," guidance v1.1, 2024.
[3] F. Miao and W. Holmes, Guidance for Generative AI in Education and Research. Paris: UNESCO, 2023.
[4] M. Bearman, J. Ryan, and R. Ajjawi, "Discourses of artificial intelligence in higher education: A critical literature review," Higher Education, vol. 86, no. 2, pp. 369–385, 2023, doi: 10.1007/s10734-022-00937-2.

[5] A. Dignan, Brave New Work: Are You Ready to Reinvent Your Organisation? London: Penguin, 2019.
[6] A. C. Edmondson, The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth. Hoboken, NJ: Wiley, 2019.
[7] H. Chang, F. W. Ngunjiri, and K.-A. C. Hernandez, Collaborative Autoethnography. New York: Routledge, 2013.
[8] K.-A. C. Hernandez, H. Chang, and F. W. Ngunjiri, "Collaborative autoethnography as multivocal, relational, and democratic research: Opportunities, challenges, and aspirations," a/b: Auto/Biography Studies, vol. 32, no. 2, pp. 251–254, 2017, doi: 10.1080/08989575.2017.1288892.
[9] C. Nerantzi, J. Palfreyman, and M. Pavlopoulou, "A reflective conversation on assessment in the age of generative AI using the traffic light system at the University of Leeds," draft article, 21 Oct. 2025.
[10] C. Nerantzi, "Living in the uncomfort zone or towards a creativity manifesto," in Encouraging Imagination and Creativity in Higher Education: Personal Manifestos, N. Jackson and J. Willis, Eds. Creative Academic, 2019.
[11] C. Nerantzi, F. Z. Chai, A. Poradowska, and M. Pavlopoulou, "Reimagining a postgraduate module: Injecting problem-based learning and normalising the use of GenAI for learning," in Proc. 13th Int. Conf. Open Distance Learning, Patra, Greece, 2025, pp. 103-120, doi: 10.12681/icodl.8935.
[12] C. Nerantzi, Z. Chai, A. Poradowska, and M. Pavlopoulou, "We are the future: Educators and students as co-creators in conversation," Irish J. Academic Practice, vol. 14, no. 1, Art. 2, 2026, doi: 10.21427/bmrg-6e24.
[13] British Educational Research Association (BERA), Ethical Guidelines for Educational Research, 4th ed. London: BERA, 2018.
[14] N. K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research," BMC Medical Research Methodology, vol. 13, no. 117, 2013, doi: 10.1186/1471-2288-13-117.
[15] P. Sahlberg, "Trust: the missing link in educational change," Journal of Educational Change, 2026, doi: 10.1007/s10833 026-09568-1.
[16] F. Miao and M. Cukurova, AI Competency Framework for Teachers. Paris: UNESCO, 2024.
[17] M. Hauck, E. Moore, and C. Wright, "A framework for the learning and teaching of critical AI literacy skills," version 0.1, The Open University, 2025.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

From Automation to Judgment: Evaluating an AI-Augmented Monte Carlo Instructional Architecture for Responsible Managerial Decision-Making

Roberto Gómez Tobías, Giovanni María Baldini

(email. roberto.gt@tec.mx) and (email. gbaldini@tec.mx)

International Business Department, Tecnológico de Monterrey Monterrey, Mexico

Article DOI: 10.68286/samyrad2026.363

ABSTRACT

Generative artificial intelligence can help business students produce analyses, but learning also requires them to examine uncertainty and justify decisions. This study evaluates whether adding structured risk modeling and reflection to a common AI-supported managerial task is associated with stronger responsible judgment. Responsible Augmented Business Judgment (RABJ) means using AI for analytical support while retaining accountability for assumptions, uncertainty, stakeholder consequences, and ethically defensible decisions. A mixed-methods quasi-experimental study involved 120 undergraduates in 24 teams across four sections. Both conditions shared the case, roles, faculty guidance, and executive recommendation. The full intervention additionally used Monte Carlo simulation and a matrix for reflecting on present and future rational and emotional considerations. Evidence included a 16-item pre-post survey, team level performance rubrics, expert content-validation evidence, and qualitative coding of anonymized reflection summaries. Baseline-adjusted ANCOVA favored the full intervention on the overall composite, B = 0.557, 95% CI [0.482, 0.632], p < .001; the strongest adjusted difference occurred in metacognition and emotional foresight, B = 0.793. Team performance showed the same directional pattern (3.47 vs. 2.93), with high interrater reliability, ICC(2,k) = .954. Findings suggest that structured uncertainty analysis and reflection can support accountable use of AI in business education.

KEYWORDS Artificial intelligence, Monte Carlo simulation, responsible managerial judgment, business education, instructional design, ethical decision-making, uncertainty

How to cite: Gómez Tobías, R., & Baldini, G. M. (2026). From automation to judgment: Evaluating an AI-augmented Monte Carlo instructional architecture for responsible managerial decision-making. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.363

Introduction

This study asks whether business students make more responsible decisions when AI-supported analysis includes explicit risk modeling and structured reflection. Generative artificial intelligence has rapidly entered higher education as a tool for explanation, writing support, feedback, ideation, and analysis. Large language models can expand access to information, increase task efficiency, and support

personalized learning processes, but their educational impact remains contingent on how they are integrated into instructional design rather than on the availability of the tool alone [1]. Systematic reviews of artificial intelligence in higher education have also noted that many applications remain technology driven, with insufficient attention to educators, pedagogy, assessment, and the learning processes through which students transform AI-supported activity into durable competence [2], [3]. Recent experimental evidence from secondary-school mathematics also shows that learning consequences depend on how AI support is designed, with pedagogical safeguards mitigating the disadvantages of unrestricted assistance [4].

This issue is particularly relevant in business education. Managerial decision-making requires more than rapid information production or fluent analytical narratives. Students must learn to reason under uncertainty, examine assumptions, interpret risk, evaluate stakeholder consequences, and defend decisions that remain responsible when outcomes are ambiguous. Recent work on generative AI and critical thinking suggests that AI can both support and obscure higher-order cognition, depending on whether students are required to evaluate, challenge, and justify the outputs they receive [5]. Evidence from business schools further suggests that unguided generative AI use may produce uneven effects across students and may alter the distribution of cognitive effort, especially when students use AI-generated outputs as substitutes for analytical monitoring rather than as objects for critical evaluation [6].

The present study is grounded in the view that AI should complement human judgment rather than function as an oracle or autonomous decision-maker. Human-AI collaboration research suggests that hybrid systems are most valuable when AI contributes computational reach and pattern generation while humans retain interpretive responsibility, contextual awareness, and accountability [7]. Human-factors research also cautions that automation can produce misuse, disuse, overreliance, and complacency when users fail to monitor outputs or accept recommendations without sufficient scrutiny [8], [9]. In business education, the instructional challenge is therefore not simply to permit AI use, but to preserve active monitoring and human responsibility within the decision process.

One way to address this challenge is to introduce structured cognitive friction into the human-AI decision loop. In this study, cognitive friction is not treated as an obstacle to learning, but as a deliberate pedagogical mechanism that slows down automatic acceptance, requires validation of assumptions, and redirects attention toward uncertainty, ethics, and consequences. Fan et al. reported stronger essay improvement with ChatGPT but no significant between-group differences in knowledge gain or transfer [10]. This distinction motivates routines that require students to inspect AI-supported reasoning. Monte Carlo simulation serves this purpose by transforming single-point estimates and deterministic recommendations into probabilistic risk maps. This is especially relevant for business education because stochastic simulation can help students understand variability, sensitivity, and risk exposure in decisions involving demand, cost, implementation, and stakeholder impact [11]. More broadly, decision research has long shown that judgments under risk are shaped by biases, reference points, and perceptions of loss and gain, making explicit probabilistic reasoning essential for responsible managerial choice [12].

The second mechanism is a temporal-emotional decision matrix developed for this study. The matrix requires students to examine four dimensions before issuing an executive recommendation: present emotional pressures, present rational arguments, future emotional consequences, and future rational consequences. This structure builds on experiential learning, where competence develops through action, reflection, conceptualization, and renewed experimentation [13]. It also aligns with learning-oriented assessment, where feedback, criteria, and performance tasks are designed not only to measure learning but to shape it [14]. From a metacognitive perspective, the matrix functions as a cognitive forcing tool because it requires students to monitor their own reasoning, identify emotional and cognitive biases, and regulate their decision process before acting [15]. In addition, its stakeholder orientation reflects the idea that business decisions should be evaluated not only by economic feasibility, but also by their consequences for actors affected by the organization [16].

To capture this integrated capability, the study proposes the construct of Responsible Augmented Business Judgment (RABJ). RABJ refers to the learner's capacity to use generative AI as an analytical support system while retaining human accountability for validating assumptions, modeling uncertainty, evaluating stakeholder consequences, regulating cognitive and emotional biases, and defending ethically responsible managerial decisions under ambiguity. This construct differs from AI literacy because it focuses not only on tool use, but on accountable managerial application. It differs from general critical thinking because it is situated in business decisions involving uncertainty, AI-generated advice, and stakeholder consequences. It also differs from business analytics because it treats quantitative analysis as one component of responsible judgment rather than as the endpoint of decision-making.

This study evaluates whether a structured instructional architecture that augments a common AI-supported managerial task with Monte Carlo risk modeling and temporal-emotional reflection is associated with stronger RABJ-related outcomes than AI-supported traditional strategic analysis. RQ1 asks whether the full architecture is associated with stronger gains in the overall RABJ-related composite than the active comparison condition. RQ2 asks which prespecified dimensions show the strongest baseline-adjusted differences between conditions. RQ3 asks whether team performance and qualitative evidence converge with the survey findings. The study makes three contributions: conceptually, it proposes RABJ as a provisional multidimensional applied capability; pedagogically, it operationalizes productive cognitive friction within the human AI decision loop; and empirically, it evaluates the added value of structured uncertainty modeling and reflective accountability against an active AI-supported comparison.

Background and related work

This study builds on the premise that the educational value of generative AI depends on how students are required to interact with it. In business education, AI can support rapid information synthesis, scenario generation, and analytical drafting, but these affordances may also reduce the need for

students to formulate assumptions, test alternatives, and monitor the quality of their own reasoning. For this reason, the proposed intervention does not treat AI access as the instructional innovation itself. Instead, AI is embedded within a structured decision architecture designed to preserve human accountability while expanding analytical capacity.

The first theoretical mechanism is human-AI complementarity. From this perspective, AI contributes computational fluency, rapid synthesis, and alternative generation, while the learner remains responsible for interpretation, contextual judgment, and final decision accountability [7]. This distinction is central to RABJ because the target outcome is not AI literacy alone, but the ability to use AI outputs as objects of evaluation. Students must determine whether AI-generated claims are plausible, incomplete, biased, or insufficiently sensitive to risk and stakeholder implications. This design directly addresses concerns about automation bias and over reliance, where users may accept automated recommendations without adequate monitoring or independent verification [8], [9].

The second mechanism is probabilistic grounding through Monte Carlo simulation. Managerial recommendations generated by AI often appear as coherent narratives, but coherence does not guarantee robustness under uncertainty. Monte Carlo simulation introduces stochastic variation into the decision process by requiring students to model input uncertainty, observe outcome distributions, and interpret risk exposure rather than relying on single-point estimates [11]. In the proposed architecture, Monte Carlo simulation functions as a corrective to deterministic reasoning. It transforms the decision task from “Which recommendation sounds best?” into “Which recommendation remains defensible across plausible scenarios?”

The third mechanism is temporal-emotional metacognition. The temporal-emotional decision matrix requires students to pause before issuing a final recommendation and examine their decision from four angles: present rational arguments, present emotional pressures, future rational consequences, and future emotional consequences. This structure translates metacognitive monitoring into an applied business decision routine, motivated by concerns about dependence on AI during self-regulated learning [10]. Students are required to identify assumptions, emotional triggers, stakeholder consequences, and possible future regret before finalizing the executive recommendation. In this sense, the matrix operates as a cognitive forcing function because it intentionally interrupts premature closure and redirects attention toward reflective judgment [15].

The fourth mechanism is faculty scaffolding and rubric-based accountability. The intervention was not designed as unsupervised AI use. Faculty guidance, structured prompts, performance criteria, and post-task reflection were used to calibrate standards of evidence and decision quality. This is consistent with learning-oriented assessment, where assessment criteria and feedback are part of the learning process rather than only a measurement device [14]. The rubric therefore evaluated not only whether a student reached a recommendation, but whether the decision incorporated uncertainty, AI critique, stakeholder reasoning, ethical defensibility, and future-oriented reflection.

Fig. 1 summarizes the instructional architecture. Role based managerial scenarios and a generative AI tool serve as learning inputs. Monte Carlo modeling, the temporal-emotional decision matrix, and faculty-mediated rubric design operate as instructional mechanisms. The intended outcome is Responsible Augmented Business Judgment, measured through survey outcomes, expert-validated instruments, rubric-based performance assessment, and qualitative coding of student reflections.

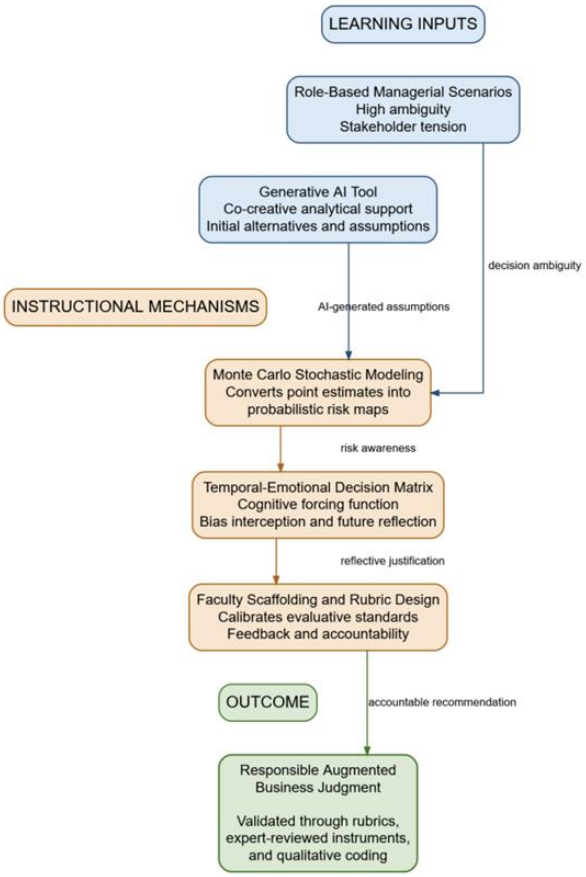


Fig.1 Pedagogical architecture for Responsible Augmented Business Judgment.

Pedagogical Architecture and Methods

Research Design and Participants

A mixed-methods quasi-experimental classroom study was conducted with 120 undergraduate students at a private university in Mexico. Students were nested within 24 teams of five, which were nested within four intact academic sections; teams were the units of condition allocation. Condition allocation was balanced within section blocks, with three teams in the active comparison and three in the full intervention in each section, yielding 12 teams and 60 students per condition. This blocked classroom structure preserved the instructional ecology while reducing section-level imbalance [17]. Both conditions used the same managerial case, role structure, faculty guidance, executive deliverable, and common performance criteria.

All 120 students completed the pre- and post-intervention survey measures. Team-level performance evidence consisted of 24 final recommendations, and anonymized individual reflections were retained for descriptive triangulation [18]. Participation in the research component was voluntary and based on written informed consent under institutional procedures for minimal-risk educational research. Students could decline or withdraw; research records were kept confidential, de-identified before analysis,

and reported only in aggregate. Separate formal committee approval was not required under the applicable institutional procedure.

The active comparison was deliberately substantive rather than passive: both conditions used generative AI and produced an executive recommendation. Across sections, the sequence followed the same broad order—pretest, common case framing and AI-supported diagnosis, condition-specific analysis, final recommendation, posttest, and individual reflection. The experimental contrast therefore concerned the structured uncertainty and reflection mechanisms added in the full intervention rather than simple exposure to AI.

Instructional Conditions

Both conditions worked with the same managerial case and role structure. In the active comparison, students used generative AI to develop a strategic and financial diagnosis of the case and formulate an executive recommendation. Their work focused on examining the plausibility of assumptions, considering stakeholder implications, and justifying the recommendation through conventional strategic and financial reasoning. This condition included faculty guidance, the common assessment criteria, and individual reflection, but did not require stochastic simulation or the structured temporal-emotional matrix. The full intervention retained these elements and added a structured Monte Carlo phase in which key uncertain assumptions were translated into stochastic inputs and interpreted through outcome distributions, downside exposure, sensitivity, and robustness rather than single-point estimates [11]. Students then completed the investigator-developed temporal-emotional decision matrix, which required them to reconsider the recommendation across present emotional pressures, present rational evidence, future emotional consequences, and future rational consequences. The matrix functioned as a forcing routine before the final recommendation, consistent with the intervention's metacognitive and stakeholder rationale [15], [16]. The comparison therefore evaluates Monte Carlo modeling and structured temporal-emotional reflection as a combined addition to the shared AI-supported task; the design does not separate their individual contributions or interaction.

Measures

The study used three outcome evidence sources and one measurement-quality source. First, an investigator-developed 16-item pre-post survey measured four prespecified domains: probabilistic reasoning and risk mitigation (5 items), ethical and stakeholder reasoning (4), critical AI use (3), and metacognition with emotional foresight (4).

Items used a five-point Likert scale. Domain scores were calculated as the mean of their constituent items, and the overall 16-item mean was treated as a provisional composite summary rather than evidence of a unidimensional latent construct. Second, a 10-criterion neutral performance rubric assessed the common team recommendation without requiring use of intervention-specific tools; criteria covered problem framing, evidence use, assumption transparency, uncertainty, alternative comparison, risk mitigation, stakeholder impact, ethical accountability, temporal consequences, and integrated recommendation quality.

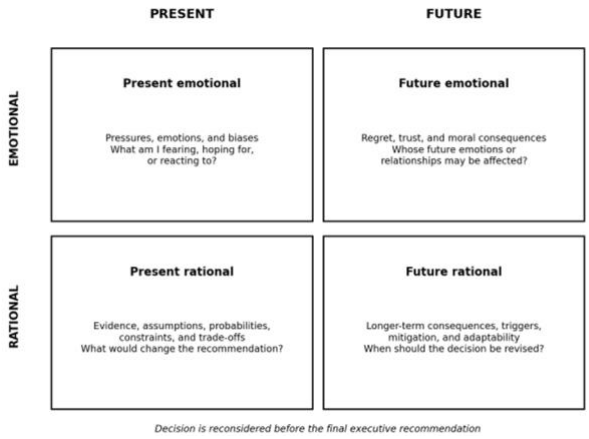


Fig. 2. Temporal-emotional decision matrix used in the full intervention. Students reconsidered the decision across present/future and emotional/rational perspectives before finalizing the executive recommendation.

Two evaluators scored each artifact, and the two-rater mean was used for performance comparison. Third, anonymized reflections were coded for uncertainty reasoning, critical AI use, stakeholder reasoning, ethical reasoning, emotional foresight, and metacognition; only prompts common to both conditions were eligible for between condition comparison. Separately, seven faculty judges reviewed the survey items and rubric criteria for relevance and clarity to provide content-validity evidence [19], [20].

Data Analysis

The analysis followed a triangulated mixed-methods strategy [18]. Descriptive statistics were computed for pre-, post-, and gain scores, and baseline equivalence was examined before intervention effects were tested. Baseline-adjusted ANCOVA was the primary survey analysis, with post-intervention scores as dependent variables, condition as the predictor, and the corresponding pre-intervention score as covariate. Survey models used individual students as the analysis units and did not adjust standard errors for dependence within teams or sections. Gain-score comparisons were retained as a complementary sensitivity description, and Hedges' g summarized between-condition mean gain differences standardized by the pooled within-condition standard deviation of individual gains, with a small-sample correction [21]. Because the four domain outcomes were prespecified and correlated, interpretation emphasizes effect magnitudes and 95% confidence intervals alongside p-values rather than treating each significance test as an independent discovery. Content validity was examined using S-CVI/Ave and Aiken's V [19], [20]. Team performance was interpreted at the team level; ICC(2,k), a two-way random effects, Fig. 2. Temporal-emotional decision matrix used in the full intervention. Students reconsidered the decision across present/future and emotional/rational perspectives before finalizing the executive recommendation. C. Measures The study used three outcome evidence sources and one measurement-quality source. First, an investigator-developed 16 item pre-post survey measured four prespecified domains: probabilistic reasoning and risk mitigation (5 items), ethical and stakeholder reasoning (4), critical AI use (3), and metacognition with emotional foresight (4). Items used a five-point Likert scale. Domain scores were calculated as the mean of their constituent items, and the overall 16-item mean was treated as a provisional composite summary rather than evidence of a unidimensional latent construct. Second, a 10-

criterion neutral performance rubric assessed the common team recommendation without requiring use of intervention-specific tools; criteria covered problem framing, evidence use, assumption transparency, uncertainty, alternative comparison, risk mitigation, stakeholder impact, ethical accountability, temporal consequences, and integrated recommendation quality. Two evaluators scored each artifact, and the two-rater mean was used for performance comparison. Third, anonymized reflections were coded for uncertainty reasoning, critical AI use, stakeholder reasoning, ethical reasoning, emotional foresight, and metacognition; only prompts common to both conditions were eligible for between condition comparison. Separately, seven faculty judges reviewed the survey items and rubric criteria for relevance and clarity to provide content-validity evidence [19], [20]. D. Data Analysis The analysis followed a triangulated mixed-methods strategy [18]. Descriptive statistics were computed for pre-, post-, and gain scores, and baseline equivalence was examined before intervention effects were tested. Baseline-adjusted ANCOVA was the primary survey analysis, with post-intervention scores as dependent variables, condition as the predictor, and the corresponding pre-intervention score as covariate. Survey models used individual students as the analysis units and did not adjust standard errors for dependence within teams or sections. Gain-score comparisons were retained as a complementary sensitivity description, and Hedges' g summarized between condition mean gain differences standardized by the pooled within-condition standard deviation of individual gains, with a absolute-agreement, average-measures coefficient, quantified rater agreement and was not treated as a test of the condition difference [22]. This rater-agreement coefficient does not quantify within-team dependence among survey responses. Qualitative evidence was used for descriptive triangulation rather than an independent causal test, and only aggregate coded patterns are reported. Given the intact section and team ecology, student-level ANCOVA estimates are interpreted as baseline adjusted classroom associations rather than component-specific causal effects.

Table 1. Instructional Conditions and Evidence Sources

Component	AC	FI	Implementation / evidence
AI-supported strategic analysis	Yes	Yes	Classroom task
Executive recommendation	Yes	Yes	Final team output
Role-based managerial scenario	Yes	Yes	Common case activity
Monte Carlo risk modeling	No	Yes	Simulation output
Temporal-emotional matrix	No	Yes	Reflection protocol
Faculty guidance	Yes	Yes	Common instructional support
Rubric-based performance assessment	Yes	Yes	Two-rater team scores
Individual reflection	Yes	Yes	Common + condition-specific prompts
Expert content validation	NA	NA	Seven faculty judges
Qualitative reflection coding	Yes	Yes	Descriptive triangulation

a. AC = active comparison; FI = full intervention; NA = not applicable. Both conditions shared the case, role structure, AI access, guidance, and assessment; FI added Monte Carlo and temporal-emotional reflection.

Results and findings

Data Integrity and Measurement Evidence

The analytic dataset contained 120 complete student records, with 60 students in each condition and no missing pre- or post-survey values. Team-level performance data comprised 24 final recommendations, 12 per condition. No duplicate student identifiers or out-of-range Likert responses were detected.

Expert review supported the content validity of the measurement system. Across the 16 survey items and 10 neutral rubric criteria, S CVI/Ave was .984 for relevance and .989 for clarity; mean Aiken's V was .870 and .828, respectively. Internal consistency was interpreted cautiously because RABJ was conceptualized as a multidimensional applied capability rather than a single reflective trait. The overall post-intervention composite showed acceptable reliability, $\alpha = .786$, whereas pre-intervention reliability was lower, $\alpha = .400$. The baseline coefficient therefore limits interpretation of the overall pretest composite and increases the importance of domain-level and non-survey evidence.

Pre and Post Survey Outcomes

Survey outcomes were analyzed using the four prespecified domain means and the overall 16-item composite described above. Results are reported using both gain comparisons and baseline-adjusted contrasts so that descriptive change and covariate-adjusted differences can be distinguished.

Table 2 summarizes the main survey outcomes. Both conditions improved from pre to post measurement, but gains were consistently larger in the full intervention. The overall composite increased by 0.407 in the active comparison and by 0.898 in the full intervention, a gain difference of 0.491. The largest gain difference appeared in metacognition and emotional foresight, followed by probabilistic reasoning and risk mitigation.

Baseline-adjusted ANCOVA was used as the primary survey analysis because small pre-intervention differences were present in some outcomes. The full intervention was associated with higher adjusted post-intervention scores across all prespecified domains. The adjusted difference on the overall composite was $B = 0.557$, 95% CI [0.482, 0.632], $p < .001$. The largest adjusted difference was observed for metacognition and emotional foresight, $B = 0.793$, 95% CI [0.652, 0.934], $p < .001$; probabilistic reasoning and risk mitigation also showed a strong adjusted difference, $B = 0.569$, 95% CI [0.426, 0.712], $p < .001$. Fig. 3 visualizes the baseline-adjusted condition contrasts across the five survey outcomes.

Performance-Based Evidence

Team-level performance evidence converged directionally with the survey results. Across 24 final recommendations, the full intervention obtained a higher two-rater mean rubric score than the active comparison (3.47 vs. 2.93). Agreement between evaluators was high for the overall rubric score, ICC(2,k) = .954. Because ICC quantifies rater agreement rather than the significance of the condition contrast, the between-condition rubric difference is interpreted here as

convergent performance evidence rather than as a separate inferential confirmation.

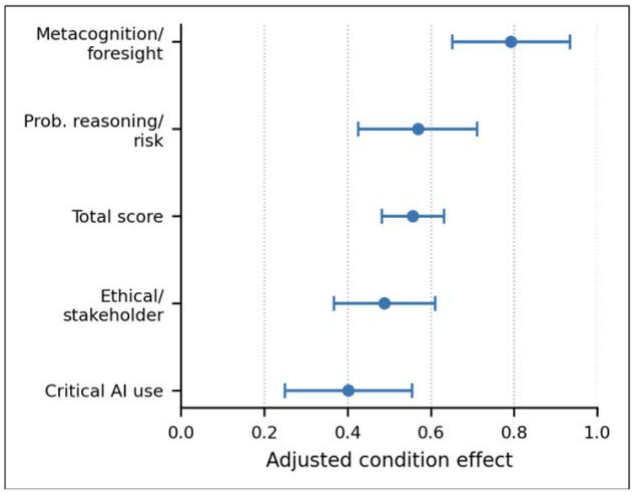


Fig. 3. Baseline-adjusted ANCOVA effects favoring the full intervention across RABJ outcomes. Error bars represent 95% confidence intervals.

Table 2. Main Survey Outcomes by Condition

Outcome	AC Δ	FI Δ	g	Adj. B [95% CI]
Prob. reasoning / risk	0.450	0.913	1.040	0.569 [0.426, 0.712]
Ethical / stakeholder	0.388	0.758	0.751	0.489 [0.367, 0.610]
Critical AI use	0.550	0.828	0.470	0.402 [0.249, 0.555]
Metacog. / foresight	0.267	1.071	1.481	0.793 [0.652, 0.934]
Overall composite	0.407	0.898	1.951	0.557 [0.482, 0.632]

a. AC = active comparison; FI = full intervention; Δ = pre-post mean change; g = Hedges' g for between-condition gain differences; Adj. B = unstandardized baseline-adjusted posttest contrast. b. All adjusted effects were significant at $p < .001$. g and Adj. B represent different estimands and are not expected to be numerically equivalent.

Qualitative Triangulation and Evidence Integration

Qualitative coding of anonymized reflections showed stronger full-intervention patterns of uncertainty awareness, critical AI use, stakeholder sensitivity, ethical reasoning, emotional foresight, and metacognitive reflection. To reduce prompt-induced bias, only prompts common to both conditions were used for direct condition comparisons; intervention-specific reflection prompts were treated as descriptive process evidence only. The qualitative component therefore contextualizes and corroborates the quantitative pattern without functioning as an independent causal test [18].

Table 3 summarizes the inferential role of each evidence stream. The distinction is important because measurement validation, student self-report, team performance, and qualitative triangulation answer different questions and should not be treated as interchangeable evidence.

Table 3. Evidence Triangulation and Inferential Role

Evidence	Observed pattern	Interpretive role
Survey	FI larger gains; overall Adj. B = .557; all adjusted domain effects $p < .001$.	Primary student-level evidence; total composite interpreted cautiously.
Team rubric	FI = 3.47 vs. AC = 2.93; ICC(2,k) = .954.	Convergent team-level performance evidence; condition contrast descriptive.
Reflections	Stronger FI patterns across six coded themes.	Descriptive triangulation; only common prompts compared.
Expert review	S-CVI/Ave = .984/.989; Aiken's V = .870/.828.	Measurement quality only; not an intervention outcome.

Discussion

The findings suggest that generative AI is more educationally valuable when students are required to evaluate, constrain, and justify its outputs rather than use it primarily to accelerate analytical production. Because both conditions used AI, the observed pattern is more consistent with an instructional architecture effect than with a simple technology-access effect. The full intervention was associated with stronger outcomes across the prespecified survey domains and showed convergent team-performance and reflection patterns, supporting the central argument that the relevant pedagogical question is how the human-AI decision loop is structured.

The large overall gain effect (g = 1.951) should be interpreted in relation to its denominator: the pooled within condition standard deviation of individual gain scores [21]. The unstandardized gain difference was 0.491 points on the five point scale; a relatively small dispersion of gains yields a large standardized contrast. Averaging 16 items can reduce dispersion, so the composite effect is not directly comparable with effects for individual domains or differently standardized outcomes. The immediate self-report measure and its close alignment with the intervention may also contribute to the magnitude. These are plausible explanations, not demonstrated mechanisms, and the estimate should not be read as evidence of durable or generalizable causal learning gains.

The strongest adjusted difference appeared in metacognition and emotional foresight. This pattern is theoretically consistent with the temporal-emotional matrix, which required students to pause before issuing a recommendation, identify present pressures, anticipate future consequences, and connect immediate analysis with longer-term ethical and stakeholder implications. The alignment is mechanism-consistent rather than mechanism-proving: because both added components were delivered together, the design cannot attribute the observed differences to either one or test their interaction.

The probabilistic reasoning and risk-mitigation results are likewise consistent with the intended role of Monte Carlo

simulation. AI-generated recommendations can appear coherent while relying on fragile assumptions or deterministic simplifications. Requiring students to examine distributions, downside exposure, sensitivity, and decision robustness shifted the task from producing a plausible answer toward evaluating whether that answer remained defensible across uncertain scenarios, consistent with prior simulation-based decision learning in business contexts [11].

More broadly, the architecture reframes AI-related cognitive efficiency. The pedagogical objective is not to minimize cognitive effort, but to relocate it from information production toward validation, uncertainty interpretation, and accountable justification. This interpretation is consistent with concerns that automation can displace monitoring [8], [9], with evidence that generative AI can redistribute cognitive effort in business school tasks [6], and with metacognitive approaches that deliberately require learners to inspect and regulate their own reasoning [15].

Performance-based and qualitative evidence strengthened—but did not replace—the survey interpretation. The higher team rubric mean indicates that the pattern was not limited to self-report, while high interrater reliability supports scoring stability. Reflection coding provided explanatory convergence across uncertainty awareness, AI critique, stakeholder sensitivity, ethical reasoning, emotional foresight, and metacognition. At the same time, Table 3 makes the inferential boundaries explicit: rubric means are descriptive between conditions, qualitative evidence is contextual, and expert review supports measurement quality rather than intervention effectiveness.

Conceptually, RABJ should therefore be treated at this stage as a provisional multidimensional applied capability rather than a fully validated psychometric trait. The domain structure, the low pre-intervention alpha of the overall composite, and the use of multiple evidence streams all argue against relying on a single total score. The stronger interpretation rests on convergence across prespecified domains, externally rated team performance, and reflection evidence, consistent with the mixed-methods logic of complementarity and triangulation [18].

The study also has practical implications for business education. AI integration should not be limited to allowing generative tools in assignments; instructors can design sequences that require students to disclose assumptions, model uncertainty, challenge AI outputs, consider affected stakeholders, and justify decisions using explicit performance criteria. The architecture is intentionally modular: the same logic can be adapted to international business, finance, operations, entrepreneurship, sustainability, and strategy courses in which students must defend decisions under uncertainty.

Several limitations qualify these findings. The quasi experimental, single-institution design limits causal and contextual generalization, and outcomes capture immediate rather than durable change. Although condition allocation was balanced within sections, students shared team activities. The student-level ANCOVA did not adjust for this dependence; its confidence intervals may therefore be too narrow and p-values too small if residual within-team or within-section correlation is positive. The intervention bundled Monte Carlo modeling with temporal-emotional reflection and related scaffolding, so component-specific contributions cannot be isolated. The low pre-intervention reliability of the overall survey composite warrants caution, and RABJ remains a provisional framework rather than a fully validated latent construct. Team rubric differences are descriptive between conditions, while qualitative evidence is non-inferential and constrained to comparable prompts. Replication with cluster-aware models, independent samples, delayed retention measures, and component-level tests is therefore needed.

Conclusion

This study advances a practical model for integrating generative AI into business education without reducing learning to faster output production. In a common AI-supported managerial task, adding Monte Carlo risk modeling and temporal-emotional reflection was associated with stronger baseline-adjusted survey outcomes and convergent-team performance and reflection evidence. The central contribution is therefore not AI as a tool in itself, but an instructional architecture that keeps students responsible for assumptions, uncertainty, stakeholder consequences, and the defensibility of final decisions. Future research should replicate the architecture across institutions and disciplines, use cluster-aware and longitudinal designs, and isolate the contribution of individual mechanisms. RABJ is best viewed here as a provisional multidimensional capability that provides a testable framework for designing and evaluating accountable human-AI judgment in management education.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the International Business Department, School of Business, Tecnológico de Monterrey, Mexico, in support of this work. The authors also express sincere appreciation to the participating students for their engagement, collaboration, and thoughtful reflections, which contributed meaningfully to this research.

References

[1] E. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education,” Learning and Individual Differences, vol. 103,. Art. 102274, 2023, doi: 10.1016/j.lindif.2023.102274

[2] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education: Where are the educators?,” International Journal of Educational Technology in Higher Education, vol. 16, Art. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[3] H. Crompton and D. Burke, “Artificial intelligence in higher education: The state of the field,” International Journal of Educational Technology in Higher Education, vol. 20, Art. 22, 2023, doi: 10.1186/s41239-023-003928.

[4] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, “Generative AI without guardrails can harm learning: Evidence from high school mathematics,” Proceedings of the National Academy of Sciences, vol. 122, no. 26, Art. e2422633122, 2025, doi: 10.1073/pnas.2422633122.

[5] C. Gonsalves, “Generative AI’s impact on critical thinking: Revisiting Bloom’s taxonomy,” Journal of Marketing Education, early access, 2024, doi: 10.1177/02734753241305980.

[6] C. Bergenholtz, O. Vuculescu, F. Günzel-Jensen, and L. Frederiksen, “Leveling up or leveling down? The impact of generative

AI on student performance in business schools," Academy of Management Learning & Education, early access, 2026, doi: 10.5465/amle.2025.0029.
[7] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister, "Hybrid intelligence," Business & Information Systems Engineering, vol. 61, no. 5, pp. 637–643, 2019, doi: 10.1007/s12599-019-00595-2.
[8] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," Human Factors, vol. 39, no. 2, pp. 230–253, 1997, doi: 10.1518/001872097778543886.
[9] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," Human Factors, vol. 52, no. 3, pp. 381–410, 2010, doi: 10.1177/0018720810376055.
[10] Y. Fan et al., "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance," British Journal of Educational Technology, vol. 56, no. 2, pp. 489-530, 2025, doi: 10.1111/bjet.13544.
[11] A. Pritchard, D. F. Taylor, and M. Belford, "Teaching data-driven decision making for inventory analysis with Monte Carlo simulation," Decision Sciences Journal of Innovative Education, early access, 2024, doi: 10.1111/dsji.12328.
[12] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," Econometrica, vol. 47, no. 2, pp. 263–291, 1979, doi: 10.2307/1914185.
[13] D. A. Kolb, Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ, USA: Prentice Hall, 1984.
[14] D. Carless, "Learning-oriented assessment: Conceptual bases and practical implications," Innovations in Education and Teaching International, vol. 44, no. 1, pp. 57–66, 2007, doi: 10.1080/14703290601081332.
[15] J. H. Flavell, "Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry," American Psychologist, vol. 34, no. 10, pp. 906–911, 1979, doi: 10.1037/0003-066X.34.10.906.
[16] R. E. Freeman, Strategic Management: A Stakeholder Approach. Boston, MA, USA: Pitman, 1984.
[17] W. R. Shadish, T. D. Cook, and D. T. Campbell, Experimental and Quasi Experimental Designs for Generalized Causal Inference. Boston, MA, USA: Houghton Mifflin, 2002.
[18] J. W. Creswell and V. L. Plano Clark, Designing and Conducting Mixed Methods Research, 3rd ed. Thousand Oaks, CA, USA: SAGE, 2018.
[19] D. F. Polit, C. T. Beck, and S. V. Owen, "Is the CVI an acceptable indicator of content validity? Appraisal and recommendations," Research in Nursing & Health, vol. 30, no. 4, pp. 459–467, 2007, doi: 10.1002/nur.20199.
[20] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," Educational and Psychological Measurement, vol. 45, no. 1, pp. 131–142, 1985, doi: 10.1177/0013164485451012.
[21] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," Frontiers in Psychology, vol. 4, Art. 863, 2013, doi: 10.3389/fpsyg.2013.00863.
[22] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," Journal of Chiropractic Medicine, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

When Immersive Learning Works: AI- and VR-Supported Feedback for Competency-Sensitive Skill Development in Higher Education

María del Pilar Gómez Carreño and Roberto Gómez Tobías
Tecnológico de Monterrey, Mexico

Article DOI: 10.68286/samyrad2026.365

ABSTRACT

Artificial intelligence (AI) and virtual reality (VR) may support different learning outcomes unevenly. This study examines competency-technology fit in an undergraduate business course through a randomized classroom study with individual allocation. A digital wheel assigned 34 students to control and experimental conditions (17 each). Both conditions received lectures, case activities, project preparation, and faculty guidance. The experimental condition added three 15-minute sessions using VirtualSpeech and Meta Quest 3 headsets for familiarization, rehearsal, and a final presentation. The course instructor assessed organizational behavior, human capital processes, and social intelligence individually using a competency-based rubric; satisfaction, engagement, and perceived learning were student-reported. Post-intervention ANOVA favored the experimental condition for organizational behavior, $F(1, 32) = 12.71$, $p = .0012$, and nominally for social intelligence, $F(1, 32) = 4.57$, $p = .0402$, but not human capital processes, $p = 1.0000$. Only organizational behavior remained significant after Holm correction within the competency family (adjusted $p = .0035$). No student-perception outcome survived correction within its family. Both groups improved on the rubric-based competency scores. The small sample and exploratory comparison across competencies warrant caution about effect magnitude and mechanism. The results motivate further tests of whether immersive feedback is especially useful for competencies requiring situated, observable performance; they do not establish competency-technology fit as a causal mechanism.

KEYWORDS Artificial intelligence, virtual reality, immersive learning, competency-based education, organizational behavior, social intelligence, formative feedback.

How to cite: Gómez Carreño, M. del P., & Gómez Tobías, R. (2026). When immersive learning works: AI- and VR-supported feedback for competency-sensitive skill development in higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.365

Introduction

Artificial intelligence (AI) and virtual reality (VR) have become prominent tools in higher education because they can make learning more interactive, adaptive, and performance oriented. AI can provide rapid feedback, personalized prompts, and analytic support, while VR can place learners in simulated environments where they can rehearse professional behaviors without the cost or risk of live organizational settings [1]-[5]. For business education,

this combination is especially promising because many managerial competencies are not merely declarative. They require interpretation, communication, social perception, judgment, and behavioral performance under conditions that resemble professional practice.

At the same time, the educational value of immersive technologies should not be assumed. The relevant question is not only whether AI and VR improve learning, but which competencies are likely to benefit from them and why. Competency-based education requires alignment among the competency being developed, the learning task, the feedback mechanism, the assessment evidence, and the form of faculty mediation [6], [7]. A technology-rich activity can therefore be pedagogically weak when the technology does not match the structure of the target competency, even if students find the experience novel or engaging.

Recent empirical findings make this distinction relevant to business education. Hubbard [8] compared written, video, and VR business cases: VR improved visual recall but reduced numerical recall, indicating outcome-specific trade-offs. In presentation training, Sichterman et al. [9] compared AI-only feedback with teacher-supported AI feedback after immersive practice and detected no significant between-condition differences. That finding does not establish equivalence, but it cautions against assuming that an additional instructional component necessarily improves every outcome. Together, these studies motivate closer attention to the learning target and the feedback arrangement.

The three course competencies offer a theoretically relevant comparison within a common instructional setting. Organizational behavior requires interpreting how a proposal affects workplace behavior and organizational wellbeing. Social intelligence is expressed through persuasive, audience-sensitive communication and the ability to elicit commitment. Human capital processes emphasizes the justification and integration of talent-management procedures. These are differences in the dominant demands of the assessed tasks, rather than claims that any competency is purely behavioral or purely conceptual. Immersive presentation practice may make the first two demands more directly rehearsable, while cases and guided analysis may already support much of the third.

The remaining question is how the affordances of immersive feedback align with different competencies assessed within the same business course. Competency-technology fit adds an outcome-specific proposition to experiential learning, feedback, and technology-integration accounts: the same instructional technology may offer different incremental benefits depending on what students must demonstrate and what its feedback makes observable. The present randomized classroom study examines that proposition through three faculty-rated competencies and a separate set of student-perception outcomes.

Three research questions guide the analysis. RQ1 asks whether the experimental condition produced stronger post-intervention competency outcomes than the control condition. RQ2 asks whether those differences varied across competencies in a pattern consistent with competency-technology fit. RQ3 asks whether student perceptions followed the same pattern as competency outcomes or revealed a more nuanced relationship among satisfaction, engagement, and perceived learning.

The theoretical proposition is that immersive rehearsal and formative feedback may be more useful when learners can practice, observe, and revise the behavior targeted by assessment. Structured conceptual or procedural demands may receive less additional support from presentation rehearsal. We use this proposition to interpret an exploratory pattern across outcomes, not as a confirmatory test of an independently prespecified hierarchy of competency fit.

Conceptual background and competency-technology fit model

Competency-sensitive immersive learning

Immersive learning is most defensible when the target outcome requires action in context. Constructivist and experiential learning perspectives emphasize that competence develops through active experience, feedback, reflection, and repeated performance [10], [11]. VR can support this process by creating controlled practice environments, while AI can accelerate formative feedback by identifying observable features of performance and offering suggestions for improvement. This combination is especially relevant when students must rehearse communication, leadership, negotiation, presentation, or socially situated judgment.

A competency-sensitive approach requires distinguishing technology affordances from competency requirements. VR offers presence, simulation, and repeated rehearsal; AI offers feedback, personalization, and rapid analytic response. These affordances are strongest when the competency can be practiced through observable behavior and when feedback can be linked to performance indicators. They may be less decisive for competencies that depend primarily on conceptual classification, procedural knowledge, or content mastery already supported by cases, lectures, worked examples, and guided analysis.

Feedback, faculty mediation, and assessment alignment

Feedback is one of the strongest mechanisms through which learning environments shape performance, but feedback is most effective when it is timely, specific, understandable, and connected to criteria that students can act on [12]. In immersive AI-supported environments, automated feedback can increase the frequency of practice-feedback cycles. Yet automated feedback should not be treated as a complete judgment of competence. Faculty mediation remains necessary to interpret feedback, connect it to disciplinary standards, challenge superficial responses, and ensure that platform-generated indicators are translated into meaningful learning decisions.

The Technological Pedagogical Content Knowledge framework situates technology within pedagogy and disciplinary content [13]. Here, immersive practice was embedded in project preparation and faculty guidance. Automated feedback introduced an additional source of performance information, while the course instructor remained responsible for interpreting disciplinary expectations and assessing competencies. This arrangement is part of the instructional condition being evaluated.

Competency-technology fit as a design proposition

Competency-technology fit is defined here through four related dimensions: behavioral observability, or whether the target performance can be seen or heard; interpersonal situatedness, or dependence on an audience or social context; rehearsability, or the opportunity to practice the target action; and feedback responsiveness, or whether feedback can guide a concrete revision. These dimensions connect competency demands to the specific affordances of immersive presentation practice. They extend general accounts of learning through experience and feedback by identifying why the same activity might support some assessed outcomes more directly than others.

Table 1 applies these dimensions to the three competencies and juxtaposes the proposed fit with the observed outcome pattern. The fit classifications function here as theory-informed interpretive judgments; the present analysis does not establish prospective confirmation of a competency hierarchy. Accordingly, the table provides an exploratory explanation to be tested prospectively. The study does not manipulate competency fit or estimate a formal treatment-by-competency interaction.

Competency fit and assessment evidence

Competency-technology fit also has an assessment implication. A technology may create an engaging experience yet still fail to produce evidence that corresponds to the intended competency. For outcomes centered on social interaction or behavioral performance, assessment should capture what students can enact, adjust, and justify in context. For more structured procedural outcomes, written analyses, process accuracy, or concept application may provide a closer representation of competence. The assessment method therefore participates in the fit relationship rather than serving as a neutral measurement step.

This perspective is consistent with feedback research emphasizing that learners benefit when performance criteria are visible and actionable [12]. In the present study, the same logic connects rehearsal to competency evidence: immersive practice is educationally meaningful only when feedback points students toward the behaviors and interpretations that the course intends to develop. This reinforces the human role in the model because faculty must decide which indicators are instructionally relevant and which platform outputs are merely proxies for more complex professional performance.

Table 1. Competency-affordance alignment and interpretive fit

Competency	Learning demand	Proposed fit	Observed evidence
Organizational behavior	Situated behavioral interpretation	High	Robust; p=.0012
Human capital processes	Structured procedural/conceptual	Mid-low	No difference; p=1.000
Social intelligence	Social interpretation and interaction	High	Nominal only; p=.0402

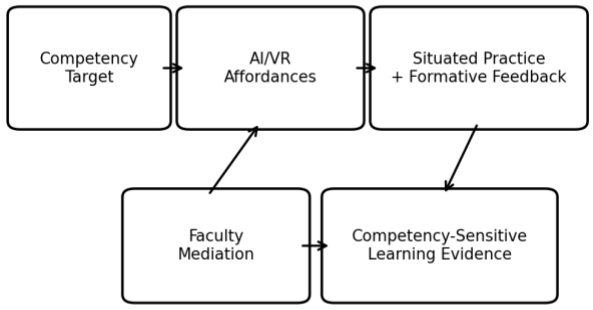


Fig. 1. Proposed competency-sensitive alignment model for AI- and VR-supported formative feedback.

Fig. 1 summarizes the proposed relationship. A target competency defines the learning demand; AI and VR contribute simulation, rehearsal, and feedback; situated practice creates learning opportunities; and faculty

mediation preserves disciplinary meaning and assessment responsibility. The proposed selective benefit of this alignment remains a question for prospective testing.

Research design and methods

Research design, context, and participants

The study used a randomized classroom design with individual allocation [14] in an undergraduate course on organizational behavior and human capital processes at a private university in Mexico. The analytic sample comprised 34 students, with 17 in each condition. Students had previously formed project teams through Canvas, but these teams were not the allocation unit. A freely available online digital wheel randomly assigned individual students to the experimental or control condition. Each student presented their project individually, and the course instructor assessed competencies at the individual level. Instruction and the evaluated presentations were conducted in Spanish.

The course lasted five weeks, with four scheduled classroom hours per week (20 hours in total). The sample reflects the available classroom cohort, rather than a random sample of the wider student population. Pre-intervention ratings preceded the intervention; post-intervention competency assessment and student-perception measurement followed the instructional sequence. The analysis therefore concerns immediate course-related outcomes in a single classroom context.

Instructional conditions and faculty mediation

Both conditions shared lectures, case activities, project preparation, faculty guidance, and pre/post measurement. The experimental condition used VirtualSpeech with Meta Quest 3 headsets during three 15-minute sessions embedded in regular classes. The first session introduced the headset and interface; the second involved presentation rehearsal with virtual avatars; and the third comprised the final presentation. Thus, the sequence covered 45 scheduled minutes across familiarization, rehearsal, and final presentation. During the corresponding periods, control students engaged in face-to-face preparation and presentation rehearsal. Table 2 summarizes the conditions.

The virtual avatar provided formative feedback, followed by an automated performance evaluation. The results interface displayed content, eye contact, filler words, speaking pace, body language, and listenability indicators, together with an overall score and presentation duration. For example, speaking pace was expressed in words per minute, filler words as a count, and eye contact as a percentage. These indicators provided feedback on presentation performance. They were not the competency outcome scores analyzed in the study; those scores were assigned by the course instructor using the rubric.

Both groups worked within the same scheduled course hours and corresponding preparation periods. The conditions differed in the medium and source of feedback: immersive avatar-supported feedback was added in the experimental condition, while faculty guidance remained part of both. Equal scheduled time does not establish an identical amount of individual rehearsal or feedback. The comparison therefore evaluates the integrated instructional condition, without separating the contributions of immersion, automated feedback, and faculty mediation.

Table 2. Instructional conditions and evidence sources

Component	Control	Experimental	Evidence
Lectures/case activities	Yes	Yes	Course activities
Face-to-face project preparation	Yes	Yes	Final project
AI-enabled immersive rehearsal	No	Yes	Platform practice
Automated formative feedback	No	Yes	Platform feedback
Simulated professional scenarios	No	Yes	Immersive activity
Faculty mediation	Yes	Yes	Guided instruction
Pre/post competency ratings	Yes	Yes	Instructor rubric

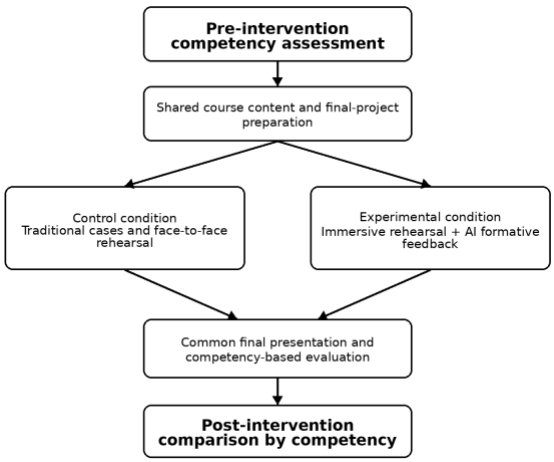


Fig. 2. Classroom sequence with individual random allocation and condition-specific learning activities.

Measures and competency operationalization

Competency outcomes were individual performance evaluations made by the course instructor responsible for the challenge, using the competency-based rubric. Its five performance levels were Destacado (outstanding), Avanzado (advanced), Sólido (solid), Básico (basic), and

Incipiente (emerging), from highest to lowest. The analytic dataset contained one instructor-assigned score for each competency at each measurement occasion on a 1–5 scale, with higher scores indicating stronger performance. The three competencies were analyzed separately; no overall competency index was constructed.

The rubric described two criteria for each competency. Organizational behavior concerned the clarity of the proposed improvement and justification of its contribution to organizational wellbeing and performance, including relevant evidence or indicators. Human capital processes concerned a well-grounded talent-management proposal and integration of prior learning and current trends. Social intelligence was operationalized through persuasive communication adapted to the managerial audience and a conclusion that encouraged commitment and collaborative action. These criteria assess the application of the course competencies within an individual project presentation; they do not constitute an exhaustive measure of each broader construct.

Satisfaction, engagement, and perceived learning were collected through student surveys, separately from faculty assessment. Each was represented by its corresponding course-specific item on a five-point agreement scale (1 = strongly disagree; 5 = strongly agree). For example, the pre-intervention engagement item asked whether students felt motivated and committed to course activities and materials; the perceived-learning item asked whether their learning in the course had improved. The post-intervention items concerned the instructional experience. These indicators were analyzed individually, without creating a multi-item perception scale.

The course instructor applied the rubric as the sole assessor; inter-rater agreement is therefore not applicable to this assessment design. Internal-consistency coefficients are likewise not applicable to individual single-item perception indicators. The recorded rubric-based ratings constitute the competency outcomes analyzed in Tables 3, 4, and 7. Results are reported at group level without student identifiers.

Data analysis and interpretive strategy

Paired t tests examined within-condition pre/post competency changes, and one-way ANOVA compared post-intervention outcomes between groups (df1 = 1, df2 = 32). Means, sample standard deviations, and pointwise 95% t-based confidence intervals (CIs) are reported for both groups at each occasion. Post-intervention mean-difference CIs use the pooled-variance model underlying the ANOVA. Holm adjustment was applied separately to the three competency comparisons and the three perception comparisons, using unrounded p values; the six paired competency tests also received a separate Holm sensitivity check. Within-group effect sizes use d_{av}: the pre-minus-post mean difference divided by the square root of the average of the pre and post variances [15]. Negative d_{av} values therefore indicate improvement. Partial eta squared (ηp^2) summarizes each post-intervention ANOVA.

The five-point ratings were treated as numerical scores. Confidence intervals are unadjusted for multiplicity and describe sampling precision under these models. Holm correction limits familywise false-positive risk within each specified family, but does not remove the imprecision of a small sample or the possibility that statistically selected effects overstate their population magnitude. Nominal findings that do not survive correction are treated as exploratory. Post-intervention ANOVA does not adjust for baseline differences; this constraint informs interpretation.

Results

Competency outcomes

Table 3 reports competency scores before and after the intervention. At baseline, experimental-group means exceeded control means by 0.12 points for organizational behavior and human capital processes and by 0.24 points for social intelligence. These differences provide context for the post-intervention comparisons.

Table 3. Descriptive statistics for competency ratings

Outcome / group	Pre M (SD) [95% CI]	Post M (SD) [95% CI]
Organizational behavior Control	3.24 (0.66) [2.89, 3.58]	4.18 (0.39) [3.97, 4.38]
Organizational behavior Experimental	3.35 (0.70) [2.99, 3.71]	4.71 (0.47) [4.46, 4.95]
Human capital processes Control	3.29 (0.69) [2.94, 3.65]	4.47 (0.51) [4.21, 4.74]
Human capital processes Experimental	3.41 (0.51) [3.15, 3.67]	4.47 (0.51) [4.21, 4.74]
Social intelligence Control	3.24 (0.66) [2.89, 3.58]	4.35 (0.49) [4.10, 4.61]
Social intelligence Experimental	3.47 (0.72) [3.10, 3.84]	4.71 (0.47) [4.46, 4.95]

Note. n = 17 per group at each occasion. M = mean; SD = sample standard deviation. Brackets show unadjusted 95% t-based confidence intervals for the mean. Scores range from 1 to 5.

Table 4 reports the post-intervention competency ANOVAs. Organizational behavior favored the experimental condition, $F(1, 32) = 12.71$, nominal $p = .0012$, $\eta p^2 = .284$, and remained significant after Holm correction (adjusted $p = .0035$). Its post-intervention mean difference was 0.53 points (95% CI [0.23, 0.83]). Social intelligence showed a nominal difference, $F(1, 32) = 4.57$, $p = .0402$, $\eta p^2 = .125$, but did not survive correction (adjusted $p = .0805$). Human capital processes showed no detected difference,

$F(1, 32) = 0.00$, $p = 1.0000$, $\eta p^2 = .000$. Only organizational behavior therefore provides a multiplicity-robust between-condition result in this sample.

Table 4. Anova results for post-intervention competency outcomes.

Outcome	F	p (nominal)	Interpretation
Organizational behavior	12.71	.0012	Holm-robust; ηp^2=.284
Human capital processes	0.00	1.0000	No difference; ηp^2=.000
Social intelligence	4.57	.0402	Suggestive; ηp^2=.125

Student perceptions

Table 5 presents the pre/post student-perception descriptives. These ratings describe the students' instructional experience and are distinct from the instructor-assessed competency outcomes.

Table 5. Descriptive statistics for student perceptions.

Outcome / group	Pre M (SD) [95% CI]	Post M (SD) [95% CI]
Satisfaction Control	3.41 (0.51) [3.15, 3.67]	4.29 (0.47) [4.05, 4.54]
Satisfaction Experimental	3.41 (0.51) [3.15, 3.67]	4.53 (0.51) [4.26, 4.79]
Engagement Control	3.35 (0.79) [2.95, 3.76]	4.29 (0.59) [3.99, 4.60]
Engagement Experimental	3.47 (0.62) [3.15, 3.79]	4.65 (0.49) [4.39, 4.90]
Perceived learning Control	3.59 (0.62) [3.27, 3.91]	4.47 (0.51) [4.21, 4.74]
Perceived learning Experimental	3.59 (0.62) [3.27, 3.91]	4.82 (0.39) [4.62, 5.03]

Note. n = 17 per group at each occasion. M = mean; SD = sample standard deviation. Brackets show unadjusted 95% t-based confidence intervals for the mean. Scores range from 1 to 5. Model-based intervals are not truncated to the response range.

Table 6 summarizes post-intervention perception ANOVAs. Perceived learning nominally favored the experimental condition, $F(1, 32) = 5.05$, $p = .0316$, $\eta p^2 = .136$. Neither engagement, $F(1, 32) = 3.60$, $p = .0668$, $\eta p^2 = .101$, nor satisfaction, $F(1, 32) = 1.94$, $p = .1733$, $\eta p^2 = .057$, met the nominal .05 threshold. Holm-adjusted p values were .0948 for perceived learning, .1337 for engagement, and .1733 for satisfaction. No perception outcome survived correction; the perceived-learning result is exploratory.

Table 6. Anova results for post-intervention perceptions

Perception	F	p (nominal)	Interpretation
Satisfaction	1.94	.1733	No difference; ηp^2=.057
Engagement	3.60	.0668	Not significant; ηp^2=.101
Perceived learning	5.05	.0316	Suggestive; ηp^2=.136

Pre-post gains within groups

Paired t tests showed increased competency ratings in both conditions. Table 7 reports the t statistics and d_{av} values; their negative signs reflect pre-minus-post coding. All six paired comparisons remained below .001 after Holm correction across these tests. The absolute d_{av} values ranged from 1.72 to 2.27. These large standardized changes reflect both the observed score increases and the relatively small dispersion of five-point ratings. They describe within-condition change and do not estimate the incremental treatment effect.

Both instructional conditions were accompanied by improved rubric-based performance. The between-condition evidence is more selective: organizational behavior favored the experimental condition after multiplicity adjustment, social intelligence provided a nominal signal, and human capital processes showed no detected difference. Improvement within each group should not be interpreted as proof of superiority of either instructional approach.

Table 7. Pre-post competency changes within groups

Group	Competency	t	p	d_{av}
Control	Organizational behavior	-5.19	<.001	-1.72
Control	Human capital processes	-7.63	<.001	-1.94
Control	Social intelligence	-5.37	<.001	-1.91
Experimental	Organizational behavior	-7.95	<.001	-2.27
Experimental	Human capital processes	-5.28	<.001	-2.07
Experimental	Social intelligence	-9.06	<.001	-2.04

Note. df = 16 for each paired test. d_{av} uses the pre-minus-post difference standardized by the square root of the average pre/post variances; negative values indicate higher post scores.

Integrative competency-fit pattern

The observed pattern is consistent with a competency-sensitive interpretation, but remains preliminary. Organizational behavior produced the strongest between-condition evidence. Social intelligence showed a smaller nominal effect, while human capital processes showed no detected advantage. Because the same course and instructor provided the setting for all three outcomes, this contrast motivates further investigation of how presentation rehearsal supports different task demands.

A significant result for one competency and a nonsignificant result for another do not by themselves establish that the treatment effects differ. Competency type was not manipulated, and no formal treatment-by-competency interaction was estimated. The pattern

therefore motivates a prospective test of the proposed fit dimensions rather than establishing competency-technology fit as a causal mechanism.

Research-question synthesis

RQ1 receives selective support: only organizational behavior showed a between-condition advantage robust to Holm correction. RQ2 yields an exploratory pattern that is compatible with the proposed fit interpretation, without demonstrating effect heterogeneity or mechanism. RQ3 indicates that no student-perception outcome survived multiplicity correction. The findings support further testing of a conditional proposition rather than a general superiority claim for immersive technology.

Discussion

From technology effect to conditional effectiveness

The findings motivate a competency-sensitive interpretation of AI- and VR-supported formative feedback, with the strongest evidence for organizational behavior. The assessed task required students to explain and justify behavioral improvements in an organizational context. Social intelligence showed a nominal effect in the same direction, but it did not survive Holm correction. Human capital processes showed no detected difference; this is compatible with, but does not establish, a smaller incremental benefit for more structured learning demands.

Competency-technology fit directs attention to the relation among affordances, task demands, feedback, and assessment. In this study, presentation practice made some communication and behavioral features observable, while the instructor assessed their disciplinary relevance. A plausible interpretation is that this alignment supported the organizational-behavior task. Establishing that interpretation as a mechanism would require process measures and direct tests of fit, beyond the present outcome comparisons.

The null result for human capital processes is therefore not a residual finding to be minimized. Together with the strong organizational-behavior effect and the weaker social-intelligence signal, it creates the contrast that makes the study theoretically informative. The evidence does not justify claiming that competency-technology fit has been causally demonstrated; competency type was not manipulated and the sample is small. It does, however, support a more precise testable proposition than a generic technology-effect account: the incremental educational value of AI and VR may depend on what learners are being asked to become competent at doing.

Human mediation as part of the instructional architecture

Faculty mediation remained part of both conditions. The platform introduced avatar-delivered feedback and automated performance indicators, while the instructor retained responsibility for course goals, disciplinary expectations, and competency assessment. This division of roles is consistent with TPACK's emphasis on integrating technology with pedagogy and content [13]. Because the study did not vary teacher support independently, it cannot estimate the separate contribution of faculty mediation.

This human mediation is especially important when AI-generated feedback concerns socially situated performance. Automated indicators can identify observable patterns, but they do not exhaust the meaning of organizational behavior or social intelligence. Faculty judgment is required to connect performance signals to context, professional norms, ethical considerations, and the learning objectives of the course. From this perspective, the relevant innovation is a human-AI instructional architecture in which automated feedback increases practice opportunities without transferring educational accountability to the system.

Perceived learning, satisfaction, and productive effort

The perception results add an important qualification. Perceived learning favored the experimental condition at the nominal level, but the effect did not remain statistically significant after Holm correction; satisfaction and engagement did not meet the nominal significance threshold. These results caution against treating satisfaction or self-reported engagement as proxies for educational effectiveness. Immersive learning can introduce adaptation demands, unfamiliar interaction patterns, and additional effort even when students perceive value in the experience. Prior work similarly emphasizes that emotion, cognitive response, and engagement in VR are intertwined with instructional design rather than automatic consequences of immersion [16].

For instructors, the implication is not to maximize friction or difficulty, but to distinguish productive effort from technical burden. Practice should require students to use feedback, revise performance, and reflect on evidence. Technical complexity that does not serve these processes should be minimized. This distinction is particularly relevant as institutions adopt AI-enabled systems: ease of use matters, but the educational objective is not frictionless completion; it is meaningful rehearsal and improvement.

Hybrid rather than replacement logic

The findings suggest a provisional hybrid design in which conventional instruction supports conceptual framing and guided analysis, while immersive feedback supports rehearsal of observable performance. This is a pedagogical implication to be tested, rather than an empirically established optimal sequence. The study evaluates one integrated classroom implementation and does not compare alternative sequencing or replacement strategies.

Selective adoption may be more defensible than placing every outcome inside an immersive environment. A competency map can identify tasks for which presentation practice makes relevant behavior visible and feedback actionable. Communication, social interpretation, and leadership-related performance are plausible candidates; more structured tasks may require different supports. These are design proposals derived from the framework, not additional outcomes demonstrated by the present sample.

Toward a research agenda for competency-sensitive immersive learning

The competency-sensitive perspective suggests a research agenda that goes beyond comparing one technology-enhanced group with one conventional group. Future studies can classify competencies according to theoretically relevant attributes such as observability, social situatedness, ambiguity, feedback responsiveness, and dependence on repeated rehearsal. Those attributes can then be modeled as moderators of the incremental benefit produced by immersive AI-supported practice. Such designs would permit the field to test whether competency-technology fit explains variation across outcomes more effectively than broad labels such as 'soft skills' or 'technology-enhanced learning.'

A second research direction concerns the decomposition of the instructional architecture. The present study evaluates an integrated condition in which immersion, AI feedback, and faculty mediation operate together. Subsequent research could compare immersive rehearsal with and without AI feedback, AI feedback with and without immersion, or different levels of faculty scaffolding. This would help distinguish the affordance of the medium from the instructional value of feedback and human mediation, while retaining the ecological validity of authentic classroom implementation.

Limitations and future research

The sample was small (n = 34; 17 per condition), drawn from a single course, and assessed immediately after the intervention. Effect estimates are therefore imprecise and may differ substantially in larger samples. Large standardized effects should be interpreted alongside the raw-score differences and confidence intervals, particularly because the outcomes were five-point ratings closely aligned with the course tasks. Small sample size does not itself change a test's nominal significance threshold, but unstable estimates and selection of significant results can exaggerate apparent effect magnitude. Holm correction addresses specified families of comparisons, not these broader sources of uncertainty.

Individual random allocation supports the group comparison, but chance baseline differences remain possible in a small class. The experimental group had somewhat higher initial competency means, and the post-intervention ANOVA does not adjust for those differences. Future tests should prespecify baseline-adjusted models and formal comparisons of treatment effects across competencies. The integrated condition also does not separate immersion, automated feedback, and instructional mediation. Equal scheduled class time cannot establish equal individual practice or feedback exposure. The social-intelligence and perceived-learning findings remain nominal signals because neither survived its family's Holm correction.

The assessment design used one course instructor and separate single-item perception indicators. Replication should include independent performance assessment and broader multi-item measures of student perceptions, together with multiple instructors, larger cohorts, and delayed follow-up. Process evidence on rehearsal, uptake of avatar feedback, and subsequent revisions would help distinguish the proposed fit mechanism from alternative explanations. Prior VR experience, digital self-efficacy, and task complexity are also relevant candidates for future measurement.

Practical implications for AI-supported higher education

For educational leaders, the practical implication is to avoid technology-first adoption. AI and VR investments should begin with a competency map. If the target outcome requires performance, social interaction, communication, behavioral interpretation, or repeated rehearsal in context, immersive practice with AI-supported feedback is pedagogically defensible. If the target outcome is primarily procedural or conceptually structured, AI and VR may still be useful, but different supports may be more appropriate and more economical.

For faculty, the central design principle is alignment. Learning tasks should require students to use feedback, not merely receive it. Assessment should capture observable performance and disciplinary criteria rather

than rely primarily on satisfaction. Automated suggestions should be translated into discipline-specific improvement through instructor mediation. Under these conditions, AI and VR can support a responsible form of educational innovation in which technology expands opportunities for practice while human educators retain responsibility for meaning, standards, and assessment.

For curriculum designers, the study suggests a practical sequence: establish conceptual foundations through conventional instruction; identify competencies that require enactment or social interpretation; introduce immersive rehearsal where it can make performance visible; use AI feedback to accelerate formative cycles; and return to faculty-mediated assessment for contextual judgment. This sequence operationalizes competency-sensitive adoption and avoids the assumption that every learning objective benefits from the same technological treatment.

At the program level, competency-sensitive mapping can also improve adoption decisions. Rather than purchasing immersive technologies and subsequently searching for uses, schools can identify high-value competency targets first, specify the evidence needed to demonstrate improvement, and then select technologies whose affordances match those requirements. This reverses the usual adoption sequence and provides a more defensible basis for scaling AI and VR across curricula.

Conclusion

This randomized classroom study examined AI- and VR-supported feedback in an undergraduate business course. Both conditions showed improved rubric-based competency ratings. Organizational behavior was the only post-intervention outcome that favored the experimental condition after Holm correction; social intelligence and perceived learning showed nominal signals, while human capital processes showed no detected difference. These results offer preliminary evidence motivating prospective tests of competency-technology fit. They do not establish a causal fit mechanism or demonstrate that the treatment effects differ across competencies. The practical proposition is to align immersive rehearsal and feedback with specific, observable task demands while retaining faculty responsibility for assessment.

The study contributes a concrete classroom comparison and a testable account of why immersive feedback might support learning outcomes differently. Larger studies with direct tests of the proposed fit dimensions are needed to determine where this instructional approach offers a reproducible advantage.

References

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education - where are the educators?" Int. J. Educ. Technol. High. Educ., vol. 16, no. 39, 2019.

[2] H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," Int. J. Educ. Technol. High. Educ., vol. 20, no. 22, 2023.

[3] D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson, "Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design," J. Comput. Educ., vol. 8, pp. 1-32, 2021, doi: 10.1007/s40692-020-00169-2.

[4] B. M. Kyaw et al., "Virtual reality for health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration," J. Med. Internet Res., vol. 21, no. 1, e12959, 2019.

[5] C. Dede, "Immersive interfaces for engagement and learning," Science, vol. 323, no. 5910, pp. 66-69, 2009.

[6] D. Carlgren, "Competency-based curriculum transition: A conceptual framework," J. Competency-Based Educ., vol. 6, no. 2, e1229, 2021, doi: 10.1002/cbe2.1229.

[7] C. Rivers, S. Gibson, E. Contreras, T. Livingston, and P. Hanson, "Competency-based education: An evolutionary higher education business model," J. Competency-Based Educ., vol. 4, no. 1, e1179, 2019, doi: 10.1002/cbe2.1179.

[8] T. D. Hubbard, "Transforming business education: The impact of virtual reality on learning outcomes in case studies," Front. Virtual Real., vol. 6, Art. no. 1461690, 2025, doi: 10.3389/frvir.2025.1461690.

[9] B. Sichterman et al., "Can AI feedback stand alone for fostering students' presentation performance? A comparison of AI-only versus teacher-supported AI feedback," Interactive Learning Environments, vol. 34, no. 3, pp. 1280–1297, 2026, doi: 10.1080/10494820.2025.2521338.

[10] D. A. Kolb, Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ, USA: Prentice Hall, 1984.

[11] D. H. Jonassen, "Instructional design models for well-structured and ill-structured problem-solving learning outcomes," Educ. Technol. Res. Dev., vol. 45, no. 1, pp. 65-94, 1997.

[12] J. Hattie and H. Timperley, "The power of feedback," Rev. Educ. Res., vol. 77, no. 1, pp. 81-112, 2007.

[13] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge," Teach. Coll. Rec., vol. 108, no. 6, pp. 1017-1054, 2006.

[14] W. R. Shadish, T. D. Cook, and D. T. Campbell, Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA, USA: Houghton Mifflin, 2002.

[15] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum, 1988.

[16] M. Vesisenaho et al., "Virtual reality in education: Focus on the role of emotions and physiological reactivity," J. Virtual Worlds Res., vol. 12, no. 1, 2019, doi: 10.4101/jvwr.v12i1.7329.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

AI-Supported Learning and Job Readiness in Vocational Tourism Education: A Comparative Study of Two Albanian Schools

Albjona Selimi[1,2], Ardita Boriçi[2]
(e-mail: albjonah@outlook.it), (e-mail: ardita.borici@unishk.edu.al)

[1]Vocational High School, Shkodër, Albania
[2]University of Shkodër "Luigj Gurakuqi", Albania

Article DOI: https://doi.org/10.68286/samyrad2026.366

ABSTRACT

Artificial intelligence is entering classrooms rapidly and is increasingly promoted as a way to personalize learning and prepare young people for work. In vocational education and training (VET), where the explicit goal is employability, this promise is especially significant. Tourism and hospitality are a major employer in Albania, and the quality of vocational graduates entering this sector has direct consequences for the local labour market. Understanding whether AI-supported learning genuinely helps vocational students feel ready for work is therefore both an educational and an economic question. Despite growing enthusiasm for AI in education, it remains unclear whether AI-supported learning actually improves vocational students' perceived job readiness, and how its effect compares with that of traditional work-based learning. Even less is known about whether these relationships hold across different institutions or vary by school context. Evidence from Albanian vocational tourism education, in particular, is largely absent from the literature. The study uses a quantitative, cross-sectional design. Survey data are collected from tourism and hospitality students in two Albanian vocational schools — "Hamdi Bushati" in Shkodër and "Kolin Gjoka" in Lezhë — and analyzed using multiple linear regression. The dependent variable is self-assessed job readiness (1–5 Likert scale). The main predictors are students' use of AI-supported learning tools and the quality of their internship / work-based learning (both 1–5), together with a school dummy and a school × AI interaction term; gender, age, grade and prior work experience are included as controls. Four hypotheses test whether AI-supported learning (H1) and internship quality (H2) positively predict job readiness, whether readiness differs between the two schools (H3), and whether the effect of AI-supported learning differs by school (H4). As an ongoing study, full results are pending data collection. The analysis is expected to show that greater use of AI-supported learning tools and higher-quality work-based learning are each positively associated with perceived job readiness (H1, H2). Comparing the two schools will indicate whether readiness levels differ by institution (H3) and whether the contribution of AI-supported learning is stronger in one setting than the other (H4). Such patterns would suggest that AI tools complement, rather than replace, high-quality practical placements, and that local school context shapes how much students benefit from them. By isolating the contribution of AI-supported learning alongside work-based learning in a real VET setting, the study aims to clarify how AI can strengthen employability in vocational tourism education.

KEYWORDS vocational education and training, artificial intelligence, AI-supported learning, job readiness, employability

How to cite: Selimi, A., & Borici, A. (2026). AI-supported learning and job readiness in vocational tourism education: A comparative study of two Albanian schools. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.366

Introduction

Artificial intelligence is becoming part of everyday learning. Students now use generative AI, chatbots, and other digital tools to ask questions, clarify difficult ideas, practice tasks, and receive feedback outside the traditional classroom. These tools can make learning more flexible and responsive, but their growing presence also raises a practical question for educators: does using AI actually help students feel better prepared for work? This question is especially important in vocational education and training (VET), where learning is closely tied to occupational competence and the transition from school to employment. Vocational students need more than theoretical knowledge. They are also expected to develop practical skills, digital confidence, communication abilities, and an understanding of real workplace demands. In tourism and hospitality, this combination matters greatly because the sector is becoming more digital while still depending heavily on human interaction, service quality, teamwork, and hands-on experience. Albania offers a particularly relevant setting for exploring these issues. Tourism and hospitality are important sources of employment, and vocational schools have a direct role in preparing young people for jobs in the sector. Yet there is still little evidence on how AI-supported learning relates to job readiness among Albanian vocational students. At the same time, the literature on AI in education and the literature on internships and work-based learning have often developed separately. This leaves an important question unanswered: does AI-supported learning add something to students' perceived job readiness when the value of practical workplace experience is also taken into account?

To address this question, the study compares tourism and hospitality students from "Hamdi Bushati" Vocational High School in Shkodër and "Kolin Gjoka" Vocational High School in Lezhë. It asks whether AI-supported learning predicts perceived job readiness, whether the quality of internship and work-based learning also matters, whether students at the two schools differ in their average readiness, and whether the relationship between AI-supported learning and readiness changes across school contexts.

The study makes three main contributions. First, it brings AI-supported learning into a vocational tourism context and links it directly with students' perceptions of job readiness. Second, it considers AI-supported learning together with internship quality, recognizing that digital learning and practical experience may reinforce rather than replace one another. Third, by comparing two schools, it considers whether institutional context may shape the benefits students associate with AI-supported learning.

Literature Review

AI-Supported Learning in Vocational Education and Training

AI is opening new possibilities for teaching and learning by making it easier to provide personalized explanations, immediate feedback, problem-solving support, and opportunities for repeated practice. These features can be especially valuable in VET, where learning is expected to lead to usable occupational skills rather than remain only theoretical. Human Capital Theory offers a useful way to understand this connection. Becker (1993) views education, training, and work experience as investments that build knowledge and skills and, in turn, improve productivity and labour-market value. From this perspective, AI-supported learning can be seen as another learning resource within the broader process of human-capital development. It may strengthen cognitive and digital skills, while internships provide the applied and occupation-specific experience that develops in real workplaces. Research on AI in education points to clear opportunities, including more personalized learning, faster access to assistance, and support for learning achievement. In vocational education, therefore, the key issue is not simply whether students have access to AI, but how they use it: for example, to understand vocational content, work through job-related problems, or practise realistic professional situations. Recent VET research similarly stresses that infrastructure, teacher capability, curriculum integration, and responsible use all influence whether AI becomes educationally meaningful (Prasetya et al., 2025).

AI-Supported Learning in Tourism and Hospitality Education

Tourism and hospitality education is being reshaped by the wider digital transformation of the sector. AI-supported customer service, recommendation systems, automated booking, data analytics, and generative AI are increasingly part of tourism workplaces. As a result, schools and universities must prepare students for two realities at once: the enduring importance of service, communication, and operational skills, and the growing need to work confidently with digital technologies. Recent studies suggest that ChatGPT and related tools can support learning in tourism education, although the benefits depend partly on students' digital literacy and on how AI is incorporated into learning activities (Dalgıç et al., 2024). Research using the Technology Acceptance Model (TAM) also shows that students' responses to ChatGPT are influenced by whether they see the technology as useful and easy to use (Zhu et al., 2024). TAM (Davis, 1989) helps explain why students adopt a technology, but adoption alone does not tell us whether meaningful learning has taken place. A student may use AI often without developing the judgement needed to evaluate its answers or apply them appropriately in a professional setting. For this reason, recent tourism and hospitality research has increasingly emphasized AI literacy, critical evaluation, and thoughtful curriculum integration rather than simple exposure to new tools (Başer et al., 2025; Fathy et al., 2025). This distinction is central to the present study, which focuses on AI use that supports vocational understanding and job-related practice.

These changes also matter for employability. As technology alters tourism jobs, graduates are likely to need a mix of digital, interpersonal, operational, and problem-solving skills. Recent work on hospitality graduates and tourism employability therefore treats technological capability as part of a broader set of competencies needed to adapt to changing workplaces (El Hajal, 2025; Ibourk & El Alami, 2026). Evidence from Albania is beginning to reflect the same concerns. Peqini, Toma, and Borici (2025), studying Tourism and Computer Science students at the University of Shkodra "Luigj Gurakuqi," examined perceptions of AI alongside labour-market readiness and career orientation. Their work draws attention to the place of AI-related competencies in

students' preparation for an increasingly digital labour market. Although their sample consists of university rather than vocational students, the study provides useful local evidence that AI is already becoming part of the conversation about employability and career preparation in Albania.

The wider Western Balkan context adds another layer to this discussion. Borici, Peqini, and Bashi (2025) review recent work on the economic, ethical, and regulatory implications of AI across the region, with particular attention to Albania. Their review stresses that technological progress needs to be accompanied by investment in workforce education, digital skills, infrastructure, and appropriate institutional frameworks. For vocational education, this matters because students are preparing to enter workplaces that are themselves changing through digitalization and AI.

Changes in the nature of work also underline the importance of preparing people to understand AI rather than simply use it. Borici and Xhaferi (2025), writing about algorithmic management and digital labour in Southeastern Europe, discuss how AI can influence task allocation, performance evaluation, surveillance, and compensation. Their analysis highlights the need for digital literacy and AI-related training, while also drawing attention to transparency, worker autonomy, and inequality. Although their focus is not vocational tourism education, the study helps place the present research within a broader regional shift toward technology-mediated work.

Internship and Work-Based Learning as Drivers of Job Readiness

Even as AI becomes more prominent, practical workplace experience remains one of the defining features of vocational education. Internships and other forms of work-based learning give students the chance to apply what they have learned, deal with real problems, develop professional routines, and interact with supervisors, colleagues, and customers. In tourism and hospitality, where service and practical performance are central, these experiences are difficult to reproduce fully in a classroom.

Kolb's (1984) Experiential Learning Theory helps explain why these experiences matter. Learning develops through a cycle of concrete experience, reflection, conceptual understanding, and active experimentation. A good internship can bring these stages together: students perform real tasks, receive feedback, reflect on what happened, and adjust their behaviour. However, simply having an internship is not enough. What students are asked to do, the guidance they receive, and the relevance of the placement to their field can determine how much they actually learn.

This distinction is supported by previous research. Pianda et al. (2024), in a bibliometric and systematic review, identify internship experience as an important contributor to vocational students' employability. In hospitality education, Alharethi et al. (2025) likewise report a positive relationship between internship experience and work readiness. Together, these studies suggest that the quality of workplace learning deserves attention in its own right rather than being reduced to a simple yes-or-no measure of internship participation.

Evidence from Northern Albania points in the same direction. Borici, Selimi, and Shkambi (2026), studying 96 students, examined work experience, work-study compatibility, and institutional integration as predictors of competency development. Work experience emerged as the strongest positive predictor, followed by institutional integration, while work-study compatibility was not significant once the other factors were considered. The finding is relevant here because it suggests that direct engagement with work and structured institutional support may matter more for competency development than perceived alignment alone.

Taken together, the international and Albanian evidence suggests that work-based learning is most useful when it is a meaningful educational experience: students need relevant tasks, opportunities to practise, mentoring, and institutional support (Alharethi et al., 2025; Borici et al., 2026; Pianda et al., 2024).

Seen in this way, AI-supported learning and work-based learning do not need to compete for space in vocational education. They can serve different but complementary purposes. AI can help students explore ideas, obtain explanations, and practise digitally, while internships expose them to the pace, expectations, relationships, and practical demands of real work. The empirical question is whether each makes an independent contribution to how ready students feel for employment.

Perceived Job Readiness, Employability, and Self-Efficacy

In this study, job readiness refers to students' own perceptions of how prepared they are to enter employment; it is not an objective measure of workplace performance. The construct captures confidence in finding a job, beliefs about possessing labour-market-relevant skills, the perceived value of schooling for employment, expectations that employers will value those skills, and an overall sense of preparedness for work.

This emphasis on perceived readiness connects closely with self-efficacy. Bandura (1997) describes self-efficacy as people's beliefs about their ability to organize and carry out the actions needed to achieve particular outcomes. Such beliefs can grow through successful experience, observation, feedback, and social interaction. Social Cognitive Career Theory (SCCT) extends this idea to educational and career choices by linking learning experiences with career self-efficacy and outcome expectations (Lent et al., 1994). Lent et al. (1999) further show how this perspective can help explain the transition from school to work. Both AI-supported learning and internships may strengthen perceived readiness through these pathways. AI can give students repeated opportunities to solve problems and receive feedback, while internships offer direct experience of performing tasks in authentic workplaces. Still, confidence should not be confused with demonstrated employability. Students may feel ready without yet possessing all skill employers expect, and the reverse may also be true. This distinction is particularly important in research on generative AI, where ease of access to answers can increase confidence without necessarily producing equivalent gains in professional judgement.

Institutional Context and the AI-Readiness Relationship

AI does not enter every school in the same way. Access to technology, teacher confidence, curriculum design, industry links, pedagogical practice, and local norms around AI use can all shape students' experiences. TAM suggests that perceived usefulness and ease of use influence adoption (Davis, 1989), but the educational value of adoption is also likely to depend on the environment in which it occurs. This means that two schools could report similar average levels of job readiness while arriving there through different pathways. In one school, structured use of AI may be closely connected with students' confidence; in another, teacher support, workplace partnerships, or other learning experiences may matter more. This possibility motivates the study's comparison of average readiness as well as its test of the AI × School interaction.

Research Gap and Hypotheses

Taken together, the literature leaves a clear opening for the present study. Research on AI-supported learning is expanding quickly, tourism and hospitality education is paying greater attention to digital competencies, and a substantial body of work already demonstrates the value of internships and work-based learning. Yet these strands are not often examined together. Much of the recent AI literature also concentrates on higher education, leaving secondary-level vocational students comparatively underrepresented. Albanian and regional research has started to address pieces of this picture. Peqini et al. (2025) connect AI perceptions with labour-market readiness and career orientation among university students. Borici, Peqini, and Bashi (2025) place AI within the wider Western Balkan discussion of skills, infrastructure, ethics, and regulation. Borici and Xhaferi (2025) examine the changing nature of AI-mediated work, while Borici et al. (2026) show the importance of work experience and institutional integration for competency development. What remains largely unexplored is how AI-supported learning and work-based learning operate together in shaping perceived job readiness among secondary-level vocational tourism students. The present study brings these strands together. It treats AI-supported learning and internship quality as complementary learning experiences, examines their relationships with perceived job readiness in the same model, and asks whether the AI-readiness relationship differs across two vocational-school contexts.

H1. AI-supported learning use positively predicts vocational students' perceived job readiness.

H2. Internship/work-based learning quality positively predicts vocational students' perceived job readiness.

H3. Perceived job readiness differs between "Hamdi Bushati" and "Kolin Gjoka" students.

H4. The relationship between AI-supported learning and perceived job readiness differs by school.

Research methodology

Research Design, Sample, and Data Status

This study uses a quantitative, cross-sectional comparative design. The analytic dataset contains 316 observations, divided equally between "Hamdi Bushati" Vocational High School in Shkodër (n = 158) and "Kolin Gjoka" Vocational High School in Lezhë (n = 158). The model examines perceived job readiness as the dependent variable, AI-supported learning and internship/work-based learning quality as the principal predictors, school as a contextual variable, and an AI × School interaction. Gender, age, grade, and prior work/internship experience are included as controls. An important data-status qualification must accompany the current analysis. Participants were recruited from all the classes except the first year, because the first year students are not registered in vocational practices. The Hamdi Bushati subsample contains real responses for demographics and job-readiness items from the original school survey; the AI-supported learning and internship-quality items were not included in that original instrument and were therefore modelled synthetically for pilot-testing the analysis. The current Kolin Gjoka subsample is also simulated pilot data. Accordingly, the statistical results in this manuscript demonstrate the planned analytical structure and should not be interpreted as final empirical evidence from both schools. Further empirical validation would require administering the finalised questionnaire at both institutions and conducting the corresponding analyses using the primary data collected.

Respondent Profile

Table 1 summarizes the characteristics of the analytic sample. The sample is predominantly male (76.3%). Most respondents are aged 17–18 years (60.8%), and 80.4% report prior work or internship experience.

Table 1. Profile of respondents

Variable	n	%
School: Hamdi Bushati (Shkodër)	158	50.0
School: Kolin Gjoka (Lezhë)	158	50.0
Gender: Male	241	76.3
Gender: Female	75	23.7
Age: 15–16	47	14.9
Age: 17–18	192	60.8
Age: 19+	77	24.4
Grade: Class 11	134	42.4
Grade: Class 12	60	19.0
Grade: Class 13	122	38.6
Prior work/internship experience: Yes	254	80.4
Prior work/internship experience: No	62	19.6

Measures

AI-supported learning use was measured using three five-point Likert items (1 = strongly disagree; 5 = strongly agree): use of AI tools/apps to learn vocational content, AI tools helping students understand difficult concepts, and use of AI to practise job-related scenarios. Internship/work-based learning quality was measured with three five-point items addressing mentor guidance, relevance of tasks to the student's professional profile, and real opportunities to develop practical skills. Perceived job readiness was measured with five five-point items covering confidence in finding a job, possession of skills required by the market, perceived labour-market advantage from schooling, belief that employers would value the student's skills, and overall

preparedness for work. School was dummy-coded with Kolin Gjoka = 1. The interaction term AI × School was included to test whether the relationship between AI-supported learning and perceived job readiness differed by institution. Gender, age, grade, and prior work/internship experience were included as controls.

Analytical Strategy

The analysis proceeded in four stages. First, the sample profile was summarized descriptively. Second, exploratory factor analysis and Cronbach's alpha were used to assess the structure and internal consistency of the three multi-item constructs. Third, simple regressions tested H1 and H2, and an independent-samples comparison tested H3. Fourth, a hierarchical multiple regression including AI-supported learning, internship quality, school, the AI × School interaction, and the control variables tested H4. School-specific regressions were used to interpret the interaction.

Results

Factor Analysis and Reliability

Exploratory factor analysis using principal-axis extraction with varimax rotation indicated that the eleven Likert items were suitable for factor analysis. The Kaiser-Meyer-Olkin measure was .873 and Bartlett's test of sphericity was significant, χ^2 = 1815.2, p < .001. A three-factor structure corresponding to AI-supported learning, internship quality, and perceived job readiness emerged, with all reported primary loadings above .70. Internal consistency was good to excellent: α = .834 for AI-supported learning, α = .851 for internship quality, and α = .902 for perceived job readiness.

Table 2: Measurement-scale summary

Construct	Item	Loading	Cronbach's α
AI-supported learning	AI1. Use AI tools/apps to learn vocational content	.745	.834
	AI2. AI tools help understand difficult concepts	.774	
	AI3. Use AI to practise job-related scenarios	.725	
Internship/work-based learning quality	IQ1. Mentor guided me well	.782	.851
	IQ2. Tasks relevant to my profile	.771	
	IQ3. Real opportunities to develop skills	.797	
Perceived job readiness	WR1. Confident I can find a job	.799	.902
	WR2. Have the skills the market requires	.800	
	WR3. School gives me an advantage	.720	
	WR4. Employers would value my skills	.753	
	WR5. I feel prepared for work	.779	

Hypothesis Tests

To test H1, perceived job readiness was regressed on AI-supported learning. The relationship was positive and statistically significant, F(1, 314) = 73.611, p < .001, b = 0.359, r = .436, R^2 = .190. Thus, greater AI-supported learning use was associated with higher perceived job readiness, supporting H1. To test H2, perceived job readiness was regressed on internship/work-based learning quality. The relationship was also positive and significant, F(1, 314) = 41.243, p < .001, b = 0.277, r = .341, R^2 = .116. H2 was therefore supported. For H3, mean perceived job readiness was compared between schools. Hamdi Bushati students reported M = 4.02 (SD = 0.86), while Kolin Gjoka students reported M = 4.10 (SD = 0.61). The difference was not statistically significant (p = .390). Accordingly, H3 was not supported. Because the current dataset is partly simulated, this school comparison remains provisional and must be re-estimated using complete primary data. To test H4, a hierarchical multiple regression was estimated with perceived job readiness as the outcome and AI-supported learning, internship quality, school, AI × School, gender, age, grade, and prior work experience as predictors. The full model was significant, F(8, 307) = 30.81, p < .001, R^2 = .445, adjusted R^2 = .431. AI-supported learning (b = 0.645, p < .001) and internship quality (b = 0.331, p < .001) remained positive predictors. The AI × School interaction was negative and significant (b = -0.273, p = .002), supporting H4. School-specific regressions clarified the interaction. The AI-readiness slope was stronger at Hamdi Bushati (b = 0.745, R^2 = .341, p < .001) than at Kolin Gjoka (b = 0.356, R^2 = .270, p < .001). None of the demographic or experience controls reached statistical significance.

Table 3: Hierarchical regression predicting perceived job readiness

Predictor	b	SE	t	p
Intercept	-0.218	0.507	-0.43	.668
AI-supported learning	0.645	0.068	9.45	< .001
Internship quality	0.331	0.041	8.05	< .001
School (Kolin Gjoka = 1)	1.755	0.305	5.75	< .001
AI × School	-0.273	0.085	-3.20	.002
Gender (male = 1)	0.083	0.076	1.09	.276
Age (ordinal)	-0.034	0.053	-0.64	.526
Grade (ordinal)	0.040	0.037	1.09	.279
Work experience (yes = 1)	-0.020	0.080	-0.25	.800

Discussion

The findings offer provisional support for three of the four hypotheses and, more importantly, suggest that digital and experiential learning may both matter for how vocational students see their transition into work. AI-supported learning was positively associated with perceived job readiness. This is consistent with research showing that AI tools can provide individualized explanations, feedback, and learning support (Labadze et al., 2023), as well as tourism-education evidence linking ChatGPT-supported learning with educational outcomes (Dalgıç et al., 2024). It also fits with TAM: students

are more likely to engage with technologies they consider useful and manageable (Davis, 1989; Zhu et al., 2024).

Internship quality also showed an independent positive relationship with readiness. This result is consistent with Experiential Learning Theory (Kolb, 1984) and with studies showing that authentic workplace experience, relevant tasks, and mentoring can strengthen employability and work readiness (Alharethi et al., 2025; Pianda et al., 2024). It also echoes evidence from Northern Albania, where work experience and institutional integration were important predictors of competency development (Borici et al., 2026). The fact that AI-supported learning and internship quality remain relevant in the same model supports a complementary view: technology may enrich learning, but it does not remove the need for meaningful contact with real workplaces. The absence of a significant difference in average readiness between the two schools is also informative. Students can arrive at similar levels of confidence through different educational experiences. The significant AI × School interaction points to this possibility: the association between AI-supported learning and readiness was stronger at Hamdi Bushati than at Kolin Gjoka. Rather than suggesting that one school is simply 'better,' the result raises questions about what may differ behind the averages—such as teacher guidance, access to technology, curriculum integration, or the way students are encouraged to use AI. Those explanations remain hypotheses rather than findings from the present data. The study does not directly measure teacher AI competence, digital infrastructure, school-industry partnerships, or institutional policies on AI. Future research should measure these features explicitly before attributing the interaction to any particular school characteristic.

Limitations and future research

The most important limitation concerns the status of the data. The AI-supported learning and internship-quality variables for the Hamdi Bushati subsample were synthetically modelled because they were absent from the original survey, and the current Kolin Gjoka subsample is simulated pilot data. The reported coefficients, factor structure, and hypothesis tests therefore represent a rehearsal of the planned analytical pipeline rather than final evidence from fully observed primary data. This limitation must be resolved before the study is presented as a completed empirical comparison. After full primary data collection, the same measurement and regression procedures should be repeated and the stability of the three-factor structure should be reassessed. If sample size permits, confirmatory factor analysis could provide a stronger test of construct validity. Additional limitations will remain even with complete primary data. The cross-sectional design cannot establish causality; self-reported measures are vulnerable to common-method bias and social-desirability effects; and perceived readiness is not equivalent to employment outcomes. Future research should follow graduates longitudinally, incorporate employer assessments or objective competency measures, distinguish among specific AI applications, and examine institutional moderators such as teacher AI competence, infrastructure, and curriculum integration.

Conclusion

This study brings together three issues that are becoming increasingly important in vocational tourism education: AI-supported learning, the quality of workplace experience, and students' confidence about entering employment. Within the current pilot analytic dataset, both AI-supported learning and internship quality are positively associated with perceived job readiness. Average readiness is similar across the two schools, but the strength of the AI-readiness relationship differs between them, suggesting that institutional context may matter. The practical message is not that AI should replace traditional vocational training. The more useful interpretation is that the two forms of learning can complement one another. AI can give students additional opportunities to ask questions, practise, and develop digital confidence; high-quality internships place them in real situations where they must apply knowledge, communicate, adapt, and learn from others. Preparing students for a changing tourism labour market is likely to require both. At the same time, the present conclusions must remain provisional. The finalized questionnaire needs to be administered in full at both schools, and the analyses should then be repeated using complete primary data. That step is essential before the reported relationships are used to draw firm institutional or policy conclusions.

Acknowledgements

The authors express their gratitude to everyone who participated in this study.

Declaration

ClaudeAI was used during the writing process, to ensure good quality of grammar and translation.

References

[1] Alharethi, T., Awan, M. I., & Saleem, M. S. (2025). Investigating the impact of internship experience on work readiness among hospitality program students: The moderating role of student mindfulness. Frontiers in Education, 10, Article 1497721. https://doi.org/10.3389/feduc.2025.1497721

[2] Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.

[3] Borici, A., & Xhaferi, S. (2025). Algorithmic management and gender inequalities in digital work: Challenges and opportunities for a fairer economy in Southeastern Europe. Paper presented at the International Conference "Platform Work in South East Europe – Intersectionality, Feminist Futures, and the Gold Economy," University of Shkodra "Luigj Gurakuqi," Shkodër, Albania.

[4] Borici, A., Peqini, V., & Bashi, E. (2025). Artificial intelligence in the Western Balkans: Economic impact, ethical challenges, and regulatory frameworks. Journal of Information Systems Engineering and Management.

[5] Başer, M. Y., Kozak, M., & Büyükbeşe, T. (2025). Integration of emerging technologies in tourism and hospitality curriculum: An international perspective. Journal of Hospitality, Leisure, Sport & Tourism Education, 36, Article 100546. https://doi.org/10.1016/j.jhlste.2025.100546

[6] Becker, G. S. (1993). Human capital: A theoretical and empirical analysis, with special reference to education (3rd ed.). University of Chicago Press.

[7] Borici, A., Selimi, A., & Shkambi, S. (2026). Understanding competency development: The role of work experience, work-study compatibility, and institutional integration in Northern Albania. International Scientific Journal Monte, 12(1).

[8] Dalgıç, A., Yaşar, E., & Demir, M. (2024). ChatGPT and learning outcomes in tourism education: The role of digital literacy and individualized learning. Journal of Hospitality, Leisure, Sport & Tourism Education, 34, Article 100481. https://doi.org/10.1016/j.jhlste.2024.100481

[9] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

[10] El Hajal, G. (2025). Transitions into the futures: AI disruption and resilience in hospitality graduates. Research in Hospitality Management, 15(1), 20–22. https://doi.org/10.1080/22243534.2024.2438709

[11] Fathy, E. A., Zidan, H. A. K. Y., El Sayed, A. H., & Fouad, A. M. (2025). Integrating ChatGPT in tourism and hospitality education: A systematic and bibliometric analysis of research trends, applications, and implications. Tourism and Hospitality Research. Advance online publication. https://doi.org/10.1177/14673584251357510

[12] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Pearson.

[13] Ibourk, A., & El Alami, S. (2026). Graduate employability in tourism: Recruitment practices, skills, and the role of digitalisation and AI in Marrakech. Societies, 16(2), Article 58. https://doi.org/10.3390/soc16020058

[14] Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice-Hall.

[15] Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. Journal of Vocational Behavior, 45(1), 79–122. https://doi.org/10.1006/j.jvbe.1994.1027

[16] Lent, R. W., Hackett, G., & Brown, S. D. (1999). A social cognitive view of school-to-work transition. The Career Development Quarterly, 47(4), 297–311. https://doi.org/10.1002/j.2161-0045.1999.tb00739.

[17] Peqini, V., Toma, V., & Borici, A. (2025). The role of artificial intelligence in transforming the tourism labor market: Evidence from student perspectives and career aspirations at “Luigj Gurakuqi” University. Paper presented at the Shkodra International Economic Conference (SHKODRA IEC2025), Shkodër, Albania.

[18] Pianda, D., Hilmiana, Widianto, S., & Sartika, D. (2024). The impact of internship experience on the employability of vocational students: A bibliometric and systematic review. Cogent Business & Management, 11(1), Article 2386465. https://doi.org/10.1080/23311975.2024.2386465

[19] Prasetya, F., Fortuna, A., Samala, A. D., Latifa, D. K., Andriani, W., Gusti, U. A., Raihan, M., Criollo-C, S., Kaya, D., & Cabanillas García, J. L. (2025). Harnessing artificial intelligence to revolutionize vocational education: Emerging trends, challenges, and contributions to SDGs 2030. Social Sciences & Humanities Open, 11, Article 101401. https://doi.org/10.1016/j.ssaho.2025.101401

[20] Zhu, C. Z. G., Hall, C. M., Fong, L. H. N., Lin, F., & Naderi Koupaei, S. (2024). Examining the effects of ChatGPT on tourism and hospitality student responses through integrating technology acceptance model. International Journal of Tourism Research, 26(4), Article e2727. https://doi.org/10.1002/jtr.2727

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Comparing AI Interaction Logs and Classroom Analytics for Detecting Conceptual Difficulties in Statistics

Fulya Kula
(e-mail: f.kula@utwente.nl)
Applied Mathematics Department, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, The Netherlands

Article DOI: https://doi.org/10.68286/samyrad2026.370

ABSTRACT
Generative AI tools are increasingly used by students of statistics, but it is unclear whether the questions students ask them reveal the conceptual difficulties that classroom assessment already detects. Classroom polling surfaces a misconception only when a question happens to probe it, so a source available throughout a course could in principle complement it. This exploratory case study asks which topics generate the most AI interactions, which show the highest classroom-response error rates, and how far the two sources agree. We compared 728 AI-log items (162 conversations, 58 students) from a curriculum-bounded platform with 18 Wooclap items from three tutorials in a second-year statistics course at a Dutch university, with final-exam results as a third reference point, using a shared six-topic coding scheme. AI-log volume did not align with Wooclap error rates: among the three topics present in both sources, Type I/II errors and statistical power had the fewest AI conversations (4) but the highest error rate (57.4%), and point estimation, with the highest exam error rate among comparable items, had no Wooclap item. The comparison is descriptive and limited to one course. It suggests that AI-log volume shows where attention concentrates rather than where students struggle, and that the two sources are best read together.

KEYWORDS generative AI, statistics education, interaction logs, classroom response systems, conceptual difficulties

How to cite: Kula, F. (2026). Comparing AI interaction logs and classroom analytics for detecting conceptual difficulties in statistics. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.370

Introduction

Generative AI tools are increasingly available to students studying statistics, but whether the questions students choose to ask such a tool reveal anything about their conceptual understanding, rather than simply serving as a convenient homework aid, is not yet well established. Statistics education research has documented a range of persistent misconceptions in how students reason about inference [1]. One well-studied example concerns confidence intervals: Hoekstra et al. [2] show that misinterpreting a confidence interval as a probability statement about the population parameter persists even among trained researchers, not only introductory students. Findings of this kind give a new data source something concrete to test itself against: if the topics students raise with an AI tool point toward misconceptions already documented in this literature, that convergence is reassuring; if the two diverge, the divergence itself says something about what each method can and cannot detect.

The three tutorial sessions examined in this study were structured around think-pair-share, a cooperative-learning

technique in which students first consider a question individually, then discuss it with a partner, before a wider class discussion [3], [4]. The technique has been applied specifically in statistics instruction, where it forms part of a broader set of active-learning methods shown to improve engagement over traditional lecturing [5]. More broadly, across disciplines, recent work has also questioned how essential the final "share" step is to the technique's benefit, and has called for research into alternative ways of closing a think-pair cycle [6]. The present course's tutorials adapted the format accordingly: alongside discussion with a partner and the class, students could also consult the AI platform directly, a variant we refer to as think-pair-and-AI-share. Within the same sessions, Wooclap, an interactive classroom response system, was used to poll students on the same material: students answer a question on their own device after scanning a QR code, and the instructor then displays the class-wide response distribution and the correct answer, which anchors discussion during the share phase of each tutorial. This built on a longer tradition in which structured peer discussion measurably improves performance on in-class concept questions [7], and on more recent evidence that Wooclap specifically supports engagement and achievement in comparable higher-education settings [8], [9].

What such classroom-based methods cannot easily capture is a misconception a student never has the opportunity to reveal, because no polling question that particular session happened to probe it. An AI tool that a student can query at any point, on any topic, in principle fills exactly that gap. Whether it does so in practice, generating conceptually meaningful questions rather than off-task use, is less established. A semester-long study of a physics course chatbot showed that recovering interpretable patterns of student reasoning from raw conversational logs is possible, but that doing so requires deliberate methodological work [10]. The AI log examined here spans a comparably extended period, and this study faces the same underlying challenge: showing that the signal recovered from student-authored prompts reflects genuine conceptual engagement, not an artifact of how the data happened to be logged.

This paper compares AI interaction logs and Wooclap classroom response data collected within the same second-year undergraduate statistics course, both generated in parallel within the same three think-pair-and-AI-share tutorial sessions, Tutorials I, II, and III. It addresses three questions: which statistical topics generate the most AI interactions, which topics show the highest rates of incorrect Wooclap responses, and to what extent the two sources agree on where student difficulty concentrates. Final exam records, including two items whose scenarios were numerically identical to specific Wooclap questions, provide a third, independent comparison point. Reaching this comparison required substantial preliminary data-cleaning work, documented in the Methodology: the raw AI-log export required structural reconstruction before it could be parsed at all, contained interactions from a non-student test account, and included conversations scoped to courses other than the one under study. These steps are reported in detail because similar problems are likely to come up in future studies using AI interaction logs. The paper first reviews what is known about documented misconceptions in statistical inference, classroom response systems, and AI interaction logs and generative AI tools in statistics education, before describing the course context and data sources, reporting results for each of the three questions above, and discussing what the alignment and divergence between data sources suggests, together with the limitations of a design in which AI-tool use was voluntary rather than assigned.

Literature review

Documented Misconceptions in Statistics Education

Statistics education research has documented a range of persistent misconceptions in how students reason about data and inference [1]. One well-studied example concerns confidence intervals: Hoekstra et al. [2] show that misinterpreting a confidence interval as a probability statement about the population parameter persists even among trained researchers, not only introductory students. A second example concerns point estimation and sampling-distribution reasoning. A review of research on statistical sampling instruction finds that students, with or without instructional simulations, often fail to grasp that a sampling distribution describes the spread of a statistic across many samples, not a single sample, a distinction made harder because both are commonly visualized the same way, as a histogram [11]. A separate study notes that misconceptions research of this kind has relied on a fairly narrow set of methods, mainly concept inventories and a smaller number of think-aloud interview studies, and argues that other methods for surfacing student reasoning remain underused [12]. This gap is directly relevant here: an AI interaction log is one such underused method, and this study examines whether it can surface the same kinds of documented difficulty.

Classroom Response Systems and Think-Pair-Share in Statistics Instruction

The tutorials examined in this study were built around think-pair-share, a cooperative-learning technique introduced by Lyman [3], [4], in which students consider a question individually, discuss it with a partner, and then take part in wider class discussion. The technique has been applied specifically in statistics instruction as part of a broader set of active-learning methods shown to improve engagement over traditional lecturing [5]. More recent work has questioned how essential the final "share" step is to the technique's benefit and has called for research into alternatives [6]. Smith et al. [7] provide foundational evidence that structured peer discussion, paired with instructor explanation, measurably improves performance on in-class concept questions, establishing the pedagogical logic behind polling-based tutorials of the kind used here. Wooclap specifically has been shown to support engagement and achievement in comparable higher-education settings, including economics [8] and chemical engineering [9] courses. This literature establishes that classroom response data reliably surfaces topic-level error rates in real time, but it does not typically compare those error rates against an independent, self-directed data source such as AI tool use, which is the comparison this study attempts.

AI Interaction Logs and Generative AI Tools in Statistics Education

Whether AI interaction logs reveal genuine conceptual engagement, rather than mostly off-task or administrative use, is not yet well established. A semester-long study of a physics course chatbot showed that recovering interpretable patterns of student reasoning from raw conversational logs is possible but requires deliberate methodological work, in that case an unsupervised clustering pipeline [10]. Within statistics specifically, GenAI's role has so far been discussed mainly through student self-report and course-design studies rather than through analysis of behavioral log data. A survey-based study spanning a full semester found that students rate ChatGPT favorably across several dimensions of academic support, with writing assistance and exam preparation emerging as the strongest predictors of perceived overall

effectiveness [13]. A study of project-based statistics courses found that students report favorable views of course projects generally, and that many had independently incorporated ChatGPT into their coursework, describing perceived benefits, though this use was student-initiated rather than part of the course design [14].

Ellis and Slade [15] outline considerations for using ChatGPT to enhance statistics and data science instruction. Working with actual student writing samples rather than self-report, one study analyzes a corpus of over 1,600 undergraduate data-analysis reports from 2021 to 2025 and finds that students' writing style has shifted measurably since large language models became widely available, becoming both more similar to LLM-generated text and, at the same time, more similar to expert statistical writing [16]. A related study shows that LLM-generated and human-generated statistical writing nonetheless remain stylistically distinguishable, informing how instructors might guide LLM use in writing-based assessments [17]. None of this work, to our knowledge, has used the content of AI interaction logs themselves as a diagnostic signal for topic-level conceptual difficulty in a statistics course, cross-checked against an independent classroom data source. That is the gap this study addresses.

Methodology

Context and Participants

This study was conducted in a second-year undergraduate statistics course at a university in the Netherlands. Around 400 students were registered, and 265 completed the final exam. The course used two tools together: Alexandria [18] (since renamed Odea), a curriculum-bounded generative AI platform restricted to course materials, and Wooclap.

Three tutorials ran as think-pair-and-AI-share sessions, combining both tools. Tutorial I came after descriptive statistics had been introduced and covered point estimation and confidence intervals, with 45 Wooclap participants. Building on this, Tutorial II covered basic concepts of hypothesis testing, with 27 participants. Tutorial III followed and covered two sample hypothesis tests, with 16 participants. Later course content, including normality assessment, categorical data analysis, and linear regression, came after these three tutorials and was not part of any think-pair-and-AI-share session.

Ethical approval was granted by the university's faculty ethics committee. Students opted in at first login to Alexandria and at the start of each Wooclap session.

Data Sources

AI interaction logs.

Alexandria exported timestamped question-answer pairs, item type (free-text query, or a request to explain highlighted text or an image), conversation grouping, and course scope. Three cleaning steps came before analysis. One account showed a usage pattern inconsistent with genuine student use: the same prompt repeated 138 times, starting before the course's first lecture. This account was removed. The exports also included interactions from at least three other modules, so records were restricted to those scoped to this course. A further six conversations referenced a different course or module by name in their actual content, not just their scope tag, and were excluded too. The resulting dataset held 728 interaction items across 162 conversations from 58 students, spanning the full course period.

Wooclap responses.

These data came from the same three think-pair-and-AI-share tutorials described above, Tutorials I, II, and III, producing 18 usable items once one consent or demographic item per session was set aside: 7 multiple-choice items on confidence intervals (Tutorial I), 8 on hypothesis-testing fundamentals, Type I and II errors, and statistical power (Tutorial II), and 3 open-response items on two-sample hypothesis testing (Tutorial III). Response counts per item ranged from 12 to 45, depending on attendance. The three open-response items were not machine-graded and are treated separately from the rest. Table 1 shows two representative items, one from Tutorial I and one from Tutorial III, to illustrate the item format; these two also reappear later as the items whose scenarios were reused verbatim on the final exam.

Exam records.

Final exam grades for all 265 students covered 11 questions and 43 points total. The exam was written and closed book: no electronic devices or internet access were permitted, and students were limited to a simple, non-graphing calculator. The exam had four versions with identical content and point structure; only the ordering of multiple-choice options and the wording of distractors differed, so scores are directly comparable across versions. Two of the eleven questions matched Wooclap content exactly, reusing the same scenario, one on confidence intervals from Tutorial I, and one on two-sample procedures from Tutorial III. These two gave an exact-match comparison point in addition to the broader topic-level comparison.

Table 1. Example WOOCLAP items

Session	Topic	Item text
Tutorial I	Confidence intervals	A tech company is developing a new wearable device. The device's older model had a well-documented population standard deviation σ = 2.0 hours for battery lifetime. However, after updates to the hardware and firmware, engineers suspect this value may no longer apply. A tester measures the battery lifetime of 18 prototypes and obtains: mean = 204.3 hours, sd = 2.1 hours. Which of the following correctly expresses the 95% confidence interval for the true mean battery lifetime?
Tutorial III	Two-sample procedures	A cybersecurity team is analyzing phishing-click rates in two departments of a company: Department A: 3 of 40 employees clicked the phishing link. Department B: 12 of 200 employees clicked the phishing link. You want to test whether the phishing-click rate in Department A differs from that in Department B, at significance level α = 0.05.

Topic Coding

All three sources were coded against a shared set of topics, built from what the three tutorials actually covered: descriptive statistics, point estimation, confidence intervals, hypothesis-testing fundamentals, Type I and II errors and statistical power, and two-sample procedures. Later syllabus content was left out of the coding scheme and out of cross-source comparison, since neither tool could have shaped performance on material students had not yet seen.

AI-log items were coded at the conversation level, not the individual turn. A conversation's topic came from the first turn

with a topic-specific keyword match, and every later turn in that conversation inherited the same code. Many follow-up turns, things like "help me answer part a," carry no topical keyword of their own but still belong to the conversation that opened with one. Coding turn by turn left about half of all items with no topic match; coding conversation by conversation cut that to under a fifth. At the conversation level, however, 77 of the 162 conversations (47.5%) matched no topic keyword at any point and were coded "OTHER". A manual read suggests most of these were not conceptual (see Results), but some may have used vocabulary the keyword scheme did not capture. Keyword lists were built from the vocabulary specific to each of the six tutorial topics and are not included here in full; the approach can be shared on request.

Wooclap items were coded individually, since each one was a single self-contained scenario. Exam items were mapped to topics by reading their content directly rather than by keyword matching, since there were only 11 of them and several combined more than one topic. Three questions mixed topics under one score and were dropped from topic-level comparison: one combining nine true/false statements across descriptive statistics, confidence intervals, and Type I/II errors; one combining confidence-interval construction with hypothesis formulation and a statistical power calculation; and one combining a variance-comparison test with a sample-size calculation. Three further questions covered later course content (normality assessment, categorical data analysis, and linear regression) and fell outside the coding scheme. The remaining five questions each addressed a single topic, point estimation, hypothesis testing, two-sample comparison (twice, across two different questions), and confidence intervals, and were kept for the comparison. All topic coding, the keyword matching applied to the AI log and the manual mapping of exam questions to topics, was carried out by the author; no second coder independently verified these assignments.

Analytic Procedure

Three questions were addressed descriptively. Which topics generate the most AI interactions was answered by ranking topics on conversation count and item count, with conversation count treated as primary, since item counts are sensitive to a handful of long conversations (median 2 items per conversation, maximum 57). Which topics show the highest Wooclap error rates was answered by aggregating respondent-level correctness within each topic across the 15 graded multiple-choice items; the 3 open-response items were read qualitatively instead. How far the two sources align was addressed descriptively too: topics were ranked by AI-log conversation volume next to their Wooclap error rate and, where a matching single-topic exam item existed, their exam error rate. No correlation coefficient was computed, given only six topic categories to work with. The two exam items with an identical Wooclap scenario were treated as a separate, higher-confidence comparison rather than folded into the broader topic-level one.

Limitations of the Data

Three things about the data are worth flagging before the Results. First, the AI log covers the whole course, not just the three tutorials; only 89 of 728 items fall within the three tutorial moments. Most AI-log activity is independent study outside these sessions, not in-session use directly comparable to a Wooclap snapshot. Second, course-scope tags in Alexandria are not a perfect guide to content, since a few tagged conversations turned out to reference other coursework by name. The exclusion described above catches the clear cases, but subtler off-topic content without an explicit course name could still remain. Third, because a conversation's topic is set by its first keyword match and carried forward, a conversation that opens on one topic and later shifts to another is coded entirely under the opening topic, so topic drift within a conversation is not captured.

Results

Which statistical topics generate the most AI interactions?

Table 2 breaks down the 728 retained AI-log items (162 conversations, 58 students) by topic; the percentage column expresses each topic's share of the 85 conversations that received a topic code, excluding the 77 conversations coded "OTHER".

Table 2. AI-log volume by topic

Topic	Conversations	Items	% of coded conversations
Descriptive statistics	33	231	38.8%
Point estimation	18	122	21.2%
Confidence intervals	15	79	17.6%
Hypothesis-testing fundamentals	12	121	14.1%
Type I/II errors and statistical power	4	35	4.7%
Two-sample procedures	3	12	3.5%
"OTHER"	77	128	—

Descriptive statistics drew the most AI-log activity, followed by point estimation, confidence intervals, and hypothesis-testing fundamentals. Type I/II errors and statistical power and two-sample procedures trailed well behind. A large share of conversations, 77 of 162, matched no topic keyword and were coded "OTHER". A manual read of this category showed it was predominantly course logistics, platform-navigation questions, and requests to explain a highlighted passage or image; a smaller share could not be ruled out as conceptual content using vocabulary the keyword scheme did not capture, though a precise breakdown was not retained. Item counts per conversation varied a lot too, from 1 to 57, with a median of 2. This spread means item counts overweight a handful of long conversations, so the conversation counts in Table II are the more reliable guide to where attention concentrated. Descriptive statistics and point estimation, the two earliest topics on the syllabus, together made up 60% of coded activity. Students seem to have leaned on the tool more heavily early in the course.

Which topics show the highest rates of incorrect Wooclap responses?

Error rates were high across all three topics with graded Wooclap items. Type I/II errors and statistical power had the highest error rate, 57.4% across 61 respondent-answers on 4 items. Hypothesis-testing fundamentals followed closely at 56.1% across 66 respondent-answers on 4 items. Confidence intervals came in at 50.4% across 119 respondent-answers on 7 items, including the example item shown in Table I. The three two-sample items were open-response and not machine-graded, so they sit outside this comparison; their content comes back later, via the matching exam item. Within confidence intervals, error rates ranged widely by item, from 12.5% on a margin-of-error formula question to 66.7% on a question asking students to interpret a confidence interval as

a probability statement about the population mean. The topic-level average hides this spread: the highest error rate in the topic belonged to the item on interpreting a confidence interval as a probability statement, although a single item cannot show that this one misconception drives the topic-level rate.

How far do the two sources agree?

Before comparing the three sources directly, it is worth repeating that AI-log volume reported here accumulates across the full course period, while the Wooclap and matching exam figures reflect three specific tutorial sessions; the comparison that follows therefore sets a cumulative, full-term measure against session-specific snapshots, not two measures taken over the same window. Table 3 lines up AI-log volume, Wooclap error rate, and exam error rate by topic. The exam column uses five exam questions that each addressed a single topic cleanly; three further exam questions mixed several topics under one score and are left out here, as explained in Methodology. Descriptive statistics, despite generating the highest AI-log volume of any topic, is not shown in Table 3: unlike point estimation, it has neither a matching Wooclap item nor a single-topic exam question, so no comparison point exists for it in either source, not because it was judged less relevant.

Table 3. AI-log volume, Wooclap error rate, and exam error rate

Topic	AI conversations	Wooclap error rate (n)	Exam error rate (n)
Point estimation	18	— (no matching Wooclap item)	70.4% (n = 265)
Confidence intervals	15	50.4% (n = 119)	38.4% (n = 265), exact scenario match
Hypothesis-testing fundamentals	12	56.1% (n = 66)	25.2% (n = 265)
Type I/II errors & statistical power	4	57.4% (n = 61)	—
Two-sample procedures	3	N/A (open response)	12.5% / 40.1% (n = 265 each), exact scenario match

Two things stand out. First, point estimation had the second-highest AI-log volume of any topic and the highest exam error rate of the four topics with exam data in Table III, but no Wooclap item across the three sessions touched it at all. The topic students engaged with most in the AI tool and struggled with most on the exam was invisible to the classroom-response data, simply because none of the three sessions happened to include a question on it. Second, two-sample procedures had the lowest AI-log volume of any topic (only 3 conversations, so this comparison should be read cautiously), yet its two matching exam items told very different stories: 12.5% error on one two-sample scenario, 40.1% error on the other, the same scenario also used in a Wooclap session. Low AI-tool engagement did not mean uniformly strong performance on that topic.

For the three topics present in both the AI log and Wooclap, confidence intervals, hypothesis-testing fundamentals, Type I and II errors and statistical power, more AI-log conversations did not mean a lower Wooclap error rate. If anything the opposite held: the topic with the fewest AI-log conversations, Type I and II errors and statistical power (4 conversations, the smallest of the three), had the highest Wooclap error rate of the three; given that small base, this specific comparison should be read cautiously. AI-log volume, on its own, does not tell you where the difficulty is.

Where an exam question used the exact same scenario as a Wooclap item, the exam error rate for the confidence-interval item was lower than the corresponding Wooclap figure. On the confidence-interval item shown in Table I, the exam error rate for the identical scenario (38.4%) was lower than the confidence-interval topic average in the tutorial (50.4%), though a topic average is not directly comparable to a single item. This fits with continued exposure between the tutorial and the exam, but this study cannot separate the AI tool's specific role in that improvement from ordinary revision or other exam preparation.

Discussion

Convergence and Divergence Between the Two Sources

The Introduction posed a simple test: if the topics students raised with the AI tool pointed toward misconceptions already documented in the statistics education literature, that would be reassuring; if the two diverged, the divergence itself would be informative. This was not framed as a fourth research question, it is a lens for interpreting the three results already reported, not a separate finding to report on its own. Under that lens, the answer is divergence. For confidence intervals, hypothesis-testing fundamentals, and Type I and II errors and statistical power, the only three topics present in both datasets, the ranking by AI-log conversation count ran opposite to the ranking by Wooclap error rate. The topic with the fewest AI-log conversations, Type I and II errors and statistical power, had the highest Wooclap error rate of the three.

Several explanations seem plausible, and the data collected here cannot distinguish between them. Students who struggled most with Type I and II error and statistical power reasoning may not have recognized their own difficulty clearly enough to ask about it. It is equally possible that attendance varied across the three tutorials, or that fewer students had the AI tool open during that particular exercise, so low volume on this topic may reflect opportunity rather than confidence. Alternatively, confidence-interval content, despite a lower classroom-response error rate, may have drawn more follow-up questions because its correct interpretation is subtle enough that a single classroom session leaves residual uncertainty; Hoekstra et al. [2] found that confidence-interval misinterpretation persists even among trained researchers, which fits this third explanation. A broader asymmetry sits behind all three: Wooclap participation was compulsory and tied to the three tutorial sessions, while AI-tool use was voluntary and open across the whole course, so the two sources were never sampling the same decision to engage in the first place, itself a plausible driver of the divergence, separate from topic difficulty. What the data do show is that AI-log volume and classroom-response error rate measure different things, volume reflects inquiry frequency, error rate reflects incorrect responses, and the one cannot be treated as an unproblematic stand-in for the other, a conclusion only visible because the Wooclap comparison existed at all.

Point Estimation

Point estimation generated the second-highest AI-log conversation count of any topic, and on the exam, the highest error rate of the five single-topic questions. The relevant exam question asked students to identify the value of a scaling constant that made a sample mean an unbiased estimator, and the value that minimized mean squared error (MSE), a

common measure of estimator accuracy that combines both bias and variability. No Wooclap item across the three sessions touched point estimation at all, since Wooclap items were written session by session to match each tutorial's planned content, and none of the three happened to include one on estimator bias or MSE.

This gap matches a pattern already documented in the statistics education literature. A review of research on statistical sampling instruction finds that students, with or without instructional simulations, often fail to grasp that a sampling distribution describes the spread of a statistic across many samples, not a single sample [11]. A separate study notes that student misconceptions in this area have historically been studied through a narrow set of methods, mainly concept inventories and a smaller number of think-aloud studies, leaving open questions about how such difficulties surface in less structured settings [12]. An AI interaction log is one such less structured setting. Here, point estimation drew substantial AI-log activity but had no Wooclap item, so a Wooclap-only design would have given no classroom-response signal on it. Whether that activity reflects the estimator-reasoning difficulties described in this literature cannot be determined from volume alone.

The two exact scenario matches between the exam and the Wooclap sessions add a further point of comparison. One was the confidence-interval item shown in Table I. The other was a two-sample scenario that appeared in both a Tutorial III Wooclap item and, later, on the exam. For the confidence-interval item the exam error rate was lower than the tutorial topic average, while for the two-sample item no graded Wooclap rate exists to compare with. This is consistent with continued exposure between the tutorial and the exam, but this study's design cannot separate the AI tool's specific role in that improvement from ordinary revision, lecture reinforcement, or studying with classmates.

Practical Lessons for Future Work

The dataset built for this study needed considerably more preprocessing than the raw export initially suggested. This belongs in the Discussion rather than the Methodology because the point here is not what this study did, that is already described in the Methodology, but what the experience implies for anyone attempting similar work in future: these are exactly the problems worth checking for before treating an AI-log export as analysis-ready. The raw export required reconstruction of the underlying file structure before any of it could be read. One account behaved nothing like a student, repeating a single prompt 138 times starting before the course had begun, and needed to be removed outright. Course-scope metadata missed content from at least three other modules and, on closer reading, a handful of conversations that named a different course directly in their text despite carrying the correct scope tag. Existing chatbot-log studies describe some preprocessing, typically isolating student-authored text from bot responses, but a broken export, unreliable scope tags, and a non-student account are problems that literature has not yet described.

Limitations

A few limitations qualify what can be claimed here. The two datasets were not collected under comparable conditions: Wooclap participation was compulsory and tied to three specific sessions, while AI-tool use was voluntary and spread across the whole course, with only 89 of 728 retained items falling within the three tutorial moments. Students who chose to use the AI tool are not necessarily representative of the class, and motivation, prior ability, or study habits could plausibly explain both AI-tool use and exam performance. Individual students could not be linked by identity across the AI log, Wooclap, and exam records, so every comparison in this study is aggregate and topic level only, not traceable to a single student's path across all three sources. Topic and tutorial session are also fully confounded in this design: estimation and confidence intervals were covered only in Tutorial I, hypothesis testing and Type I/II errors only in Tutorial II, and two-sample procedures only in Tutorial III, so topic-level differences cannot be separated from differences in attendance or cohort composition across the three sessions. Sample sizes were also small: Wooclap items drew between 12 and 45 responses, and with 12 responses the 95% margin of error on an error rate can reach roughly ±28 percentage points, so many item-level differences lie within measurement noise. The topics with only 3 or 4 AI-log conversations (two-sample procedures, and Type I and II errors and statistical power) are especially fragile. The topic-coding scheme relies on keyword matching at the conversation level, which cannot distinguish surface vocabulary from real conceptual content. The keyword scheme was not validated against human coding of a sample of conversations, so its accuracy is unknown and topic counts may undercount conversations that used vocabulary the scheme did not capture. All topic coding, the keyword matching applied to the AI log and the manual mapping of exam questions to topics, was carried out solely by the author, with no second coder independently verifying these assignments. The comparison between sources was kept descriptive rather than inferential, given only six topic categories, so the patterns reported here should be read as a starting point for further testing, not as confirmed. Three exam questions with genuinely relevant content could not be used in the comparison because their scores mixed several constructs into one number. This remains a single course at a single institution, with 58 students in the AI-log sample after cleaning, so how far these specific topic-level findings generalize elsewhere is untested.

Future Research

The practical takeaway for anyone teaching a similar course is that an AI interaction log and a classroom response system appear to capture different things, and relying on only one may leave a topic like point estimation undetected that the other would have caught. A practical starting point is to treat AI-log volume as a signal of where student attention concentrates, not of where students are struggling, and to read it alongside classroom-response error rates rather than in place of them; where the two disagree, as they did here for point estimation, that disagreement is itself the signal worth following up on. As a working rule for instructors, the two signals can be read together in four cases. High AI-log volume with a high classroom error rate marks a topic worth revisiting. High volume with a low error rate may mark a subtle topic that prompts follow-up questions. Low volume with a high error rate, as for Type I and II errors and statistical power here, may mark a difficulty students are not voicing, and calls for a targeted check. Low volume with a low error rate calls for no action. Where one source has no data on a topic, as for point estimation here, the gap itself is a prompt to add a classroom item. This rule is a hypothesis for future testing, not a finding of this study. For researchers working with AI interaction logs specifically, the preprocessing steps documented here, verifying course scope, checking for non-student accounts, reconstructing broken exports, deserve more explicit attention than they currently receive. A larger sample of AI-tool users would let the topic-level comparison attempted here be tested formally, rather than only described. A version of this study restricted strictly to the three tutorial windows, rather than the full course period, would isolate in-session AI use from the much larger volume of self-directed use captured in this dataset, and would be a natural next step for statistics education research building on this design.

Conclusion

This study compared AI interaction logs, classroom polling data, and final exam records within the same second-year undergraduate statistics course, and found that the three sources point to different, only partly overlapping, pictures of where students struggle. AI-log volume did not consistently align with Wooclap error rates for the topics present in both sources, and point estimation, the topic with the highest exam error rate among directly comparable items, had no Wooclap item at all across three tutorial sessions. These findings are descriptive rather than confirmatory, drawn from a single course, a voluntary AI-tool sample of 58 students, and a six-category topic scheme, so they should be read as a starting point for further work rather than an established result. Within those boundaries, the practical implication is that no single data source captured the full range of student difficulty in this course, and that reading AI interaction logs alongside existing classroom analytics may help to flag topics that either source alone would miss.

Acknowledgment

This research was partially supported by the "Innovate and Share it EEMCS 2025" grant, University of Twente.

Conflict of Interest

The author declares no conflict of interest. Neither the author, the instructional team of the course, nor the University of Twente had any financial, advisory, governance, or developmental relationship with Alexandria or its developers at the time of the study or during the data collection and analysis reported here. A licence agreement between the University of Twente and the platform's developers was established after the study concluded.

References

[1] A. E. Castro Sotos, S. Vanhoof, W. Van den Noortgate, and P. Onghena, "Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education," Educational Research Review, vol. 2, no. 2, pp. 98–113, 2007. https://doi.org/10.1016/j.edurev.2007.04.001

[2] R. Hoekstra, R. D. Morey, J. N. Rouder, and E.-J. Wagenmakers, "Robust misinterpretation of confidence intervals," Psychonomic Bulletin & Review, vol. 21, no. 5, pp. 1157–1164, 2014. https://doi.org/10.3758/s13423-013-0572-3

[3] F. Lyman, "The responsive classroom discussion: The inclusion of all students," Mainstreaming Digest, University of Maryland, College Park, MD, pp. 109–113, 1981.

[4] J. McTighe and F. T. Lyman, "Cueing thinking in the classroom: The promise of theory-embedded tools," Educational Leadership, vol. 45, no. 7, pp. 18–24, 1988.

[5] A. R. Corbo and D. G. G. Sasaki, "Using active learning methodologies in an introductory statistics course," Statistics Education Research Journal, vol. 20, no. 2, Art. 2, 2021. https://doi.org/10.52041/serj.v20i2.113

[6] K. M. Cooper, J. N. Schinske, and K. D. Tanner, "Reconsidering the share of a think-pair-share: Emerging limitations, alternatives, and opportunities for research," CBE—Life Sciences Education, vol. 20, no. 1, fe1, 2021. https://doi.org/10.1187/cbe.20-08-0200

[7] M. K. Smith et al., "Combining peer discussion with instructor explanation increases student learning from in-class concept questions," CBE—Life Sciences Education, 2011. https://doi.org/10.1187/cbe.10-08-0101

[8] B. Catala, D. Muñoz-Higueras, and J. Sanjuán, "Student Response Systems in higher education: A comparative analysis using Wooclap platform in economic courses," Journal of Management and Business Education, 2024. https://journaljmbe.com/article/view/6398

[9] I. Moreno-Medina, M. Peñas-Garzón, C. Belver, and J. Bedia, "Wooclap for improving student achievement and motivation in the Chemical Engineering Degree," Education for Chemical Engineers, vol. 45, pp. 11–18, 2023. https://doi.org/10.1016/j.ece.2023.07.003

[10] A. Dange and R. E. Lopez, "Chatbot Conversations in Physics Education: Using Artificial Intelligence to Analyze Student Reasoning through Computational Grounded Theory," arXiv:2603.04616, 2026. https://arxiv.org/pdf/2603.04616

[11] S. Gok and R. L. Goldstone, "How do students reason about statistical sampling with computer simulations? An integrative review from a grounded cognition perspective," Cognitive Research: Principles and Implications, vol. 9, Art. 33, 2024. https://doi.org/10.1186/s41235-024-00561-x

[12] A. Reinhart, C. Evans, A. Luby, J. Orellana, M. Meyer, J. Wieczorek, P. Elliott, P. Burckhardt, and R. Nugent, "Think-aloud interviews: A tool for exploring student statistical reasoning," Journal of Statistics and Data Science Education, vol. 30, no. 2, pp. 100–113, 2022. https://doi.org/10.1080/26939169.2022.2063209

[13] A. Al-Abri, "Exploring ChatGPT as a virtual tutor: A multi-dimensional analysis of large language models in academic support," Education and Information Technologies, vol. 30, pp. 17447–17482, 2025. https://doi.org/10.1007/s10639-025-13484-x

[14] L. Al Labadi and A. Ly, "Enhancing statistics education through project-based learning (PBL) and the emergence of ChatGPT," Teaching Statistics, vol. 47, no. 3, pp. 200–218, 2025. https://doi.org/10.1111/test.12405

[15] A. R. Ellis and E. Slade, "A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education," Journal of Statistics and Data Science Education, vol. 31, no. 2, pp. 128–133, 2023. https://doi.org/10.1080/26939169.2023.2223609

[16] S. Colando, E. Franke, G. Weinberg, and A. Reinhart, "Analyzing Students' Statistics Writing Before and After the Emergence of Large Language Models," arXiv:2606.22735, 2026. https://arxiv.org/pdf/2606.22735

[17] L. S. DeLuca, A. Reinhart, G. Weinberg, M. Laudenbach, S. Miller, and D. W. Brown, "Developing Students' Statistical Expertise Through Writing in the Age of AI," Journal of Statistics and Data Science Education, vol. 33, no. 3, pp. 266–278, 2025. https://doi.org/10.1080/26939169.2025.2497547

[18] Odea (formerly Library of Alexandria). [Online]. Available: https://odea.education/ (accessed Oct. 9, 2026).

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

An XAI Pipeline for Gender Bias Detection in Advertising Images

Fernando Martínez-Gil[1], Consolación Gil[1], Ainhoa Sánchez[1], Julio Gómez[1], Marjan Alavi [2]

[1] University of Almería, CIAIMBITAL, Almería, Spain

[2] McMaster University, Hamilton, Canada

Article DOI: 10.68286/samyrad2026.380

ABSTRACT

Gender bias remains pervasive in advertising, but large-scale assessment is challenging because manual coding is labour-intensive and automated methods may lack transparency and expert validation. This study presents an Explainable AI (XAI)-based pipeline to validate the Gender Bias Density Index (IDS) and examine its pedagogical potential. A corpus of 2,295 fashion and luxury advertising images was automatically coded using the locally deployed gemma4:26b model and evaluated against expert coding through agreement metrics, gradient boosting, permutation importance, and SHAP (SHapley Additive exPlanations). Results showed robust agreement for visually explicit variables but lower reliability for context-dependent features such as body agency and facial expression. Overall IDS agreement was moderate (ICC = 0.58), while the regression model achieved R^2 = 0.96 on internal test data and R^2 = 0.86 against expert coding. SHAP analysis showed that the most discriminative features broadly align with the theoretical structure of the IDS. The pipeline therefore supports scalable, interpretable bias screening for industry and provides a reproducible case study for teaching critical AI literacy in higher education.

KEYWORDS Explainable AI, advertising, gender bias, higher education, agency, facial expression

How to cite: Martínez-Gil, F., Gil, C., Sánchez, A., Gómez, J., & Alavi, M. (2026). An XAI pipeline for gender bias detection in advertising images. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.380

Introduction

Gender bias in advertising has been studied for decades, with strong evidence that advertising media reproduce and reinforce stereotypes through physical attributes, social roles and visual codes. Structured instruments such as the Gender Bias Density Index (IDS) [1] have been developed to operationalise these patterns systematically. Gender asymmetries have also been documented in competence-related representations [2] and in the visual prominence of women and men in online images [3].

However, analysing gender bias at scale remains challenging. Manual content analysis is time-consuming and depends partly on the interpretive criteria of individual coders, while automated approaches require validation to determine whether they capture the intended social constructs reliably. This creates a need for methods that combine large-scale automatic analysis with expert validation and model interpretability.

Explainable Artificial Intelligence (XAI) provides a suitable framework for this purpose. Unlike conventional predictive models, which produce a score without explaining how it is obtained, tools such as SHAP make it possible to identify which specific visual variables increase or reduce the detected bias in each image. This transparency is valuable both for validating the measurement instrument and for understanding how automated systems operationalise socially meaningful variables.

This paper presents a validation pipeline for the IDS based on machine learning and XAI, implemented with open-source models running locally to ensure data privacy and reproducibility. The study is guided by three research

questions and a transversal pedagogical objective. RQ1: To what extent can the IDS variables be reliably coded by a locally deployed generative vision model, and which variables require human interpretive judgement? RQ2: To what extent does the IDS capture patterns that remain consistent across coding sources, so that a model trained on AI-coded data can predict the index computed from expert coding? RQ3: Which visual features empirically drive the IDS in the corpus, and how do they compare with the theoretical weights defined in the coding manual? As a transversal objective, the pipeline is structured so that each stage can be reused as a teaching unit for critical AI literacy in higher education. It provides a complete case study in which students can observe how a socially grounded construct, gender bias in advertising, is translated into operational variables, quantified through a structured index, evaluated through human-AI agreement, and finally interpreted using explainable machine learning techniques.

The paper is organised as follows: Section II reviews related work on advertising bias analysis and explainable AI; Section III describes the pipeline methodology; Section IV presents and discusses the results; and Section V sets out the conclusions and future work.

Background and related work

Gender bias can be defined as the tendency, conscious or unconscious, to favour one gender over another, producing inequalities across social, professional, educational and technological domains [2]. In advertising and digital content, its presence is well documented: men are associated with roles of power or action while women are given passive or decorative roles, systematic biases appear in the type of product advertised, and when both genders share the frame, the male figure tends to occupy the visual focus [3].

Automatic detection of these biases has developed mainly for text. Natural language processing techniques and supervised classifiers have proved effective in identifying biased language in job advertisements [4] or detecting stereotypes in news text [5]. Comparing lexicon-based approaches with transformer models shows that supervised learning, and BERT (Bidirectional Encoder Representations from Transformers) in particular, produces more accurate results, though it requires larger training datasets [6]. Automatic image analysis has moved more slowly, partly because of the greater technical complexity of computer vision and partly because visual bias patterns are harder to operationalise than textual ones. A further difficulty is the opacity of conventional predictive models, which produce a score without explaining how they reach it, limiting their usefulness as tools for auditing or validating measurement instruments.

Recent research on generative AI has made it possible to characterise more precisely which types of visual bias are reproduced automatically: European white phenotypes as the default, slim and athletic body types, straight hair as the female beauty standard, underrepresentation of older people especially women, systematic absence of visible disability, and gender asymmetries in facial expressions [7]–[9]. These dimensions match the variables captured by the IDS, ethnicity, body type, hair type, apparent age, disability, facial expression, which provides a solid empirical grounding for the instrument.

However, although recent studies of image-generative AI confirm visual representational biases, they do not establish whether a structured visual-bias index can be reliably automated or interpreted at the feature level. XAI research demonstrates the value of such feature-level explanations in other domains [10]. The present study bridges these strands by combining expert validation, predictive modelling and XAI within a single IDS-based framework.

Methods or analytical approach

This study adopts a quantitative, multi-stage validation design combining human–AI agreement assessment, predictive modelling and post-hoc explainability. This design was selected because validating a composite index requires evidence at three complementary levels, each corresponding to one research question: coding reliability against expert judgement (RQ1), internal and external predictive consistency (RQ2), and feature-level interpretability against the theoretical weighting (RQ3). A descriptive analysis of the IDS distribution across the corpus provides context for all three.

Dataset and inclusion criteria

The images used in this study come from the social media posts of internationally recognised fashion and luxury brands. The initial dataset contained 4,355 images, which were filtered using the following inclusion criteria: colour image, clear presence of one or two people with an apparent age of 18 or older, unambiguous gender identification, and representation of real people (excluding cartoons, caricatures or unidentifiable figures). After this filtering process, the dataset was fixed at 2,295 valid images. The distribution by gender composition is as follows: 1,459 images with one woman, 637 with one man, 89 with two women and 97 with two men. A further 13 images contained two people of whom only one was clearly visible and gender-identifiable; these were coded as single-person images using the formula for the visible person's gender. This distribution reflects the predominance of female figures in the advertising corpus, consistent with what has been documented in the literature [9].

The Gender Bias Density Index (IDS)

The Spanish acronym IDS (Índice de Densidad de Sesgo) is retained because it is the name used in the coding manual. The IDS is a measurement instrument developed by the research group of the project in [1] to quantify the level of gender bias in advertising images. The index brings together binary variables organised into four dimensions: biological biases (body type, hair type and hair color, skin conditions, disability representation), demographic biases (apparent age group, ethnicity), cultural biases (body agency, facial expression, visual focus, body exposure, spatial hierarchy) and socioeconomic biases (activity type).

Each variable is binarised as presence (1) or absence (0) of bias and is assigned a weight that differs according to the gender of the person depicted. The index takes the form of a normalised weighted sum, producing a continuous value in the range [0, 1], where 0 means no bias and 1 means maximum bias. Since the composition of people in the image determines both the applicable variables and the corresponding weights, the IDS is operationalised in four

specific formulas: image with one woman, one man, two women, or two men. For brevity, only the formulas for images with one woman (1) and one man (2) are shown here. The full specification, including the remaining formulas, the coded variables and their parameterisation, is available at [1].

Image with one woman:

$$IDS_i^F = \frac{\sum_{j \in J_F} x_{ij}^F w_j^F}{\sum_{j \in J_F} w_j^F} \qquad (1)$$

Image with one man:

$$\mathbf{IDS_i^M} = \frac{\sum_{j \in J_M} \mathbf{x_{ij}^M}\, \mathbf{w_j^M}}{\sum_{j \in J_M} \mathbf{w_j^M}} \qquad (2)$$

where:

- x_{ij}^F or x_{ij}^M = 1 if the image shows bias j, 0 otherwise
- w_{ij}^F or w_j^M = weight of bias j for female or male

Automatic feature extraction using local generative AI

The 2,295 images were coded using the generative AI model gemma4:26b, running locally on the University of Almería's own infrastructure. Using local models removes dependence on external paid services and ensures data privacy, which matters when the dataset includes images of real people.

Two servers were used. The first is a testing server, used for preliminary prompt-validation tests: a VPS with 8 vCPUs and 16 GB of RAM equipped with an NVIDIA RTX 2000E Ada Generation GPU (16 GB). The second is a production server, used for large-scale processing: it is equipped with two Intel® Xeon® Gold 5220 processors, 256 GB of RAM, and two GPUs, which serve the models. Both servers run an Ollama server that receives requests remotely from the application; the results reported in this article were obtained on the production server.

The prompt designed for feature extraction instructs the model to apply the IDS coding manual protocol, returning for each image and each identified person the categorical values of each variable and the corresponding binary value; the IDS is then calculated by applying the weighted formula described in Section III.B. The prompt includes explicit coherence rules that link categorical values to binary values, for example, if hair type takes the value “straight”, the corresponding binary variable must be 1, reducing the likelihood of internal contradictions in the model’s output.

Automatic coding validation: intercoder agreement

To assess the reliability of the automatic coding, a subsample of 85 individual images, 55 female and 30 male, was selected and manually coded by two experts. The experts first coded the images independently, following the IDS coding manual and without discussing their decisions. This independent round made it possible to identify variables with higher and lower levels of interpretive agreement. Discrepancies were then reviewed jointly and resolved by consensus, producing the final human expert coding used as the reference criterion for comparison with the AI output. Images with two people were excluded from this analysis because assigning biases to each subject separately introduces additional interpretive complexity.

Agreement between the automatic (AI) coding and the human coding was measured using Cohen’s Kappa [11], [12], calculated per variable and per gender on the binary presence/absence-of-bias values. The Kappa formula is:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \qquad (3)$$

where P_o is the observed proportion of agreement and P_e is the proportion expected by chance, based on the marginal frequencies of each coder. The denominator ($1 - P_e$) normalises the index so that $\kappa = 0$ means chance-level agreement and $\kappa = 1$ means perfect agreement.

Interpreting Kappa requires looking at the value together with P_o and the confusion matrix, because a low Kappa can result from methodologically different causes. The analysis identified four qualitative profiles:

- Profile 1. Robust agreement ($\kappa \geq 0.50$): genuine agreement backed by sufficient variation in the data. Indicates that the variable can be reliably automated.
- Profile 2. Corpus homogeneity with bias: low Kappa with high P_o and P_e because the distribution of values is so homogeneous that chance agreement is already high. This does not indicate AI failure; it means that the bias is systematically present in nearly all images in the corpus.
- Profile 3. Real AI failure: low Kappa with moderate P_e and systematic false negatives in the confusion matrix. The AI does not detect the bias that the expert identifies, usually because the variable requires interpreting the narrative context of the advertisement.
- Profile 4. Corpus homogeneity without bias: all values are 0 for both coders. The type of bias captured by that variable does not appear in this corpus.

The full classification of variables by profile and gender is presented in Section IV.B and in Fig. 2 and Fig. 3

Machine learning model and IDS validation

The empirical validation of the IDS was designed as a two-phase process. In the first phase, a regression model based on gradient boosting (HistGradientBoostingRegressor from scikit-learn) was trained using as input variables the categorical values of the 12 IDS variables returned by the AI for each person, together with the number of people and the gender of each person, and the IDS calculated from that same coding as the target variable. Nominal variables were one-hot encoded (one binary column per category), apparent age group was ordinally encoded and disability representation was encoded as a single binary column; missing values were imputed with the most frequent category. To ensure independence of the external validation, the 99 images reserved for human coding (the 85 individual images of

Section III.D plus 14 two-person images) were excluded from this phase, leaving a modelling set of 2,196 images. This set was split 80% for training and 20% for testing, with strict separation of the two sets before any transformation to prevent data leakage. Model performance was measured using the coefficient of determination (R^2) and mean absolute error (MAE). The model hyperparameters were: L2 regularisation: 10.0, learning rate: 0.2, maximum iterations: 300, maximum leaf nodes: 15, minimum samples per leaf: 20.

In the second phase, the trained model was applied without retraining to the 99 images coded manually by human experts (the gold set), including the two-person images, since the IDS formula applies equally to those cases. The expert-coded categorical values were transformed with the encoders fitted on the AI-coded data, so that categories absent from training (e.g. the experts' "other" or "not applicable" codes) yield all-zero indicator columns. The target variable was the IDS calculated from the human coding [1]. This phase constitutes the external validation of the index: a high R^2 against the expert criterion shows that the IDS captures objectively detectable patterns that go beyond the coding source, and the gap between the internal (Phase 1) and external (Phase 2) R^2 reflects how far the expert coding departs from the AI coding, linking model validation to the agreement results of Section III.D.

Two complementary techniques were used for model interpretability: permutation importance, which quantifies each variable's contribution to the prediction by measuring the drop in R^2 when its values are randomly shuffled in the test set; and SHAP, which, unlike permutation importance, allows us to examine both the average importance of each variable and the direction of its effect — whether the presence of a category (indicator value = 1) tends to increase or decrease the predicted IDS. This dual capacity, magnitude and direction, makes it possible to compare the empirical weights learned by the model with the theoretical weights in the IDS manual, and to assess how far the model's behaviour matches the theoretical design of the instrument.

Pedagogical use of the validation pipeline

From a pedagogical perspective, the pipeline was designed as a modular workflow transferable to teaching in artificial intelligence, digital literacy and social bias analysis. Each stage can be used as an instructional unit: the coding manual introduces the operationalisation of abstract social constructs; the automatic coding stage illustrates the possibilities and limits of generative AI for image interpretation; the reliability analysis shows why algorithmic outputs must be validated against human judgement; and the XAI stage lets students examine which variables drive the predictions and whether they match the theoretical design of the instrument.

The pipeline thus supports active learning activities in which students compare human and AI codings, interpret confusion matrices and discuss whether model errors reflect technical limitations, corpus homogeneity or genuinely ambiguous social meanings, developing critical AI literacy through a concrete and reproducible bias detection task.

Results and Discussion

Distribution of the IDS in the corpus

The IDS calculated automatically for the 2,295 images shows a mean of 0.443 (SD = 0.114) and a range from 0.0 to 0.92. The histogram (Fig. 1) reveals an irregular shape with two frequency peaks around 0.45 and 0.55, and a progressive decline towards the extremes. Although the IDS is expressed as a continuous value in the range [0, 1], it is calculated from a weighted sum of binary variables. Therefore, local accumulations around specific values are expected, since images with similar combinations of activated bias indicators may obtain similar IDS scores. The low proportion of values below 0.2 or above 0.75 indicates that the advertising corpus tends towards moderate levels of bias, with neither neutral nor strongly biased images predominating.

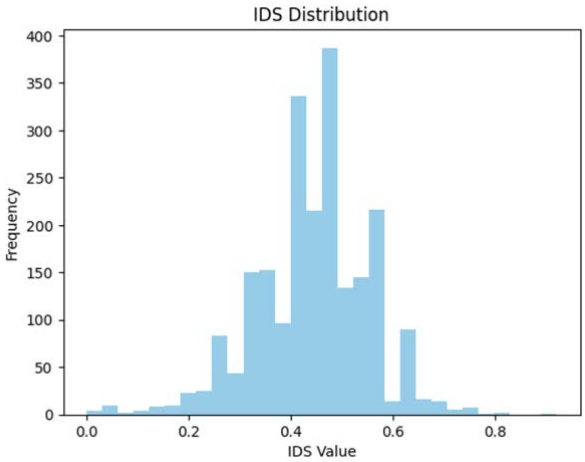


Fig. 1. Histogram showing the distribution of the Gender Bias Density Index (IDS).

Images at the extremes of the index illustrate the discriminative capacity of the IDS. The image with the highest detected bias (IDS = 0.92) is an Adidas post featuring a young woman with a European white phenotype, slim build, straight blonde hair, blemish-free skin, an unstable or forced body posture, tight-fitting clothing, a camera focus on the body and a self-care role: a combination of indicators that activates nearly all the biases in the female IDS formula. At the other end, the image with IDS = 0.0 belongs to the Hermès brand and shows two women, one with an Afro-descendant phenotype, natural curly hair and a serious expression, and another whose characterisation is not applicable due to low visibility, resulting in no detectable biases under the coding manual.

Intercoder agreement and IDS concordance: reliability of automatic coding (RQ1)

The Kappa analysis results, shown in Fig. 2, give a mean of κ = 0.397 for female images (N = 55) and κ = 0.529 for male images (N = 30), excluding trivial variables. This moderate overall value conceals, however, meaningful qualitative differences between variables that the profile analysis can unpack.

Profile 1: Robust agreement. The variables that exceed the $\kappa \geq 0.50$ threshold are: ethnicity (κ = 0.855, Po = 92.7%, Pe = 50.0% for female; κ = 0.864, Po = 93.3%, Pe = 50.9% for male), body exposure (κ = 0.595, Po = 80.0%, Pe = 50.6% for female), hair color (κ = 0.614, Po = 90.9%, Pe = 76.4% for female; κ = 0.516, Po = 90.0%, Pe = 79.3% for male) and activity type for female (κ = 1.000, Po = 100%, Pe = 93.0%). For male images, hair type (κ = 0.706, Po = 86.7%, Pe = 54.7%), disability (κ = 1.000, Po = 100%, Pe = 87.6%) and visual focus (κ = 1.000, Po = 100%, Pe = 93.6%) are also in this profile. In all these

variables Po clearly exceeds Pe, confirming that the agreement is genuine and not an artefact of the corpus distribution. These are visually and objectively detectable features: ethnicity, body framing or activity type present unambiguous cues for the vision model. The contrast between hair type for male images (P1, κ = 0.706, Po = 86.7%, Pe = 54.7%) and female images (P3, κ = 0.350, Po = 65.5%, Pe = 46.9%) shows how the same variable can have very different reliability depending on gender, possibly because the distinction between straight and wavy hair is more ambiguous in the female fashion and luxury corpus. Fig. 2 shows this profile in green and Fig. 3(a) shows the confusion matrix for ethnicity.

Profile 2: Corpus homogeneity with bias. The variables skin conditions (κ = 0.151, Po = 85.5%, Pe = 82.9% for female; κ = 0.286, Po = 76.7%, Pe = 67.3% for male), disability for female (κ = 0.486, Po = 96.4%, Pe = 92.9%), apparent age (κ = 0.347, Po = 85.5%, Pe = 77.7% for female; κ = 0.474, Po = 93.3%, Pe = 87.3% for male), visual focus for female (κ = −0.031, Po = 89.1%, Pe = 89.4%), body agency for male (κ = −0.035, Po = 93.3%, Pe = 93.6%) and activity type for male (κ = 0.000, Po = 96.7%, Pe = 96.7%) share the defining feature of Profile 2: Pe is close to or equal to Po, so Kappa is not measuring AI failure but corpus homogeneity. The luxury and fashion advertising analysed systematically shows perfect skin, absence of disability, young or middle-aged subjects and men in direct product-showcase roles, meaning these variables barely distinguish between images. This result is itself a finding: the biases in Profile 2 cannot be quantified by Kappa precisely because they are always present in the corpus. Fig. 2 shows this profile in orange and Fig. 3(b) shows the confusion matrix for skin conditions.

Profile 3: Real AI failure. The variables body type for female (κ = 0.189, Po = 67.3%, Pe = 59.7%), hair type for female (κ = 0.350, Po = 65.5%, Pe = 46.9%), body agency for female (κ = −0.012, Po = 72.7%, Pe = 73.1%) and facial expression (κ = 0.217, Po = 80.0%, Pe = 74.5% for female; κ = 0.483, Po = 73.3%, Pe = 48.4% for male) differ from Profile 2 in that Pe is moderate and Po does not exceed it significantly, revealing a real disagreement and not a homogeneity effect. The confusion matrices show systematic false negatives: the AI does not detect the bias that the expert identifies. The most extreme case is body agency for female (κ ≈ 0, Po = 72.7%, Pe = 73.1%), where the AI performs no better than chance at distinguishing a decorative female posture from a functional one — a distinction that requires reading the full narrative context of the advertisement. The combined weight of these variables in the female IDS formula (body agency = 4, facial expression = 4, hair type = 3, body type = 3) accounts for 28% of the index, with a direct impact on external validation (Section IV.C). The gender asymmetry in facial expression, better detected for male images (κ = 0.483, Po = 73.3%, Pe = 48.4%) than for female (κ = 0.217, Po = 80.0%, Pe = 74.5%), reflects the fact that male biased expressions (serious/authoritative) are more objectively observable than female ones (seductive/vulnerable), which carry a heavier cultural load. For male images, with Po exceeding Pe by 25 points, it falls just below the Profile 1 threshold and reflects moderate rather than failed agreement. Fig. 2 shows this profile in red and Fig. 3(c) shows the confusion matrix for body agency.

Profile 4: Corpus homogeneity without bias. Body type for male (trivial κ, Po = 100%, Pe = 100%; all values are zero because the instrument does not code male slimness as a bias indicator) and body exposure for male (trivial κ, Po = 100%, Pe = 100%; body exposure is absent in male luxury advertising) show trivial agreement because the bias simply does not appear in this corpus. The fact that body objectification and slimness as an aesthetic ideal are biases that are absent for men but present for women is itself a finding about the structural asymmetry of gender bias in the advertising analysed. Fig. 2 shows this profile in blue and Fig. 3(d) shows the confusion matrix for body type.

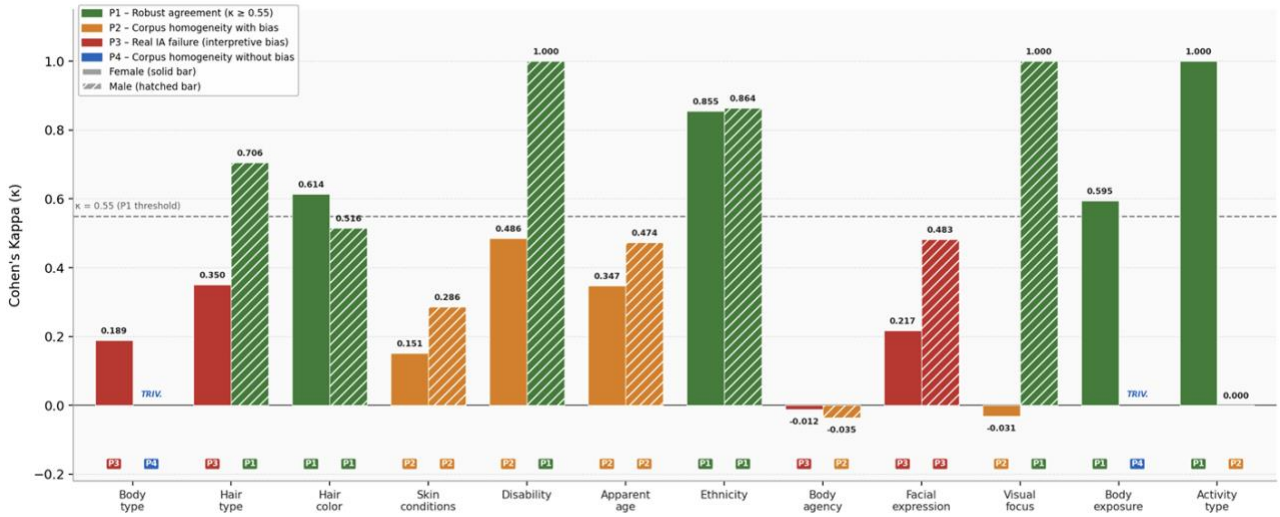


Fig. 2. Cohen's Kappa intercoder agreement between AI and human coding, by variable and gender (female N = 55, male N = 30).

Pedagogically, the four-profile interpretation shows students that a low Kappa does not always mean model failure: it may reflect corpus homogeneity or the model's inability to read culturally loaded cues, as the confusion matrices in Fig. 3 make concrete by distinguishing which biases are visually explicit, which are inferred from context, and which the model misses.

Variable-by-variable Kappa analysis gives only a partial picture of reliability, as it evaluates individual binary components that are sensitive to class imbalance in the corpus. To assess whether the automatic coding produces an overall IDS equivalent to the expert's, three complementary analyses were carried out on the N = 85 individual images (N = 55 female, N = 30 male).

The Intraclass Correlation Coefficient (ICC), a standard psychometric tool for validating composite scales [13], gave ICC(A,1) = 0.582 (95% CI = [0.42, 0.71], p < 0.001) overall; 0.489 ([0.26, 0.67]) for female images; and 0.593 ([0.25, 0.79]) for male images. According to Koo and Li's criteria [13] (below 0.50 = poor, 0.50–0.75 = moderate, above 0.75 = good), reliability is moderate overall and for male images, and at the boundary between poor and moderate for female images, where the Profile 3 variables concentrate; this is acceptable for an instrument that includes partly subjective interpretive variables. The closeness of ICC(A,1) = 0.582 and ICC(C,1) = 0.587 confirms that the systematic bias between the two coding sources is minimal, so the instrument can be applied automatically, with acceptable limitations, for comparative analyses at scale.

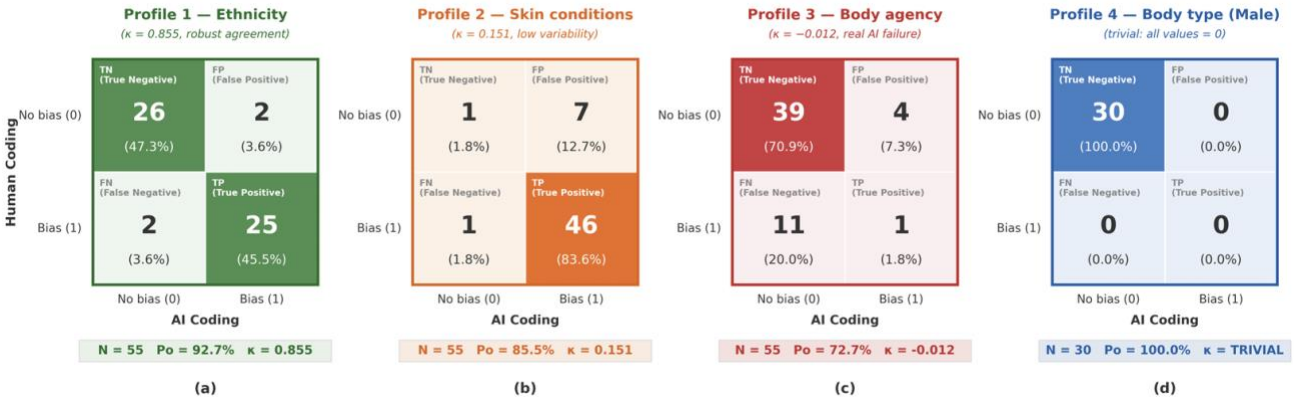


Fig. 3. Representative confusion matrices for the four agreement profiles between AI and human coding (absolute values and percentages).

To evaluate the practical image-by-image error, Bland-Altman analysis was used [14]. The overall mean bias is +0.020 (95% CI = [−0.003, +0.043]), equivalent to just 2 points out of 100, with no practical relevance. The limits of agreement (LoA = [−0.191, +0.231]) indicate that empirically, 92.9% of the images fell within these limits between −19 and +23 points out of 100. A statistically significant proportional error was found for female images (r = −0.376, p = 0.005): the AI tends to overestimate the IDS for low-bias images and underestimate it for high-bias ones, a pattern consistent with the greater weight of Profile 3 variables in images with higher IDS values. For male images the bias is slightly larger (+0.039) but constant (p = 0.161). Decomposed by variable, it originates mainly in skin conditions and facial expression, where the AI codes bias more often than the expert (93% vs 70% and 73% vs 47% of male images, respectively), whereas hair color and apparent age, whose binarisation rules differ between genders, contribute negligibly.

Spearman correlation was used for image ranking because it is more robust to the non-normal distribution of the IDS and to the presence of outliers. The results (ρ = 0.623 overall; ρ = 0.480 for female; ρ = 0.649 for male; p < 0.001 in all cases) indicate moderate ordinal concordance. The AI correctly identifies which images in a campaign carry the highest bias load in most cases, which is the most practically relevant use of the instrument for communication teams.

The analysis of systematic error (IDS_AI − IDS_Expert) by brand reveals a methodologically significant pattern. Luxury brands with a stronger narrative component show negative errors, the AI underestimates the IDS, possibly because the bias is expressed through staging and mise-en-scène codes that go beyond the vision model's capacity: Dior (error = −0.046), Chanel (−0.039). By contrast, sports and accessible brands show positive errors where the bias attributes are more directly observable: Adidas (+0.056), Nike (+0.042). The most extreme case is Rolex (+0.109), where the AI overestimates the IDS in both female (+0.137) and male (+0.083) images: this brand favours mature subjects, whose visible signs of age the expert codes as non-idealised attributes, whereas the AI tends to code them as unblemished skin, youthful appearance and serious expression. This pattern suggests that the reliability of automatic coding is not uniform across the corpus but varies systematically with the advertising genre of the brand, which is a specific limitation of the instrument that should be taken into account when interpreting results at the level of individual campaigns.

Empirical validation of the IDS using gradient boosting (RQ2)

Phase 1: Internal consistency

The HistGradientBoostingRegressor trained on the AI-coded binary features from the 2,196 images achieved R^2 = 0.99 on the training set (1,757 images, 80%) and R^2 = 0.96 on the test set (439 images, 20%), with MAE = 0.004 and MAE = 0.010 respectively. The three-percentage-point difference between training and test is expected and does not indicate overfitting, but rather solid generalisation to unseen data. The test MAE of 0.010 corresponds to about one percentage point on the IDS scale. This shows that the model reproduces the logic of the IDS without accessing the formula, validating the internal consistency of the instrument.

Phase 2: External validation against human criterion

When the trained model was applied to the 99 manually coded images, R^2 = 0.86 and MAE = 0.031, equivalent to a mean error of 3.1 percentage points on the index scale. The 10-percentage-point drop from the Phase 1 test R^2 (0.96) reflects combinations of categories that are rare or absent in the AI-coded training data: the experts make systematic use of the "other" and "not applicable" codes (e.g. in body agency, facial expression, body exposure and activity type), which the AI seldom returns, and binarise some variables differently (male body type is never coded as bias; serious male expressions less often). These are the discrepancies

documented by the Kappa profiles in Section IV.B, so the external R^2 is consistent with the moderate agreement between coding sources (ICC(A,1) = 0.58) while confirming that the model has learned the logic of the index rather than the idiosyncrasies of the AI coding.

In particular, body agency (weight = 4) and facial expression (weight = 4), the Profile 3 variables with the largest AI–human disagreements, make up 16% of the female index, so their coding differences have a proportional impact on the IDS; this link between the reliability analysis and the model validation is one of the contributions of this work.

Variable importance: coherence with the theoretical framework (RQ3)

The permutation importance analysis on the test set identified five variables as the most influential in IDS prediction: gender_b (1.107), body_exposure_a (0.277), ethnicity_a (0.273), gender_a (0.197) and hair_type_a (0.101). The dominant role of gender_b is methodologically expected: this variable encodes whether the image contains one or two people, determining which IDS formula to apply. Permuting its value forces the model to use the wrong formula, causing the largest drop in performance (a value above 1 indicates that the permuted R^2 becomes negative). This confirms that the model has learned the compositional structure of the index rather than simply individual variable patterns.

Among the substantive variables, body_exposure_a ranks second; its most influential category groups forms of body exposure other than tight clothing or partial nudity, reflecting that body exposure type is highly discriminative between images with different bias levels. Ethnicity ranks third, consistent with its theoretical weight (weight = 4 for women, 3 for men) and the high prevalence of European white phenotypes in the corpus. Gender_a appears in fourth place and, like gender_b, determines which IDS formula to apply. Fifth place goes to hair_type_a despite having a lower theoretical weight (weight = 3 for women, 2 for men), suggesting that its high frequency in the corpus amplifies its discriminative capacity.

The SHAP analysis (Fig. 4) confirms and nuances these results. In the global beeswarm plot, red dots indicate images where the category is present (indicator value = 1): when they are concentrated to the right of the axis, the presence of that category increases the predicted IDS. The categories ethnicity_a_white_european, hair_type_a_straight and body_type_a_thin show this pattern clearly and consistently in the global analysis. By contrast, the red dots of body_exposure_a_other are concentrated to the left of the axis: this category groups forms of body exposure other than tight clothing or partial nudity, i.e. the non-biased value of the variable, so its presence lowers the predicted IDS. The model has thus learned the direction of each category, in agreement with the design of the instrument.

It is worth noting that variables with high theoretical weight but low variability in the corpus, such as disability and skin conditions, classified in Profile 2 in the Kappa analysis (Section IV.B), are absent from the top of both analyses. This is not a flaw in the instrument but a direct consequence of corpus homogeneity: when a variable takes the value 1 in more than 90% of images, it contributes to the IDS in a constant way but has no capacity to discriminate between images. Taken together, the permutation importance and SHAP analyses suggest that part of the theoretical weighting in the IDS reflects biases that are systematically present throughout the luxury and fashion advertising corpus, rather than features that differentiate more from less biased images within it.

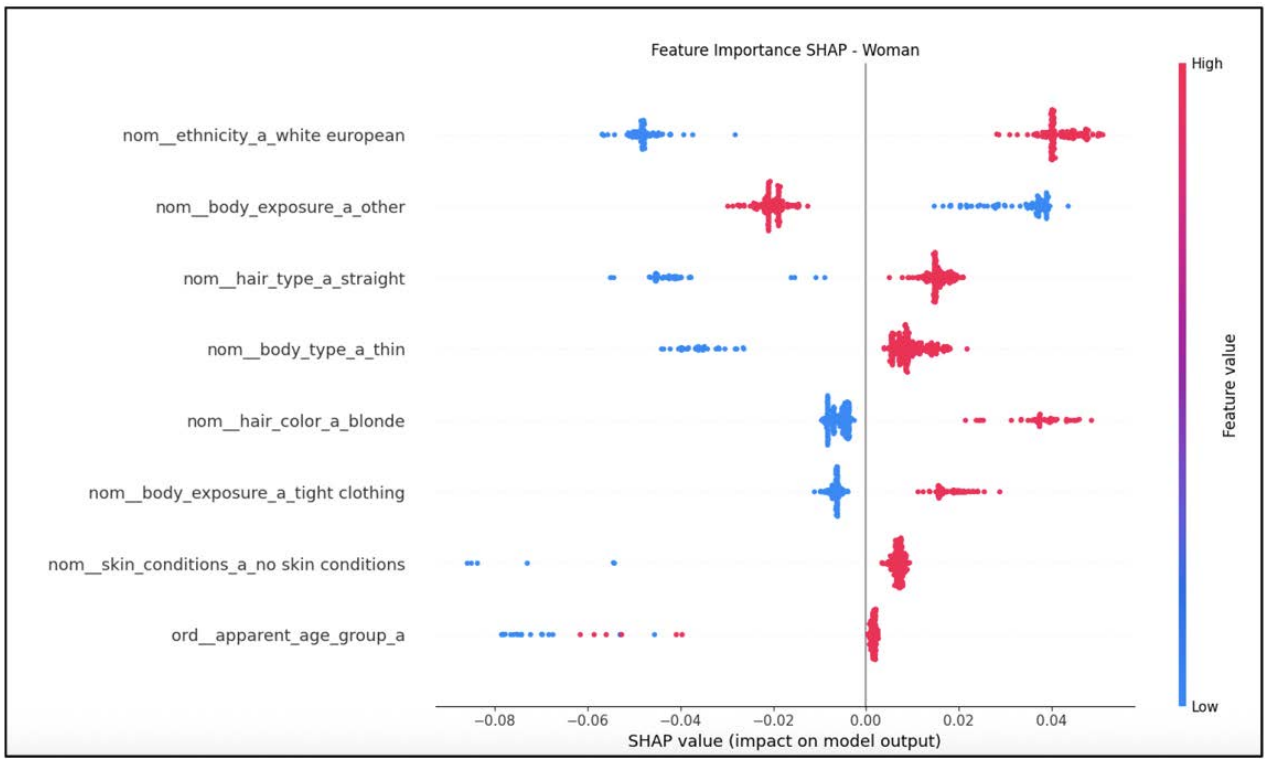


Fig. 4. SHAP analysis of the main features driving IDS predictions (global beeswarm plot).

For teaching purposes, SHAP lets students contrast the manual's theoretical weights with the empirical importances and discuss why high-weight variables can lack discriminative power in a homogeneous corpus.

Practical implications

For industry, the results have three practical implications. Communication and brand teams could use the pipeline as a pre-publication screening tool, as the AI correctly ranks the most biased images in a campaign in most cases (ρ = 0.62). Profile 1 variables can be reliably automated, whereas Profile 3 variables, such as body agency and facial expression, should be flagged for human review. Finally, the brand-level error pattern suggests that screening thresholds may require calibration by advertising genre, since narrative luxury brands tended to be under-scored, whereas sports brands tended to be over-scored in this corpus.

Conclusion

This work has shown that empirically validating a gender bias index using machine learning and XAI is feasible, with models running locally to ensure privacy and reproducibility.

In answer to the research questions, the results show that (RQ1) visually explicit variables such as ethnicity, hair color and body exposure can be reliably automated, whereas context-dependent variables such as body agency and facial expression still require human judgement; (RQ2) the IDS captures patterns consistent across coding sources, with a model trained on AI-coded data predicting the expert-coded index with R^2 = 0.86; and (RQ3) the features that empirically drive the index partially coincide with those of highest theoretical weight, the main discrepancies arising from variables that are near-universal in the corpus and therefore lack discriminative power.

The intercoder analysis reveals four agreement profiles with different methodological implications. Variables with robust agreement (Profile 1: ethnicity, hair color, body exposure and, for men, hair type) can be directly automated. Profile 2 variables (skin conditions, disability and apparent age) show low Kappa not because the AI fails but because the bias is so systematically present in the luxury advertising analysed that it does not discriminate between images: the low Kappa itself is a finding about the nature of the corpus. Profile 3 variables (body agency and facial expression for female) expose a real limit of the vision model in interpreting the narrative context of the advertisement, and their combined weight in the female IDS formula (16%) is consistent with the drop in R^2 in the external validation (R^2 = 0.86).

The SHAP analysis confirms that the variables with the greatest discriminative power broadly match those with the highest theoretical weights in the instrument, with the exception of Profile 2 variables, whose theoretical importance does not translate into empirical importance within this corpus. The brand-level error analysis further shows that the reliability of automatic coding varies systematically between narrative luxury brands and sports brands, which is a specific limitation to bear in mind when applying the instrument at the level of individual campaigns.

From an educational perspective, the main contribution of the work is the design of a replicable XAI-based learning scenario for social bias detection. The pipeline allows students to engage with the full cycle of responsible AI analysis: defining a socially meaningful construct, operationalising it through variables and weights, comparing human and automatic coding, evaluating reliability, training predictive models and interpreting their decisions through XAI. This makes the IDS validation process a useful pedagogical resource for promoting critical AI literacy, particularly in contexts where students need to understand both the technical potential and the social limitations of automated decision-making systems.

Future work will therefore not only focus on improving the predictive pipeline, but also on adapting it into structured teaching activities for higher education. These activities may include guided coding exercises, human-AI comparison tasks, interpretation of Kappa profiles and confusion matrices, SHAP-based model explanation, and counterfactual redesign of biased images using DiCE (Diverse Counterfactual Explanations) to generate image-level recommendations, completing the pipeline from bias detection to bias correction. Such activities would make it possible to evaluate the pipeline not only as a validation method, but also as an educational intervention for teaching explainable and socially responsible AI.

Declarations

Funding: This work is supported by Project “21-ID25, Femvertising, AI and Body Diversity: Towards Inclusive and Transformative Advertising”, Instituto de la Mujer (Ministry of Equality, Spain).

AI use disclosure: AI was used for translation assistance and English language editing.

Acknowledgement

The authors thank Ivan González for his help collecting images from social media, the members of the HUM-1022 and TIC-221 UAL research groups for their collaboration in developing the IDS and their participation as expert coders, and the anonymous reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

References

[1] IDS (Índice de Densidad de Sesgo). IDS Coding Manual for Gender Bias Detection. Project 21-ID25 (UAL) [Online]. Available: https://www.ualtech.es/projects/gender_bias_detection/#top. [Accessed: Aug. 21, 2026]

[2] R. Poveda-Bautista, J. A. Diego-Mas, H. Gonzalez-Urango, and C. Corona-Sobrino, “Conscious and unconscious gender bias in competence evaluations: Mental representations of project managers,” Complexity, vol. 2025, 2025, doi: 10.1155/cplx/7974362.

[3] D. Guilbeault, S. Delecourt, T. Hull, B. S. Desikan, M. Chu, and E. Nadler, “Online images amplify gender bias,” Nature, vol. 626, no. 8001, pp. 1049–1055, 2024, doi: 10.1038/s41586-024-07068-x.

[4] R. Frissen, K. J. Adebayo, and R. Nanda, “A machine learning approach to recognize bias and discrimination in job advertisements,” AI & Society, 2022, doi: 10.1007/s00146-022 01574-0.

[5] C. H. M. Landa, Implementación de NLP y análisis temático para la detección de sesgos de género. Tlaquepaque, Jalisco, México: Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO), 2024.

[6] J. Cryan, S. Tang, X. Zhang, M. Metzger, H. Zheng, and B. Y. Zhao, "Detecting gender stereotypes: Lexicon vs. supervised learning methods," in Proc. 2020 CHI Conf. Human Factors in Computing Systems, pp. 1–11, 2020, doi: 10.1145/3313831.3376488.

[7] J. Shuford, "Examining ethical aspects of AI: Addressing bias and equity in the discipline," Journal of Artificial Intelligence General Science (JAIGS), vol. 3, no. 1, pp. 262–280, 2024, doi: 10.60087/jaigs.v3i1.119.

[8] L. Sun, M. Wei, Y. Sun, Y. J. Suh, L. Shen, and S. Yang, "Smiling women pitching down: Auditing representational and presentational gender biases in image-generative AI," Journal of Computer Mediated Communication, vol. 29, no. 1, Art. no. zmad045, 2024, doi: 10.1093/jcmc/zmad045.

[9] S. L. Grau and Y. C. Zotos, "Gender stereotypes in advertising: A review of current research," International Journal of Advertising, vol. 35, pp. 761–770, 2016, doi: 10.1080/02650487.2016.1203556.

[10] S. Nejadshamsi, C. H. Chu, K. McGilton, X. Li, C. E. Ronquillo, and S. Abbasgholizadeh-Rahimi, "Evaluation and improvement of algorithmic fairness for COVID-19 severity classification using explainable artificial intelligence-based bias mitigation," JAMIA Open, vol. 9, no. 1, Art. no. ooaf171, 2026, doi: 10.1093/jamiaopen/ooaf171.

[11] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37 46, 1960, doi: 10.1177/001316446002000104.

[12] K. A. Neuendorf, The Content Analysis Guidebook, 2nd ed. Thousand Oaks, CA, USA: Sage, 2016.

[13] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," Journal of Chiropractic Medicine, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[14] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," The Lancet, vol. 327, no. 8476, pp. 307–310, 1986, doi: 10.1016/S0140 6736(86)90837-8.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Toward AI-Assisted Early Academic Risk Detection in Introductory Physics Courses: Interpretable Profiling among First-Year Engineering Students

Claudia Lizette Garay-Rondero[1] (claudia.garay@itsta.edu.mx), Seyed Behbood Issa-Zadeh[2] (Behbood.Issa-Zadeh@unigib.edu.gi), Fabiola Sánchez Galván[3] (fabiola.sanchez@itsta.edu.mx), Adriana Erika Martínez-Cantón[4] (amartine@tec.mx)

[1]Tecnológico Nacional de México, Instituto Tecnológico Superior de Tantoyuca, Veracruz, Mexico
[2]School of Maritime and Science, University of Gibraltar, Campus Europa Point, Gibraltar, UK
[3]Tecnológico Nacional de México, Instituto Tecnológico Superior de Tantoyuca, Veracruz, México
[4]School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, México

Article DOI: https://doi.org/10.68286/samyrad2026.384

ABSTRACT

Early academic risk detection in first-year engineering education requires approaches that are not only predictive, but also interpretable and actionable for instructors, tutors, and student-support teams. This study examines how multidimensional student data can be used to generate interpretable academic-risk profiles among first-year engineering students enrolled in introductory physics courses. Data were collected from 691 students at a private multi-campus university in Mexico across three academic cohorts. A baseline questionnaire covering academic preparation, perceived previous knowledge, study habits, well-being, academic status, contextual indicators, and subject relevance was linked to first-month course grades using course-based codes and de-identified before analysis. The analysis combined exploratory explanatory modeling and Chi-square Automatic Interaction Detection to identify variables and profile patterns associated with early academic performance and first-month academic risk. The results showed that early performance was associated with prior high-school average, perceived sufficiency of previous knowledge, academic status, homework completion, sleeping hours, time management, scholarship context, cohort-related variation, and perceived subject relevance. The study synthesized four interpretable academic-risk profiles: preparation gap, self-regulation difficulty, contextual compound risk, and motivation-relevance risk. The study also proposes a conceptual AI-assisted early-support framework connecting student indicators, interpretable profiles, traffic-light decision support, AI-assisted tutoring mechanisms, human-in-the-loop governance, and follow-up recalibration. The findings provide an empirical basis for future AI-assisted early-alert strategies, while emphasizing that the proposed framework is conceptual and requires validation with recent cohorts before operational implementation.

KEYWORDS Academic risk detection; first-year engineering; interpretable machine learning; learning analytics; CHAID; structural equation modeling; AI-assisted support.

How to cite: Garay-Rondero, C. L., Issa-Zadeh, S. B., Sánchez Galván, F., & Martínez-Cantón, A. E. (2026). Toward AI-assisted early academic risk detection in introductory physics courses: Interpretable profiling among first-year engineering students. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.384

Introduction

First-year engineering education represents a critical transition period in which students must adapt to new academic expectations, increased workload, independent study routines, and demanding foundational courses. Introductory physics courses are particularly important in this context because they provide conceptual and analytical foundations for later engineering subjects. However, early underperformance in these courses may signal difficulties related not only to academic preparation, but also to study habits, self-regulation, well-being, motivation, and institutional context.

Universities increasingly use learning analytics and artificial intelligence to identify students who may be academically at risk. These approaches can support early-warning systems, academic advising, tutoring, and student-support interventions. Nevertheless, many predictive models focus mainly on classification accuracy and provide limited explanation of why a student or group has been identified as at risk. In educational contexts, this lack of interpretability is problematic because risk labels may influence academic decisions, support allocation, and student perceptions. Therefore, early academic-risk detection should not only predict low performance, but also generate transparent and actionable information for instructors, tutors, and support teams.

This issue is especially relevant in foundational physics courses for first-year engineering students. Low early grades may result from different combinations of factors. For example, one student may struggle because of insufficient prior academic preparation, while another may struggle with time management, sleep, motivation, or academic status. Treating all students with low performance as a single homogeneous group may lead to generic interventions that do not respond to their specific needs. Interpretable profiling can help address this limitation by identifying different patterns of academic vulnerability and linking them to differentiated forms of support.

Building on this need, the present study examines how multidimensional student data can be used to generate interpretable academic-risk profiles among first-year engineering students enrolled in introductory physics courses. The study uses data from 691 students at a private multi-campus university in Mexico, collected across three academic cohorts. The study linked baseline questionnaire data from Physics I and Introduction to Physics to first-month course grades using course-based codes, and the dataset was de-identified before analysis in *Physics I* and *Introduction to Physics*. The analysis combines exploratory explanatory modeling and Chi-square Automatic Interaction Detection to identify variables and profiles associated with early course performance and first-month academic risk.

The study is guided by two research questions:

RQ1: Which academic, behavioral, contextual, motivational, and well-being-related variables are associated with early academic performance in introductory physics courses among first-year engineering students?

RQ2: How can interpretable profiling identify combinations of student characteristics associated with first-month pass/fail classification in introductory physics courses?

The contribution of this paper is threefold. First, it identifies multidimensional variables associated with early academic performance in introductory physics courses. Second, it synthesizes these variables into interpretable academic-risk profiles: preparation gap, self-regulation difficulty, contextual compound risk, and motivation-relevance risk. Third, it proposes a conceptual AI-assisted early-support framework in which interpretable profiles can inform traffic-light alerts, tutoring recommendations, and human-in-the-loop academic support. The study does not claim to deploy or validate an operational AI system; rather, it provides an empirical and interpretable basis for future AI-assisted early-alert strategies in first-year engineering education.

The remainder of this paper is organized as follows. Section II presents the theoretical background, focusing on early academic risk in first-year engineering education, AI-enabled learning analytics and early alert systems, and the role of interpretable risk profiling for actionable support. Section III describes the methodology, including the study context, instrument, variables, data preparation, and analytical strategy. Section IV reports the results and synthesizes the CHAID-based interpretable academic-risk profiles. Section V discusses the implications of the findings and introduces the conceptual AI-assisted early-support framework. Finally, Section VI presents the conclusions, methodological considerations, and future research directions.

Theoretical background

Early Academic Risk in First-Year Engineering Education

First-year engineering education is a critical transition period in which students must adapt to new academic expectations, higher workload, independent study routines, and foundational disciplinary requirements. Courses such as mathematics and physics often play a central role in this transition because they provide the conceptual and analytical basis for later engineering subjects. Previous studies have shown that early achievement, prior academic preparation, institutional conditions, and students' capacity to adapt to university life influence persistence and success in engineering programs [1]. In this context, academic risk should not be understood only as an individual deficit, but as the result of interactions among preparation, study behavior, well-being, motivation, and institutional support.

Recent research also highlights the importance of looking beyond grades alone. Engineering students may experience workload pressure, academic stress, and well-being challenges that affect their learning experience and performance [2]. At the same time, students who experience early academic failure may still persist when adequate support structures are available, suggesting that early risk detection should be connected to timely and differentiated support rather than used only for classification [3]. This is especially relevant in first-year physics and mathematics courses, where early underperformance can provide an important signal for academic support and progression monitoring. Studies focused on engineering courses have therefore argued for risk-based approaches that identify students who may require support before academic difficulties become harder to reverse [4], [5].

AI-Enabled Learning Analytics and Early-Alert Systems

Learning analytics and artificial intelligence have increasingly been used to identify students exposed to academic risk and to support early decision-making in higher education. Early-warning systems can help institutions detect students who may be struggling during the first weeks of study and provide evidence for academic advising, tutoring, or other forms of support when combined with appropriate institutional processes. Recent work on first-year academic performance prediction shows the potential of early-warning models to support institutional monitoring and timely action [6]. Similarly, studies on course-level prediction indicate that academic performance can be modeled using student data to identify risk patterns in specific learning contexts [5].

However, the educational value of AI-enabled learning analytics depends on more than model accuracy. Recent reviews show that many AI-based educational tools and dashboards still face limitations related to real-world implementation, evaluation quality, privacy, bias, transparency, and weak integration between prediction and intervention [7]–[9]. In practice, a system that only assigns a risk score may be insufficient for instructors, tutors, or student-support teams, because they need to understand the nature of the risk and select an appropriate response. Therefore, early alert systems require approaches that translate student data into meaningful and actionable information.

Interpretable Risk Profiling for Actionable Support

The need for explainable and interpretable approaches has become increasingly important in educational data mining and learning analytics. Predictive models may classify students as at risk, but complex models do not always show why a particular student or group has been flagged. This is problematic in education, where risk labels can influence support decisions and where transparency is important for trust, fairness, and responsible use. Recent studies on explainable learning analytics, course-success prediction, AI-based learning tools, and AI-supported academic support suggest that educational AI should combine prediction, interpretability, feedback, and pedagogical action rather than rely solely on model accuracy [8], [10]– [12].

Explainable artificial intelligence has been used to improve the interpretation of academic performance models by identifying the factors that contribute to success or failure. Ujkani et al. [12], for instance, used machine learning and deep learning techniques to predict course completion and identify at-risk students, applying SHAP to explain the factors behind model outputs. Their results showed that student engagement and registration timing were important factors for performance, illustrating how XAI can support more targeted educational interventions. Nevertheless, for many educational settings, highly complex explanations may still be difficult to translate into practice. Interpretable models, such as decision trees, remain useful because they can generate transparent rules and profiles that are easier for tutors and support teams to understand.

Building on this literature, the present study focuses on interpretable academic-risk profiling rather than opaque prediction. It combines exploratory explanatory modeling with CHAID-based decision-tree profiling to identify combinations of academic, behavioral, contextual, motivational, and well-being-related variables associated with early performance in introductory physics courses. This approach responds to the need for learning analytics that are not only data-driven, but also transparent and actionable for tutors and student-support teams. The study therefore provides an empirical basis for future AI-assisted early-support strategies, while avoiding claims of fully validated prediction or automated intervention.

Methodology

The study is positioned as an empirical learning analytics contribution. It combines explanatory modeling and interpretable CHAID-based profiling to identify academic, behavioral, contextual, motivational, and well-being-related indicators associated with early course performance. The resulting profiles are intended to inform future AI-assisted early-alert and support strategies, but this study did not implement or evaluate an operational AI system.

Instrument Development and Study Context

The research instrument was developed through an exploratory phase conducted during the August–December 2016 and January–May 2017 academic periods with 100 undergraduate engineering students. This phase was used for item generation, refinement, and identification of possible factors associated with failure in first-year physics courses. It did not involve an experimental intervention or manipulation.

The main field study was conducted at a private multi-campus university in Mexico. The sample included 691 first-year engineering students enrolled in common

introductory physics courses: *Physics I* and *Introduction to Physics*. Data were collected across three academic cohorts: August–December 2017, January–May 2018, and August–December 2018. The two courses were analyzed together because they represent foundational physics experiences in the first-year engineering curriculum and because the original analysis found insufficient evidence to treat their score distributions as substantially different.

Although the dataset corresponds to pre-pandemic cohorts, the aim of this conference paper is not to claim direct operational deployment in current AI systems. Instead, the study uses an existing multidimensional dataset to examine how interpretable academic-risk profiling can provide an empirical basis for future AI-assisted early-alert models. Such models should be re-estimated and validated with more recent cohorts before institutional implementation.

Instrument, Variables and Outcome

Data were collected using a baseline questionnaire administered at the beginning of the course, before the first academic evaluation. The instrument comprised 39 questions and 49 coded items covering academic preparation, perceived prior knowledge, physical well-being, family and social context, academic development, previous educational experience, social development, organization and study habits, and perceived relevance of physics to the engineering career.

Questionnaire responses were linked to first-month course grades using course-based codes, and the dataset was de-identified before analysis. The first-month grade was used as the early academic performance outcome. For the profiling analysis, early academic risk was operationalized through low early performance and first-month pass/fail classification based on predefined grade thresholds, rather than institutional dropout.

The main variables later retained in the explanatory and profiling stages included semester, scholarship type, academic status, prior high-school grade average, perceived sufficiency of previous knowledge, homework completion during the senior year of high school, sleeping hours, self-evaluation of time management, and perceived importance of the subject for the student's career.

Data Preparation and Analytical Strategy

Data preparation included scale verification, consistency checks, missing-data screening, category harmonization, and recoding of categorical variables when required for statistical analysis. The final analytical sample included 334 students from *Introduction to Physics* and 357 students from *Physics I*, resulting in a combined sample of 691 students.

Given the heterogeneous nature of the questionnaire data, association measures were selected according to the measurement level of each variable. Spearman's rho was applied to ordinal or non-normally distributed variables; polychoric and tetrachoric coefficients were used for ordinal–ordinal and binary–binary associations, respectively; biserial or point-biserial coefficients were considered for relationships involving a continuous outcome and a dichotomous variable; and Cramér's V was used for nominal categorical associations. This strategy preserved the structure of the original variables rather than forcing all questionnaire items into a single statistical format.

The analytical strategy followed four stages. First, descriptive statistics and association analyses were used to examine relationships between student characteristics and early course grades. Second, exploratory path analysis /SEM-based explanatory modelling was applied as a procedure to identify variables associated with early academic performance [13]. Third, significant and theoretically meaningful variables were analyzed using Chi-square Automatic Interaction Detection to generate course non-passing risk [14]. Fourth, Shapiro–Wilk, Mann–Whitney U, and Kruskal–Wallis tests were used to examine normality and grade differences across relevant groups.

The SEM and CHAID results were interpreted as exploratory and explanatory evidence. No claim is made that the resulting tree constitutes an externally validated predictive model.

Figure 1 summarizes the analytical workflow used in the study. The process begins with the baseline questionnaire and continues through data cleaning, association analysis, exploratory SEM, CHAID-based profiling, non-parametric group comparisons, and the synthesis of interpretable academic-risk profiles.

Students received an explanation of the purpose of the study and provided informed consent. Participation was voluntary and did not affect course evaluation. The questionnaire did not request directly identifying information such as student name, student ID, or email address. The original project was reviewed by the Institutional Research and Ethics Review Committee of Tecnológico de Monterrey. Course-based codes were used to link questionnaire responses with first-month grades, and the dataset was de-identified before analysis.

Results and Interpretable Risk Profiles

Overview of Early Academic Performance

The first stage of the analysis examined the distribution of first-month grades in *Physics I* and *Introduction to Physics*. The overall mean grade was 72.92, with a standard deviation of 17.86 and a coefficient of variation of 24.49%. The Shapiro–Wilk test indicated that grades were not normally distributed ($W = 0.95221$, $p \leq 0.05$), supporting the use of non-parametric tests for group comparisons.

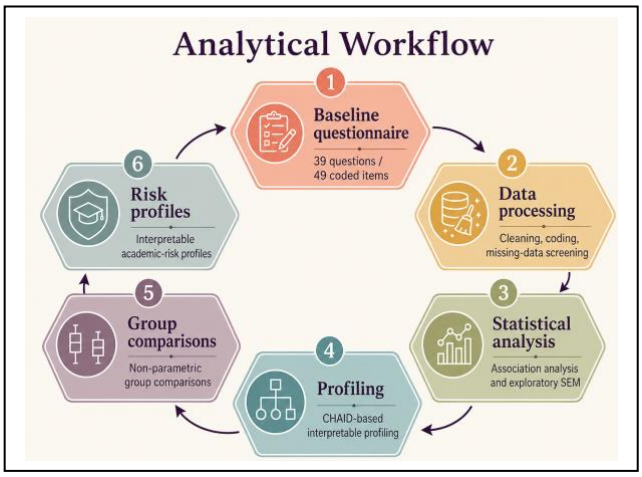


Fig. 1. Methodology for the interpretable academic-risk profiling.

These results show that early course performance varied substantially across the sample. Since the first-month grade was used as the outcome variable, the analysis focused on early academic performance, low early performance, and first-month academic risk, rather than institutional dropout. This distinction is important because the objective of the study was to generate interpretable risk profiles that could inform early academic support in foundational physics courses.

Variables Retained for Risk Profiling

The exploratory modeling process retained nine variables associated with early academic performance: semester, scholarship type, academic status, prior high-school grade average, perceived sufficiency of previous knowledge, homework completion during the senior year of high school, sleeping hours, self-evaluation of time management, and perceived importance of the subject for the student's career. These variables were retained in the final explanatory model and later used to support the interpretation of CHAID-based risk profiles. The model was re-evaluated from eleven initially significant variables to nine final explanatory variables.

Together, Table 1 and the group comparisons address RQ1 by identifying the variables associated with early academic performance. Students in the August–December 2017 cohort obtained the lowest mean grade (69.80), while students in January–May 2018 and August–December 2018 obtained mean grades of 73.96 and 75.63, respectively. Academic status also differentiated early performance, with conditional-status students obtaining a lower mean grade (63.28) than academic-support students (71.29) and regular students (74.07).

Preparation-related differences were also observed: students with a prior high-school average of 80 or below obtained a mean grade of 60.64, compared with 73.98 for those above 80, and students reporting insufficient previous knowledge obtained a lower mean grade (68.32) than those reporting sufficient knowledge (76.83).

Table 1. Main variables retained for interpretable academic-risk profiling

Variable	Evidence	Risk-profiling interpretation
Semester (V1)	Kruskal-Wallis = 12.444, p = 0.0019; SEM p = 0.003	Cohort-related variation
Scholarship type (V13)	SEM p = 0.011	Contextual compound risk
Academic status (V14)	Kruskal-Wallis = 15.339, p = 0.0004; SEM p = 0.010	Prior academic vulnerability
Prior high-school average (V19)	Mann-Whitney U = 10302, p < 0.05; SEM p < 0.001	Preparation gap
Perceived previous knowledge (V20)	Mann-Whitney U = 42739, p <= 0.05; SEM p < 0.001	Perceived preparation gap
Homework completion (V23)	Kruskal-Wallis = 31.579, p <= 0.05; SEM p = 0.022	Prior study-habit risk
Sleeping hours (V30)	Kruskal-Wallis = 19.087, p = 0.0007; SEM p = 0.020	Well-being-related risk
Time management (V47)	Kruskal-Wallis = 49.413, p <= 0.05; SEM p < 0.001	Self-regulation risk
Subject relevance (V49)	Kruskal-Wallis = 19.876, p <= 0.05; SEM p = 0.004	Motivation-relevance risk

Note. SEM = Structural Equation Modeling. The Results section discusses detailed group patterns

Behavioral and well-being-related indicators also showed relevant differences. Students reporting 61–80% homework completion in senior high school obtained a lower mean grade (63.86) than those reporting 81–100% completion (74.84). Students sleeping fewer than three hours had a low mean grade (51.38), while those sleeping between eight and ten hours had the highest mean grade (77.16). Although the group reporting more than ten hours of sleep had an even lower mean grade, this subgroup was very small and should not be overinterpreted. Time management followed a similar pattern, with students reporting bad organization obtaining lower mean grades (62.00) than those reporting excellent organization (81.77). Finally, perceived subject relevance was also associated with early performance: students who strongly disagreed that the subject was important obtained a mean grade of 64.10, compared with 75.11 among those who strongly agreed.

Scholarship type showed differentiated descriptive patterns and was retained in the SEM model. Students with sports scholarships, SEDENA scholarships, and no scholarship showed lower mean grades than several other scholarship groups; however, these subgroup patterns should be interpreted cautiously because they are context-specific and may reflect institutional conditions, selection effects, or additional unmeasured factors.

CHAID-Based Interpretable Risk Profiles

The CHAID analysis identified how selected explanatory variables combined in relation to pass/fail course

outcomes. In the original analysis, the response variable was the student grade, which was dichotomized into passed and failed categories using two alternative thresholds. Under Criterion 1, students with a grade of 75 or above were classified as passing, while those below 75 were classified as failing. Under Criterion 2, students with a grade of 70 or above were classified as passing, while those below 70 were classified as failing. In both criteria, the CHAID results showed that non-passing profiles were not explained by a single variable, but by combinations of contextual, preparation-related, behavioral, and cohort-related indicators.

Under Criterion 1, the analysis identified failure-associated profiles involving students with SEDENA or sports scholarships, students without scholarships, and students reporting insufficient prior preparation, in combination with injury status and cohort period. Additional failure-associated profiles involved students with other scholarship types who had failed subjects in high school, or students without previous failed subjects but with a high-school average of 80 or below. Under Criterion 2, similar patterns were observed, again combining scholarship type and insufficient preparation, but also including low frequency of nutritious meals before the first daily activity, previous failed subjects, cohort period, and prior high-school average.

The two thresholds were used as sensitivity criteria to examine whether the identified risk patterns remained consistent under alternative definitions of early non-passing status. The 70-point threshold reflects a more permissive early-performance criterion, whereas the 75-point threshold represents a stricter criterion for identifying students requiring early academic support.

These findings address RQ2 by showing how interpretable machine-learning profiling can identify combinations of student characteristics associated with pass/fail course outcomes.

For this research paper, the CHAID outputs were synthesized into four interpretable academic-risk profiles. These profiles should be interpreted as evidence-informed analytical labels rather than mutually exclusive operational categories. They synthesize evidence from the explanatory model, group comparisons, and CHAID terminal patterns. Therefore, a student could show indicators associated with more than one profile, and the profiles should not be used as automatic assignment rules without further validation. Instead, they provide a transparent way to summarize different combinations of risk indicators that may require different types of support:

- Preparation gap: students whose risk is mainly associated with lower prior achievement, insufficient perceived previous knowledge, or previous failed subjects identified in the CHAID profiles.

- Self-regulation difficulty: students whose risk is linked to weaker study habits, lower homework completion, or weaker time-management self-evaluation.

- Contextual compound risk: students whose risk appears through combinations of academic status, scholarship context, cohort-related variation, or other institutional/contextual indicators.

- Motivation-relevance risk: students whose lower performance is associated with weaker perceived relevance of physics to their engineering career.

Figure 2 summarizes these four interpretable academic-risk profiles and the main indicator types associated with each. Rather than reproducing the full CHAID trees, the figure provides a more accessible representation of the patterns identified in the exploratory profiling stage.

This synthesis preserves the empirical basis of the CHAID analysis while making the findings more actionable for tutors, instructors, and student-support teams. The profiles should be interpreted as exploratory and interpretable representations of academic-risk patterns, rather than as fixed or deterministic student categories.

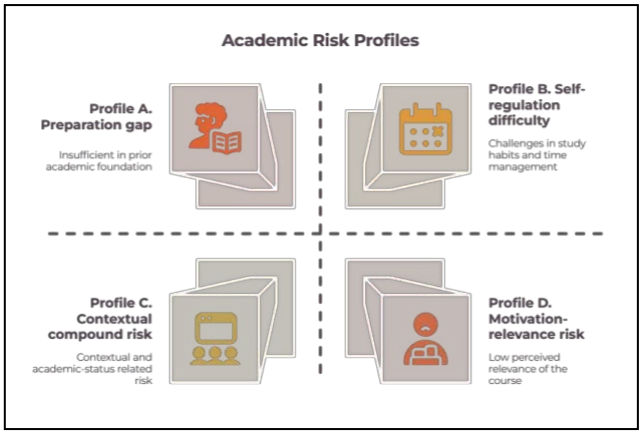


Fig. 2. Interpretable academic-risk profiles derived from explanatory and CHAID-based analysis.

Actionable Profile Summary

Table 2 translates the interpretable profiles into possible future support mechanisms. These actions are proposed as examples of how the profiles could inform AI-assisted or tutoring-oriented support. They were not implemented or evaluated in the present study.

These support mechanisms are proposed as future implementation options derived from the profiles; the present study did not implement or evaluate their effectiveness. In this sense, the study's contribution is not an operational AI system, but an interpretable empirical basis for designing future AI-assisted early-alert and academic-support strategies.

Table 2. Interpretable risk profiles and possible AI-assisted support

Risk profile	Main indicators	Possible AI-assisted / tutoring support
Preparation gap	Prior high-school average; perceived previous knowledge	Adaptive prerequisite review; subject-specialized chatbot tutor; targeted tutoring
Self-regulation difficulty	Homework completion; time management	Student-facing dashboard; SRL conversational agent; weekly study-plan nudges
Contextual compound risk	Scholarship context; academic status; cohort-related variation	Proactive advising; human-in-the-loop early alert; academic support referral
Motivation-relevance risk	Low perceived importance of physics	Career-linked examples; personalized feedback; chatbot-guided relevance prompts

Discussion

The results show that early academic performance in introductory physics courses was associated with a multidimensional set of variables rather than with a single isolated predictor. Prior academic preparation, perceived sufficiency of previous knowledge, academic status, study habits, sleep, time management, scholarship context, cohort-related variation, and perceived subject relevance were associated with early performance and informed the profiling process. This supports the view that early academic risk in first-year engineering education should be understood as the interaction of academic, behavioral, contextual, motivational, and well-being-related dimensions.

A key finding is the relevance of prior preparation. Students with lower high-school averages and those who reported insufficient previous knowledge obtained lower first-month grades. This suggests that early risk detection in foundational physics should include information about prerequisite knowledge and prior academic trajectories, not only current course performance. In practical terms, preparation-related indicators may help identify students who require diagnostic review, prerequisite reinforcement, or targeted tutoring before the course becomes more demanding.

The results also highlight the importance of self-regulation. Homework completion and time-management self-evaluation were associated with early academic performance. This finding points to a type of risk that may be addressed through structured academic support rather than only through content remediation. Students with self-regulation difficulties may benefit from study-planning tools, progress dashboards, weekly nudges, or conversational agents designed to support goal setting and reflection. This is consistent with recent work on AI-powered learning analytics dashboards, AI-supported learning tools, student-facing dashboards for self-regulated learning, and conversational agents for self-regulated learning in higher education, which emphasizes the need to connect analytics with feedback, monitoring, and learner agency [7], [8], [15], [16].

Well-being-related indicators also appeared in the explanatory and profiling process. Sleeping hours were associated with early grades, suggesting that academic risk may be linked to students' physical routines and capacity to sustain attention, study effort, and recovery. This does not imply a causal relationship. Rather, it reinforces the need to avoid purely academic interpretations of underperformance. For tutors and academic advisors, early-alert systems should support careful and human-centered conversations with students instead of automatically assigning deterministic risk labels.

The scholarship-related findings require particular caution. Scholarship context was retained in the explanatory model and showed differentiated descriptive patterns, but these patterns are institution-specific and may reflect selection effects, program conditions, or additional unmeasured factors. Therefore, scholarship context should not be used as a deterministic risk marker. Instead, it should be considered as one contextual signal within a broader human-in-the-loop advising process. This distinction matters for responsible learning analytics, especially when risk models include variables linked to student background or institutional status.

In response to RQ2, the CHAID-based synthesis contributes by translating statistical results into four interpretable academic-risk profiles: i) preparation gap, ii) self-regulation difficulty, iii) contextual compound risk, and iv) motivation-relevance risk. These profiles make the results more actionable for tutors and student-support teams. Two students with similar low grades may require different forms of support: one may need prerequisite reinforcement, while another may need time-management guidance, motivational support, or academic advising. This is the main value of interpretable profiling compared with opaque prediction: it supports differentiated action rather than a generic at-risk label.

The proposed profiles can also inform future AI-assisted academic support. Recent studies on explainable learning analytics, course-success prediction, AI-based learning tools, and AI-supported academic support suggest that educational AI should combine prediction, interpretability, feedback, and pedagogical action rather than rely only on model accuracy [8], [10]–[12]. In this context, the profiles identified in this study could inform a conceptual early-support framework that connects the data required for risk detection with interpretable profiling, traffic-light decision support, AI-assisted tutoring mechanisms, and human-in-the-loop academic governance.

Figure 3 presents a conceptual framework for translating interpretable academic-risk profiles into future AI-assisted support processes. The framework begins with multidimensional data inputs, including academic preparation, self-regulation, well-being,

contextual indicators, and motivation. These inputs are then processed through an interpretable risk-profiling layer informed by SEM-selected variables and CHAID-based profiles. The resulting profiles can support a traffic-light decision layer in which green indicates regular monitoring, amber indicates emerging risk, and red indicates priority support.

The framework also shows how traffic-light status could be connected to AI-assisted support mechanisms, such as adaptive prerequisite review, student-facing dashboards, subject-specialized chatbot tutoring, self-regulated learning conversational agents, and proactive advising prompts. However, these mechanisms should not operate as fully automated decision systems. The human-in-the-loop governance layer is therefore central to the framework, as tutors, advisors, and student-support teams would validate alerts, interpret student needs, refer students to appropriate support, and ensure ethical safeguards [17], [18].

Finally, the framework includes expected benefits and recalibration. Potential benefits include earlier support, differentiated intervention, and greater student agency. Recalibration emphasizes that academic-risk profiling should be dynamic rather than fixed: new grades, engagement data, and student feedback should update the profile and adjust the support provided. Therefore, Figure 3 should be interpreted as a design-oriented framework for future research and institutional experimentation, not as an implemented or evaluated AI system.

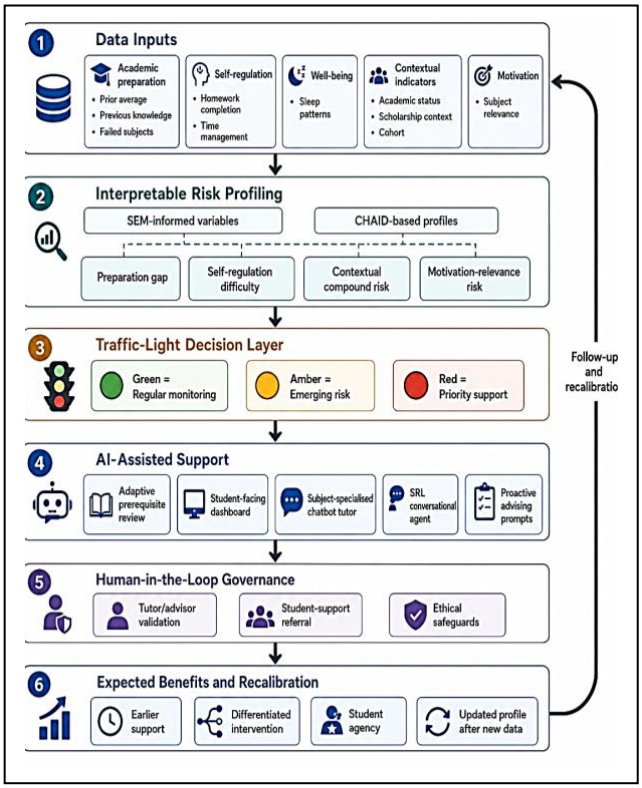


Fig. 3. Conceptual AI-assisted early-support framework derived from interpretable academic-risk profiles.

These tools are proposed only as future implementation options derived from the profiles; their effectiveness was not implemented or evaluated in the present study. Any future AI-assisted early-alert system based on these profiles should be validated with more recent cohorts, tested for fairness and calibration, and evaluated through implementation research before use in institutional decision-making [18].

Overall, this study's contribution is not the deployment of an AI system, but the development of an empirical and interpretable basis for future AI-assisted early support. By combining exploratory explanatory modeling and CHAID-based profiling, the study provides a transparent way to identify different types of academic vulnerability in first-year engineering students enrolled in introductory physics courses. This approach meets the need for learning analytics that are not only technically accurate but also understandable, actionable, and suitable for responsible educational decision-making.

Conclusions

This study examined how multidimensional student data can generate interpretable academic-risk profiles among first-year engineering students enrolled in introductory physics courses. In relation to RQ1, the results showed that early academic performance was associated with academic preparation, perceived sufficiency of previous knowledge, academic status, study habits, sleeping hours, time management, scholarship context, cohort-related variation, and perceived subject relevance. These findings reinforce the idea that early academic risk should not be interpreted as a single-variable problem, but as the interaction of academic, behavioral, contextual, motivational, and well-being-related dimensions.

In relation to RQ2, the study's main contribution is an interpretable academic-risk profiling approach that combines exploratory explanatory modeling with CHAID-based decision-tree analysis. The resulting profiles- preparation gap, self-regulation difficulty, contextual compound risk, and motivation-relevance risk- provide a transparent way to understand different forms of early academic vulnerability. This matters for tutors, instructors, and student-support teams because students with similar low grades may need different types of support.

The study also proposed a conceptual AI-assisted early-support framework based on the identified profiles. This framework links student indicators, interpretable risk dimensions, traffic-light decision support, AI-assisted support mechanisms, human-in-the-loop governance, and follow-up recalibration. However, the framework should be interpreted as a future design proposal rather than as an implemented or evaluated system. The study does not claim to have operationally deployed an AI-based early-warning system.

These findings should be interpreted in light of some methodological considerations. The dataset covers three pre-pandemic cohorts from 2017–2018 at a private

multi-campus university in Mexico, which may limit transferability to current cohorts and other institutional contexts. The study used observational and self-reported data and did not include fine-grained digital learning traces such as learning management system activity, submission patterns, or platform engagement. In addition, the authors used SEM as an exploratory explanatory procedure, and the CHAID profiles were not externally validated. Therefore, the profiles should not be interpreted as fixed student categories or as a fully deployable predictive model.

Future research should re-estimate and validate the profiles using more recent and diverse cohorts, integrate questionnaire-based indicators with behavioral learning analytics data, and compare alternative interpretable machine-learning methods. Further work should also evaluate calibration, fairness, privacy, transparency, and the educational effectiveness of AI-assisted support mechanisms such as dashboards, chatbot tutors, adaptive prerequisite review, self-regulated learning agents, and early human-in-the-loop alerts. In this way, interpretable academic-risk profiling can move from exploratory analysis toward responsible, evidence-based early support in first-year engineering education.

Acknowledgment

The authors thank Tecnológico de Monterrey, Mexico, for supporting the academic context in which this research was developed, as well as the School of Maritime Science, University of Gibraltar. The authors also thank the students and academic staff who contributed to data collection.

References

[1] S. Amin, K. E. Rambo-Hernandez, B. A. Pedersen, C. S. Burnett, B. P. Nepal, and N. V. Mendoza Diaz, “First-year engineering student persistence amidst COVID-19 interruptions: Evidence from two diverse institutions in the United States,” Cogent Education, vol. 11, no. 1, Art. no. 2413810, 2024, doi: 10.1080/2331186X.2024.2413810.

[2] J. E. Tait, L. A. Alexander, E. I. Hancock, and J. Bisset, “Interventions to support the mental health and wellbeing of engineering students: A scoping review,” European Journal of Engineering Education, vol. 49, no. 1, pp. 45–69, 2024, doi: 10.1080/03043797.2023.2217658.

[3] A. C. Emberley, D. S. Choi, T. Williams, and M. C. Loui, “Engineering survivors: Students who persisted through academic failures,” Journal of Engineering Education, vol. 113, no. 1, pp. 12–29, 2024, doi: 10.1002/jee.20564.

[4] S. M. C. Pinheiro, K. P. Oliveira-Esquerre, M. A. F. Martins, and R. Oliveira, “Student performance in first-year math and physics courses as predictor of student dropout in engineering programs,” International Journal of Engineering Education, vol. 37, no. 2, pp. 471–481, 2021.

[5] V. J. Abuchar, J. De La Hoz, C. Vieira, and C. Arteta, “Risk-based student performance prediction model for engineering courses,” Computer Applications in Engineering Education, vol. 32, no. 5, Art. no. e22757, 2024, doi: 10.1002/cae.22757.

[6] C. R. Çırak, H. Akıllı, and Y. Ekinci, “Development of an early warning system for higher education institutions by predicting first-year student academic performance,” Higher Education Quarterly, vol. 78, no. 4, Art. no. e12539, 2024, doi: 10.1111/hequ.12539.

[7] L. Cabral, R. Pinto, and G. Gonçalves, “AI-powered learning analytics dashboards: A systematic review of applications, techniques, and research gaps,” Discover Education, vol. 4, Art. no. 525, 2025, doi: 10.1007/s44217-025-00964-y.

[8] J. Luo, C. Zheng, J. Yin, and H. H. Teo, “Design and assessment of AI-based learning tools in higher education: A systematic review,” International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 42, 2025, doi: 10.1186/s41239-025-00540-2.

[9] M. Á. Rodríguez-Ortiz, P. C. Santana-Mancilla, and L. E. Anido-Rifón, “Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges,” Applied Sciences, vol. 15, no. 15, Art. no. 8679, 2025, doi: 10.3390/app15158679.

[10] S. Gunasekara and M. Saarela, “Systematic literature review on explainable learning analytics and educational data mining,” IEEE Access, vol. 13, pp. 214387–214408, 2025, doi: 10.1109/ACCESS.2025.3643645.

[11] E. Tiukhova, P. Vemuri, N. López Flores, A. S. Islind, M. Óskarsdóttir, S. Poelmans, B. Baesens, and M. Snoeck, “Explainable learning analytics: Assessing the stability of student success prediction models by means of explainable AI,” Decision Support Systems, vol. 182, Art. no. 114229, 2024, doi: 10.1016/j.dss.2024.114229.

[12] B. Ujkani, D. Minkovska, and N. Hinov, “Course success prediction and early identification of at-risk students using explainable artificial intelligence,” Electronics, vol. 13, no. 21, Art. no. 4157, 2024, doi: 10.3390/electronics13214157.

[13] R. B. Kline, Principles and Practice of Structural Equation Modeling, 4th ed. New York, NY, USA: Guilford Press, 2016.

[14] G. V. Kass, “An exploratory technique for investigating large quantities of categorical data,” Applied Statistics, vol. 29, no. 2, pp. 119–127, 1980, doi: 10.2307/2986296.

[15] L. de Vreugd, A. van Leeuwen, R. Jansen, and M. van der Schaaf, “Learning analytics dashboard design and evaluation to support student self-regulation of study behavior,” Journal of Learning Analytics, vol. 11, no. 3, pp. 249–262, 2024, doi: 10.18608/jla.2024.8529.

[16] J. W. Lai, W. Qiu, M. Thway, L. Zhang, N. B. Jamil, C. L. Su, S. S. H. Ng, and F. S. Lim, “Leveraging process-action epistemic network analysis to illuminate student self-regulated learning with a Socratic chatbot,” Journal of Learning Analytics, vol. 12, no. 1, pp. 32–49, 2025, doi: 10.18608/jla.2025.8549.

[17] S. Buckingham Shum, R. Martínez-Maldonado, Y. Dimitriadis, and P. Santos, “Human-centered learning analytics: 2019–24,” British Journal of Educational Technology, vol. 55, no. 3, pp. 755–768, 2024, doi: 10.1111/bjet.13442.

[18] C. Patterson, E. York, D. Maxham, R. Molina, and P. Mabrey, “Applying a responsible innovation framework in developing an equitable early alert system: A case study,” Journal of Learning Analytics, vol. 10, no. 1, pp. 24–36, 2023, doi: 10.18608/jla.2023.7795.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Comparing AI-Generated and Teacher-Generated Lessons in Moodle

Sima Darabi

Department of management and marketing, Bratislava University of economics and management, Slovakia

Article DOI: https://doi.org/10.68286/samyrad2026.399

ABSTRACT

This study examines the pedagogical quality and student acceptance of AI-generated lesson materials produced by Moodle's integrated AI tool. It compares an AI-generated moodle course with a teacher-generated course on the same computer science topic, using the Technological Pedagogical Content Knowledge (TPACK) framework to evaluate the AI-generated lesson against the teacher-generated one. The main research question asks to what extent AI-generated lesson texts differ from teacher-generated lesson materials in their integration of technological, pedagogical, and content knowledge. A comparative qualitative content analysis based on Mayring's deductive–inductive approach was applied as the methodology. CK, PK, and TK served as deductive main categories, while subcategories were developed from students' open-ended questionnaire responses. The study involved 14 students, including 10 primary participants and 4 supplementary validation participants; a second teacher reviewed the questionnaire and analytical categories. The findings show that students perceived the AI-generated lessons as concise, clear, logically structured, and useful for revision or examination preparation. However, they considered teacher-generated materials more detailed, context-sensitive, and suitable for first-time learning. In the pedagogical dimension, AI-generated materials lacked sufficient scaffolding, practical examples, and real-world relevance. In the content dimension, they were accurate but superficial rather than in-depth. In the technological dimension, AI-generated images were often blurry, unreadable, and misleading, reducing students' trust in the materials. The study concludes that Moodle-integrated AI can support teachers in creating summaries, revision materials, and initial lesson drafts, but does not yet replace teachers' expertise in pedagogical design, subject depth, contextual adaptation, and reliable multimedia development.

KEYWORDS AI in education, moodle AI tool, TPACK

How to cite: Darabi, S. (2026). Comparing AI-generated and teacher-generated lessons in Moodle. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.399

Introduction

Artificial intelligence (AI) and Massive Open Online Courses (MOOCs) are increasingly reshaping the international educational landscape. As noted by Reich and Ruipérez-Valiente (2019), MOOCs enable universities, schools, companies, and public institutions to offer digital learning opportunities to large numbers of learners regardless of their location. They have expanded access to lectures, assessments, resources, and learning communities beyond traditional face-to-face classrooms. However, the authors emphasize that access alone does not automatically ensure educational quality or successful learning outcomes [1]. According to Kasneci et al. (2023) Generative AI tools have further increased the relevance of AI for online learning. Platforms such as ChatGPT, Google Gemini, Claude, and LMS-integrated AI tools can generate texts, questions, feedback, and images. In Moodle-based courses, such tools can support the creation and revision of learning material directly within the course environment. This may help teachers develop content more efficiently, especially in large-scale online courses [2]. As highlighted by Zawacki-Richter et al. (2019), these opportunities and challenges become particularly significant when AI is

embedded directly in a learning management system (LMS) such as Moodle. In this context, AI-generated material becomes part of a structured course that includes learning objectives, resources, exercises, assessments, and communication. Therefore, the key question is not only whether AI can generate a text or image, but also whether its output supports the instructional purpose of the course. AI-generated materials should be accurate, understandable, accessible, relevant to the curriculum, and appropriate for learners' prior knowledge and learning needs [3]. To address these requirements systematically, the next section discusses the methodological aspects of digital learning through the Technological Pedagogical Content Knowledge (TPACK) framework. This framework provides a suitable basis for evaluating how AI-generated materials can be integrated effectively into Moodle-based learning environments.

TPACK Framework

To evaluate whether these requirements are fulfilled, the Technological Pedagogical Content Knowledge (TPACK) framework provides an appropriate theoretical perspective. Building on Lee Shulman's earlier model of pedagogical content knowledge (PCK), Mishra and Koehler (2006) developed the TPACK framework, arguing that meaningful technology integration requires a situated and context-sensitive knowledge base rather than technical skill alone. Since its introduction, TPACK has become one of the most influential frameworks for analyzing technology-enhanced teaching and for evaluating the quality of digital lesson design [4]. Regarding AI-generated learning materials, content knowledge is particularly important for assessing the quality and instructional suitability of the output. As noted by Shulman (1986), content knowledge concerns the correctness, depth, and organization of subject matter. Accordingly, AI-generated lessons should be evaluated according to whether they provide sufficient technical detail, clear explanations, relevant examples, and logically connected concepts [5].

Research Gap

According to Le Tan et al. (2025), the use of artificial intelligence (AI) in education is expanding rapidly, ranging from intelligent tutoring systems and adaptive learning systems to generative tools that support teachers and students. Their systematic review indicates that AI can facilitate personalization and learning analytics; however, it also highlights the need for further research on pedagogical quality and ethical use in authentic educational settings [6]. As reported by Zhu et al. (2026), teachers recognize the potential of AI tools in education but remain cautious about issues of control, responsibility, and alignment with their own didactic beliefs. Their findings suggest that acceptance of AI-generated materials depends strongly on perceived didactic fit, transparency, and teachers' ability to adapt and critically evaluate AI-generated outputs [7]. Similarly, as noted by Stadler et al. (2020), research on AI in education has often focused on adaptive learning, automated assessment, and stand-alone tutoring systems. In contrast, the integration of AI into mainstream learning management systems (LMSs), such as Moodle, has received comparatively limited detailed attention [8]. As Zawacki-Richter et al. (2019) noted, AI applications in education include intelligent tutoring systems, learning analytics, adaptive learning systems, and generative AI tools. These applications can support teachers in preparing instructional materials Nevertheless, research indicates that the use of AI in higher education has often focused more on technological possibilities than on teachers' pedagogical roles and classroom-related needs [3]. Finally, as demonstrated by Koraishi and Karatepe (2025), AI-generated materials can effectively summarize and restate key points but tend to underperform in areas requiring nuanced pedagogical judgement. These areas include the formulation of clear learning objectives and the adaptation of content to actual classroom conditions [9].Taken together, these findings reveal a research gap concerning the pedagogical quality, instructional design, and acceptance of AI-generated materials created directly within LMS environments such as Moodle.

The Aim of the Study

The identified research gap gives rise to three interrelated needs:

Need to characterize didactic quality differences: As highlighted by Henderson et al. (2025), studies comparing generative AI outputs with teacher-generated inputs demonstrate that, although AI can provide rapid and extensive feedback or content, its outputs are often perceived as less context-sensitive, less reliable, and less closely aligned with course-specific learning objectives. Teacher-generated materials, by contrast, typically reflect a deeper understanding of curriculum requirements, students' backgrounds, and classroom culture. These factors influence the selection, organization, and sequencing of lesson content [10].

Need to measure teacher perceptions and acceptance: Research by Fan et al. (2025) suggests that teachers' beliefs, values, and perceived self-efficacy strongly influence whether and how they adopt AI tools in their teaching practices. Research on AI literacy and competence frameworks further indicates that acceptance is shaped by perceived usefulness, perceived ease of use, ethical concerns, and consistency with professional identity. Therefore, teachers may be more likely to accept AI-generated content when it supports, rather than undermines, their didactic autonomy [11]. Although teachers' beliefs, values, perceived self-efficacy, and professional autonomy may influence their adoption of AI tools in teaching practices, teacher acceptance is not the focus of the present study. Instead, this study concentrates on students' perceptions and acceptance of AI-generated lesson content in Moodle. Need for validated instruments and comparative methodology: As noted by Barcelona and Dela Cruz (2025), research on the

development and validation of scales for AI-related constructs in education emphasizes the importance of rigorous instrument development. This process combines qualitative exploration with psychometric testing to ensure instrument reliability and validity [12].

Based on the findings of these studies, the present research is motivated to conduct a systematic comparison between AI-generated and teacher-generated lessons addressing the same topic within a Moodle-based course. Specifically, the study seeks to analyze differences in didactic structure and instructional methods, assess the perceived pedagogical and methodological quality of AI-generated lesson materials, and investigate teachers' and students' acceptance of AI-produced resources. Accordingly, the aim of this study is to analyze differences in the didactic structure and instructional methods of AI-generated and teacher-generated lesson materials. Furthermore, it aims to assess the perceived pedagogical and methodological quality of AI-generated materials and investigate teachers' and students' acceptance of these AI-produced resources

Research Questions

Based on the identified research gap and research needs, the following research questions were developed. The main research question is grounded in the Technological Pedagogical Content Knowledge (TPACK) framework. As argued by Ning et al. (2024), effective technology-integrated teaching does not result from treating technological, pedagogical, and content knowledge as separate competencies. Rather, it depends on the dynamic interaction among these domains and on an understanding of how they intersect within a specific teaching and learning context. This perspective is particularly relevant to the use of AI-generated lesson materials in Moodle, as their educational quality depends not only on the technology used but also on how appropriately content, pedagogy, and technology are integrated [13].

Research Question: To what extent do AI-generated lesson texts produced by Moodle's integrated AI tool differ from teacher-generated lesson texts in their integration of technological, pedagogical, and content knowledge (TPACK), and how do students perceive and accept these AI-generated materials?

This research question therefore examines whether AI-generated lesson material demonstrate an appropriate integration of the three central TPACK domains—technological knowledge, pedagogical knowledge, and content knowledge—in comparison with teacher-generated texts. It further investigates students' perceptions and acceptance of these AI-generated materials within the Moodle-based learning environment.

Research Question a: What qualitative differences exist between AI-generated and teacher-generated lesson materials in how they integrate content knowledge, pedagogical knowledge, and technological knowledge, as conceptualized by the TPACK framework?

Research Question b: How do students evaluate the pedagogical adequacy and usability of AI-generated lesson material in Moodle, particularly regarding their perceived integration of content, pedagogy, and technology?

Research Question C: which TPACK-related factors do students associate with their acceptance of AI-generated material?

The research questions are addressed through a comparative qualitative content analysis based on Mayring's deductive–inductive approach. The three core domains of the TPACK framework—content knowledge, pedagogical knowledge, and technological knowledge—serve as the deductive main categories. Additional subcategories are developed inductively through the comparison of AI-generated and teacher-generated course materials, as well as through the analysis of students' responses to open-ended questionnaire questions. This approach combines a theory-driven analysis of the lesson materials with insights derived from students' actual learning experiences.

Relevant Literature

The Technological Pedagogical Content Knowledge (TPACK) framework provides the theoretical foundation for examining effective technology integration in teaching. In the present study, it is used to compare AI-generated and teacher-generated moodle lesson texts. Rather than evaluating AI-generated material only in terms of its technical production, the framework makes it possible to examine whether the material integrates accurate subject content, suitable pedagogical strategies, and appropriate technological features in a coherent manner. Valtonen et al. (2023) conceptualize TPACK as the interaction of technological, pedagogical, and content knowledge, rather than the isolated application of technological skills. This framework is relevant to the present study because the educational quality of AI-generated moodle lesson texts depends not only on their technical generation, but also on their capacity to support learning and represent disciplinary content accurately [14]. Ali and Hawk (2024) emphasize that TPACK is context-dependent because teachers' instructional decisions must account for learners' needs, disciplinary requirements, available technologies, and institutional conditions. Their mixed-method study further indicates that teachers' knowledge of students' cultural backgrounds is associated with their perceptions of TPACK, highlighting the importance of learner context in technology-supported instructional design [15]. Engin et al. (2023) demonstrate that the development of TPACK requires more than familiarity with digital tools. Lesson planning, microteaching, reflection, and the revision of technology-enhanced learning activities can support teachers in connecting technological choices with pedagogical and subject-specific decisions [16].

The following figure illustrates the TPACK framework, which describes the integration of teachers' technological knowledge (TK), pedagogical knowledge (PK), and content knowledge (CK).

Koehler and Mishra (2009) describe TPACK as the meaningful integration of technological, pedagogical, and content knowledge within a specific teaching context. In this study, these three domains provide the analytical basis for comparing AI-generated and teacher-generated moodle lesson texts: pedagogical knowledge is used to assess explanations, examples, scaffolding, and learner support; content knowledge is used to examine accuracy, depth, and subject-specific relevance; and technological knowledge is used to evaluate the suitability and reliability of Moodle-integrated AI functions and generated outputs [17].

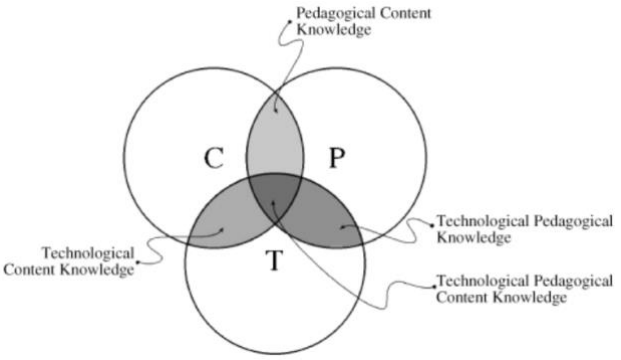


Fig. 1 Technological Pedagogical Content Knowledge (Koehler & Mishra, 2009; Source: Koehler & Mishra, 2008)

Pedagogical Knowledge

Valtonen et al. (2023) show that teachers' technological, pedagogical, and content-related knowledge may develop at different rates. This finding indicates that the ability to use a digital tool does not automatically ensure the pedagogical capacity to structure learning, provide scaffolding, or support meaningful student engagement. For the present study, pedagogical knowledge is therefore relevant for assessing whether AI-generated moodle lesson texts offer clear explanations, appropriate learning sequences, concrete examples, and adequate support for learners [14]. Abouelenein and Selim (2024) found that digital interventions such as interviews, reflective reports, and video simulations can support the development of pre-service teachers' TPACK. Their findings suggest that the pedagogical value of technology depends on teachers' ability to reflect on, evaluate, and adapt instructional decisions to learners' needs. This is relevant to AI-generated moodle materials because teachers must evaluate and revise AI-generated texts when explanations, examples, or scaffolding are insufficient [18]. Mekheimer (2025) argues that effective technology integration in higher education requires the combined consideration of technological knowledge, pedagogical knowledge, and content knowledge. Pedagogical knowledge is particularly important because it guides the selection of teaching strategies, learning activities, feedback methods, and assessment practices that make technology-supported learning meaningful [19]. Simon et al. (2025) highlight that learning-management-system use in higher education should be evaluated in relation to teachers' and students' educational experiences. moodle functions should therefore be selected and organised to support pedagogical purposes, such as interaction, task completion, feedback, and learner progress, rather than being used solely because they are technically available [20]. Koraishi and Karatepe (2025) compare AI-generated and teacher-generated summaries, providing a closely related basis for the present study. Their comparison shows that the quality of AI-generated educational text should not be evaluated only in terms of efficient content generation or broad coverage; it should also be assessed in relation to pedagogical depth, contextual relevance, and suitability for learners. The present study extends this comparison by examining AI-generated and teacher-generated moodle lesson texts rather than summaries alone [9]. Henderson et al. (2025) compare generative-AI and teacher feedback from students' perspectives, focusing on usefulness and trustworthiness. Their study is relevant because it demonstrates that students' evaluations of AI-generated material may depend not only on whether the content is available or useful, but also on whether it is perceived as trustworthy, contextually appropriate, and educationally supportive. The present study expands this perspective from feedback to moodle lesson texts embedded in a course environment [10].

Content Knowledge

Velander et al. (2024) reports that teachers' AI-related content knowledge may be affected by misconceptions and incomplete understanding, particularly when knowledge is acquired through informal or incidental learning. Their findings emphasize the importance of accurate subject knowledge when teachers introduce complex digital and AI-related topics in educational settings. In the present study, content knowledge is therefore used to assess whether AI-generated moodle lesson texts provide accurate, sufficiently detailed, and relevant subject content [21]. Mailizar et al. (2026) found that generative AI can support the development of mathematics teachers' TPACK, particularly in relation to pedagogical strategies. However, their findings also indicate that teachers require adequate subject knowledge to evaluate, adapt, and use AI-generated resources appropriately in their lessons. This supports the assumption that teacher expertise remains necessary to review the accuracy, depth, and educational appropriateness of AI-generated moodle lesson texts [22]. Amiruddin et al. (2026) emphasize that technology-integrated teaching competence requires subject-specific approaches. Content knowledge should therefore not be understood as general factual information alone; rather, it must be connected to the concepts, practices, and epistemological characteristics of the relevant discipline. This perspective guides the present study's assessment of whether AI-generated moodle lesson

texts provide content that is appropriate for the specific course context [23]. Koraishi and Karatepe (2025) demonstrate the importance of examining AI-generated texts in relation to teacher-generated material. Their comparison of summaries suggests that factual coverage alone is not sufficient for educational quality; the depth, contextualization, and relevance of the content must also be considered. This supports the present study's focus on students' perceptions of conceptual depth, technical detail, and the need for additional teacher-generated explanations in moodle lesson texts [9].

Technological Knowledge

Aji et al. (2026) link teachers' readiness for AI integration to AI-specific TPACK and attitudes towards AI. Their findings suggest that technological knowledge includes understanding AI tools' functions, limitations, and appropriate educational applications, as well as the ability to evaluate whether their outputs are suitable for teaching and learning. In the present study, this dimension is relevant for examining the quality and educational suitability of AI-generated moodle lesson texts and visual materials [24]. Joshi (2026) uses the Intelligent-TPACK framework to highlight the relationship between AI knowledge, pedagogical strategies, and ethical decision-making in university teaching. This perspective indicates that technological knowledge in AI-supported learning involves not only operational competence but also critical awareness of responsible use, ethical implications, and the limitations of AI-generated outputs [25]. Ergo et al. (2026) identify institutional and professional factors that influence the implementation of TPACK in teacher-education institutions. Their findings suggest that technological knowledge develops more effectively when educators have access to relevant technologies, structured support, and opportunities to apply digital tools in authentic teaching contexts. This is relevant to Moodle-based AI integration because the availability of AI functions alone does not ensure meaningful learning; teachers must also be able to configure, evaluate, and integrate them appropriately into course design [26]. Henderson et al. (2025) show that the evaluation of AI-generated educational support should include students' perceptions of trustworthiness as well as usefulness. Their comparison of AI and teacher feedback indicates that technological quality is not limited to speed or availability, but also includes the reliability, transparency, and contextual appropriateness of AI-generated outputs. This perspective is applied in the present study when evaluating students' perceptions of Moodle-integrated AI-generated materials [10].

AI in Education

Artificial intelligence (AI) is increasingly shaping educational practice through adaptive learning systems, intelligent tutoring, chatbots, learning analytics, automated assessment, and generative tools. These applications can personalize learning, support learner engagement, and assist teachers with instructional tasks and the creation of teaching materials [27] [6]. In LMS environments such as Moodle, AI functions can provide more interactive and adaptive support for students and instructors. Generative AI may also reduce teachers' planning workload by helping them create lesson plans, examples, and resources. Nevertheless, teachers remain concerned about ethical issues, loss of control, and whether AI-generated content reflects appropriate pedagogical reasoning and curricular objectives [28] [29]. Therefore, AI-supported educational tools should be evaluated not only for technical accuracy but also for their didactic quality, alignment with learning outcomes, integration into instructional workflows, and support for teacher agency. Although research has extensively examined adaptive systems and stand-alone AI applications, comparatively little is known about AI text-generation modules embedded in mainstream LMSs and their use in everyday course development [8][30]. TPACK offers an established framework for analyzing the integration of technology, pedagogy, and content in lesson design [4].

Research Methodology

This study adopts Philipp Mayring's approach to qualitative content analysis as its methodological foundation. Mayring's qualitative content analysis is a rule-guided, systematic procedure in which text material is examined through clearly defined units of analysis, and categories are developed either inductively from the data or deductively from existing theory, then applied using explicit coding rules to ensure transparency and reliability [31]. According to Barcelona et al. (2025), in the context of Moodle's AI integration, there is a need for a research design that uses systematic content analysis to compare AI- and teacher-generated texts, combined with a validated questionnaire or mixed-methods protocol to measure teachers' acceptance and didactic evaluations [12]. Figure 2 illustrates the steps of Mayring's qualitative content analysis.

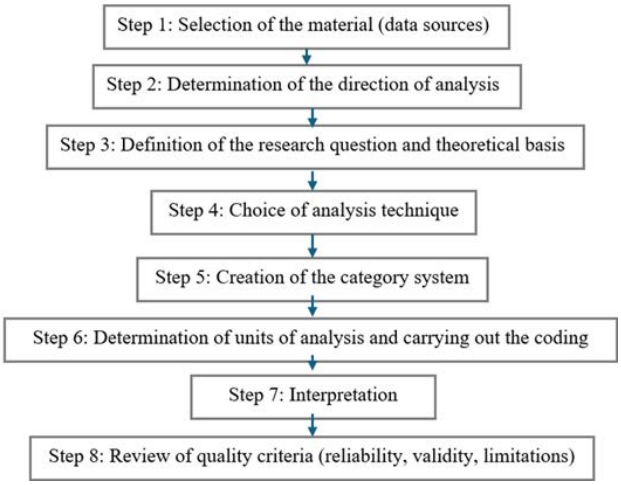


Fig. 2 Steps of qualitative content analysis, adapted from Mayring (2019)

The following subsections describe the study in terms of Mayring's eight procedural steps:

Data Sources:

From moodle version 4.5 onward, administrators can integrate AI services, such as OpenAI in moodle. Figure 3 shows an AI-generated course created by the author in Moodle. The empirical material comprised two versions of a computer science course implemented in moodle at SRH Neckargemünd, Germany in June 2026. The first version was AI-generated: all textual content, images, and exercises were created using Moodle's integrated AI tool. The second version was teacher-generated and developed by a single instructor, drawing on her subject knowledge, existing teaching materials, and self-designed exercises.

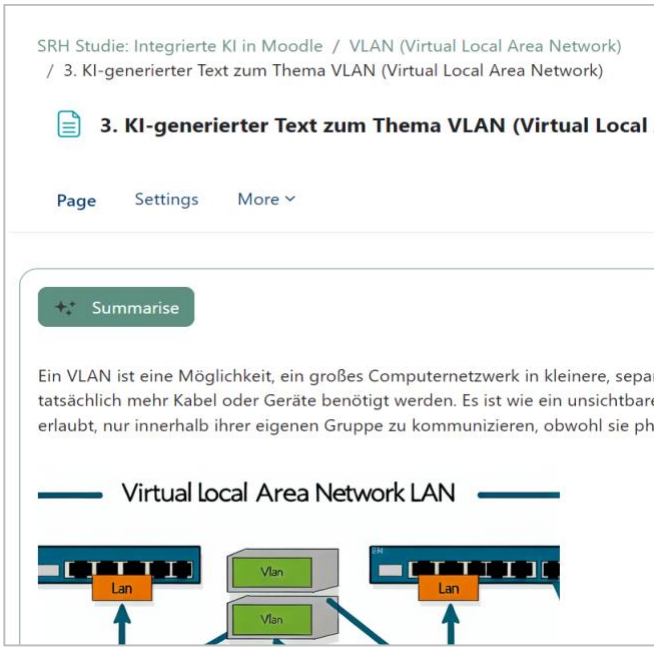


Fig. 3 AI-generated course created by the author in moodle

Both course versions were completed by 10 students (8 male and 2 female). After completing the courses, the students filled in a questionnaire evaluating the didactic quality, trustworthiness, and acceptance of the materials. To examine the consistency of responses across a different instructional context, the same questionnaire was also administered to an additional class of four students taught by a different teacher. The questionnaire consisted of three sections, each containing four open-ended questions. It was designed to address the research question by exploring students' experiences of AI-supported learning materials, their perceptions of the didactic quality of AI-generated content, and their comparisons between AI-generated and teacher-generated materials.

- Section 1: Learning Experience with the AI-Supported Course: This section explored students' overall experiences with the AI-supported moodle course. Participants were asked to describe their general impressions of the course, identify the aspects they found most helpful for learning, and indicate whether any AI-generated content was confusing or unclear. The section also examined whether the structure and organization of the AI-generated lessons supported or hindered their understanding.
- Section 2: Perceived Didactic Quality**:** This section investigated students' perceptions of the didactic quality of the AI-generated materials. Participants were asked to assess how effectively the materials explained the selected topic, such as switching, and whether the lessons included clear learning objectives. They were also asked how clearly these objectives were communicated or recognizable. In addition, participants evaluated whether the content followed a logical instructional sequence and identified any elements they believed were missing from a learning perspective.
- Section 3: Comparison with Teacher-Generated Materials: This section compared the AI-generated materials with teacher-generated lessons. Participants were asked to evaluate the two types of material, particularly in terms of clarity, depth, and structure. They were also asked which type of material supported their learning more effectively and why. Finally, participants were invited to identify situations in which they would prefer AI-generated materials and situations in which they would prefer teacher-generated materials.

Data Analysis: Link the TPACK framework to the Category System

The study applies comparative qualitative content analysis in a mixed deductive–inductive design. The direction of analysis is content structuring: predefined main categories derived from theory (TPACK) structure the material, while subcategories are refined inductively from students' questionnaire responses. A mixed-methods design was chosen because it combines theoretical grounding with attention to students' actual learning experiences. A purely deductive approach would only assess whether the lessons formally address the three domains of TPACK, without capturing how students perceived and experienced the materials. In contrast, a purely inductive approach could reduce the connection to established theory and limit comparability with previous TPACK research. Therefore, the deductive stage uses the established TPACK categories as a framework for coding, while the inductive stage further refines these categories based on students' perceptions of clarity, trust, and acceptance. In line with the TPACK framework, three main deductive categories were defined for the content analysis: Content Knowledge (CK), Pedagogical Knowledge (PK), and Technological Knowledge (TK). After coding the lesson texts and analyzing the students' questionnaire

responses, several inductive subcategories emerged, which can be systematically assigned to these three domains.

First, the use of deductive main categories (CK, PK, TK) is theoretically justified, as these dimensions are well-established components of the TPACK model and provide appropriate analytical lens for examining technology-enhanced teaching. Applying these predefined categories ensures theoretical alignment and construct validity, as the analysis directly reflects the conceptual framework underpinning the study.

Three deductive main categories were defined:

- Pedagogical Knowledge (PK)
- Content Knowledge (CK)
- Technological Knowledge (TK)

Second, the development of inductive subcategories from the empirical material allows for a data-driven refinement of these broad domains. By deriving subcategories from students' questionnaire responses and lesson analyses, the study captures nuanced perceptions that may not be fully anticipated by the theoretical model. This enhances the content validity and sensitivity of the analysis, as participants' perspectives are systematically integrated rather than forced into rigid predefined structures.

Within these, inductive subcategories were developed from the empirical material:

PK subcategories:

- Perceived clarity, brevity, and accessibility of texts
- Deficits in depth, examples, and scaffolding
- Recognition of basic learning objectives
- Importance of examples, use-cases, and real-world anchoring

CK subcategories:

- Perception of superficiality and missing technical detail
- AI as concise summary vs. teacher as in-depth explainer

TK subcategories:

- Visual (in)effectiveness and quality of AI-generated images
- Deficient visualization and multimodality
- Perceived limitations of the specific Moodle-integrated AI model
- Differences in clarity, multimodality, and adaptivity

These subcategories reflect students' nuanced perceptions and allow the TPACK framework to be empirically examined and, where necessary, adjusted in light of actual learner evaluations.

Category System

According to Byrne (2022), a deductive research approach is a theory-driven, "top-down" method in which the researcher derives propositions or hypotheses from an established theoretical framework. In contrast, an inductive research approach is a data-driven, "bottom-up" method in which researchers derive codes and categories from participants' accounts [32]. In this research, the deductive categories were guided by the TPACK framework, while each category consisted of inductively derived subcategories that emerged from students' questionnaire responses.

Deductive Main Category 1: Pedagogical Knowledge:

Deductive Category 1 consists of four inductive subcategories:

Inductive Subcategory 1: Perceived clarity, brevity, and accessibility of texts: Several students described the AI-generated texts as clear, concise, structured, and accessible. The material was perceived as suitable for obtaining a basic overview and, in some cases, for learners with little prior knowledge. Students also appreciated the logical organization of the content and the possibility of learning at their own pace.

Inductive Subcategory 2: Deficits in depth, examples, and scaffolding: Although students valued the concise presentation, many considered the AI-generated materials too superficial for comprehensive learning. They reported missing concrete examples, step-by-step explanations, detailed guidance, and sufficient depth. Therefore, the materials were perceived as more appropriate for revision or introductory orientation than as a complete resource for students who were new to the topic.

Inductive Subcategory 3: Recognition of basic learning objectives and logical sequencing: Students identified basic learning objectives in the AI-generated lessons, including an introduction to switching, VLANs, and MAC address tables. The sequence of the content was generally perceived as logical, typically moving from basic definitions to explanations, visualizations, and examples. However, some students believed that the limited scope of the materials prevented the learning objectives from being fully developed.

Inductive Subcategory 4: Importance of examples, use-cases, and real-world anchoring: Students repeatedly emphasized the instructional value of concrete examples, everyday applications, and metaphors. Examples were considered important because they enabled learners to connect abstract technical concepts with practical situations. However, students also noted that examples were insufficiently detailed or inconsistently provided in the AI-generated materials.

Deductive Main Category 2: Content Knowledge:

Deductive category 2 consists of two inductive subcategories:

Inductive Subcategory 1: Perception of superficiality and missing technical detail: Students frequently perceived the AI-generated content as factually basic but insufficiently detailed. They particularly criticized the absence of technical processes, detailed explanations, and procedural descriptions necessary for a full understanding of switching and VLANs. The materials were therefore considered inadequate as a stand-alone learning source for complex technical topics.

Inductive Subcategory 2: AI as a concise summary versus the teacher as an in-depth explainer: Students commonly differentiated between the functions of AI-generated and teacher-generated learning materials. AI-generated material was perceived as concise, structured, and useful for revision, preparation, or gaining a quick overview. Teacher-generated content was perceived as more detailed, more informative, more adaptable to learners' prior knowledge, and more appropriate for developing a deeper conceptual understanding.

Deductive Main Category 3: Technological Knowledge:

Deductive category 3 consists of four inductive subcategories:

Inductive Subcategory 1: Visual effectiveness and quality of AI-generated images: The visual quality of the AI-generated images was a major concern for students. Images were frequently described as blurred, unreadable, misleading, or technically incorrect. Rather than supporting understanding, the visuals often cause confusion, particularly in relation to the MAC address table and VLAN diagrams.

Inductive Subcategory 2: Deficient visualization and multimodality: Students considered visualization essential for understanding technical processes. They perceived the AI-generated materials as lacking coherent visual support, while teacher-generated videos were seen as more effective because they integrated animations, readable graphics, spoken explanations, and step-by-step demonstrations.

Inductive Subcategory 3: Perceived limitations of the specific Moodle-integrated AI model: Several students differentiated between their evaluation of the particular Moodle-integrated AI tool and their views on AI more generally. They attributed many shortcomings—especially poor images, limited content, and low trust—to the perceived outdatedness or technical limitations of the specific system used in the study rather than to all current AI technologies.

Inductive Subcategory 4: Differences in clarity, multimodality, and adaptivity: Students perceived differences between AI-generated and teacher-generated materials regarding clarity, multimedia design, and responsiveness to individual learner needs. Although AI-generated materials were described as structured and efficient, teacher-generated videos were generally seen as clearer, more visually coherent, more detailed, and better adapted to learners' needs. Students also recognized the potential of interactive AI to provide individualized learning support when learners can ask follow-up questions.

Review of quality criteria (reliability, validity, limitations)

According to Klock et al. (2021), the quality criteria of empirical research refer to three core standards—objectivity, reliability, and validity—used to evaluate the rigor of a research instrument. Objectivity means that the results are independent of the researcher and is achieved through objectivity in implementation, evaluation, and interpretation. Reliability refers to the consistency or precision of measurement: an instrument is reliable if it produces the same results under the same conditions, regardless of what it measures. Validity concerns whether an instrument actually measures what it claims to measure and includes content, criterion-related, and construct validity [33]. To enhance the reliability and validity of the instrument, the same course was implemented in an additional class consisting of four students and conducted by a different teacher. The same questionnaire was administered in this setting to examine whether the results remained consistent across different participants and instructional contexts.

Reliability: In the study, the similar response patterns in the main study and the additional class provide preliminary evidence of consistency across groups: both groups showed conditional acceptance and similar concerns about depth, examples, scaffolding, and teacher support.

Validity: The results support validity because the instrument differentiated meaningful aspects of students' perceptions of AI-generated moodle materials: pedagogical adequacy, content depth and accuracy, technological quality, usability, and acceptance. These categories align with the intended TPACK-related dimensions, suggesting that the instrument captured the construct under investigation.

Objectivity: Objectivity was supported using clearly defined categories and systematic coding rules; however, it would be strengthened by independent second-coder review or an inter-coder agreement assessment.

Findings

Quantitative Result

Overall Student Perception of AI-Generated moodle Materials:

In the main study, the sample consisted of N=10 students. An additional class of N=4 students, taught by

a different teacher, participated to provide additional evidence regarding the reliability and validity of the instrument. Figure 4 shows that most students in the main study (n=6) and in the additional class (n=3) accepted the AI-generated moodle materials only conditionally. They considered the materials useful for gaining a basic overview, reviewing key concepts, or supplementing other learning resources, but not as a complete substitute for teacher-generated instruction. A further n=3 students in the main study and n=1 student in the additional class evaluated the materials mainly positively. In contrast, n=1 student in the main study expressed predominantly critical or low acceptance.

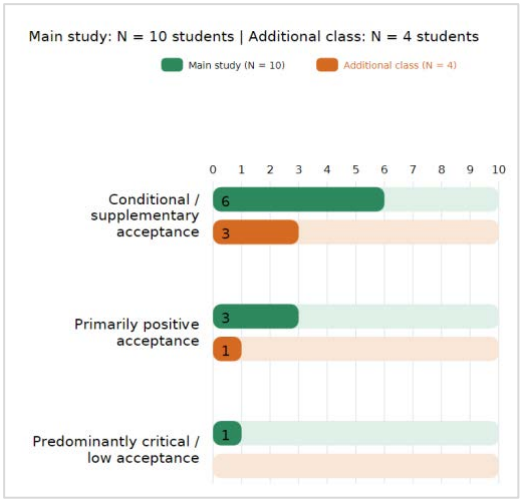


Fig. 4 Overall Student Perception of AI-Generated moodle Materials

Dominant Perceived Difference Between AI- and Teacher-Generated Materials

Figure 5 indicates that n= 6 students in the main study and n=2 students in the additional class perceived teacher-generated materials as superior in terms of conceptual depth, concrete examples, and visual support. A further n=3 students in the main study and n=2 students in the additional class viewed AI-generated and teacher-generated materials as complementary. They considered AI-generated materials useful for revision, whereas teacher-generated materials were preferred for initial learning and deeper understanding. Only n=1 student in the main study preferred the AI-generated material, although this student also acknowledged that the teacher-generated materials provided greater depth.

Pedagogical Adequacy and Usability of AI-Generated moodle Lessons:

Figure 6 shows that n=6 students in the main study and n=2 students in the additional class regarded the AI-generated moodle lessons as useful mainly for introductory orientation, review, or basic concepts. A further n=3 students in the main study and n=2 students in the additional class found the lessons generally clear and usable but identified limitations, including insufficient examples, scaffolding, technical detail, or visual explanations. One student in the main study n=1 considered the materials insufficient for independent learning or for learners with limited prior knowledge. No student in the additional class expressed this view.

Overall, these findings suggest qualitative differences in the integration of the TPACK framework.

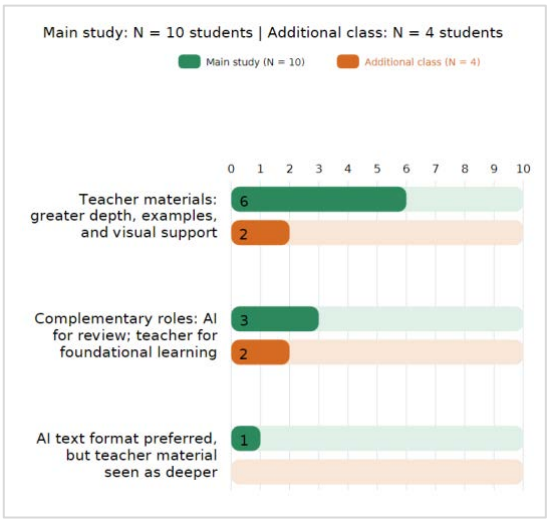


Fig. 5. Dominant Perceived Difference Between AI- and Teacher-Generated Materials

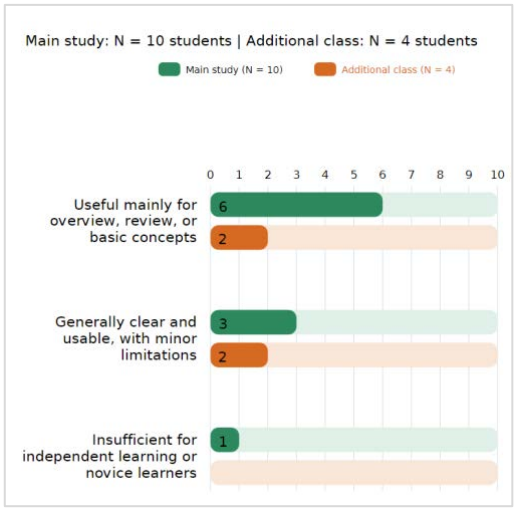


Fig. 6 Pedagogical Adequacy and Usability of AI-Generated moodle Lessons

Dominant TPACK-Related Factor Shaping Acceptance

Figure 7 shows that pedagogical factors were most frequently associated with students' acceptance of AI-generated materials, identified by five students in the main study N=10 and all four students in the additional class N=4. These factors included adequate depth, concrete examples, scaffolding, and teacher support. Technological factors were important for four students in the main study, particularly readable and accurate visualisations, the quality of the Moodle-integrated AI model, and opportunities for interaction. No technological factors were identified by students in the

additional class. Content-related factors—such as accuracy, technical detail, and the need to verify information—were identified by one student in the main study and by no students in the additional class. Overall, acceptance appears to depend primarily on the pedagogical and, particularly in the main study, technological quality of AI-generated lesson materials rather than on content-related issues alone. Because the study used a small qualitative sample (N=10; additional class: N=4), these factors should be described as associated with acceptance rather than as statistically tested predictors.

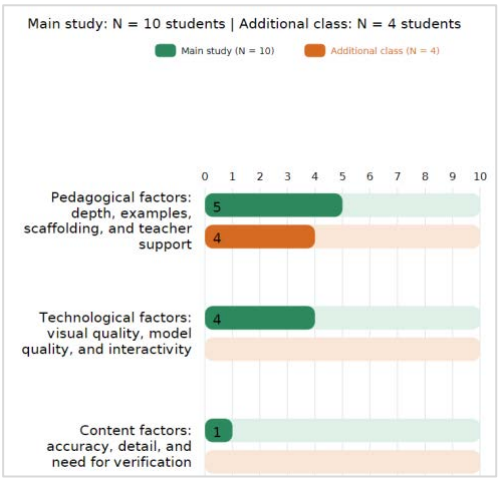


Fig. 7 Dominant TPACK-Related Factor Shaping Acceptance

A. *Qualitative Result*

In "**pedagogical knowledge**" Students generally perceived the AI-supported lessons as clear, short, and well structured. One student stated that *"the clear and precise structure supports my learning"*. Another student observed that the text began with "a general definition" and then gradually moved into more detail, which made the topic understandable even for learners with little prior knowledge. However, several students criticized the limited depth, examples, and scaffolding. One student described the AI materials as "more like a somewhat more detailed definition". Another student stated that "more concrete examples" were needed, especially to understand how a switch works in practice. One student also missed the teacher's *"facial expressions and gestures to support the teacher"*. Thus, the AI lessons were mainly perceived as useful for an initial overview or revision, rather than for comprehensive first-time learning.

In "**Content knowledge**" Students mostly considered the AI-generated materials to be concise summaries of the main points. One student stated that AI content was *"a summary of the most important points"*. Another explained that the materials offered *"a rough insight"* into Switching and VLANs but did not provide enough information to fully understand the topics without prior knowledge. Several students stated that technical details were missing. One student criticized the lack of *"detailed step-by-step processes",* for example regarding broadcast behavior. Another student noted that the teacher video clearly showed *"how a switch builds the MAC address table and how packets travel from A to B"*. In contrast, the AI material was perceived as more suitable for reviewing content, gaining a quick overview, or preparing for examinations. One student stated: *"for exam preparation, I prefer AI content*".

In "**Technological knowledge**" the most frequent criticism concerned the AI-generated images. Students described them as *"blurry"*, *"not readable"*, and *"misleading"*. One student stated that the MAC address table contained *"unreadable, hallucinated characters"*. Students compared the visual problems with the teacher-generated videos. One student stated that the videos offered "*much more depth, clarity, and error-free animations*". Another student explained that the combination of clear pictures, videos, animations, and an explanatory voice created *"a very good learning structure"*. Consequently, students' trust in the AI-generated content was limited. One student stated: *"I do not trust AI content without reservations"* and explained that AI-generated information should be checked against reliable sources.

The findings support the argument of Ning et al. (2024) that effective technology-integrated teaching results from the dynamic interaction of technological, pedagogical, and content knowledge within a specific learning context [13].

Although students evaluated the AI lessons as clear and well structured, the limited depth of content, lack of concrete examples, and poor quality of the AI-generated images reduced their learning value. This shows that pedagogical clarity alone cannot compensate for insufficient content detail or technically weak visualizations. Similarly, accurate content alone is not enough when it is not supported by suitable explanations, examples, and reliable technological tools. Therefore, the findings confirm the central assumption of the TPACK framework: meaningful AI-supported learning requires the integrated application of Pedagogical Knowledge, Content Knowledge, and Technological Knowledge rather than the separate consideration of these domains.

Discussion

Pedagogical Knowledge

The present findings show that students generally perceived the AI-generated moodle lessons as clear, concise, and logically structured. However, they also considered the materials insufficiently detailed for comprehensive first-time learning, particularly because concrete examples, step-by-step explanations, and scaffolding were often missing. Therefore, students primarily viewed the AI-generated content as useful for introductory orientation, revision, or supplementary learning rather than as a replacement for teacher-generated instruction. As highlighted by Qian (2025), research on pedagogical applications of generative AI identifies both instructional opportunities and

pedagogical challenges. The review shows that the educational value of generative AI depends on how it is integrated into teaching and learning, rather than on the mere availability of the technology [34]. This supports the present finding that clear and concise AI-generated texts are not automatically sufficient for meaningful learning when they do not include adequate explanation, task design, examples, and learner support. As explained by Walter (2024), effective AI-supported learning requires pedagogical scaffolding, including clear instructions, prompts, feedback, reflection, modelling, and direct support from teachers. This is consistent with students' comments in the present study, which emphasized the need for more practical examples, guided explanations, and teacher support [35]. The findings therefore suggest that AI-generated moodle texts should be pedagogically revised by teachers before being used with learners who have limited prior knowledge.

Content Knowledge

Regarding Content Knowledge, students perceived the AI-generated materials as concise summaries that communicated basic concepts of switching, VLANs, and MAC address tables. Nevertheless, several students reported that the texts did not provide sufficient technical detail, procedural explanation, or subject-specific depth. They especially missed explanations of technical processes, such as how switches create MAC address tables or how packets and broadcasts are handled. Teacher-generated videos were perceived as more useful because they provided detailed explanations and demonstrated these processes step by step. As reported by Albadarin et al. (2024), empirical research on ChatGPT in education shows that learners may find AI-generated answers helpful for explanations, summaries, and personalized learning support. However, the review also identifies concerns related to inaccurate, misleading, or insufficiently satisfactory responses [36]. This supports the present finding that AI-generated text can provide a useful overview but requires critical evaluation when learners need accurate and detailed subject knowledge.

As found by Luo et al. (2025), instructional designers use generative AI for tasks such as brainstorming, creating learning resources, and streamlining design processes, but they also report concerns about inaccurate, biased, and low-quality outputs. The study stresses that AI-generated learning materials require critical review and revision to ensure that they are accurate and aligned with learning objectives. This corresponds with the current study, where students perceived teacher-generated materials as more detailed, reliable, and appropriate for developing a deeper understanding of technical networking concepts [37].

Technological Knowledge

The technological findings demonstrate that the quality of the Moodle-integrated AI tool strongly influenced students' evaluations. The most common criticism concerned AI-generated images that were blurry, unreadable, misleading, or technically incorrect. By comparison, teacher-generated videos were perceived as more effective because they combined readable graphics, animations, spoken explanation, and step-by-step demonstrations. These results indicate that technological knowledge includes not only knowing how to use AI tools, but also evaluating whether their outputs are reliable, understandable, and appropriate for the specific learning topic. As demonstrated by Cooper and Tang (2024), AI-generated educational imagery must be critically evaluated because it may contain visual inaccuracies and can fail to represent scientific concepts appropriately. Their study of generative-AI imagery in science education highlights that teachers need to assess visual outputs before using them in instruction [38]. This supports the present finding that poor-quality AI-generated diagrams, such as unclear MAC address tables and misleading VLAN visualizations, can hinder rather than support students' understanding. As observed by Baidoo-Anu et al. (2024), higher-education students recognize potential learning benefits of generative AI but also report concerns related to excessive reliance, academic risks, and insufficient training for safe and effective use. The authors argue that institutions should provide clear guidance and strategies for responsible AI use [39]. This aligns with the present study, in which students expressed limited trust in AI-generated content and stressed the need to verify AI output against reliable sources.

In summary, the results indicate that AI-generated moodle materials can support learning when used for short explanations, initial orientation, and revision. However, their value was reduced when pedagogical scaffolding was limited, technical content lacked depth, or AI-generated visuals were inaccurate and difficult to understand. Based on the findings of Qian (2025), Albadarin et al. (2024), and Cooper and Tang (2024), AI should not be treated as an autonomous substitute for teacher-designed instruction. Instead, teachers should use AI-generated moodle content as a draft or supplementary resource that requires pedagogical adaptation, subject-matter verification, and technological quality control. This interpretation supports the central assumption of TPACK: meaningful AI-supported learning depends on the integrated application of Pedagogical Knowledge, Content Knowledge, and Technological Knowledge [36].

Implications

The central implication of this study is that AI-generated lessons in moodle are promising as concise supplementary materials, but they do not yet achieve the full integration of pedagogical, content, and technological knowledge that characterizes high-quality, teacher-designed lessons. Therefore, AI should complement rather than replace human instructional design. In teaching practice, instructors can use AI-generated materials for revision, quick overviews, exam preparation, brief summaries, or alternative explanations; however, teacher-designed lessons

should remain the core resource, particularly for first-time learning and complex topics. Teachers should also critically evaluate AI-generated visuals and replace inaccurate or misleading images where necessary. From a theoretical perspective, the findings indicate that current AI-generated lessons only partially fulfil TPACK expectations, particularly regarding content depth and technological quality. Future research should examine AI–teacher co-design approaches, include larger and more diverse student samples, and investigate improved AI integration in LMS environments to determine how AI can better support full TPACK integration in everyday teaching practice.

Limitations and Future direction

This study has several limitations that should be considered when interpreting its findings. First, the sample size was small, consisting of 10 students and 4 additional students involved in the reliability check. Consequently, the findings cannot be statistically generalized to a wider student population. Future research should involve a substantially larger sample across multiple classes to enable more robust statistical analysis and improve the generalizability of the results. Second, the study was conducted at a single institution and within one geographical location. Educational practices, students' backgrounds, technological infrastructure, and institutional policies may differ across regions and countries. Therefore, the findings may not be transferable to other geographical, cultural, or institutional contexts. Future studies should replicate research across institutions in different regions or countries to examine whether similar patterns emerge. Third, the participants were drawn from only one academic discipline and one specific course context, namely computer science. Students from other degree programs may have different learning needs, levels of prior knowledge, attitudes towards AI-generated materials, and expectations of teacher-generated content. Future research should include students from diverse disciplines, such as business, education, engineering, social sciences, and humanities, as well as different course levels. Finally, the teacher-generated lesson was designed by one instructor and may reflect that instructor's individual teaching style, pedagogical knowledge, experience, and familiarity with the subject. Thus, the observed differences between teacher-generated and AI-generated content may partly be attributable to individual instructor characteristics rather than to teacher-generated content in general. Future studies should involve several instructors who independently develop lessons for the same topic, allowing comparisons across teaching styles and improving the representativeness of the teacher-generated materials.

Conclusions

This study found that AI-generated moodle lesson materials were mainly accepted as supplementary resources rather than as a replacement for teacher-generated instruction. Students considered the AI-generated texts useful for short overviews, revision, and examination preparation, but they preferred teacher-generated materials for first-time learning and complex technical topics. Although the AI lessons were perceived as clear, concise, and well structured, students identified limitations in pedagogical support, content depth, practical examples, and visual quality. Unclear or inaccurate AI-generated images reduced students' understanding and trust in the materials. Overall, the findings support the TPACK framework by showing that effective AI-supported learning requires the integrated consideration of Pedagogical Knowledge, Content Knowledge, and Technological Knowledge. AI-generated moodle materials should therefore be reviewed and improved by teachers before use. Teachers should verify content accuracy, add examples and scaffolding, adapt the materials to learners' needs, and replace low-quality visuals. AI can support instructional design, but teacher expertise

remains essential for ensuring meaningful and reliable learning experiences.

References

[1] J. Reich and J. A. Ruipérez-Valiente, "The MOOC pivot," Science (New York, N.Y.), vol. 363, no. 6423, pp. 130–131, 2019, doi: 10.1126/science.aav7958.
[2] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, p. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.
[3] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," Int J Educ Technol High Educ, vol. 16, no. 1, 2019, doi: 10.1186/s41239-019-0171-0.
[4] P. Mishra and M. J. Koehler, "Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge," Teachers College Record: The Voice of Scholarship in Education, vol. 108, no. 6, pp. 1017–1054, 2006, doi: 10.1111/j.1467-9620.2006.00684.x.
[5] L. S. SHULMAN, "Those Who Understand: Knowledge Growth in Teaching," Educational Researcher, vol. 15, no. 2, pp. 4–14, 1986, doi: 10.3102/0013189X015002004.
[6] Y. Le Tan, S. Hu, D. J. Yeo, and K. H. Cheong, "Artificial intelligence-enabled adaptive learning platforms: A review," Computers and Education: Artificial Intelligence, vol. 9, p. 100429, 2025, doi: 10.1016/j.caeai.2025.100429.
[7] Z. Zhu, Q. Gan, and P. Duan, "Art and design teachers' acceptance of AI-generated content for assisted tutoring: an extended TAM-TPACK framework," Humanit Soc Sci Commun, vol. 13, no. 1, 2026, doi: 10.1057/s41599-026-06692-4.
[8] M. Stadler, K. Herborn, M. Mustafić, and S. Greiff, "The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks," Computers & Education, vol. 157, p. 103964, 2020, doi: 10.1016/j.compedu.2020.103964.
[9] O. Koraishi and Ç. Karatepe, "Minds vs machines: A comparative study of AI and teacher-generated summaries in ELT," TLTL, vol. 7, no. 1, p. 1796, 2025, doi: 10.29140/tltl.v7n1.1796.
[10] M. Henderson et al., "Comparing Generative AI and teacher feedback: student perceptions of usefulness and trustworthiness," Assessment & Evaluation in Higher Education, pp. 1–16, 2025, doi: 10.1080/02602938.2025.2502582.
[11] J. Fan, H. Wang, and X. Gu, "Development and validation of a framework and scale for primary and secondary school teachers' data-artificial intelligent competence," Frontiers in psychology, vol. 16, p. 1756893, 2025, doi: 10.3389/fpsyg.2025.1756893.
[12] A. Barcelona and S. R. Dela Cruz, "Development and validation of a scale measuring students' use of generative artificial intelligence tools," IJERE, vol. 14, no. 5, p. 3612, 2025, doi: 10.11591/IJERE.V14I5.34809.

[13] Y. Ning, C. Zhang, B. Xu, Y. Zhou, and T. T. Wijaya, "Teachers' AI-TPACK: Exploring the Relationship between Knowledge Elements," Sustainability, vol. 16, no. 3, p. 978, 2024, doi: 10.3390/su16030978.
[14] T. Valtonen et al., "Emerging imbalance in the development of TPACK - A challenge for teacher training," Educ Inf Technol, vol. 28, no. 5, pp. 5363–5383, 2023, doi: 10.1007/s10639-022-11426-5.
[15] S. S. Ali and N. A. Hawk, "Examining cultural background as context and in-service teachers' perception of TPACK: A mixed-method study," Educ Inf Technol, vol. 29, no. 3, pp. 3547–3570, 2024, doi: 10.1007/s10639-023-11939-7.
[16] R. Aldemir Engin, D. Karakuş, and M. L. Niess, "TPACK development model for pre-service mathematics teachers," Educ Inf Technol, vol. 28, no. 4, pp. 4769–4794, 2023, doi: 10.1007/s10639-022-11381-1.
[17] P. Mishra, M. Koehler, and J. Harris, "Teachers' Technological Pedagogical Content Knowledge and Learning Activity Types," Journal of Research on Technology in Education, vol. 41, no. 4, pp. 393–416, 2009, doi: 10.1080/15391523.2009.10782536.
[18] Y. A. M. Abouelenein and S. A. S. Selim, "Impact of digital interventions on the development of TPACK: Interviews, reports, and video simulation among pre-service teachers," Educ Inf Technol, vol. 29, no. 11, pp. 1–40, 2024, doi: 10.1007/s10639-023-12422-z.
[19] M. A. Mekheimer, "Effective technology integration in higher education: a mixed-methods study of professional development," Educ Inf Technol, vol. 30, no. 17, pp. 25013–25058, 2025, doi: 10.1007/s10639-025-13750-y.
[20] P. D. Simon, J. Jiang, L. K. Fryer, R. B. King, and C. E. Frondozo, "An Assessment of Learning Management System Use in Higher Education: Perspectives from a Comprehensive Sample of Teachers and Students," Tech Know Learn, vol. 30, no. 2, pp. 741–767, 2025, doi: 10.1007/s10758-024-09734-5.
[21] J. Velander, M. A. Taiye, N. Otero, and M. Milrad, "Artificial Intelligence in K-12 Education: eliciting and reflecting on Swedish teachers' understanding of AI and its implications for teaching & learning," Educ Inf Technol, vol. 29, no. 4, pp. 4085–4105, 2024, doi: 10.1007/s10639-023-11990-4.
[22] M. Mailizar, M. Hidayat, M. Syukri, and D. Wijayanti, "Exploring the impact of generative AI on pre-service mathematics teacher TPACK: a mixed-method study," Discov Educ, vol. 5, no. 1, 2026, doi: 10.1007/s44217-026-01432-x.
[23] M. Z. B. Amiruddin, N. Winarno, A. Samsudin, A. Suhandi, B. Costu, and K. Fenyvesi, "How to Improve TPACK in Science Education: Systematic Literature Review of Empirical Studies," J Sci Educ Technol, 2026, doi: 10.1007/s10956-026-10349-x.
[24] S. D. Aji et al., "Assessing Teachers' Readiness for AI Integration: Insights from AI-Specific TPACK and Attitudes Toward AI," Asia-Pacific Edu Res, vol. 35, no. 4, pp. 921–931, 2026, doi: 10.1007/s40299-026-01093-2.
[25] H. Joshi, "Integrating trust and satisfaction into the UTAUT model to predict Chatbot adoption – A comparison between Gen-Z and Millennials," International Journal of Information Management Data Insights, vol. 5, no. 1, p. 100332, 2025, doi: 10.1016/j.jjimei.2025.100332.
[26] M. M. Ergo, A. T. Bishaw, and B. M. Yimer, "Exploring the underlying factors influencing the implementation of the TPACK framework in Ethiopian teacher education institutions," Discov Educ, vol. 5, no. 1, 2026, doi: 10.1007/s44217-026-01690-9.
[27] P. Baudier, G. Kondrateva, and C. Ammi, "Can blockchain enhance motivation to donate: The moderating impact of religion on donors' behavior in the USA's charity organizations," Technological Forecasting and Social Change, vol. 191, p. 122524, 2023, doi: 10.1016/j.techfore.2023.122524.
[28] N. O. Awuor, C. Weng, E. Piedad, JR, and R. Militar, "Teamwork competency and satisfaction in online group project-based engineering course: The cross-level moderating effect of collective efficacy and flipped instruction," Computers & Education, vol. 176, p. 104357, 2022, doi: 10.1016/j.compedu.2021.104357.
[29] P. Kulju, E. K. Hämäläinen, M. Mäkinen, E. Räikkönen, and C. Kiili, "Pre-service teachers evaluating online texts about learning styles: there is room for improvement in justifying the credibility," Front. Educ., vol. 9, 2024, Art. no. 1451002, doi: 10.3389/feduc.2024.1451002.
[30] C. S. Chai, D. Yu, R. B. King, and Y. Zhou, "Development and Validation of the Artificial Intelligence Learning Intention Scale (AILIS) for University Students," Sage Open, vol. 14, no. 2, 2024, Art. no. 21582440241242188, doi: 10.1177/21582440241242188.
[31] P. Mayring, "Qualitative Inhaltsanalyse – Abgrenzungen, Spielarten, Weiterentwicklungen," (Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, Vol 20, No 3 (2019): Qualitative Content Analysis I), 2019, doi: 10.17169/fqs-20.3.3343.
[32] D. Byrne, "A worked example of Braun and Clarke's approach to reflexive thematic analysis," Qual Quant, vol. 56, no. 3, pp. 1391–1412, 2022, doi: 10.1007/s11135-021-01182-y.
[33] H. Klock, R. Wess, H.-S. Siller, and G. Greefrath, "Test Quality: in Reading Motivation and Reading Engagement," Springer/Scopus, pp. 77–84, 2021, doi: 10.1007/978-3-030-78071-5_4.
[34] Y. Qian, "Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field," TechTrends, vol. 69, no. 5, pp. 1105–1120, 2025, doi: 10.1007/s11528-025-01100-1.
[35] Y. Walter, "Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education," Int J Educ Technol High Educ, vol. 21, no. 1, 2024, doi: 10.1186/s41239-024-00448-3.
[36] Y. Albadarin, M. Saqr, N. Pope, and M. Tukiainen, "A systematic literature review of empirical research on ChatGPT in education," Discov Educ, vol. 3, no. 1, 2024, doi: 10.1007/s44217-024-00138-2.
[37] T. Luo, P. S. Muljana, X. Ren, and D. Young, "Exploring instructional designers' utilization and perspectives on generative AI tools: A mixed methods study," Education Tech Research Dev, vol. 73, no. 2, pp. 741–766, 2025, doi: 10.1007/s11423-024-10437-y.
[38] G. Cooper and K.-S. Tang, "Pixels and Pedagogy: Examining Science Education Imagery by Generative Artificial Intelligence," J Sci Educ Technol, vol. 33, no. 4, pp. 556–568, 2024, doi: 10.1007/s10956-024-10104-0.
[39] D. Baidoo-Anu, D. Asamoah, I. Amoako, and I. Mahama, "Exploring student perspectives on generative artificial intelligence in higher education learning," Discov Educ, vol. 3, no. 1, 2024, doi: 10.1007/s44217-024-00173-z.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Digital Participation to AI Competence: Challenges for Equitable GenAI Adoption in Mexican Higher Education

Anabel López Chávez[1] and Claudia Camacho-Zuñiga[2]

[1]Tecnologico de Monterrey, Pachuca, Mexico, where she is currently pursuing a master's degree in education
[2]Institute for the Future of Education and the School of Engineering and Sciences, Tecnologico de Monterrey, Mexico

Article DOI: 10.68286/samyrad2026.402

ABSTRACT
This study examines whether existing digital skills in Mexico provide the conditions necessary for the meaningful adoption of Generative Artificial Intelligence (GenAI) in higher education. Using a descriptive and exploratory approach, the study analyzes data from the 2024 National Survey on the Availability and Use of Information Technologies in Households (ENDUTIH) and contextualizes these findings with recent evidence from the Digital Education Council's Survey on AI in Higher Education in Latin America 2026. The analysis focuses on patterns of digital participation, including communication, information searching, educational use of the internet, and more specialized digital activities, as well as differences in internet use across age groups. The findings show that digital participation is widespread in Mexico, particularly in communication, information searching, and educational activities, while more specialized forms of digital engagement remain considerably less common. At the same time, evidence from higher education in Latin America indicates widespread GenAI adoption alongside important gaps in AI understanding and the critical evaluation of AI-generated content. Taken together, the findings suggest that digital access and participation should not be considered equivalent to digital or AI competence. Meaningful and equitable GenAI adoption in higher education therefore depends not only on access to digital technologies, but also on the development of digital skills and AI literacy.

KEYWORDS Digital skills; Generative artificial intelligence; AI literacy; Higher education; Digital inequality; Mexico.

How to cite: López Chávez, A., & Camacho-Zuñiga, C. (2026). Digital participation to AI competence: Challenges for equitable GenAI adoption in Mexican higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.402

Introduction

The rapid expansion of Generative Artificial Intelligence (GenAI) is transforming how information is accessed, produced, and used in educational contexts. In this study, AI is used as a broad term when referring to artificial intelligence generally or to terminology used by the source data, while GenAI specifically refers to generative artificial intelligence systems that produce content such as text, images, or other outputs. The analysis focuses specifically on GenAI adoption in higher education. Tools such as ChatGPT, Gemini, and Copilot are increasingly incorporated into academic practices, supporting information retrieval, writing, problem-solving, research, and content generation [1]. In higher education, these technologies create new opportunities for learning and knowledge production while raising questions about

accessibility, responsible use, and educational inequality [2], [3].

However, access to digital technologies and the skills required to use them effectively are not equally distributed. Having an internet connection or digital device does not necessarily imply the competencies needed to navigate digital environments, evaluate information, or use emerging technologies meaningfully. This distinction becomes particularly relevant with GenAI, whose effective use requires users to formulate queries, evaluate generated information, identify potential errors or biases, and make informed decisions about its application.

Evidence from Mexico illustrates these differences. Data from the 2024 National Survey on the Availability and Use of Information Technologies in Households (ENDUTIH) [4] show that while basic digital activities are widespread among internet users, more specialized digital activities are considerably less common. Communication services are used by 93.0% and information searching by 88.2%, compared with 37.3% who report downloading software or applications and 35.8% who report buying products or services online. At the same time, 81.3% report using the internet for educational or training purposes, highlighting the importance of digital environments in learning.

These conditions acquire greater significance as higher education moves toward increasingly AI-mediated learning environments. The Digital Education Council's Survey on AI in Higher Education in Latin America 2026 reports that 92% of surveyed students have used or currently use AI in their learning. However, high adoption coexists with important gaps in AI literacy: 45% of students report only a basic level of understanding of AI, while more than half report difficulties in critically evaluating AI-generated content.

Against this background, this paper asks:

What do existing patterns of digital participation and selected digital activities in Mexico suggest about the conditions for meaningful GenAI adoption in higher education?

The study examines patterns of digital participation and selected digital activities using ENDUTIH 2024 and contrasts these findings with recent evidence on AI use and literacy among higher education students in Latin America. It argues that digital access alone does not constitute readiness for GenAI; meaningful and equitable adoption also requires the development of digital and AI-specific competencies.

Conceptual Framework: From Digital Inequality to AI Literacy

The concept of the digital divide has evolved as digital technologies have become increasingly embedded in educational, economic, and social life. Early approaches focused primarily on differences in access to technological infrastructure and connectivity. However, access alone provides an incomplete explanation of how individuals benefit from digital technologies. Digital inequalities can also emerge from differences in skills, patterns of use, and the capacity to translate technological resources into meaningful educational and social opportunities [3], [5]. Thus, digital inclusion should be understood not only in terms of whether people can connect to digital environments, but also in terms of what they are able to do once connected.

This distinction is particularly relevant in the context of GenAI. Effective interaction with systems such as ChatGPT, Gemini, and Copilot requires more than access to an AI platform. Users must formulate appropriate instructions, interpret generated information, evaluate its reliability, recognize potential inaccuracies or biases, and make informed decisions about how AI-generated content should be used. UNESCO's guidance [2] therefore emphasizes that the integration of GenAI into education should promote inclusion, equity, human agency, and responsible use rather than assume that technological availability will automatically produce educational benefits.

Recent research in higher education further supports the distinction between AI adoption and AI literacy. Beckman et al. [8] identify different profiles of university students according to their knowledge, confidence, and intended use of GenAI, highlighting the relevance of existing digital inequalities to the emerging "GenAI divide." Similarly, Wang, Bai, and An [9] identify a knowledge-practice gap in AI literacy among college students, suggesting that frequent interaction with GenAI does not necessarily translate into the competencies required for responsible and effective use. These findings indicate that familiarity or frequency of use should not be treated as equivalent to AI literacy.

Based on this literature, this study conceptualizes the relationship between digital skills and GenAI adoption through three analytically distinct dimensions: digital access, digital skills, and AI literacy. Digital access refers to the material conditions required to participate in digital environments, including connectivity and access to appropriate devices [3]. Digital skills refer to the competencies required to perform increasingly complex activities using digital technologies [6]. AI literacy represents a further level of capability involving the understanding, critical evaluation, and responsible use of AI systems [8]. In this study, meaningful GenAI adoption refers to the ability to engage with generative AI in ways that involve purposeful use, critical evaluation of AI-generated outputs, recognition of potential inaccuracies or biases, and informed decisions about how generated content is incorporated into educational activities. This concept is treated as an analytical framework rather than as a directly measured outcome in the datasets used in this study.

This framework allows ENDUTIH 2024 indicators to be interpreted as contextual indicators of digital participation that may be relevant to the conditions for GenAI adoption, rather than as direct measures of AI use or readiness. ENDUTIH does not measure whether respondents use ChatGPT, Gemini, or other GenAI systems; instead, it provides evidence on the distribution of digital access and patterns of digital participation in Mexico. The study therefore does not assume that individuals with stronger digital skills necessarily use AI. Rather, it examines whether the digital conditions that may facilitate AI adoption are themselves unevenly distributed. To complement this analysis, evidence from the Digital Education Council's Survey on AI in Higher Education in Latin America 2026 is used to contextualize these conditions alongside recent evidence on AI use and AI literacy among higher education populations in the region.

Accordingly, this study argues that access to digital technologies should be considered an important starting point for GenAI adoption, but not a sufficient condition for its effective use. The development of digital skills and AI literacy may influence how individuals engage with these technologies, particularly in educational settings. This perspective provides the basis for examining whether existing differences in digital skills in Mexico may represent an important consideration for the equitable incorporation of GenAI into higher education.

Data and Methods

This study uses a descriptive and exploratory approach based on two complementary sources. The first is the 2024 National

Survey on the Availability and Use of Information Technologies in Households (ENDUTIH), conducted by the National Institute of Statistics and Geography (INEGI). ENDUTIH provides nationally representative information on access to and use of information and communication technologies in Mexico.

The analysis focuses on indicators of digital participation and internet use, including information searching, online communication, downloading software or applications, buying products or services online, and educational uses of the internet. These indicators capture different forms of engagement with digital technologies, allowing us to distinguish between more common and more specialized forms of digital participation while providing only indirect evidence about the digital skills that may support these activities. Additional information on age, educational attainment, socioeconomic conditions, and access to digital devices is used to contextualize differences in digital capabilities.

The analysis is descriptive and does not seek to estimate a causal relationship between digital skills and GenAI adoption. This distinction is important because ENDUTIH 2024 does not directly measure the use of GenAI applications such as ChatGPT, Gemini, or Copilot. Its indicators are therefore interpreted as conditions that may facilitate or constrain participation in increasingly digital and AI-mediated environments, rather than as direct measures of AI adoption or AI literacy.

These sources address different populations and research purposes. ENDUTIH provides nationally representative evidence for the Mexican population and includes indicators of internet use and digital activities, but it does not specifically target higher education students or measure GenAI use. In contrast, the Digital Education Council survey focuses specifically on higher education and GenAI, but its participating institutions are distributed across Latin American countries and do not provide a Mexico-specific estimate. The two sources are therefore used as complementary evidence rather than as directly comparable measures of the same population.

The second source is the Digital Education Council's Survey on AI in Higher Education in Latin America 2026. The survey collected responses from 22,941 students and 7,319 faculty members across 29 higher education institutions in Latin America and provides information on AI use, adoption patterns, and AI literacy among students and faculty.

The two sources are not statistically combined. Instead, the DEC survey is used to contextualize and contrast the findings from ENDUTIH with recent evidence on the use and understanding of AI in higher education. The analysis first describes patterns of digital participation and selected digital activities in Mexico and then considers how these patterns relate to the competencies increasingly required to engage with GenAI in educational settings.

This approach allows us to examine whether widespread digital participation necessarily implies readiness to use GenAI effectively.

Results

Digital activities and educational use

The ENDUTIH 2024 data reveal a marked differentiation in the intensity and type of digital participation among internet users in Mexico. Communication is the most widespread activity, reported by 93.0% of users, followed by information searching (88.2%) and educational or training activities (81.3%). In contrast, substantially smaller proportions report downloading software or applications (37.3%) or buying products and services online (35.8%) (Fig. 1).

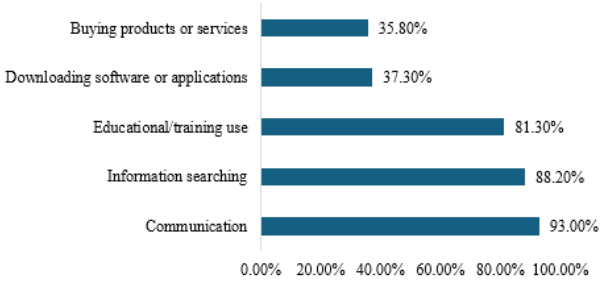


Fig. 1 Selected digital activities and internet uses among internet users in Mexico, 2024. Data from [4].

The distribution reveals a substantial gap between widespread forms of digital participation and activities that involve more specialized interaction with digital technologies. The difference between communication and online purchasing alone is 57.2 percentage points, while educational use remains considerably closer to communication and information searching than to the more specialized activities.

This pattern is relevant to the study because it suggests that digital participation in Mexico is not a uniform phenomenon. Individuals may be highly active in digital environments without necessarily engaging in the broader range of activities that characterize more intensive or diversified technology use. Educational use is particularly relevant in this respect: although 81.3% of internet users report using the internet for education or training, this figure does not indicate whether users possess the competencies required to evaluate, produce, or transform digital information.

Age differences in digital participation

Digital participation also varies substantially across age groups. Internet use is particularly high among younger populations, reaching 97.0% among people aged 18–24 and 95.1% among those aged 25–34. The proportion remains above 90% among people aged 35–44 (92.3%), but declines to 83.8% among those aged 45–54 and 71.0% among those aged 55–64. Among people aged 65 and over, internet use falls to 42.1% (Fig. 2).

The distribution shows that internet participation remains high across younger and middle-aged groups but declines sharply among the oldest population. The difference between the youngest and oldest groups reaches 54.9 percentage points. This pattern indicates that digital participation remains strongly differentiated by age, even within a context of widespread internet use.

Digital activities in the context of GenAI

The ENDUTIH findings can be considered alongside recent evidence on AI use in higher education. The Digital Education Council's Survey on AI in Higher Education in Latin America 2026 reports that 92% of surveyed students have used or currently use AI in their learning, while 67% of AI users report using these tools daily or weekly. At the same time, 45% of students remain at basic stages of AI understanding, and more than half report difficulties in critically evaluating AI-generated content.

The contrast between these findings is particularly relevant. ENDUTIH shows that digital participation is already widespread in Mexico, with high levels of communication, information searching, and educational use of the internet. However, participation in other digital activities is considerably lower. The Digital Education Council survey, meanwhile,

shows that GenAI has rapidly entered higher education practices in Latin America, even as important differences remain in students' understanding and critical evaluation of AI-generated content.

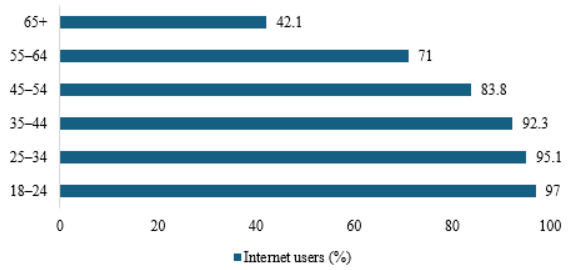


Fig. 2 Internet users by age group in Mexico, 2024 Data from [4].

Taken together, these findings suggest that high levels of digital participation and high levels of AI adoption do not necessarily imply equivalent levels of digital or AI competence. The Mexican data provide a broader picture of the digital environment in which emerging technologies are being introduced, while the Latin American higher education evidence illustrates the extent to which AI is already being incorporated into academic practices.

Because the two sources differ in population, sampling design, geographic scope, and objectives, their findings are not treated as a statistical association. Rather, the comparison provides complementary evidence for exploring what existing patterns of digital participation in Mexico may suggest about the conditions for meaningful GenAI adoption in higher education. The findings are therefore interpreted in light of this broader question.

Discussion

The results suggest that the relationship between digital participation, digital skills, and GenAI adoption in higher education cannot be understood simply as a question of technological access. ENDUTIH 2024 shows that digital participation is widespread in Mexico, particularly in communication, information searching, and educational uses of the internet. However, participation in other potentially advanced digital activities is considerably less common. This distinction is consistent with broader approaches to digital equity, which emphasize that inequalities persist not only in access to technology but also in the ability to use digital resources effectively and meaningfully [3], [4], [10]. Access is therefore necessary for digital participation, but it should not be assumed to produce advanced digital competence.

This distinction becomes particularly relevant with GenAI. Effective interaction with AI systems requires users to formulate requests, interpret outputs, assess information quality, identify potential inaccuracies, and make informed decisions about the use of generated content. Digital skills can therefore provide an important foundation for AI literacy, although the two should not be treated as equivalent [10], [11]. The educational challenge is thus to help students progress from technological use toward more critical and purposeful engagement with digital tools.

The contrast with the Digital Education Council's 2026 evidence further illustrates this distinction. AI use among higher education students in Latin America is already widespread, with 92% of surveyed students reporting that they have used or currently use GenAI in their learning. At the same time, 45% remain at basic stages of GenAI understanding, while more than half experience difficulties critically evaluating AI-generated content. Although these data cannot establish that frequent use leads to competence, the coexistence of widespread adoption and uneven AI literacy is consistent with the distinction identified in the conceptual framework between AI use and AI literacy [8], and shows that access and use should not be treated as indicators of competence.

From an equity perspective, this finding has important implications for higher education institutions. Providing access to AI platforms or technological infrastructure may not be sufficient if students have different levels of digital competence. Higher education institutions can play a pivotal role by transforming students' exposure to digital technologies into opportunities to develop more advanced forms of competence. Institutional strategies should therefore address not only access, but also the development of the skills needed to use AI critically and effectively. Digital literacy initiatives, AI literacy training, academic guidance, and learning activities requiring students to evaluate AI-generated information could form part of a broader strategy for equitable AI integration. This approach is consistent with the emphasis placed by Miao and Holmes [2] on inclusion, equity, human agency, and responsible use in the integration of GenAI into education.

At the institutional level, these findings suggest several practical areas for action. Higher education institutions can incorporate AI literacy activities into existing curricula, provide guidance on evaluating AI-generated information, and develop institutional resources that support responsible use of GenAI in academic work. Institutions can also offer differentiated training according to students' previous levels of digital experience, rather than assuming a homogeneous level of digital competence. These measures can help translate access to AI technologies into more critical and purposeful academic engagement.

The results also contribute to the emerging discussion of a "GenAI divide" in higher education. Beckman et al. [8] identify differences in students' knowledge, confidence, and intended use of GenAI that reflect broader patterns of digital inequality. From this perspective, the relevant issue is not simply whether students can access GenAI, but whether they have the capacities to turn that access into meaningful academic use. GenAI adoption may therefore advance without producing equivalent levels of AI literacy, potentially creating a new dimension of inequality within higher education.

In emerging economies, the role of institutions is demanding and challenging. Beyond broadening access to intelligent tools, higher education institutions must create educational conditions for mobilizing students from digital consumption toward critical evaluation, knowledge creation, and responsible participation.

The comparison presented here has important source-specific limitations. ENDUTIH 2024 provides nationally representative evidence on digital access and selected internet activities in Mexico, but it does not directly measure GenAI use or AI literacy. Its indicators therefore cannot establish whether individuals with stronger digital skills actually adopt GenAI or use it more effectively. In addition, the survey indicators used in this study capture selected forms of digital participation and do not provide a comprehensive measure of digital competence.

The Digital Education Council survey also has important limitations. Its findings are based on participating higher education institutions and should not be interpreted as representative of all higher education students in Latin America or of the Mexican higher education population. Moreover, the survey focuses specifically on AI use, adoption

patterns, and AI literacy, which makes its measures conceptually different from the digital participation indicators available in ENDUTIH. The two sources therefore cannot be statistically combined or used to establish a causal relationship. Their value in this study is complementary: ENDUTIH provides evidence on the broader digital conditions in Mexico, while the Digital Education Council survey provides contextual evidence on GenAI use and AI literacy in higher education.

Taken together, these findings can be interpreted as supporting a broader conception of digital inclusion in higher education. Access remains necessary, but it is insufficient if students lack the capacities to critically evaluate outcomes, create with, and make responsible use of GenAI. The challenge for higher education institutions is demanding; it is not only to ensure that students can access GenAI, but also to ensure that they can understand, evaluate, and use it meaningfully. This shifts the focus from technological availability alone toward the educational processes that enable students to benefit from GenAI and other emerging AI technologies.

Conclusions

The findings suggest that digital competence may constitute an important enabling condition for meaningful GenAI adoption in higher education; however, access and basic digital participation alone should not be assumed to indicate readiness for effective AI use. ENDUTIH 2024 shows that digital engagement is widespread in Mexico, particularly in communication, information searching, and educational activities, while more specialized digital activities remain considerably less common. This indicates that digital participation does not necessarily translate into the capacities required to engage effectively with increasingly complex technologies.

The comparison with recent evidence from higher education in Latin America reinforces this distinction. Although AI use among surveyed university students is already widespread, important gaps remain in AI understanding and critical evaluation. The expansion of GenAI may therefore be advancing faster than the development of the competencies required for its responsible and critical use.

The main contribution of this study is to distinguish between digital access, digital skills, AI literacy, and AI adoption. These dimensions are related but should not be treated as interchangeable. Access provides the conditions for participation, digital skills support the effective use of digital environments, and AI literacy involves the ability to understand, evaluate, and critically engage with AI systems and their outputs.

For higher education institutions, this means that equitable AI integration should extend beyond providing access to platforms and technological infrastructure. Institutional strategies should also strengthen digital and AI literacy, particularly among students with different levels of previous technological experience. Developing critical evaluation skills and the ability to make informed decisions about GenAI use should therefore become part of broader strategies for AI integration in higher education.

Finally, this study is limited by the fact that ENDUTIH 2024 does not directly measure GenAI adoption or AI literacy. The relationship identified here should therefore be understood as an analytical proposition rather than a causal finding. Future research should combine nationally representative measures of digital skills with specific data on GenAI use and literacy among university students in Mexico. Such research could help determine whether existing digital inequalities are reproduced in the emerging GenAI landscape and identify institutional strategies that can support a more equitable adoption of intelligent technologies.

AI Use Disclosure

During the preparation of this work, the authors used OpenAI's ChatGPT for translation and language editing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1] H. Rong and F. Subhan, "Encuesta sobre la IA en la Educación Superior en América Latina 2026 [Survey on AI in Higher Education in Latin America 2026]," Digital Education Council, 2026.

[2] F. Miao and W. Holmes, "Guidance for generative AI in education and research," UNESCO, 2023, doi: 10.54675/EWZM9535.

[3] F. Gottschalk and C. Weise, "Digital equity and inclusion in education: An overview of practice and policy in OECD countries," OECD Education Working Papers, no. 299, OECD Publishing, 2023, doi: 10.1787/7cb15030-en.

[4] Instituto Nacional de Estadística y Geografía, "Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2024 [National Survey on the Availability and Use of Information Technologies in Households (ENDUTIH) 2024]," 2024. [Online]. Available: https://www.inegi.org.mx/programas/endutih/2024/

[5] D. Crovi Druetta, "Sociedad de la información y el conocimiento. Entre el optimismo y la desesperanza [Information Society and Knowledge: Between Optimism and Despair]," Revista Mexicana de Ciencias Políticas y Sociales, vol. 45, no. 185, pp. 13–33, 2002.

[6] UNESCO Institute for Statistics, "A global framework of reference on digital literacy skills for indicator 4.4.2," UNESCO, 2018.

[7] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. 2020 CHI Conf. Human Factors in Computing Systems, 2020, pp. 1–16, doi: 10.1145/3313831.3376727.

[8] K. Beckman, T. Apps, S. K. Howard, C. Rogerson, A. Rogerson, and J. Tondeur, "The GenAI divide among university students: A call for action," The Internet and Higher Education, vol. 67, Art. no. 101036, 2025, doi: 10.1016/j.iheduc.2025.101036.

[9] J. Wang, B. Bai, and Q. An, "Enhancing college students' AI literacy through generative AI use: A mixed-methods investigation," Frontiers in Psychology, vol. 17, Art. no. 1728785, 2026, doi: 10.3389/fpsyg.2026.1728785.

[10] Y. Cheng, "Students' experiences of generative AI in higher education: Understanding digital inequalities through a sociomaterial lens," in Proc. Int. Conf. Networked Learning, 2026, vol. 15, doi: 10.54337/nlc.v15.10884.

[11] M. A. Matjie, A. Nethavhani, and M. Matlakala, "AI and the digital divide in education," Front. Comput. Sci., vol. 8, 2026, Art. no. 1759027, doi: 10.3389/fcomp.2026.1759027.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

The Illusion of Prompting: A Discourse Analysis of Epistemic Engagement in Student–AI Dialogue

Kiran Qasim Ali[1,2]

[1]Institute for Educational Development, Aga Khan University, Pakistan

[2]Department of Integrated Studies in Education, McGill University, Canada

Article DOI: 10.68286/samyrad2026.405

ABSTRACT

The rapid integration of generative artificial intelligence (AI) into education has intensified interest in how learners engage with AI-generated knowledge. While much research has focused on prompt engineering and output quality, less attention has been given to the epistemic work learners undertake after responses are generated. Drawing on sociocultural and dialogic perspectives on learning [1], [2], this study examines how graduate university students engaged with ChatGPT while completing a formative-assessment design task. Twenty-five student–AI transcripts were analyzed using discourse and thematic analysis, supported by an adapted Initiation–Response–Feedback (IRF) framework. Four interactional patterns were identified: Extraction, Confirmation, Negotiation, and Co-construction. Extraction involved information retrieval with limited visible evaluation; Confirmation involved presenting learner-generated ideas for AI validation; Negotiation occurred when students introduced contextual, pedagogical, or disciplinary criteria to evaluate and redirect AI-generated proposals; and Co-construction involved the progressive development and reformulation of ideas across successive turns. These patterns formed a continuum of epistemic engagement distinguished by the extent to which learners assumed responsibility for evaluating and developing generated knowledge. Neither prompt sophistication nor interactional length reliably indicated the quality of epistemic engagement. More consequential was how learners positioned themselves in relation to AI-generated content and whether that content became an object of scrutiny, judgement, and further inquiry. The findings suggest that AI literacy should extend beyond prompt competence to include evaluative judgement, dialogic engagement, and responsibility for knowledge development within human–AI interaction.

KEYWORDS Generative AI, student–AI dialogue, epistemic engagement, AI literacy, discourse analysis, formative assessment, evaluative judgement

How to cite: Ali, K. Q. (2026). The illusion of prompting: A discourse analysis of epistemic engagement in student–AI dialogue. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.405

Introduction

Generative Artificial Intelligence (GenAI) has changed how students' access, evaluate, and produce academic knowledge [3-4]. Large Language Models (LLMs) now enable students to generate explanations, summaries, and arguments with unprecedented speed and ease [4-7]. Yet this capability raises a fundamental pedagogical question: Does the ability to produce a sophisticated academic response necessarily reflect the intellectual work required to understand it? Traditionally, academic learning has required students to search for information, evaluate evidence, synthesize ideas, and construct arguments. With generative AI, however, students can produce seemingly sophisticated responses without necessarily undertaking these processes themselves. Although LLMs offer valuable opportunities for explanation, feedback, and academic support [5], [8], they also create the possibility that intellectual tasks central to

learning may be delegated to the technology rather than undertaken by the learner [9-11]. The central concern, therefore, is not merely what students can produce with AI, but how critically they engage with what it produces.

These concerns have brought renewed attention to AI literacy as an educational priority. Beyond operational proficiency, AI literacy encompasses the ability to interpret, evaluate, and use AI-generated information critically, ethically, and responsibly [8], [12], [13]. Within this broader discourse, prompt engineering has gained prominence as a means of improving the quality of interactions with GenAI [5], [14], [15]. Through the deliberate and iterative use of techniques such as role assignment, contextual framing, and output specification, learners can guide LLMs towards more accurate, relevant, and pedagogically useful outputs [16], [17]. Yet this focus primarily concerns the conditions under which information is generated, rather than learners' intellectual engagement with that information once generated. Prompt engineering may improve what AI produces, but it does not, in itself, establish whether learners critically examine its claims, evaluate its evidence, or construct their own understanding.

The distinction becomes particularly important when considering what learners actually do with AI-generated information. Research on learning has long established that meaningful learning requires more than access to information; it depends on active interpretation, evaluation, elaboration, and knowledge construction [2], [18-19]. A carefully engineered prompt may yield a coherent and persuasive response that a learner accepts without scrutiny. Conversely, a simple prompt may initiate substantial intellectual engagement when learners question claims, seek evidence, challenge assumptions, compare alternatives, or revise their understanding. The sophistication of a prompt, therefore, cannot be taken as a measure of the intellectual work undertaken by the learner. What matters is not only what students ask AI to produce, but how they engage with, challenge, and build upon its responses [8], [12-13].

Productive engagement with AI-generated responses requires more than a willingness to question them. Learners must also possess the disciplinary knowledge, critical reasoning, and relevant literacy practices needed to assess the credibility of claims, evaluate evidence, and recognize the limits of an explanation. Bearman et al. [20] conceptualize this capacity as evaluative judgement - the ability to assess the quality of work against relevant standards. Crucially, such judgements are not made in isolation but are informed by disciplinary conventions, prior knowledge, and an understanding of what constitutes credible evidence and sound reasoning within a particular field [21-22]. Research further suggests that domain specific knowledge and evaluative skills are interdependent, with each informing the exercise of the other [18], [23-24]. This has particular implications for AI-mediated learning because a superficial response may appear coherent and convincing if the learner lacks the disciplinary knowledge or evaluative capacity to identify its weaknesses.

Sociocultural perspectives further suggest that these capacities are shaped by cultural frames, interactions and social practices through which learning occurs [25], [26]. In AI-assisted learning, generative models become part of this mediation, influencing the questions learners pursue, the explanations they consider, and the judgements they make. Dialogic theories further emphasize the role of explanation, challenge, and justification in the development of understanding [1], [24]. From this perspective, analytical attention shifts from the quality of AI-generated responses to the 'interactions' through which learners engage with them [2], [27]. How learners accept, question, challenge, or revise AI-generated ideas becomes particularly significant, as these exchanges offer insight into how disciplinary knowledge and evaluative judgement are exercised in practice.

Recent studies of multi-turn interactions with GenAI have begun to examine these processes more closely. Tang and Putra [28], for example, suggest that analyzing extended exchanges can reveal patterns of reasoning and elaboration that single-turn analyzes may overlook [2], [25]. However, two limitations remain. First, much of this research examines interactions within highly structured environments, where reflective prompts, prescribed tasks, or AI-generated scaffolds are deliberately designed to encourage critique and justification. While such approaches provide valuable insights into supported forms of engagement, they offer a less complete account of how learners independently question, evaluate, and negotiate AI generated ideas in the absence of explicit scaffolding [29-30].

Second, the presence of sustained dialogue does not necessarily indicate meaningful epistemic engagement. Extended exchanges may remain largely transactional, while brief interactions can involve substantial intellectual scrutiny [1], [24], [27]. Interaction length alone, therefore, offers limited insight into the nature of learners' engagement. What requires closer examination is how learners position themselves in relation to AI-generated knowledge - whether they accept claims at face value, seek confirmation, question underlying assumptions, negotiate interpretations, or construct new understandings through dialogue. This distinction calls for greater analytical attention to the nature of discourse within AI interactions, rather than the frequency or duration of those interactions.

These gaps highlight the need to examine how students engage with AI-generated knowledge in contextually situated ways. Pakistani higher education provides a particularly relevant site for such inquiry. Although English serves as the principal medium of instruction in many universities, students often navigate academic learning through multilingual repertoires and diverse educational trajectories [47]. They routinely move between English, Urdu, and regional languages, even in formally English-medium environments [48], and many report challenges related to academic literacy and disciplinary communication in English [49], [50]. At the same time, studies of higher education in Pakistan have noted the continuing influence of transmission-oriented pedagogies

and assessment practices that may provide limited opportunities for students to critically evaluate knowledge claims [51-53]. Within such contexts, AI-generated responses may acquire authority not solely because of their accuracy but because they are presented in fluent, academically recognizable forms. Prior research suggests that users frequently overestimate the reliability of large language model outputs [55], while linguistic fluency can increase perceptions of truthfulness and credibility [56], [57]. Students may therefore face the dual challenge of evaluating both the content of AI-generated responses and the cues through which authority is communicated. Examining these processes can deepen understanding of how learners interpret, evaluate, and negotiate knowledge in AI-mediated learning environments.

To address these needs, this study examines university students' interactions with ChatGPT during a pedagogical design task, with particular attention to the forms of discourse that emerge across successive conversational turns. Through discourse analysis, the study investigates how students position themselves in relation to AI-generated knowledge, including how they accept, verify, question, challenge, and extend AI-generated ideas. Rather than assessing learning through the quality of prompts or the accuracy of AI-generated responses, the study focuses on the interactional processes through which students engage with AI as a knowledge source. In doing so, it seeks to illuminate how epistemic engagement is enacted in real-time dialogue with AI and how learners negotiate the authority and credibility of AI-generated knowledge.

Methods or analytical approach

This study employed a qualitative exploratory design to examine how university students engaged with AI-generated knowledge during multi-turn interactions with ChatGPT [31]. The study was conducted at a private English-medium university in Karachi, Pakistan and involved 25 students enrolled in a bridging semester of an MPhil Education program. Data were generated through a course-based assignment in which students used ChatGPT-4 to address a pedagogical problem involving the *'design of a formative assessment strategy for a mixed-ability secondary classroom'*. The task was selected because it required participants to draw on disciplinary knowledge, evaluative judgement, and iterative reasoning. Students were encouraged to engage in sustained interaction with ChatGPT through follow-up questions, requests for clarification or justification, challenges, and alternative lines of inquiry. No explicit guidance on evaluating or critiquing AI-generated responses was provided, allowing patterns of epistemic engagement to emerge naturally. Upon completion, participants submitted anonymized transcripts of their interactions. The final dataset comprised 25 transcripts containing approximately 300 conversational turns. Transcripts were analyzed using a combination of discourse analysis and thematic analysis informed by sociocultural and dialogic theories of learning [1-2]. An adapted Initiation Response–Feedback (IRF) framework was used to examine the sequential organization of student–AI exchanges, with particular attention to how students responded to and acted upon AI-generated content across successive turns. Analysis proceeded in two stages. First, student contributions were coded according to their epistemic function, including clarification, evidence-seeking, challenge, elaboration, revision, and acceptance. These initial codes were informed by the adapted IRF framework and the study's conceptual focus on epistemic engagement, while remaining responsive to functions evident in the data. Second, coded exchanges were compared within and across transcripts to examine how these functions combined and developed across successive turns. Through iterative comparison, recurring configurations of epistemic functions were grouped and refined into four broader patterns: *Extraction, Confirmation, Negotiation, and Co-construction*. For example, sequences dominated by requests for information, clarification, or elaboration, without subsequent evaluation of the AI response, contributed to the category of *Extraction*. In contrast, sequences in which students challenged an AI generated proposal, introduced contextual or disciplinary criteria, and prompted its revision contributed to *Negotiation*. *Co-construction* was distinguished by sequences in which successive student contributions not only evaluated AI responses but also introduced, connected, or reformulated ideas that progressively shaped the developing solution. The resulting categories therefore reflected a combined deductive–inductive process: the initial coding was conceptually informed, while the four broader patterns were developed through iterative comparison of recurring interactional configurations across the dataset. Coding and category development were conducted by the researcher using NVivo 14, with categories repeatedly checked against the transcripts and refined to ensure that they adequately represented the interactional patterns evident in the data.

Results and findings

Analysis of the transcripts revealed four recurring discourse stances: *Extraction, Confirmation, Negotiation, and Co-construction*. These patterns were not mutually exclusive; students could move between them within a single interaction. The percentages reported below therefore indicate the proportion of transcripts in which each pattern was identified and are not expected to sum to 100%.

Extraction: Treating AI as a Source of Information

In this study, *Extraction* refers to sequences in which students elicited content from ChatGPT without subsequently evaluating, challenging, contextualizing, or substantively developing it. *Extraction* sequences were identified in 13 of the 25 transcripts (52%). *Extraction* was treated as an interactional pattern rather than a fixed learner characteristic, as students frequently moved between different forms of engagement within a conversation. Within the IRF sequence, the student's Initiation (I) elicited information from ChatGPT, which was followed by the AI Response (R). The Feedback move (F) often became a new Initiation by requesting clarification, elaboration, or another example, producing

repeated I–R cycles without placing the preceding response under evaluation.

In the data, *Extraction* occurred in three forms: *immediate extraction, specified extraction*, and *extended extraction*. *Immediate extraction* involved interactions that ended following the initial AI response. *Specified extraction* involved detailed prompts that generated contextually tailored responses but were not subjected to further evaluation. *Extended extraction* involved multiple conversational turns focused on clarification or elaboration rather than scrutiny of the generated content. These variations are analytically significant because they demonstrate that *Extraction* cannot be reduced to either brief interactions or poorly formulated prompts. In some cases, students engaged only minimally with ChatGPT before concluding the interaction. In others, they invested considerable effort in specifying contextual details or extending the conversation across multiple turns. Yet across these variations, AI-generated content remained primarily a resource for information retrieval rather than an object of evaluation, comparison, or critique. The defining feature of *Extraction* was therefore not interaction length or prompt sophistication, but the absence of visible evaluative engagement. Analytically, the distinction rests on the function of the learner's subsequent moves: requests for clarification, elaboration, or additional examples extend the informational sequence, whereas questioning assumptions, identifying limitations, contesting claims, or proposing alternatives transforms the generated response into an object of evaluation.

Student: I need to design a formative assessment activity for a Grade 7 mathematics lesson on fractions. The class has around 40 students with mixed ability levels and limited classroom resources. I want something that can be done during the lesson and can help the teacher identify students who are struggling with the concept. suggest a practical strategy which would work in this context
ChatGPT: One practical option is to use mini-whiteboards or response cards during the lesson. The teacher can ask short fraction questions and all students respond at the same time. This allows the teacher to quickly see who has understood the concept and where students are having difficulty.
Student: we don't have mini whiteboards
ChatGPT: In that case, students can use small pieces of paper or their notebooks. The teacher can ask students to write their answers and hold them up together. Another option is to use an exit ticket at the end of the lesson.
Student: exit ticket how?
ChatGPT: The teacher can give students two or three short questions before they leave the class. For example:
– Which is bigger ½ or ¼ ? – Write one fraction equivalent to ½.
– Explain how you know that ½ is greater than ¼.
The teacher can review the responses and identify which students need further support.
Student: but class is 40 students so checking all will take time no?
ChatGPT: Yes, checking every response in detail may take time. The teacher does not need to mark each response formally. The responses can be scanned quickly and grouped according to common patterns. For example, the teacher might identify students who understand the concept, students who made calculation errors, and students who show a misconception about fractions.
Student: okay then what teacher do next class
ChatGPT: In the next lesson, the teacher can use the responses to plan a short follow-up activity. Students who are still confused could work with visual examples, while students who have understood the concept could attempt more challenging fraction problems.
Student: visual examples like what
ChatGPT: The teacher could draw fraction bars or circles on the board. For example, the teacher could divide one rectangle into halves and another into fourths and ask students to compare ½ and ¼. This can help students understand equivalent fractions visually.
Student: any activity students can do themselves not teacher showing
ChatGPT: Students could draw their own fraction bars in their notebooks. The teacher can give them different fractions to represent and compare. They could also work in pairs and check each other's representations.

Fig. 1. Participant Excerpt 03 - Extended extraction

The excerpt illustrates an extended form of *Extraction*. The interaction develops across multiple conversational turns as the student requests alternative suggestions, practical details, and further elaboration. At several points, the student introduces contextual considerations, including the absence of classroom resources and the challenges of managing formative assessment in a class of forty students. Notably, the student identifies a legitimate feasibility constraint in relation to the use of exit tickets yet delegates the resolution of that concern back to ChatGPT. Rather than evaluating the adequacy of the proposed solution, comparing alternatives, or weighing pedagogical trade-offs, the student treats the constraint as a prompt for further recommendation. Consequently, the interaction becomes progressively more detailed without becoming more evaluative. In addition, this example highlights an important distinction between interactional extension and epistemic evaluation. Although the exchange expands the quantity and specificity of information generated, the AI's recommendations do not become an explicit object of scrutiny. The interaction is characterised by informational refinement rather than epistemic scrutiny; the student remains engaged in shaping and refining the information provided by the AI, but not in evaluating its underlying assumptions, limitations, or adequacy.

Across *Extraction* sequences, students exercised agency in determining what information to request and whether to continue the interaction. Less evident, however, was the application of independent pedagogical, contextual, or disciplinary criteria to assess the content once generated. Although the task required judgements concerning learner variability, feasibility, and pedagogical purpose, such considerations were rarely mobilised explicitly as grounds for accepting, rejecting, adapting, or refining AI-generated recommendations. Importantly, the absence of evaluative moves should not be interpreted as evidence that students accepted responses uncritically, regarded ChatGPT as authoritative, or lacked the capacity to assess the quality of the information provided. Such evaluation may have occurred outside the recorded interaction. The claim proposed here is therefore more limited as evaluation was not made visible within the discourse itself. Extraction identifies an interactional configuration in which AI generated content is obtained, and sometimes elaborated, without becoming an explicit object of scrutiny. Nor does the presence of extraction imply that a learner occupies this position throughout an entire interaction. Because the patterns identified in this study exist along a continuum, sequences may initially take an extractive form before shifting towards *Confirmation, Negotiation,* or *Co-Construction*. Extraction is therefore best understood as a local interactional accomplishment within an unfolding interaction rather than a stable characteristic of individual learners. Its analytical significance lies in identifying moments where information generation functions as an endpoint rather than a catalyst for further epistemic work.

Confirmation: Seeking Validation of an Existing Position

Confirmation refers to instances when students brought an existing idea, judgement, or proposed strategy into the interaction and sought ChatGPT's validation of that position.

Confirmation sequences were identified in 9 of the 25 transcripts (36%). Within the IRF framework, Initiation (I) in *Confirmation* sequences typically contained a student-generated proposition or strategy, which ChatGPT evaluated in its Response (R). The student's Feedback (F) generally indicated uptake of that judgement and, in serial forms, became a new Initiation (I) in which a further or revised proposition was returned to the AI for validation. In the data, *Confirmation* occurred in three forms: direct confirmation, qualified confirmation, and serial confirmation. *Direct confirmation* involved straightforward endorsement of a student-generated proposal. *Qualified confirmation* occurred when ChatGPT endorsed a proposal while introducing conditions, limitations, or implementation considerations. *Serial confirmation* involved successive cycles in which students presented revised or additional ideas for further validation. These variations demonstrate that *Confirmation* is not defined by the complexity of the interaction but by the direction of evaluation. Unlike *Extraction*, students entered the interaction with an identifiable position of their own. However, across all three forms, evaluative authority remained oriented towards the AI. Student-generated ideas became the object of validation, while the AI's judgement was typically accepted as the basis on which adequacy was established.

Student: I need to design a formative assessment activity for an MPhil research methods class where students are learning how to develop focused and researchable research questions. I want an activity that can help me see how they are judging the quality of a research question, not just whether they can remember the criteria. what kind of activity can I use?
ChatGPT: You could give students several sample research questions and ask them to identify which ones are well focused and researchable. They could justify their choices using features such as clarity, scope, and feasibility. This would help you identify how well they understand the characteristics of a good research question.
Student: I was actually thinking something similar. I can give 3 or 4 research questions and ask them which one is better and why. but I dont want to give criteria first because then they will just use the criteria to answer
ChatGPT: That would be a useful approach. If you want to identify students' existing understanding, asking them to make an initial judgement before providing the criteria would give you better evidence of how they are currently thinking about research questions.
Student: yes and maybe first they answer individually and after that discuss in groups? because if I start with group I will not know what each student actually thinks
ChatGPT: Yes, beginning individually would allow you to capture each student's initial judgement. The group discussion could then give students an opportunity to compare their reasoning and reconsider their choices.
Student: then after discussion I can introduce the criteria and ask them to look at their first answer again. I think this will show if their thinking changed
ChatGPT: Yes. Asking students to revisit their original response after seeing the criteria would make any revision in their thinking visible. You could also ask them to briefly explain why they retained or changed their original choice.
Student: I like that. maybe they write one line like "I changed my answer because…" so I can see what actually changed not only the final answer
ChatGPT: That would strengthen the activity. A short explanation would provide evidence of the reasoning behind the revision rather than only showing that the student selected a different answer.
Student: okay so I will do individual selection with reason, then group discussion, give criteria and then they revisit original answer and explain if they changed it. this works as formative assessment?
ChatGPT: Yes. The sequence works as formative assessment because it allows you to identify students' initial understanding, observe how their reasoning develops through discussion, and examine how they apply the criteria when reconsidering their original judgement.

Fig. 2. Participant Excerpt 11 - Serial confirmation

The excerpt illustrates the most extended form of confirmation. Throughout the interaction, the student presents successive ideas and returns them to ChatGPT for assessment. The student's contributions clearly demonstrate prior knowledge of the task and an emerging understanding of formative assessment. Unlike *Extraction,* the interaction does not begin with a request for the AI to generate a solution from scratch. Instead, the student contributes substantive content and uses the AI to determine whether that content is appropriate, feasible, or educationally sound.

At first glance, such interactions appear more evaluatively sophisticated than *Extraction* because they involve learner-generated propositions rather than simple information requests. However, the evaluative work remains asymmetrical. The student seeks judgement, but primarily from the AI rather than through independent examination of the AI's reasoning. Even when ChatGPT introduces qualifications, conditions, or recommendations, these are rarely taken up as objects of inquiry in their own right. Instead, they become prompts for further validation. As a result, the interaction becomes more elaborate without substantially altering the location of evaluative authority. This distinction is important as it demonstrates that students are not relying exclusively on AI-generated content; they bring their own ideas, experiences, and disciplinary understandings into the interaction. Yet contributing an initial position and using that position to evaluate AI-generated claims remain distinct forms of epistemic engagement. The former establishes ownership of an idea, whereas the latter requires learners to subject both their own reasoning and the AI's reasoning to scrutiny.

As with *Extraction, Confirmation* should not be interpreted as evidence of uncritical reliance on AI. The transcripts cannot establish why students sought validation, how confident they were in their original proposals, or whether they subsequently verified the AI's responses using other sources. The claim advanced here is therefore more limited as student-generated ideas became visible objects of evaluation, but the AI's judgement rarely became an object of scrutiny within the interaction itself. Altogether, confirmation occupies an important position along the continuum of epistemic engagement identified in this study. It represents a movement beyond *Extraction* because learners contribute and test their own understandings within the dialogue. Nevertheless, the direction of judgement remains largely unchanged: adequacy is established primarily through AI evaluation. The transition to *Negotiation* begins when this relationship shifts and students increasingly draw upon contextual, disciplinary, and experiential knowledge to assess whether the AI's claims are acceptable to them.

Negotiation: Evaluating and Redirecting AI-Generated Content

Negotiation refers to instances in which students moved beyond requesting information or seeking validation and began evaluating AI-generated proposals against criteria derived from their own contextual, pedagogical, or disciplinary understandings. *Negotiation* sequences were identified in 7 of the 25 transcripts (28%). As with the preceding patterns, *Negotiation* was treated as an interactional position that emerged at particular moments within a conversation rather than as a fixed characteristic of either the learner or the transcript.

Within the IRF framework, in *Negotiation*, the Feedback move (F) became analytically consequential because the student evaluated the AI Response (R) against contextual, pedagogical, or disciplinary criteria introduced from outside the response itself. That Feedback frequently functioned as a new Initiation (I), requiring ChatGPT to revise, qualify, or replace its preceding proposal. In more extended forms, this produced successive I-R-F cycles in which learner judgement progressively shaped the direction of the interaction. Three broad forms of *Negotiation* were observed in the dataset. Some arose from perceived mismatches between AI-generated recommendations and classroom realities. Others focused on pedagogical adequacy, particularly whether a proposed strategy could achieve its intended educational purpose. A third form involved *recursive Negotiation*, in which students repeatedly evaluated and redirected successive AI responses across multiple turns. Although these forms differed in focus and complexity, they shared a common feature: the AI-generated response became an object of evaluation rather than a source of information or validation.

Student: I need to design a formative assessment for an undergraduate research methods class where students are learning qualitative coding. They understand what a code is but when they get actual data they mostly just pick words from the text. I want to check if they can actually interpret the data and develop meaningful codes during class. what activity can I use?

ChatGPT: You could use a collaborative coding activity. Give students a short interview excerpt and a list of possible codes. In small groups, they could apply the codes to the excerpt and then compare their choices with other groups. The discussion would allow you to identify where students are having difficulty applying codes consistently.

Student: this can work but if I give them codes already then I am checking if they can match codes, not if they can develop codes themselves. thats actually the part they struggle with

ChatGPT: In that case, you could remove the pre-prepared code list. Give all students the same short excerpt and ask them to generate two or three codes independently. They could then compare their codes in small groups and discuss why they labelled the data differently.

Student: yes but if they go directly in groups some students will just follow the person who understands coding better. I need to see their own thinking first

ChatGPT: Then you could begin with an individual coding stage. Each student could code the same excerpt independently and briefly write why they chose one of their codes. After that, they could compare their interpretations in groups.

Student: but class is around 45 students I cant really check 45 coding sheets during the activity. then I will only know later where they were struggling

ChatGPT: One option would be to ask students to submit only one code and its justification through a quick response system or on a small slip of paper. You could scan a selection of responses during class and use a few contrasting examples for discussion.

Student: I dont think slips will help because again I have to go through them. maybe I need something where responses become visible at same time. and I also dont want students names because they may hesitate if their interpretation is different

ChatGPT: In that case, you could collect the codes anonymously through a shared digital response tool and display them together. The class could examine several different codes generated from the same excerpt and discuss which interpretations are well supported by the data and why.

Student: that is closer. but I would not ask them which code is "correct" because in qualitative coding there may be more than one reasonable interpretation. I think question should be which codes are better supported by the data and what evidence they are using

ChatGPT: Yes, that would be more appropriate. The discussion could focus on the relationship between the proposed code and the evidence in the excerpt rather than treating coding as having a single correct answer. This would allow you to assess whether students can justify an interpretation rather than simply produce a label.

Fig. 3. Participant Excerpt 15 - Recursive negotiation

Figure 3 shows a clear instance of recursive negotiation. Rather than accepting the initial recommendation or simply requesting further explanation, the student repeatedly evaluates successive AI proposals against criteria drawn from their own pedagogical and disciplinary understanding. Each AI response generates a further point of scrutiny: whether the activity assesses code generation rather than matching, makes individual thinking visible, remains feasible in a large class, and reflects the interpretive nature of qualitative coding. Within the IRF sequence, the student's Feedback therefore repeatedly becomes a new Initiation, identifying a limitation in the preceding response while specifying conditions that the next proposal must address. In contrast to *Extraction* and *Confirmation*, where students sought information and validation, respectively, the direction of judgement shifts towards the learner. AI generated proposals become objects of evaluation against criteria that originate outside the AI response itself.

Negotiation was often selective rather than oppositional. Students frequently retained aspects of a proposal while questioning others. Responses such as *"That could work, but..."* acknowledged the potential value of a recommendation while opening it to revision. The issue at stake was therefore rarely whether a recommendation was simply right or wrong. More often, students considered whether it was appropriate for a particular educational context. A strategy that appeared reasonable in general terms could still be judged inadequate if it failed to account for learner variability, resource constraints, classroom realities, or the kinds of evidence needed for formative decision making. In such moments, students retained authority over the conditions under which a proposal could be regarded as feasible or educationally useful. As with the preceding patterns, the absence of *Negotiation* should not be interpreted as evidence that students lacked the capacity to evaluate AI-generated content, nor does effective engagement require learners to challenge every response they encounter.

Negotiation therefore marks an important shift along the continuum of epistemic engagement identified in this study. Whereas *Confirmation* leaves evaluative authority primarily with the AI, *Negotiation* positions learners more clearly as judges of the adequacy of AI-generated proposals. Nonetheless, the interaction remains centered on evaluating and revising proposals already on the table. Co-construction extends this process further. Rather than primarily assessing successive proposals, learners and AI become jointly involved in developing an evolving line of reasoning, with ideas progressively refined, extended, and reconfigured across turns.

Co-construction: Developing Understanding Across Turns

Co-construction refers to sequences in which the student and AI progressively develop a solution, with each turn contributing substantively to how the idea is reformulated. *Co-construction* sequences were identified in 4 of the 25 transcripts (16%). These sequences involved sustained cycles in which students introduced ideas, engaged substantively with AI-generated suggestions, reconsidered emerging problems, and progressively reformulated the assessment design. Within the adapted IRF framework, *Co-*

construction commonly appeared through recursive sequences such as I–R–F/I'–R–F/I"–R. Its defining feature, however, was not the number of turns or the recurrence of this structure. Rather, the student's Feedback/Initiation (F/I') contributed new substance to the developing solution. Ideas, constraints, experiences, and connections introduced in one turn became resources for subsequent turns. The developing design therefore changed across the interaction and could not be attributed solely to either the student's initial position or the AI's first response. This distinguishes *Co-construction* from *Negotiation*. In *Negotiation*, learner generated criteria are primarily used to evaluate and redirect an AI proposal. In *Co-construction*, such evaluation becomes part of a broader process in which the student also introduces, connects, revises, or synthesizes ideas that alter the direction of the developing solution.

Student: I need formative assessment for B.Ed class on research ethics. students know principles like consent confidentiality etc but when given actual research situation they find it difficult to decide what is ethical because sometimes there is no simple right answer. I want to see their reasoning not just if they remember principles. what can I do?

ChatGPT: You could use a short ethical dilemma. Give students a research scenario with competing concerns and ask them to decide what the researcher should do and justify their decision. You could then compare different responses in a whole-class discussion.

Student: dilemma is good. I was thinking maybe everyone should get same case first so I can see how differently they reason. but if I just ask what researcher should do they may jump to solution. maybe first ask what ethical issues they can see?

ChatGPT: That would give you two levels of evidence: whether students can recognize the ethical issues and how they use those issues to reach a judgement. You could therefore separate the activity into identifying concerns first and deciding on a course of action afterwards.

Student: yes. actually this gives me another idea. I can stop the case halfway. like first only give initial situation and they identify concerns. then give additional information which makes decision more difficult and ask if their position changes

ChatGPT: That could make the assessment more revealing because students would need to reconsider an earlier judgement when the context changes. You could ask them to record their initial position before receiving the additional information and then explain what, if anything, caused them to revise it.

Student: hmm and maybe different groups can get different second information? one group gets issue about confidentiality, and another gets issue about participant risk. then when groups compare they will see why same case can lead to different decisions

ChatGPT: Yes. That would shift the task from finding a single correct answer towards examining how ethical judgements depend on the information available and the relative weight given to different principles. The comparison between groups could become part of the formative evidence.

Student: right. then my original idea of checking if they know ethical principles is actually too limited. I think I want to see if they can identify the principle, use it in making a decision, and then reconsider decision when new information comes. maybe at end they should write what information changed their judgement and why

ChatGPT: That would give the activity a clearer progression. You would be assessing recognition of ethical concerns, application of principles, and revision of judgement in response to changing circumstances. The final explanation would also make the reasoning behind that revision visible.

Student: okay so instead of one dilemma question I now have staged case. first same incomplete case individually, identify issues and initial decision. then groups receive different additional information, reconsider decision, compare across groups, and at end each student writes what changed or did not change in their judgement and why. that is actually closer to what I want to assess

ChatGPT: Yes. The activity has developed from a single ethical-dilemma question into a staged assessment of how students identify, apply, and reconsider ethical principles as the research context changes.

Fig. 4. Participant Excerpt 07 - Co-construction

Figure 4 illustrates a clear instance of Co-construction. Across the interaction, the student does more than request information, seek validation, or identify limitations in AI-generated suggestions. Successive contributions build on preceding turns by introducing new considerations, connecting emerging ideas, and progressively refining the assessment design. Within the IRF sequence, Feedback repeatedly becomes a renewed Initiation. Unlike *Negotiation*, however, these F/I' moves do not simply identify deficiencies requiring revision. Instead, they contribute directly to the development of an evolving solution.

The cumulative nature of this process is centrally important. As the interaction develops, earlier ideas are revisited, modified, and connected to new considerations. Discussion may expand from selecting an assessment technique to broader questions of learner participation, evidence collection, feedback, feasibility, or instructional use. Interactional length alone was therefore insufficient for classification as *Co-construction*; what distinguished these exchanges was the extent to which successive contributions shaped a developing line of reasoning.

The IRF framework makes this distinction particularly visible. Similar recursive I–R–F/I'–R structures could occur across *Extraction, Confirmation, Negotiation*, and *Co-construction*, but the function of the student's Feedback/Initiation differed. In *Extraction*, F/I' primarily sought further information or elaboration. In *Confirmation*, it commonly registered uptake of the AI's judgement or introduced another proposition for validation. In *Negotiation*, it identified a limitation and established conditions for revision. In *Co-construction*, F/I' contributed directly to the emerging solution by proposing, connecting, reconsidering, or synthesizing ideas. The same structural position therefore performed different epistemic work across the interactional patterns. In addition, student-generated knowledge was also more visibly incorporated into the development of the solution during *Co-construction*. Students drew upon their understanding of formative assessment, knowledge of learners and classroom conditions, prior experience, and anticipated instructional consequences. These resources were not used only to judge whether an AI suggestion was acceptable. They were brought into relation with successive responses and used to develop the idea further. At times, this altered the AI generated proposal; at others, engagement with the AI- response prompted students to reconsider their own earlier position. *Co-construction* therefore involved development at the level of ideas rather than a succession of AI proposals followed by student evaluation.

The term *Co-construction* is used here in a deliberately limited sense. It does not imply equivalence between interaction with GenAI and collaboration between human participants, nor does it assume that student and AI contributions are comparable in status or origin. Rather, it describes an observable interactional process in which the eventual assessment design emerges through successive cycles of contribution, uptake, reconsideration, and reformulation. The analytical focus is therefore on the visible development of ideas across the interaction. *Co-construction* represents the furthest point on the continuum of epistemic engagement identified in this study. *Extraction* centres on obtaining information; *Confirmation* introduces a student position but primarily seeks its validation; and *Negotiation* uses learner-generated criteria to evaluate and

redirect AI-generated proposals. In *Co-construction*, the learner's own position also becomes open to reconsideration and development. Ideas are connected, revised, and recombined across turns, so that the assessment strategy becomes a developing solution rather than a product to be generated, validated, or corrected.

Discussion

The findings revealed substantial variation in how students engaged with AI-generated knowledge. Rather than representing fixed learner characteristics, the patterns identified in this study emerged as interactional positions that students could move between as conversations developed. Taken together, the findings suggest that the educational significance of student-AI interaction cannot be adequately inferred from prompt quality, response sophistication, or conversational length alone. More consequential is the extent to which learners assume responsibility for evaluating, questioning, and developing the knowledge generated during the interaction. Viewed in this way, the findings shift attention from the production of AI outputs to the epistemic work learners undertake in engaging with them.

Beyond Prompt Competence: Generating a Good Response Is Not Equivalent to Learning

A central implication of the findings concerns the prominence of prompt engineering within contemporary accounts of AI literacy. Previous research has shown that prompt quality influences the relevance, specificity, and usefulness of large language model outputs [14-17]. The present findings do not challenge the importance of prompt competence. Rather, they suggest that generating a high-quality response and engaging productively with that response are analytically distinct forms of activity [32]. This distinction was particularly evident in *Extraction* sequences. Students could formulate detailed prompts and obtain contextually relevant responses without subsequently examining the adequacy of what had been generated [33-35]. Conversely, relatively simple prompts sometimes gave rise to more substantive engagement when learners identified limitations, introduced contextual knowledge, or redirected the interaction. Prompt sophistication therefore provided only limited evidence of the epistemic quality of the exchange [32].

From a dialogic perspective, the educational significance of AI does not lie in response generation alone but in how learners subsequently question, evaluate, and develop the ideas those responses contain [1-2], [19], [24]. A response that appears coherent, authoritative, and contextually relevant may still require judgements concerning its accuracy, completeness, disciplinary validity, and practical usefulness [36]. Indeed, the fluency and plausibility of AI- generated content may make such scrutiny even more necessary [5], [20], [37-38]. The findings therefore support a broader conception of AI literacy in which prompt competence constitutes one component, rather than the endpoint, of effective AI use [8], [12-13], [39]. The critical educational question is not simply whether learners can obtain better responses from AI systems, but whether they retain epistemic responsibility for evaluating, questioning, and developing the knowledge those responses contain [40,41]. In this sense, prompting may shape the quality of information generated, but learning depends on how that information becomes an object of judgement, inquiry, and knowledge development.

What Learners Bring to AI Matters: From Possessing Knowledge to Mobilizing It

A second contribution of the findings concerns the role of learner knowledge within AI-mediated interaction. Across the patterns identified in this study, differences in epistemic engagement were closely associated with how students brought disciplinary, contextual, practical, and experiential knowledge into the exchange [41]. The findings suggest, however, that possessing relevant knowledge and mobilizing that knowledge as a basis for judgement are not equivalent [23]. This distinction was particularly evident in the transition from *Confirmation* to *Negotiation*. In *Confirmation* sequences, students frequently demonstrated prior understanding by proposing assessment strategies, articulating pedagogical positions, or relating classroom practices to formative-assessment principles. Yet these contributions were commonly presented to ChatGPT for validation [33], [35]. By contrast, *Negotiation* occurred when learners used their knowledge to assess AI-generated proposals. Considerations such as class size, resource constraints, learner variability, and assessment purpose became criteria through which recommendations were judged, challenged, or redirected.

The significance of prior knowledge therefore lies not simply in its possession, but in its capacity to support evaluative judgement. This aligns with Bearman et al.'s [20] conception of evaluative judgement as the ability to recognize quality and apply appropriate standards to work or information [42]. It also resonates with research demonstrating that critical evaluation depends upon the conceptual resources available to the evaluator [22], [24]. In the present study, students were better able to interrogate AI-generated proposals when they could articulate why those proposals were inadequate, incomplete, or misaligned with the instructional problem under consideration [18]. At the same time, the findings caution against equating productive AI engagement with disciplinary knowledge alone. Students drew on a wider repertoire of resources, including professional experience, knowledge of learners, contextual awareness, and practical understandings of classroom realities. What distinguished more substantive engagement was not the possession of these resources, but their mobilization through acts of comparison, critique, justification, and revision. Engagement became visible when learners brought what they knew into dialogue with what the AI had generated [19]. This interpretation is consistent with sociocultural perspectives on learning, which view knowledge not as a static possession but as something enacted through participation in meaning-making activity [1-2], [25]. From this perspective, learner

knowledge becomes educationally consequential when it is used to question assumptions, identify limitations, recognize contextual mismatches, and reshape emerging interpretations. The shifts observed within some transcripts reinforce this point as students occasionally moved from *Extraction* or *Confirmation* towards *Negotiation* when AI-generated responses conflicted with prior experience, failed to accommodate contextual realities, or introduced new issues requiring consideration. Such shifts suggest that epistemic engagement is not simply a stable learner disposition but is also shaped through interaction. The analysis indicates that productive engagement was associated not merely with students' prior knowledge or the information generated by AI, but with how learners drew upon their understanding to question, evaluate, and develop AI-generated ideas. From this perspective, the potential educational value of AI-generated responses lies not in the information they provide alone, but in the intellectual work learners undertake in examining, challenging, and reconsidering that information [18].

Multi-Turn Interaction Is Not Necessarily Dialogic Learning

The findings also have important implications for sociocultural and dialogic accounts of AI-mediated learning. Dialogic perspectives emphasize that learning develops not simply through participation in interaction, but through exchanges in which ideas are taken up, questioned, justified, challenged, and extended [1-2], [19], [24]. From this perspective, multi-turn interaction might initially appear to provide evidence of deeper engagement with AI. The present findings suggest that such an assumption is insufficient [43].

Across the four patterns identified in this study, similar interactional structures frequently performed very different epistemic functions [44]. Students could sustain extended exchanges by requesting clarification, elaboration, or additional examples while remaining primarily engaged in *Extraction*. Similarly, complete I-R-F sequences could function as *Confirmation* when feedback merely acknowledged or accepted the AI's judgement. By contrast, *Negotiation* and *Co-construction* were distinguished not by the number of conversational turns but by the work those turns performed. Feedback became epistemically consequential when it evaluated, redirected, connected, or reformulated preceding contributions. Theoretically, this distinction suggests that interactional structure and epistemic function should not be conflated [44]. A lengthy interaction is not necessarily dialogic in the educational sense, just as a brief interaction is not necessarily epistemically weak [43]. What matters is how contributions are taken up and transformed across the dialogue. Repeated requests for additional content may generate substantial interaction while leaving the learner's relationship to generated knowledge largely unchanged [18]. Conversely, relatively brief exchanges may involve considerable challenge, justification, evaluation, and revision. This finding resonates strongly with dialogic theories of learning. Mercer [1] argues that educationally productive dialogue depends upon the collective examination and justification of ideas, while Wells [2] emphasizes the progressive development of understanding through inquiry. Viewed through this lens, the educational value of AI interaction lies less in the quantity of interaction than in the extent to which generated ideas become objects of scrutiny and development. The critical issue is therefore not whether learners continue the conversation, but whether successive turns contribute to reasoning, judgement, and knowledge construction [24].

The findings also extend emerging research that examines student-AI dialogue beyond isolated prompt-response exchanges [28]. Rather than treating prompt quality, response sophistication, or turn count as proxies for learning, the present study highlights the importance of learner uptake. Whether learners elaborate, validate, challenge, compare, justify, or revise AI-generated content appears more educationally consequential than the number of conversational turns through which those actions occur [44]. Methodologically, this strengthens the case for treating unfolding discourse, rather than individual prompts or outputs, as the primary unit of analysis. It is through successive turns that epistemic engagement becomes visible, making discourse trajectories a more informative indicator of learning than interactional quantity alone.

Co-construction, but not symmetry: Who carries epistemic responsibility?

The *co-construction* sequences raise a broader theoretical question concerning the status of AI as a dialogic partner. In the most developed interactions, students' ideas changed through successive exchanges with ChatGPT. The final assessment design was neither fully specified in the student's initial contribution nor simply reproduced from the AI's first response. Instead, it emerged through cycles in which possibilities were proposed, reconsidered in relation to classroom knowledge and experience, revised, and eventually synthesized [2]. In this restricted interactional sense, describing these sequences as co-construction is warranted [45-46]. However, the findings also reveal an important asymmetry in how this development was sustained. Across *Negotiation* and *Co-construction*, substantive shifts in the direction of inquiry were predominantly initiated by students. Learners recognized when a suggestion was inadequate, introduced considerations that had been overlooked, determined which issues required further attention, and reconsidered their own emerging positions. ChatGPT responded by elaborating possibilities, generating alternatives, or revising earlier proposals, but it did not consistently assume equivalent responsibility for determining what warranted scrutiny. This distinction becomes particularly significant when viewed through sociocultural accounts of dialogue. Mercer [1] conceptualizes dialogue as a process through which participants make reasoning available for examination and use one another's contributions as resources for further thought [19]. While some student-AI exchanges displayed

these interactional qualities, responsibility for recognizing limitations, introducing new evaluative criteria, and redirecting inquiry remained largely with the learner [40]. ChatGPT could produce plausible alternatives, explanations, and revisions, but the trajectory of the interaction depended primarily on students recognizing when such moves were necessary [58].

Co-construction should therefore not be taken to imply symmetry between human and AI contributions [45]. A more precise interpretation is that ideas can develop through human-AI dialogue while responsibility for judging their adequacy remains disproportionately located with the learner [41]. The AI can contribute explanations, alternatives, representations, and revisions, but the learner remains centrally responsible for determining what should be accepted, questioned, connected, or developed further [20], [42]. Co-construction of ideas, in other words, does not necessarily entail co-distribution of epistemic responsibility. This distinction has important educational implications. The possibilities afforded by extended AI dialogue may depend partly on learners' capacity to recognize when an apparently plausible response requires further examination [32-33]. Students who cannot readily identify omissions, inconsistencies, or contextual limitations may have fewer opportunities to move the interaction beyond *Extraction* or *Confirmation* [34-35]. Supporting productive AI use may therefore require attention not only to how learners converse with AI, but also to the disciplinary and evaluative resources that enable them to assume responsibility for the direction of that dialogue [12], [39], [41], [54].

Overall, these findings suggest that a continuum of epistemic positioning may offer a useful complement to existing conceptualizations of AI literacy that emphasize technical competencies. The four discourse stances - *Extraction, Confirmation, Negotiation,* and *Co-construction* - illustrate differences in the extent to which learners assume responsibility for evaluating, questioning, and developing AI-generated knowledge. These stances do not represent fixed learner characteristics or a linear developmental sequence, as students may move between them within a single interaction. Rather, they highlight a dimension of AI literacy that neither prompt sophistication nor interaction length can adequately capture the extent to which learners exercise judgement over what constitutes credible, justified, and meaningful knowledge. The educational challenge, therefore, extends beyond teaching students to obtain better AI-generated responses to supporting their capacity to interrogate, contextualize, and develop those responses through critical inquiry.

Implications for AI Literacy and Pedagogy

At the classroom level, the findings suggest the importance of designing learning activities that make students' engagement with AI-generated content visible. Rather than assessing only whether learners can formulate effective prompts or obtain useful responses, educators might create opportunities for students to justify decisions, compare alternatives, identify limitations, and revise AI generated suggestions using disciplinary or contextual criteria. Such tasks position AI output as a starting point for intellectual work rather than its conclusion.

At the level of curriculum and teacher education, the findings suggest that AI literacy should be embedded within disciplinary learning rather than treated as a standalone technical skill. Students need opportunities to practise evaluating and revising AI-generated content in relation to the standards, evidence, and forms of reasoning valued within their field. This also has implications for teacher preparation and professional development, particularly in supporting educators to design tasks that make evaluative judgement and disciplinary reasoning visible.

At the level of AI literacy frameworks and educational policy, the findings support a broader conception of AI literacy than one centered primarily on prompt competence or technical proficiency. As GenAI becomes increasingly integrated into educational practice, greater emphasis should be placed on learners' capacity to evaluate, contextualize, and develop judgement, AI-generated disciplinary knowledge. Evaluative reasoning, and epistemic responsibility should therefore be treated as central dimensions of AI literacy rather than secondary complements to technical skill.

Limitations

Several limitations should be considered when interpreting these findings. First, the study involved 25 university students from a single institutional context in Pakistan. The relatively small, context-specific sample limits the transferability of the findings to other educational settings, disciplines, and student populations. Second, the analysis focused on a single pedagogical design task. Different tasks, disciplinary demands, or levels of familiarity with GenAI may give rise to different forms of discourse. Third, the analysis was confined to students' observable interactions with ChatGPT. Learners may have evaluated information, consulted other sources, or reconsidered their understanding without explicitly articulating these processes in the transcripts. The absence of visible critical engagement should therefore not be interpreted as evidence that no such evaluation occurred. Finally, the identification and interpretation of discourse stances involved qualitative researcher judgement, and alternative interpretations of particular exchanges remain possible. The four stances should consequently be understood as contextually situated analytical categories rather than fixed learner characteristics or a universal developmental sequence. Further research across different tasks, disciplines, and educational contexts could examine the applicability and refinement of these categories.

Conclusion

This study examined how university students engaged with ChatGPT during a pedagogical design task, with particular

attention to how they position themselves in relation to AI-generated knowledge. The findings suggest that understanding students' use of GenAI requires attention beyond the sophistication of prompts or the quality of generated responses. Of particular significance is the extent to which learners exercise evaluative judgement, question the authority of AI-generated claims, and assume responsibility for knowledge construction. By identifying the discourse stances through which these processes are enacted, the study shifts attention from AI use as a technical skill to AI engagement as an epistemic practice. This perspective extends current discussions of AI literacy by highlighting how learners negotiate, challenge, and develop knowledge in interaction with AI. For educators, the implication is not simply to teach students how to obtain better responses from AI, but to cultivate their capacity to question, evaluate, and contribute to the knowledge those responses present. Future research across different tasks and educational contexts can further examine how such forms of epistemic engagement emerge and the conditions that support them.

Acknowledgments

I thank the students who participated in this study, my colleagues for their comments on earlier drafts, and the SAMYRAD 2026 reviewers for their constructive feedback.

Declarations

Funding: This research was conducted as part of the Scholarship of Teaching and Learning (SoTL) and did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest: The authors declare no conflicts of interest.

Ethics approval and consent: Not applicable

References

[1] N. Mercer, “Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking,” Journal of Applied Linguistics, vol. 1, no. 2, pp. 137–168, 2004, doi: 10.1558/japl.v1i2.137.

[2] G. Wells, Dialogic Inquiry: Towards a Sociocultural Practice and Theory of Education. Cambridge, U.K.: Cambridge University Press, 1999.

[3] M. Bearman, J. Ryan, and R. Ajjawi, “Discourses of artificial intelligence in higher education: A critical literature review,” Higher Education, vol. 86, pp. 369–385, 2023, doi: 10.1007/s10734-022 00937-2.

[4] C. K. Y. Chan and W. Hu, “Students’ voices on generative AI: Perceptions, benefits, and challenges in higher education,” International Journal of Educational Technology in Higher Education, vol. 20, art. 43, 2023, doi: 10.1186/s41239-023-00411-8.

[5] E. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education,” Learning and Individual Differences, vol. 103, art. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

[6] Y. K. Dwivedi et al., “So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy,” International Journal of Information Management, vol. 71, art. 102642, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[7] J. Rudolph, S. Tan, and S. Tan, “ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?,” Journal of Applied Learning and Teaching, vol. 6, no. 1, pp. 342–363, 2023, doi: 10.37074/jalt.2023.6.1.9.

[8] Y. Walter, “Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education,” International Journal of Educational Technology in Higher Education, vol. 21, art. 15, 2024, doi: 10.1186/s41239-024-00448-3.

[9] A. Yusuf, N. Pervin, and M. Román-González, “Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives,” International Journal of Educational Technology in Higher Education, vol. 21, art. 21, 2024, doi: 10.1186/s41239-024-00453-6.

[10] M. Gerlich, “AI tools in society: Impacts on cognitive offloading and the future of critical thinking,” Societies, vol. 15, no. 1, art. 6, 2025, doi: 10.3390/soc15010006.

[11] M. Saqr, K. Misiejuk, and S. López-Pernas, “Human-AI collaboration or obedient and often clueless AI in instruct, serve, repeat dynamics?,” The Internet and Higher Education, vol. 70, 2026, doi: 10.48550/arXiv.2508.10919.

[12] D. Long and B. Magerko, “What is AI literacy? Competencies and design considerations,” in Proc. 2020 CHI Conf. Human Factors in Computing Systems, Honolulu, HI, USA, 2020, pp. 1–16, doi: 10.1145/3313831.3376727.

[13] D. T. K. Ng, J. K. L. Leung, K. W. S. Chu, and M. S. Qiao, “Conceptualizing AI literacy: An exploratory review,” Computers and Education: Artificial Intelligence, vol. 2, art. 100041, 2021, doi: 10.1016/j.caeai.2021.100041.

[14] E. R. Mollick and L. Mollick, “Assigning AI: Seven approaches for students, with prompts,” Wharton School Research Paper, University of Pennsylvania, 2023, doi: 10.2139/ssrn.4475995.

[15] N. Knoth, A. Tolzin, A. Janson, and J. M. Leimeister, “AI literacy and its implications for prompt engineering strategies,” Computers and Education: Artificial Intelligence, vol. 6, art. 100225, 2024, doi: 10.1016/j.caeai.2024.100225.

[16] L. Giray, “Prompt engineering with ChatGPT: A guide for academic writers,” Annals of Biomedical Engineering, vol. 51, no. 12, pp. 2629–2633, 2023, doi: 10.1007/s10439-023-03272-4.

[17] J. D. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, “Why Johnny can’t prompt: How non-AI experts try (and fail) to design LLM prompts,” in Proc. 2023 CHI Conf. Human Factors in Computing Systems, Hamburg, Germany, 2023, art. 437, pp. 1–21, doi: 10.1145/3544548.3581388.

[18] M. T. H. Chi and R. Wylie, “The ICAP framework: Linking cognitive engagement to active learning outcomes,” Educational Psychologist, vol. 49, no. 4, pp. 219–243, 2014, doi: 10.1080/00461520.2014.965823.

[19] N. Mercer and K. Littleton, Dialogue and the Development of Children’s Thinking: A Sociocultural Approach. London, U.K.: Routledge, 2007.

[20] M. Bearman, J. Tai, P. Dawson, D. Boud, and R. Ajjawi, “Developing evaluative judgement for a time of generative artificial intelligence,” Assessment & Evaluation in Higher Education, vol. 49, no. 6, pp. 893–905, 2024, doi: 10.1080/02602938.2024.2335321.

[21] D. R. Sadler, “Formative assessment and the design of instructional systems,” Instructional Science, vol. 18, no. 2, pp. 119–144, 1989.

[22] S. Wineburg and S. McGrew, "Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information," Teachers College Record, vol. 121, no. 11, pp. 1–40, 2019, doi: 10.1177/016146811912101102.
[23] J. D. Bransford, A. L. Brown, and R. R. Cocking, Eds., How People Learn: Brain, Mind, Experience, and School. Washington, DC, USA: National Academy Press, 2000.
[24] R. Alexander, A Dialogic Teaching Companion. London, U.K.: Routledge, 2020.
[25] L. S. Vygotsky, Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA, USA: Harvard University Press, 1978.
[26] J. V. Wertsch, Voices of the Mind: A Sociocultural Approach to Mediated Action. Cambridge, MA, USA: Harvard University Press, 1991.
[27] R. Wegerif, "Dialogic intelligence: Rethinking what education is for in the age of AI," Journal for the Study of Education and Development, 2026, doi: 10.1177/02103702261439739.
[28] K. S. Tang and G. B. S. Putra, "Generative AI as a dialogic partner: Enhancing multiple perspectives, reasoning, and argumentation in science education with customized chatbots," Journal of Science Education and Technology, vol. 35, pp. 128–140, 2026, doi: 10.1007/s10956-025-10240-1.
[29] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," Human Factors, vol. 52, no. 3, pp. 381–410, 2010, doi: 10.1177/0018720810376055.
[30] M. Vaccaro, A. Almaatouq, and T. W. Malone, "When combinations of humans and AI are useful: A systematic review and meta analysis," Nature Human Behaviour, vol. 8, no. 12, pp. 2293–2303, 2024, doi: 10.1038/s41562-024-02024-1.
[31] S. B. Merriam and E. J. Tisdell, Qualitative Research: A Guide to Design and Implementation, 4th ed. San Francisco, CA, USA: Jossey-Bass, 2016.
[32] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, and S. Rintel, "The metacognitive demands and opportunities of generative AI," in Proc. 2024 CHI Conf. Human Factors in Computing Systems, Honolulu, HI, USA, 2024, art. 680, pp. 1–24, doi: 10.1145/3613904.3642902.
[33] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI assisted decision-making," Proceedings of the ACM on Human Computer Interaction, vol. 5, no. CSCW1, art. 188, pp. 1–21, 2021, doi: 10.1145/3449287.
[34] L. Krupp, S. Steinert, M. Kiefer-Emmanouilidis, K. E. Avila, P. Lukowicz, J. Kuhn, S. Küchemann, and J. Karolus, "Unreflected acceptance: Investigating the negative consequences of ChatGPT assisted problem solving in physics education," in HHAI 2024: Hybrid Human AI Systems for the Social Good. Amsterdam, The Netherlands: IOS Press, 2024, pp. 199–212, doi: 10.3233/FAIA240195.
[35] S. Passi and M. Vorvoreanu, "Overreliance on AI: Literature review," Microsoft Research, Tech. Rep. MSR-TR-2022-12, Jun. 2022.
[36] M. Sharples, "Automated essay writing: An AIED opinion," International Journal of Artificial Intelligence in Education, vol. 32, no. 4, pp. 1119–1126, 2022, doi: 10.1007/s40593-022-00300-7.
[37] M. Jakesch, A. Bhat, D. Buschek, L. Zalmanson, and M. Naaman, "Co-writing with opinionated language models affects users' views," in Proc. 2023 CHI Conf. Human Factors in Computing Systems, Hamburg, Germany, 2023, art. 111, pp. 1–15, doi: 10.1145/3544548.3581196.
[38] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review," Smart Learning Environments, vol. 11, no. 1, art. 28, 2024, doi: 10.1186/s40561-024-00316-7.
[39] T. K. F. Chiu, Z. Ahmad, M. Ismailov, and I. T. Sanusi, "What are artificial intelligence literacy and competency? A comprehensive framework to support them," Computers and Education Open, vol. 6, art. 100171, 2024, doi: 10.1016/j.caeo.2024.100171.
[40] M. Bearman and R. Ajjawi, "Learning to work with the black box: Pedagogy for a world with artificial intelligence," British Journal of Educational Technology, vol. 54, no. 5, pp. 1160–1173, 2023, doi: 10.1111/bjet.13337.
[41] L. Markauskaite, R. Marrone, O. Poquet, S. Knight, R. Martinez Maldonado, S. Howard, J. Tondeur, M. De Laat, S. Buckingham Shum, D. Gašević, and G. Siemens, "Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?," Computers and Education: Artificial Intelligence, vol. 3, art. 100056, 2022, doi: 10.1016/j.caeai.2022.100056.
[42] J. Tai, R. Ajjawi, D. Boud, P. Dawson, and E. Panadero, "Developing evaluative judgement: Enabling students to make decisions about the quality of work," Higher Education, vol. 76, no. 3, pp. 467–481, 2018, doi: 10.1007/s10734-017-0220-3.
[43] C. Howe and M. Abedin, "Classroom dialogue: A systematic review across four decades of research," Cambridge Journal of Education, vol. 43, no. 3, 2013, doi: 10.1080/0305764X.2013.786024. 325–356.
[44] M. Nystrand, L. L. Wu, A. Gamoran, S. Zeiser, and D. A. Long, "Questions in time: Investigating the structure and dynamics of unfolding classroom discourse," Discourse Processes, vol. 35, no. 2, pp. 135–198, 2003, doi: 10.1207/S15326950DP3502_3.
[45] J. Roschelle and S. D. Teasley, "The construction of shared knowledge in collaborative problem solving," in Computer Supported Collaborative Learning, C. O'Malley, Ed., NATO ASI Series, vol. 128. Berlin, Germany: Springer, 1995, pp. 69–97, doi: 10.1007/978-3-642-85098-1_5.
[46] F. Miao and K. Shiohira, AI Competency Framework for Students. Paris, France: UNESCO, 2024, doi: 10.54675/JKJB9835.
[47] F. Shamim, "Trends, issues and challenges in English language education in Pakistan," Asia Pacific J. Educ., vol. 28, no. 3, pp. 235-249, 2008, doi: 10.1080/02188790802267324.
[48] M. Shah, S. Pillai, and M. Sinayah, "Translanguaging in an academic setting," Lingua, vol. 225, pp. 16–31, 2019, doi: 10.1016/j.lingua.2019.05.001.
[49] I. Akhtar and A. Nawab, "Challenges and coping strategies in EMI classrooms: A case study from Pakistan," Training, Lang. Cult., vol. 8, no. 4, pp. 64–73, 2024, doi: 10.22363/2521-442X-2024-8-4-6473.
[50] H. Irfan, "The postgraduate students and the university teachers' perceptions of the gap between policy and practice of English as the medium of instruction (EMI) at higher education in Pakistan," ELF Annu. Res. J., vol. 19, pp. 119–136, 2017.
[51] P. Hoodbhoy, "Pakistan's higher education system: What went wrong and how to fix it," Pak. Dev. Rev., vol. 48, no. 4, pt. II, pp. 581–594, 2009, doi: 10.30541/v48i4IIpp.581-594.
[52] L. Q. Warsi and A. F. Shah, "Teachers' perception of classroom assessment techniques (CATs) at higher education level," Pak. J. Soc. Sci., vol. 39, no. 1, pp. 189–199, 2021.

[53] M. Afzaal, “Status of public examinations in Pakistan: The influence of question choice on selective study and rote memorization,” Bull. Educ. Res., vol. 26, no. 1, pp. 11–24, 2004.
[54] C. Kidd and A. Birhane, “How AI can distort human beliefs,” Science, vol. 380, no. 6651, pp. 1222–1223, 2023, doi: 10.1126/science.adi0248.
[55] M. Steyvers, H. Tejeda, A. Kumar, C. Belem, S. Karny, X. Hu, L. W. Mayer, and P. Smyth, “What large language models know and what people think they know,” Nat. Mach. Intell., vol. 7, no. 2, pp. 221–231, 2025, doi: 10.1038/s42256-024-00976-7.
[56] R. Reber and N. Schwarz, “Effects of perceptual fluency on judgments of truth,” Conscious. Cogn., vol. 8, no. 3, pp. 338–342, 1999, doi: 10.1006/ccog.1999.0386.
[57] A. L. Alter and D. M. Oppenheimer, “Uniting the tribes of fluency to form a metacognitive nation,” Pers. Soc. Psychol. Rev., vol. 13, no. 3, pp. 219–235, 2009, doi: 10.1177/1088868309341564.
[58] W. Choi, H. Bak, J. An, Y. Zhang, and B. Stvilia, “College students’ credibility assessments of GenAI-generated information for academic tasks: An interview study,” J. Assoc. Inf. Sci. Technol., vol. 76, no. 6, pp. 867–883, 2025, doi: 10.1002/asi.24978.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

From AI Adoption to AI Pedagogical Transformation: Developing a Faculty AI Literacy Maturity Model for Higher Education

Fatiha Bazouche
Ohio University, Athens, OH, United States of America

Article DOI: https://doi.org/10.68286/samyrad2026.406

ABSTRACT

Faculty members across higher education are increasingly using generative artificial intelligence for instructional design, assessment, feedback, and learning support, yet many institutional AI initiatives remain focused on basic technology adoption rather than on pedagogical transformation. Existing developmental models describe how educators progress with digital technologies in general, but limited work explains how faculty move from initial AI awareness toward pedagogically grounded and critically informed practice. This conceptual paper proposes a Faculty AI Literacy Maturity Model for higher education. Building on established stage-based and maturity models, and drawing on the Technological Pedagogical Content Knowledge framework, Transformative Learning Theory, the UNESCO AI Competency Framework for Teachers, and Communities of Practice, the model describes a six-stage developmental continuum: awareness, experimentation, pedagogical integration, critical reflection, collaborative transformation, and institutional leadership. Technical competency is treated as a moderating dimension across all stages rather than as a stage in itself. The paper presents the model visually, specifies provisional indicators, evidence requirements, and progression criteria for each stage, describes graduated settings in which faculty can experiment safely, and outlines a lean, credit-bearing implementation pathway. It concludes with a staged agenda for validation addressing content validity, reliability, construct validity, cross-institutional applicability, and longitudinal progression. The model is offered as a proposed framework for future empirical development, not as an established assessment instrument.

KEYWORDS AI literacy maturity model; faculty development; AI pedagogy; higher education innovation; instructional transformation; faculty AI readiness; ethical AI integration; professional development evaluation

How to cite: Bazouche, F. (2026). From AI adoption to AI pedagogical transformation: Developing a faculty AI literacy maturity model for higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.406

Introduction

Artificial intelligence is rapidly transforming higher education teaching and learning. Faculty members are increasingly experimenting with generative AI tools for instructional design, assessment, feedback, communication, and learning support. Many institutional AI initiatives, however, remain focused on basic technology adoption rather than deeper pedagogical transformation. As AI becomes more embedded in higher education, institutions need structured approaches that help faculty move beyond tool use toward ethical, reflective, and instructionally meaningful integration.

Institutional guidance has expanded quickly, but unevenly. Recent policy analyses indicate substantial year-over-year growth in the share of institutions issuing acceptable-use guidance for AI, alongside considerable variation in how far that guidance extends beyond writing tasks into assessment and disciplinary practice [1], [2]. Guidance alone, moreover, does not equate to capability. International survey evidence points to high rates of AI use among academic staff coupled with uneven confidence in applying these tools pedagogically [3]. This gap between availability and capability is the problem the present model addresses. Work on AI in education has long argued that instructional value depends on how tools are embedded in teaching rather than on the tools themselves [4], and recent analysis positions AI literacy, prompt engineering, and critical thinking as interdependent rather than sequential competencies [5]. Systematic review of critical thinking and AI in higher education similarly reports that studies frequently leave the target construct unspecified, which makes claims of instructional benefit difficult to evaluate or compare [6], and earlier work cautioned that the value of intelligent systems in education rests on whether they demand human reasoning rather than substitute for it [7].

Although AI literacy and faculty development have gained increasing attention, limited research has examined how faculty progress from initial AI awareness and experimentation toward pedagogically grounded, critically informed, and institutionally supported AI transformation. Existing discussions often emphasize AI adoption, tool training, or policy development. Established stage-based frameworks for educator digital competence and technology adoption offer useful precedents, but, as the literature suggests, none of them describes the specific progression that generative AI demands of faculty or connects that progression to teaching innovation, critical reflection, assessment redesign, and institutional leadership.

This paper makes three contributions. First, it proposes a six-stage Faculty AI Literacy Maturity Model, represented visually in Fig. 1, and positions it against existing developmental and maturity models. Second, it offers a provisional operational layer: observable indicators, evidence requirements, and progression criteria for each stage, guidance on how faculty with different levels of technical confidence can experiment and progress, and a lean, credit-bearing implementation pathway. Third, it sets out a staged agenda for implementing and empirically validating the model. Consistent with its conceptual nature, the model is presented as a proposed framework for future empirical development rather than as an established faculty evaluation or maturity-assessment instrument.

Related work and theoretical foundations

Developmental and Maturity Models in Educational Technology

Stage-based accounts of educator change long predate generative AI. The Concerns-Based Adoption Model (CBAM) describes how educators' concerns about an innovation move from initial unconcern and information-seeking through personal and management concerns toward concerns about consequences for students, collaboration with colleagues, and refocusing, and pairs these Stages of Concern with behaviorally defined Levels of Use [8]. The European Framework for the Digital Competence of Educators (DigCompEdu) specifies six proficiency levels, from Newcomer and Explorer through Integrator and Expert to Leader and Pioneer, across six competence areas [9]. The UNESCO AI Competency Framework for Teachers organizes its competencies into three progression levels: acquire, deepen, and create [10]. At the institutional level, the e-Learning Maturity Model adapted capability-maturity logic from software process improvement to assess the quality of university e-learning processes [11]. Research on technology integration further shows that progression is constrained both by external, first-order barriers such as access, time, and support and by internal, second-order barriers such as pedagogical beliefs and confidence [12], [13].

These models supply the structural precedent for a staged account, and the present model draws on each. Its sequence parallels the movement in CBAM from self-oriented concerns toward collaboration and refocusing, its six-level granularity is comparable to DigCompEdu, in which the Explorer, Integrator, and Leader levels correspond closely to the experimentation, integration, and leadership stages proposed here, and its stages can be read against the three UNESCO progression levels (Section II-D). None of these models, however, addresses the specific demands that generative AI places on faculty. DigCompEdu treats digital competence generically, CBAM describes responses to any innovation, the e-Learning Maturity Model assesses institutional processes rather than individual practice, and the UNESCO framework, written for teachers broadly, specifies competencies at each level rather than the transitions between levels. Reviews of AI in higher education have also noted the scarcity of educator perspectives and of critical reflection on the pedagogical and ethical implications of AI [14]. The present model responds to this gap in two ways: it treats critical reflection on AI as a distinct developmental stage rather than a cross-cutting attitude, and it links individual faculty progression to the institutional conditions that enable it.

Four further bodies of work supply the model's internal logic. Technological Pedagogical Content Knowledge supplies the integration logic, Transformative Learning Theory supplies the mechanism of change, the UNESCO AI Competency Framework for Teachers supplies the competency vocabulary, and Communities of Practice supplies the social structure through which practice becomes shared.

Technological Pedagogical Content Knowledge

The TPACK framework holds that effective technology integration arises from the intersection of technological, pedagogical, and content knowledge rather than from technical skill alone [15]. Applied to AI, this positions the central question not as whether faculty can operate a generative tool but as whether they can align its use with disciplinary content and instructional purpose. Recent extensions of the framework address generative AI directly, arguing that its breadth and unpredictability strain the original construct [16]. Related work proposes Intelligent-TPACK to capture the ethical and pedagogical knowledge required to integrate AI-based tools responsibly [17], and empirical study of AI-TPACK has begun to map relationships among its knowledge components [18]. TPACK grounds the experimentation and pedagogical integration stages of the model.

Transformative Learning Theory

Mezirow describes transformative learning as beginning with a disorienting dilemma that prompts critical examination of previously unquestioned assumptions [19]. Generative AI functions as such a dilemma for many faculty, particularly where it unsettles established assessment practices. The critical reflection stage of the present model is grounded in this account.

UNESCO AI Competency Framework for Teachers

The UNESCO framework establishes a shared vocabulary for teacher AI competencies spanning human-centered mindset, ethics of AI, AI foundations and applications, AI pedagogy, and AI for professional development [10]. It sits alongside a broader literature defining AI literacy as a set of competencies for evaluating and interacting with AI systems [20], [21], which grounds the awareness stage. Read against the UNESCO progression levels, stages 1 and 2 of the present model correspond broadly to acquiring, stages 3 and 4 to deepening, and stages 5 and 6 to creating (Fig. 1). This alignment is the author's interpretation rather than part of the UNESCO framework. The present model complements the framework by describing progression between levels rather than specifying competencies at a given level.

Communities of Practice

Wenger characterizes communities of practice as groups bound by mutual engagement, joint enterprise, and a shared repertoire [22]. This account grounds the collaborative transformation stage, in which individual experimentation becomes shared departmental practice, and informs the institutional leadership stage, in which that shared practice shapes policy and culture.

Method of model development

This is a conceptual paper. The model was developed through interpretive synthesis of interdisciplinary literature rather than through systematic review, and no new empirical data were collected for this paper.

The development process followed the phase logic common to maturity-model design, which proceeds through scoping, design, population, testing, deployment, and maintenance [23], and the related requirement that a proposed model be compared with existing models before iterative development and planned evaluation [24]. This paper completes the scoping, design, and population phases. It defines the domain and audience (faculty in higher education), compares the model with existing developmental and maturity models as discussed in Related work and Theoretical Foundations, and populates each stage with orientations, institutional supports, indicators, and evidence requirements. The testing, deployment, and maintenance phases are specified as future work.

The synthesis drew on three literatures: technology integration frameworks in teacher education, adult and transformative learning theory, and emerging AI literacy competency frameworks. It was additionally informed by the author's qualitative case study of critical thinking instruction in a hybrid, AI-enhanced undergraduate education course [25]. That study found that development of critical thinking reflected deliberate instructional design more than delivery modality, and that AI functioned as a conditional scaffold whose value depended on the critical literacy of the user. Both findings shaped how the present model distinguishes tool use from pedagogical intention. The synthesis also drew on the author's related work on AI integration with preservice teachers, including a mixed-methods examination of the pedagogical, cognitive, and ethical implications of AI integration for creative and critical thinking [26], an account of promoting critical thinking in hybrid learning [27], and a case study of AI-enhanced game-based learning in hybrid education [28].

Six stages, rather than a smaller number, follow from a single criterion: each boundary must mark a qualitative shift in what a faculty member can do, not a quantitative increase in how much AI they use. Two boundaries derive from the author's Instructional Critical Thinking Model, which was itself constructed inductively from case evidence rather than imposed on it [25]. That model holds that development requires visibility before extension, since practitioners cannot deliberately extend a practice they cannot yet name. Applied to faculty AI literacy, this separates integration, where AI is used with instructional intent, from critical reflection, where the practice itself becomes nameable and therefore examinable. The remaining boundaries follow the frameworks in the Related Work and Theoretical Foundations section of this article, each of which marks a transition the others leave implicit.

The faculty AI literacy maturity model

The model conceptualizes faculty AI literacy as more than technical proficiency. It positions AI literacy as a developmental and transformative process that includes ethical awareness, pedagogical decision-making, critical thinking, assessment redesign, and collaborative innovation. Faculty are understood to move through progressive stages, beginning with awareness of AI tools and risks, proceeding through experimentation with instructional applications, and then to intentional pedagogical integration. More advanced stages involve critical reflection on AI ethics, bias, academic integrity, and student learning, followed by collaborative transformation through communities of practice and, finally, institutional leadership. Fig. 1 represents the model visually.

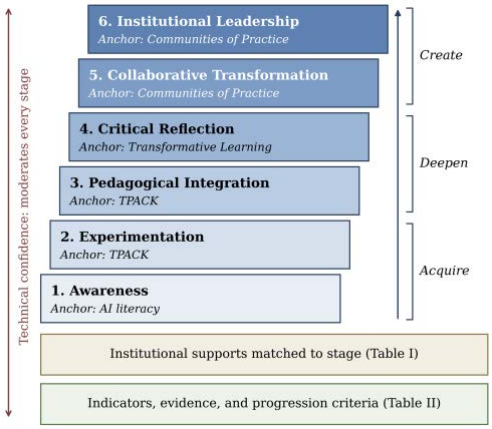


Fig. 1. The Faculty AI Literacy Maturity Model. Stages are grouped by the author against the UNESCO acquire, deepen, and create progression levels, and each stage names its anchoring framework; institutional supports (Table 1) and operational criteria (Table 2) form the base; technical confidence moderates progression across all stages.

The six stages ascend from individual adoption to collective and institutional transformation and are grouped against the three UNESCO progression levels. The theoretical framework anchoring each stage is shown within it, institutional supports and operational criteria form the base on which progression rests, and technical confidence is shown as a dimension that moderates progression at every stage rather than as a stage in itself. Table 1 summarizes the faculty orientation and institutional support associated with each stage.

Table 1. Faculty ai literacy maturity model

Stage	Faculty Orientation	Institutional Support
1. Awareness	Recognizes AI tools and associated risks; forms initial views of relevance to own teaching.	Introductory sessions; plain-language guidance on institutional policy.
2. Experimentation	Trials generative tools for instructional applications; evaluates usefulness case by case.	Sandbox environments; peer rehearsal; low-stakes pilot support.
3. Pedagogical Integration	Moves from incidental use to intentional integration into course design and assessment.	Instructional design consultation; course redesign time.
4. Critical Reflection	Examines AI ethics, bias, academic integrity, and effects on student learning.	Structured reflective forums; policy and integrity guidance.
5. Collaborative Transformation	Works with colleagues to develop shared practice through communities of practice.	Faculty learning communities; cross-disciplinary networks.
6. Institutional Leadership	Shapes departmental and institutional AI strategy, policy, and culture.	Governance roles; strategic investment in sustained development.

Progression is not assumed to be uniform or strictly linear. Faculty may occupy different stages across different courses, may return to experimentation when new tools or new courses appear, and move through the stages as much because of institutional conditions as because of individual disposition.

Stage Indicators and Progression Criteria

A maturity model becomes usable only when its stage boundaries can be applied consistently to individual faculty members. Table 2 therefore proposes, for each stage, observable indicators of practice and the evidence a faculty member would provide to demonstrate readiness to move to the next stage. The indicators are written as observable practices rather than attitudes so that they can be verified from artifacts, and each evidence requirement corresponds to the qualitative shift that defines the next boundary.

A conjunctive placement rule is proposed: a faculty member is placed at the highest stage for which the indicators are met and for which evidence exists at each preceding stage. This rule reflects the developmental claim of the model, and it prevents a faculty member from being placed at an advanced stage, for example on the basis of a governance role, without evidence of the integration and reflection that the model assumes precede it. Because practice can differ across courses, placement is made per course and summarized as a profile rather than as a single score.

Table 2. Provisional stage indicators and progression evidence

Stage	Observable Indicators	Evidence for Progression
1. Awareness	Describes what generative AI tools do and their main limitations (fabrication, bias, data privacy); knows institutional AI policy.	Brief written statement on AI relevance and risk in own discipline.
2. Experimentation	Has tried AI on at least one instructional task in a sandbox, peer rehearsal, or bounded pilot; can say what worked and what failed.	Experimentation log of tasks, outputs, problems, and judgments of usefulness.
3. Pedagogical Integration	AI use is tied to stated learning outcomes; syllabus sets AI expectations; at least one activity or assessment is redesigned.	Annotated course artifact with an alignment rationale.
4. Critical Reflection	Examines AI's effects on assessment validity, equity, integrity, and student reasoning; revises practice accordingly.	Reflective analysis grounded in student work or feedback, with documented revision.
5. Collaborative Transformation	Co-develops and shares AI-informed practice; contributes to a community of practice.	Shared resource reviewed or adopted by peers; record of community participation.
6. Institutional Leadership	Shapes departmental or institutional AI policy, curriculum, or professional development.	Documented contribution to policy, governance, or program design (terminal stage).

Self-report is likely to be adequate for screening at stages 1 and 2, but stages 3 to 6 are defined by changes in practice and should be confirmed through portfolio evidence. These indicators are provisional. They are intended as the item pool and scoring framework for the instrument development described in Evaluation Agenda and Future Work Section, not as a validated assessment.

Experimentation: From Sandbox to Live Classroom

Experimentation is the stage at which faculty first take visible risks, and where that experimentation happens matters. Three graduated settings are distinguished. The first is a sandbox: an institutionally approved, privacy-compliant environment in which faculty try tools on their own materials, such as past assignments, anonymized or synthetic student work, and draft rubrics, without students present and without student data. The second is peer rehearsal, in which an AI-supported activity is trialed with colleagues, for example within a faculty learning community, before it reaches students. The third is a bounded live pilot: a single low-stakes activity in a live class, such as an AI-enhanced game-based task of the kind examined in [28], announced to students in advance, with a non-AI alternative available and a planned fallback if the tool fails. Because mastery experiences are the most influential source of self-efficacy [29], this sequence allows faculty to accumulate successful experience before the stakes rise.

Live pilots will still produce problems: fabricated or inaccurate output, uneven student access to tools, students who outpace the instructor technically, and uncertainty about academic integrity. The model treats these not only as risks to manage but as material for later stages. An inaccurate output examined openly with students becomes an exercise in evaluative judgement [30], and a short post-session log of what happened, how the instructor responded, and what will change supplies the evidence that Table 2 requires for progression to integration and, later, the raw material for critical reflection. Institutional support at this stage therefore includes not only access to tools but explicit permission to fail in bounded ways, together with clear guidance on data privacy and academic integrity.

Technical Competency as a Moderating Dimension

Faculty arrive with different levels of technical confidence, and some approach AI with anxiety rather than curiosity. Research on technology integration identifies confidence and self-efficacy, alongside knowledge, pedagogical beliefs, and institutional culture, as central to whether teachers change practice [13]. The model treats technical competency as a moderating dimension that runs across all stages rather than as a stage in itself (Fig. 1). This choice is deliberate. If technical skill defined the stages, faculty with low technical confidence would be held at the bottom of the continuum regardless of their pedagogical expertise, and faculty with high technical confidence would be credited with a maturity they may not have.

Two implications follow. First, faculty with lower technical confidence can progress to the higher stages, because what distinguishes those stages is pedagogical and critical judgement rather than technical sophistication. An experienced teacher who uses a single tool cautiously but interrogates its effect on assessment validity is operating at stage 4. For these faculty, entry supports should reduce first-order barriers [12] through guided templates, paired mentoring with a more confident colleague, and extended sandbox time. Second, faculty with high technical confidence face a different risk: moving quickly from experimentation into widespread use while bypassing critical reflection. For these faculty, the appropriate support is structured reflection rather than further tool training. Placement should therefore record technical confidence separately, so that support can be differentiated within, as well as across, stages.

The Transition from Integration to Critical Reflection

The transition from stage 3 to stage 4 warrants particular attention, since both involve deliberate use. At stage 3, the faculty member asks whether an AI-supported activity serves the instructional purpose. At stage 4, the question turns back on the tool itself: what the model encodes, whose knowledge it represents, and what it does to the validity of the assessment it supports. Evidence from a case study of critical thinking instruction in an AI-enhanced hybrid course supports this distinction, finding that AI functioned as a conditional scaffold whose instructional value depended on whether users brought critical literacy to it [25]. Design-principles work on scaffolding critical thinking with large language models describes the stage 3 orientation [31], while scholarship on evaluative judgement in a time of generative AI [30] and on the risk of metacognitive offloading [32] describes the questions that characterize stage 4. At this stage, critical thinking is better understood as a human-AI partnership in which the instructor retains evaluative authority than as a capacity that AI either supports or replaces [33]. Case evidence from an AI-mediated undergraduate course reinforces the caution: although critical engagement appeared across student interactions with generative systems, the authors declined any causal claim absent a comparison condition [34]. The distinction also has a professional-identity dimension, since naming one's own instructional practice is bound up with how teachers understand their role [35]. The author has examined this mechanism directly, finding that Socratic questioning in AI-enhanced higher education functions to surface assumptions and make reasoning available for examination rather than to supply conclusions [36].

A lean, credit-bearing implementation pathway

The indicators in Table 2 make it possible to translate the conceptual stages into a faculty-development system. The proposed pathway is built on lean principles, which in higher education emphasize defining value from the perspective of those served, removing activities that do not contribute to that value, and improving processes through short, evidence-informed cycles [37]. Applied here, value is defined as change in faculty practice that benefits student learning, and the principal source of waste is provision mismatched to stage: introductory tool training offered to faculty who already integrate AI, or governance roles offered to faculty who have not yet experimented.

Entry by Placement

Rather than enrolling all faculty in a common sequence, the pathway begins with placement after which faculty draw on the support matched to their stage in Table 1. This allows faculty to enter at the stage their existing practice warrants, and it is consistent with evidence that teacher change often follows, rather than precedes, successful change in classroom practice [38].

Credit-Bearing Recognition

Each stage transition is associated with a micro-credential defined by three elements: the competencies addressed, mapped to the five UNESCO dimensions [10]; the evidence requirement in Table 2; and a peer or faculty-developer review of that evidence. In U.S. higher education, credentials at the earlier stages could carry recognition as documented professional-development hours, while credentials at stages 4 to 6 could be recognized in annual review and promotion and tenure dossiers, or supported through course release and stipends for facilitating communities of practice. Tying recognition to evidence of changed practice rather than to attendance keeps the credential aligned with the qualitative shifts that define the stages. Specific credit values are institutional decisions and are deliberately left open.

Implementation Cycles and Measurable Outcomes

Implementation would proceed in semester-length Plan-Do-Study-Act cycles of the kind used in improvement science [39]. Each cycle begins with placement, delivers stage-matched support, collects evidence, and reviews results before the next cycle begins. Measurable outcomes can be organized by the five levels of professional development evaluation: participants' reactions, participants' learning, organizational support and change, participants' use of new knowledge and skills, and student learning outcomes [40]. Operationally, these include the proportion of participants completing a stage transition within a cycle, the quality of submitted artifacts scored against a rubric, change in the institutional distribution of faculty across stages, and course-level indicators of student learning where AI-supported designs have been introduced. Cohort structures drawn from faculty learning community practice [41] provide the social infrastructure for stages 4 and 5.

Evaluation agenda and future work

Because the model is conceptual, its stages, boundaries, and indicators should be treated as hypotheses. Table 3 sets out a staged agenda for implementation and validation, organized around the sources of validity evidence identified in the Standards for Educational and Psychological Testing: test content, response processes, internal structure, relations to other variables, and consequences of use [42].

Two findings would count against the model as proposed. If item-response or latent class analyses showed that adjacent stages, particularly integration and critical reflection, could not be distinguished empirically, the six-stage structure would need to be collapsed or redefined. If longitudinal data showed that faculty routinely reached the collaborative or leadership stages without evidence of critical reflection, the claim of ordered progression would need to be revised. Stating these conditions in advance is intended to make the model testable rather than merely illustrative. The portfolio rubric in Phase 2 can build on prior work on measuring critical thinking in hybrid classrooms with emerging technologies, which favors evidence of reasoning processes over final products [47]. A practical first step would be a one-year pilot with a small, cross-disciplinary faculty cohort at a single institution, combining the draft instrument with stage-matched support and design-based refinement of both, before multi-site validation.

Table 3. Proposed implementation and validation agenda

Phase	Guiding Question	Proposed Methods
1. Content validation	Are stages and indicators complete, distinct, and correctly ordered?	Delphi panel of faculty developers, instructional designers, and AI-in-education researchers [43]; content validity indices.
2. Instrument development	Do faculty interpret items and evidence requirements as intended?	Items written from Table 2 indicators; cognitive interviews; pilot of the portfolio rubric.
3. Reliability and structure	Do items form a coherent scale with distinct, ordered stages?	Multi-institution survey; internal consistency; Rasch analysis of stage ordering [44]; inter-rater agreement on portfolio scoring.
4. Cross-institutional applicability	Does the model hold across institution types, disciplines, and countries?	Multi-site sampling; tests of measurement invariance across groups.
5. Longitudinal progression	Do faculty move through the stages, and in this order?	Two- to three-year panel data; latent transition analysis [45].
6. Implementation impact	Does stage-matched provision change practice and student learning?	Design-based research iterations [46]; evaluation at the five levels in [40].

Discussion and implications

Once operationalized and validated, the model is intended to be usable. A center for teaching and learning could apply it diagnostically, placing faculty at a stage and matching them to appropriately tiered programming rather than offering a single general workshop. Faculty at the awareness and experimentation stages need low-stakes exposure, safe experimentation environments, and clear institutional guidance. Faculty at the integration and reflection stages need design consultation and time for course redesign. Faculty at the collaborative and leadership stages need convening structures and governance roles rather than further training. Designing workshop provision around this kind of staged intent, rather than around tool coverage, has been explored in earlier work on professional development for critical thinking in hybrid higher education settings [48], and translated into workshop training on strategies for fostering critical thinking in hybrid and AI-enhanced learning environments [49]. A finding from that program of work bears directly on how such provision should be sequenced: a professional development workshop deepened instructors'

practice where they were already intentional, giving intuitive practice a shared vocabulary rather than introducing technique to those without one [25]. Tiered programming should therefore be designed to make existing practice nameable before attempting to extend it.

The model also clarifies where institutional effort tends to be misallocated. Development programs frequently concentrate resources at the earliest stages, where demand is most visible, while providing little structure for faculty who have already integrated AI and now face harder questions about assessment validity and disciplinary epistemology.

Limitations

The model is conceptual and has not been empirically validated. It was derived from literature synthesis rather than from observation of faculty practice, and the six stages are proposed rather than demonstrated. The indicators, evidence requirements, and progression criteria in Table 2 and the implementation pathway in discussion related to a Lean, Credit-Bearing Implementation Pathway are provisional; no instrument yet exists for reliably placing a faculty member on the continuum, and there is not yet evidence that faculty progress through the stages in the order proposed or that the stages represent distinct levels of maturity. Self-report placement may be affected by social desirability, which is one reason portfolio evidence is proposed for the higher stages. The credit-bearing pathway reflects U.S. institutional structures, and the model as a whole reflects the conditions of well-resourced higher education systems; it may transfer poorly to contexts with different infrastructure, governance, or recognition systems. The validation agenda in Section Evaluation Agenda and Future Work is designed to address these limitations.

Conclusion

Higher education institutions need developmental frameworks that support faculty AI literacy and pedagogical transformation, not only tool adoption. This paper has proposed a six-stage Faculty AI Literacy Maturity Model, positioned it against existing developmental and maturity models, and represented it visually. It has also offered a provisional operational layer of indicators, evidence requirements, and progression criteria; guidance on safe experimentation and on supporting faculty with different levels of technical confidence; a lean, credit-bearing implementation pathway; and a staged agenda for validation. The model is a proposed conceptual framework for future empirical development rather than an established assessment of faculty AI readiness. Its value will depend on whether its stages can be distinguished in practice and whether stage-matched provision helps faculty move from adopting AI toward ethical, reflective, and human-centered AI pedagogy.

Acknowledgment

Artificial intelligence (AI) tools were used solely to verify the language of this manuscript for grammar, clarity, and readability. The content, arguments, and conclusions presented here are the author's own.

Conflict of interest

The author declares no conflicts of interest.

References

[1] EDUCAUSE, "2025 EDUCAUSE AI Landscape Study," EDUCAUSE, Boulder, CO, USA, 2025.

[2] N. Azevedo, P. Robles, E. Best, and D. J. Mallinson, "Institutional policies on artificial intelligence in higher education: Frameworks and best practices for faculty," New Directions Adult Contin. Educ., vol. 2025, no. 188, pp. 70-78, 2025.

[3] UNESCO, "Survey on the use of artificial intelligence in higher education institutions," UNESCO, Paris, France, 2025.

[4] W. Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston, MA, USA: Center for Curriculum Redesign, 2019.

[5] Y. Walter, "Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education," Int. J. Educ. Technol. Higher Educ., vol. 21, art. 15, 2024, doi: 10.1186/s41239-024-00448-3.

[6] A. Salido, I. Syarif, M. S. Sitepu, Suparjan, P. R. Wana, R. Taufika, and R. Melisa, "Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies," Social Sci. Humanities Open, vol. 12, art. 101924, 2025, doi: 10.1016/j.ssaho.2025.101924.

[7] J. M. Spector and S. Ma, "Inquiry and critical thinking skills for the next generation: From artificial intelligence back to human intelligence," Smart Learn. Environ., vol. 6, no. 1, art. 8, 2019, doi: 10.1186/s40561-019-0088-z.

[8] G. E. Hall and S. M. Hord, Implementing Change: Patterns, Principles, and Potholes, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2015.

[9] C. Redecker, European Framework for the Digital Competence of Educators: DigCompEdu, Y. Punie, Ed. Luxembourg: Publications Office of the European Union, 2017, doi: 10.2760/159770.

[10] UNESCO, AI Competency Framework for Teachers. Paris, France: UNESCO, 2024.

[11] S. Marshall and G. Mitchell, "Applying SPICE to e-learning: An e-learning maturity model?" in Proc. 6th Australas. Comput. Educ. Conf. (ACE), Dunedin, New Zealand, 2004, pp. 185-191.

[12] P. A. Ertmer, "Addressing first- and second-order barriers to change: Strategies for technology integration," Educ. Technol. Res. Develop., vol. 47, no. 4, pp. 47-61, 1999, doi: 10.1007/BF02299597.

[13] P. A. Ertmer and A. T. Ottenbreit-Leftwich, "Teacher technology change: How knowledge, confidence, beliefs, and culture intersect," J. Res. Technol. Educ., vol. 42, no. 3, pp. 255-284, 2010, doi: 10.1080/15391523.2010.10782551.

[14] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education - where are the educators?" Int. J. Educ. Technol. Higher Educ., vol. 16, art. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[15] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge," Teachers College Rec., vol. 108, no. 6, pp. 1017-1054, 2006, doi: 10.1111/j.1467-9620.2006.00684.x.

[16] P. Mishra, M. Warr, and R. Islam, "TPACK in the age of ChatGPT and generative AI," J. Digit. Learn. Teacher Educ., vol. 39, no. 4, pp. 235-251, 2023, doi: 10.1080/21532974.2023.2247480.

[17] I. Celik, "Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education," Comput. Hum. Behav., vol. 138, art. 107468, 2023, doi: 10.1016/j.chb.2022.107468.

[18] Y. Ning, C. Zhang, B. Xu, Y. Zhou, and T. T. Wijaya, "Teachers' AI-TPACK: Exploring the relationship between knowledge elements," Sustainability, vol. 16, no. 3, art. 978, 2024, doi: 10.3390/su16030978.

[19] J. Mezirow, Transformative Dimensions of Adult Learning. San Francisco, CA, USA: Jossey-Bass, 1991.

[20] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. CHI Conf. Human Factors Comput. Syst., 2020, pp. 1-16, doi: 10.1145/3313831.3376727.
[21] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, "Conceptualizing AI literacy: An exploratory review," Comput. Educ.: Artif. Intell., vol. 2, art. 100041, 2021, doi: 10.1016/j.caeai.2021.100041.
[22] E. Wenger, Communities of Practice: Learning, Meaning, and Identity. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[23] T. de Bruin, R. Freeze, U. Kaulkarni, and M. Rosemann, "Understanding the main phases of developing a maturity assessment model," in Proc. 16th Australas. Conf. Inf. Syst. (ACIS), Sydney, Australia, 2005.
[24] J. Becker, R. Knackstedt, and J. Pöppelbuß, "Developing maturity models for IT management: A procedure model and its application," Bus. Inf. Syst. Eng., vol. 1, no. 3, pp. 213-222, 2009, doi: 10.1007/s12599-009-0044-5.
[25] F. Bazouche, "Promoting pre-service teachers' critical thinking skills in hybrid learning environments in the age of AI: From teachers' and students' perspectives," Ph.D. dissertation, Dept. Educational Studies, Patton College of Education, Ohio Univ., Athens, OH, USA, 2026.
[26] F. Bazouche, "Empowering preservice teachers' creative and critical thinking through AI integration: A mixed-methods examination of emerging pedagogical, cognitive, and ethical implications," Crit. Think. Lang. Learn., vol. 12, no. 1, pp. 1-24, 2025.
[27] F. Bazouche, "Promoting preservice teachers' critical thinking in hybrid learning," JALT Crit. Think. SIG J., vol. 4, no. 1, pp. 10-20, 2025.
[28] F. Bazouche, "AI-enhanced game-based learning to foster preservice teachers' critical thinking in hybrid education: A case study," in Evolution in Computational Intelligence: Proc. 13th Int. Conf. Frontiers Intell. Comput.: Theory Appl. (FICTA 2025), vol. 1, V. Bhateja, P. Patel, and V. Khattri, Eds. Cham, Switzerland: Springer, 2026, pp. 356-367.
[29] A. Bandura, Self-Efficacy: The Exercise of Control. New York, NY, USA: W. H. Freeman, 1997.
[30] M. Bearman, J. Tai, P. Dawson, D. Boud, and R. Ajjawi, "Developing evaluative judgement for a time of generative artificial intelligence," Assess. Eval. Higher Educ., vol. 49, no. 6, pp. 893-905, 2024, doi: 10.1080/02602938.2024.2335321.
[31] M. Vendrell and S.-K. Johnston, "Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education," Comput. Educ.: Artif. Intell., vol. 10, art. 100572, 2026, doi: 10.1016/j.caeai.2026.100572.
[32] Y. Fan et al., "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance," Brit. J. Educ. Technol., vol. 56, no. 2, pp. 489-530, 2024, doi: 10.1111/bjet.13544.
[33] F. Bazouche, "Redefining critical thinking as a human-AI partnership: A meta-creativity lens," in Proc. Soc. Inf. Technol. Teacher Educ. Int. Conf. (SITE), AACE, 2026, pp. 2808-2819.
[34] P. Brazao and L. Tinoca, "Artificial intelligence and critical thinking: A case study with educational chatbots," Frontiers Educ., vol. 10, art. 1630493, 2025, doi: 10.3389/feduc.2025.1630493.
[35] C. Beauchamp and L. Thomas, "Understanding teacher identity: An overview of issues in the literature and implications for teacher education," Cambridge J. Educ., vol. 39, no. 2, pp. 175-189, 2009, doi: 10.1080/03057640902902252.
[36] F. Bazouche, "The Socratic questioning method in the age of artificial intelligence (AI) in higher education," in Proc. EdMedia + Innovate Learning, T. Bastiaens, Ed. Barcelona, Spain: Assoc. Advancement Comput. Educ. (AACE), 2025, pp. 50-58.
[37] W. K. Balzer, Lean Higher Education: Increasing the Value and Performance of University Processes. New York, NY, USA: Productivity Press, 2010.
[38] T. R. Guskey, "Professional development and teacher change," Teachers Teaching: Theory Pract., vol. 8, no. 3, pp. 381-391, 2002, doi: 10.1080/135406002100000512.
[39] A. S. Bryk, L. M. Gomez, A. Grunow, and P. G. LeMahieu, Learning to Improve: How America's Schools Can Get Better at Getting Better. Cambridge, MA, USA: Harvard Education Press, 2015.
[40] T. R. Guskey, Evaluating Professional Development. Thousand Oaks, CA, USA: Corwin Press, 2000.
[41] M. D. Cox, "Introduction to faculty learning communities," New Directions Teaching Learn., vol. 2004, no. 97, pp. 5-23, 2004, doi: 10.1002/tl.129.
[42] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, Standards for Educational and Psychological Testing. Washington, DC, USA: AERA, 2014.
[43] C. Okoli and S. D. Pawlowski, "The Delphi method as a research tool: An example, design considerations and applications," Inf. Manage., vol. 42, no. 1, pp. 15-29, 2004, doi: 10.1016/j.im.2003.11.002.
[44] T. G. Bond and C. M. Fox, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, 3rd ed. New York, NY, USA: Routledge, 2015.
[45] L. M. Collins and S. T. Lanza, Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. Hoboken, NJ, USA: Wiley, 2010.
[46] S. McKenney and T. C. Reeves, Conducting Educational Design Research, 2nd ed. London, U.K.: Routledge, 2019.
[47] F. Bazouche, "Rethinking assessment in the AI age: Measuring critical thinking in hybrid classrooms with emerging technologies," in Proc. Soc. Inf. Technol. Teacher Educ. Int. Conf. (SITE), AACE, 2026, pp. 3602-3609.
[48] F. Bazouche, "Designing professional development workshop to foster critical thinking skills in hybrid learning environments in higher education," in Proc. eLmL 2025: 17th Int. Conf. Mobile, Hybrid, Online Learning, G. Hube, Ed. IARIA, 2025, pp. 20-26.
[49] F. Bazouche, "From theory to practice: Workshop training on innovative strategies to foster critical thinking skills in hybrid and AI-enhanced learning environment," in Proc. Soc. Inf. Technol. Teacher Educ. Int. Conf. (SITE), AACE, 2026, pp. 3596-3601.

SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN
SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

Enhancing Inclusive Education Through AI: Supporting Neurodivergent Students and Staff in UK Higher Education

Steph Sayan · Alexia Revueltas Roux · Sara Eftekhari · Jade Burke · Helena Paterson
School of Psychology & Neuroscience, University of Glasgow, 62 Hillhead Street, Glasgow, G12 8QB, United Kingdom

Article DOI: https://doi.org/10.68286/samyrad2026.412

ABSTRACT

Increasing numbers of students and staff in Higher Education (HE) identify as being neurodivergent. Neurodivergent individuals are often disproportionately impacted by considerable executive functioning demands in HE (e.g. planning, task initiation), resulting in cognitive overload, poor-decision making, emotional dysregulation, and stress. Artificial Intelligence (AI) tools may offer opportunities to mitigate some of the executive functioning challenges by supporting cognitive scaffolding to execute tasks in an accessible way. In a mixed methods study we explored AI tools' potential to support cognitive and executive functioning demanded in HE, by asking neurodivergent staff and students (n = 56) which AI tools they are currently using, and for what purposes. We analysed findings using the BRIEF model of executive functioning, finding that AI tools can function as forms of cognitive support for individuals experiencing executive dysfunction. We argue that such technologies may help address gaps where existing institutional resources and educational practices are insufficient to support equitable participation and success. However, while AI tools offer potential to support neurodivergent academics, the longer-term impacts of AI use are unknown. We therefore discuss potential caveats to consider when recommending use of these tools to neurodivergent academics.

KEYWORDS artificial intelligence (AI), higher education, neurodiversity, accessibility, executive function

How to cite: Sayan, S., Revueltas Roux, A., Eftekhari, S., Burke, J., & Paterson, H. (2026). Enhancing inclusive education through AI: Supporting neurodivergent students and staff in UK higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.412

Introduction

The emergence of Generative Artificial Intelligence (GenAI) with chat-like interfaces (e.g. ChatGPT, Claude, Gemini) have had extensive societal impacts, prompting wide-ranging change in Higher Education (HE). While AI tools were already being used to automate tasks such as transcription, they often needed extensive editing and correction. GenAI tools introduced Large Language Model (LLM) capabilities of generating content with unique semantic structure but based on known human knowledge and language. In HE, initial attention on AI focused on existential concerns of the future of universities, mostly centred on assessment integrity and the importance of critical engagement with AI tools [1]. Learning and teaching concerns are grounded in evidence from Higher Education Policy Institute (HEPI) surveys that

have documented a rapid behavioural shift, with students' reported use of GenAI rising from 66% in 2024 to 95% in 2026[2]. One of the institutional responses to these concerns, according to HEPI, has been the return to in-person exams as a means of containing "threat".

During assessments, under the UK Equality Act 2010, reasonable adjustments in the form of individualised support must be provided for students with a disability, to allow for equitable participation. Neurodivergent students are eligible for reasonable adjustments. Doyle [3] estimates 15–20% of the adult population has a neurodevelopmental condition (such as autism, ADHD, and dyslexia). Apperly et al. [4] found 12.97% of participants reported a self-diagnosed neurodevelopmental condition, with only 3.23% reporting a formal diagnosis. Furthermore, in 2022 out of 2,038 prospective UK undergraduates, 14.2% considered themselves to have ADHD, autism, or both [5]. In comparison, for UK HE staff, data from 2024–2025 shows a disclosure rate of ~2.3% neurodiversity related disability (~5,600 academics) [6]. Overall, these figures likely underestimate the population of neurodivergent students and staff in HE (hereafter shortened to neurodivergent academics), due to disclosure barriers such as fear of being perceived less capable than neurotypical peers, and negative prior experiences with disclosure, particularly when belonging to a minoritised cultural group [7], [8].

With large class sizes and increasingly diverse learner needs, large-scale provision of adequate support is an ongoing challenge, especially in a context of rising concerns about staff workload and wellbeing [9]. Thus incorporating AI tool use in HE can potentially support teaching and learning by improving productivity and providing individualised support for students and staff [10].

Many tasks required of individuals in HE are in high demand, yet are carried out with limited support; deadlines are commonly clustered around end-of-semester bottlenecks, requiring students and staff to juggle multiple cognitively-demanding priorities with little interim structure and fast turnarounds [11]. Resultantly, while neurodivergent and neurotypical academics alike can thrive and struggle within HE, neurodivergent individuals may be disproportionately impacted due to differences in executive functioning.

Executive function (EF) is broadly defined as the topdown, effortful cognitive control processes that support goal-directed behaviour [12]. Miyake et al. [13] divided EF into three positively correlated and distinct components: shifting, updating, and inhibiting. EF abilities have been found to rapidly develop during childhood, adolescence, and culminate in emerging adulthood [14]. EF abilities play a consequential, if invisible, role in daily activities; from switching between multiple responsibilities, planning your commute, exercising self-control like modulating one's response to a hurtful email, and for finding solutions to new problems [12]. Greater EF ability has been consistently associated with positive outcomes, like academic attainment in HE [15], and the development of social interaction skills [16].

However, it is important to caveat that EF literature has historically been framed from an implicitly normative framework, where deviations from the typically expected (i.e. neurotypical) neurodevelopmental trajectories are framed as deficits, thus failing to acknowledge the extent to which barriers arise from insufficient accommodation of natural neurotype variation. This is reflected in the language used. For example, literature referring to attentional difficulties rather than attentional differences, or using the term executive dysfunction, rather than executive variation. Taking a more strengths-based approach, Rafael et al. [17] conducted a scoping review of ADHD-related strengths in adults. Their results support a reframing of deficit-based language, demonstrating that characteristics ordinarily described as symptoms can function as strengths in particular contexts. For instance, in ADHD impulsivity can sometimes be interpreted as adaptive risk-taking, which can be an asset in competitive or entrepreneurial settings. Similarly, reframing inattention as interest-based attention may help explain why, in high interest conditions, individuals can "hyperfocus" and may sustain attention for longer, work productively under pressure, and draw on greater intrinsic motivation as a result, than when interest is low.

Supporting a strengths-based approach, Chow and Cooper [18] synthesised 26 qualitative studies and found that autistic participants embraced an autistic identity and being "wired differently", experiencing special interests as enriching to their lives. The participants identified strengths that helped them thrive in daily tasks (e.g. making processes more efficient), enhance their performance at work, get more out of the activities they enjoyed, or to help others, with recognition of those skills by others contributing to feeling valued. This reframing does not demerit the genuine struggle that arises from a mismatch between individual needs and environmental/task design. From this position, the difficulties neurodivergent academics face arises principally as an institutional design flaw rather than from inherent individual deficit [19]. Thus overall, recognising strengths allows for a more representative and inclusive reframing of neurodiversity that can help with better provision of support to neurodivergent academics.

Neurodivergent academics can be attracted to academia since it is a space that favours creativity, problem-solving, curiosity, exploration, divergent thinking and in-depth research on specific topics, in alignment with strengths that come with certain neurotypes [11], [17]. However, HE can also have hidden expectations and rigid boundaries that are often built on neurotypical expectations [20]. For example, HE often requires constant switching between tasks, whether that is to teach or to study, given "there's a lot to juggle" p. 825 [11]. For certain neurotypes, such as academics with ADHD, individuals may enjoy having options that provide a constant and varied degree of interest and problem-solving opportunity [17]. However other neurotypes, such as autistic academics, reported finding prioritising between multiple competing tasks or separating the details from the bigger picture less enjoyable and more energy consuming [21]. Similarly, lived experience work from neurodivergent academics highlights a constant feeling of taking longer to achieve similar results to peers, leading to feelings of heightened anxiety, task-paralysis, and imposter syndrome [22].

Furthermore, heightened rejection sensitivity among individuals with ADHD [23] may mean routine feedback is more readily experienced as personal criticism. Thus, the anticipation of criticism can trigger a cycle of avoidance that compounds the original difficulty. This can be true of both students handling feedback from lecturers and peers, but also for staff handling feedback from students, colleagues, and managers. These experiences matter in terms of how support is designed. Interventions must provide holistic support that addresses not just the task demand, but also account for the affective needs of the individuals, their

neurotype, and broader intersectional contexts the individual sits in.

GenAI can potentially be used as a scaffolding tool for EF. Zhao et al. [24] explored how students with disabilities in UK universities were using GenAI tools, with the most reported disability being "neurodiversity-related". The students reported using more versatile chatbots like ChatGPT, or proofreading and translation tools. Notably, as Psychology researchers, we noticed that the tasks participants reported tools performed well for in this sample were EF related. For example, clarifying objectives, providing initial structure, breaking down large tasks into manageable steps, and supporting engagement with lengthy or complex tasks. The student participants in the Zhao et al. study also talked about how tools helped them overcome struggles directly related to their disability, as well as others that are product of a disabling environment. For example, anxiety reduction when using GenAI for emotionally challenging tasks. The students also reported a desire to be more involved with University AI policy and AI literacy around ethical use, fair access and disability support use.

These AI tools can provide tailored and personalised support harnessing individual strengths, while targeting challenges for the individual and their environment [25]. In addition, these AI tools are immediately available and accessible, in part due to being increasingly embedded within technology, and their use does not require a diagnosis or even a disclosure of challenges experienced.

Our study aims to centre the lived experiences of neurodivergent academics in order to understand how they are engaging with AI tools, for both academic and nonacademic tasks. Our team consists of HE staff and students, where we co-designed our study, to bring insider perspective to the two main aspects of UK HE. The study was also designed considering some of the team's lived experience of neurodivergence. Although we acknowledge that singular views cannot be assumed as generalisation, our study still provides a dialogue between neurodivergent and neurotypical views of UK HE.

We aimed to explore staff and students' uses and attitudes towards AI tools using a mixed-methods design via an online survey. In the current paper, we focus on the quantitative analysis of familiarity with AI tools, frequency of use and type of tasks for which the tools were used, and interpret findings through an executive functioning lens.

Methods or analytical approach

Participants

Fifty-six neurodivergent (either formally diagnosed or self-diagnosed) UK staff and students who use AI tools participated in an anonymous online survey. Participants were recruited via opportunity and convenience sampling.
From 83 responses received, 61 were from participants who consented to take part and finished the full survey. Given that this data was part of a bigger project, open text responses were also screened for being AI generated using a two-reviewer procedure. Two researchers independently reviewed each response against a set of predefined red flags (e. g. use of bullet points), and responses were flagged for removal only when both reviewers agreed. Five participants were removed for this reason. The final sample of 56 participants (Female = 38; Male = 17; Non-binary = 1) were aged 18 to 60 years (M age = 28.6 years; SD age = 9.1 years), and consisted mainly of White (n = 34) and Black (n = 15) ethnic backgrounds. The majority (n = 46) of participants reported they were a student in HE. Table I shows the range of neurotypes disclosed, the three most commonly mentioned were: ADHD (n = 31), Autism (n = 20), and Dyslexia (n = 11); although these would often overlap given natural co-occurrences.

Measures

AI Tools List

We created a list capturing as many AI tools as practically possible that were available in the last quarter of 2024 and start of 2025, with a total of 25 tools identified. AI Tools were defined as software, webpages or apps that self-described as AI powered and that could be relevant for research (e.g. Elicit or Research Rabbit), academic productivity (this included tools like Grammarly, Canva AI, Research Rabbit or GenAI such as ChatGPT or Gemini), or tools that were aimed at neurodivergent individuals (e.g. goblin.tools).

Online Survey

We developed an online survey to capture how neurodivergent staff and students had been using AI tools. It was divided into three parts: Part One included items aimed at capturing how staff and students currently used and would like to use AI tools. The full list of AI tool names was presented here to ask participants about their previous knowledge and frequency of using these tools. This section of the survey aimed to capture the participant's use of AI tools in both academic and non-academic settings as part of the exploratory mapping of use.

Table 1. A count of neurotypes disclosed by participants

Neurotype[a]	n
ADHD	31
Autism	20
Dyslexia	11
OCD	5
Dyspraxia	2
Anxiety	1
Mood (unspecified)	1
Synaesthesia	1

Note. Whilst this table counts neurotypes individually, some participants disclosed multiple co-occurring neurotypes e.g. ADHD, Autism, and Dyslexia.

Part Two consisted of open-ended questions designed to capture qualitative data on the participants' perceived benefits and challenges of using AI tools, with a particular focus on their neurodivergence. In Part Three we asked participants for demographic information; including gender, age, position at their university, neurodivergent status, and ethnicity.

See supplementary materials (https://osf.io/h38rb/) for a copy of the survey, the AI tools list resource, and additional demographics information.

Data Collection

The University of Glasgow College of Medical, Veterinary and Life Sciences Ethics Committee provided ethical approval. Participants' informed consent was sought from the start, and any participation that was left unfinished was considered a voluntary withdrawal of the study and was not included in the analysis. Participation was anonymous. Email addresses for

a follow-up study were collected separately and not linked with any research data at any time.

Data collection occurred between 23 January 2025 and 15 April 2025. Upon accessing the survey and answering pre-screening questions, participants were presented with the Participant Information Sheet, Privacy Notice, and the Consent Form. After consent, they accessed Parts One of the survey. At the end, participants could also voluntarily provide an e-mail address if they wanted to take part in follow-up interviews and future research projects.

Reflexivity

The research team brought varied positionalities to the study: ARR, a queer immigrant researcher with ADHD and ND expertise, used an intersectional, decolonial lens to inform design and interpretation; SS, a neurotypical cognitive psychologist new to sustained AI use, took an exploratory stance and anchored interpretation in participants' own language via a reflexive diary; SE brought a critical realist and critical feminist positionality to the quantitative analysis, managed through collaborative team review; JB, without lived ND experience but with close ND relationships and growing AI familiarity, saw their interpretation evolve alongside this knowledge; and HP, from a neurodivergent family with extensive AI expertise, contributed established community-of-practice knowledge. Team discussion throughout analysis helped preserve diverse perspectives and avoid a homogenous interpretation of the data.

Results and findings

Pre-processing

Any identifiable or personal data revealed within the open text answers was redacted or removed from the data to avoid identification. Data was stored in a GDPR compliant server. Data wrangling, cleaning, visualising and quantitative and qualitative analyses were conducted in R v.4.6.0 [26] using mainly tidyverse v.2.0.0 [27]. Find additional details in the supplementary materials. Descriptive statistics were calculated for the demographic questions, and across the survey items, such as, participant usage of AI tools.

Tool use

We wanted to see how familiar participants were with the 25 AI tools we showed them, and whether use changed depending on their academic use, non-academic use, and which tools might be more familiar despite not being as used.

For our 56 neurodivergent participants, GenAI tools like LLMs and Generative Pre-trained Transformers (GPTs) were reported as the most used in both academic and non-academic tasks.

The most used tools were ChatGPT (n = 52) and Grammarly (n = 40). Grammarly was a pre-existing product with a web browser plug-in, and was later embedded with LLM abilities, an example of the increasing incorporation of LLMs into existing products marketed as writing support. In comparison, GenAI tools like ChatGPT do not constrain or limit the kinds of tasks a person can perform with it. We also looked at tools that participants had heard about, but not used and the top ones were Gemini (n = 22) and Copilot (n = 17).

Fig. 1 shows the most used AI tools and those that were known, but not used.

For only academic use, Grammarly was the most used tool, but overall, the most used tools can be categorised under GenAI chatbots (i.e. Copilot, Gemini or ChatGPT), proofreading/editing (e.g. Grammarly or Pro-Writing Aid), organiser and planner (i.e. Notion) or research search engines (e.g. Research Rabbit). Overall, there were similar patterns of AI tools usage for non-academic and/or academic AI tool use, with GenAI or well-known AI tools as the most popular, suggesting that for the time (early 2025), participant exposure to the AI tools may have determined usage.

Regarding frequency, AI tools were most commonly used daily or more for academic use (n=7), closely followed by 2–3 times a week as most common use (n=6). For non academic use, the AI tools were most commonly used 2–3 times a week (n=6). Overall leaning towards more frequent use, if used rather than used sparely. This suggests that for the participants that do use AI tools, it is a part of their academic and non-academic routines.

AI Use per Task

In the open-ended questions, participants identified using AI tools for a variety of tasks. Tasks were grouped based on semantic similarity. We then used a deductive theoretical approach based on the EF domain these tasks could be supplementing, using the three-factor and nine subscale model of the Behavior Rating Inventory of Executive Function (BRIEF) [28] and using the names proposed by CFA replication by Little et al, [29], which evaluates and facilitates the mapping of EF onto daily tasks. We uploaded the BRIEF coding framework and prompted Claude (Sonnet 5) to assign each response to the most appropriate category within that framework. Members of the research team then reviewed every response alongside its assigned category and moved responses to a different category if the initial assignment did not fit after expert consideration. Tasks were further refined based on our lived experience as academics in HE. Some tasks were categorised in more than one domain when task was too broad and could encompass multiple processes, while other tasks could rely on multiple EF processes, but these were allocated to the process that most likely leads the specific task. The final categorisation therefore reflects the research team's judgement, with the AI-generated categorisation serving as a first pass. We used AI to initiate this task given that the categorisation was deductive and descriptive, involving grouping by surface meaning and allocation to predefined categories, which AI has been shown to perform well [30]. Furthermore, given the study's focus on how neurodivergent people use AI as a cognitive scaffold, we considered it appropriate to reflect transparently on our own use of these tools in the research process.

The BRIEF framework showed that the majority of tasks were being used to support executive self-regulation. The metacognitive domain shows a dominance on the processes being used through AI tools, reflecting the "executive" self-regulation. Meta-cognitive support was the most reported, where AI is being used to "jump-start" a task rather than sustain it. Secondly, working memory highlights how AI seems to be used as an external "second brain" for "thoughts, research aims, arguments", while evaluating and bringing together other information.

Planning and organisation suggest that structuring can help place things in perspective as well as provide a blueprint that can then make initiating tasks less overwhelming. Task-monitoring reflects the high-attention to detail that is required

within an academic environment, from attention to citation formats. Table 2 shows the tasks allocated per EF domain and some examples from the data.

The categories comprised both academic tasks (e.g. reference search) and non-academic tasks (e.g. recipe ideas).

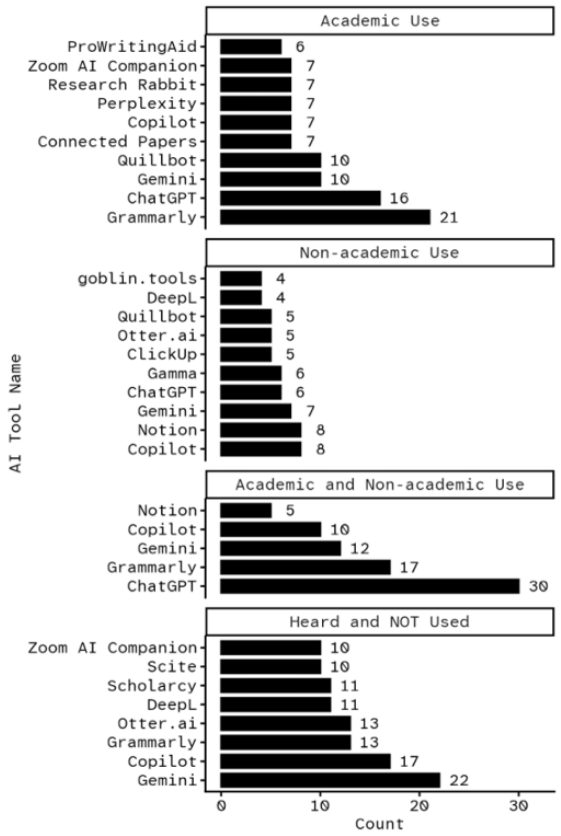


Fig. 1. Most Frequently Used AI Tools for Academic and/or Non-Academic Purposes by frequency. The Top Five AI Tools by Use Type, Tied Ranking Allowed.

The remaining identified tasks were allocated to the closest underlying process or task function unrelated to EF, showing four categories:

- *Creative Thinking*—Tasks in this category were mostly used for creating images, content and storytelling that are not associated with self-regulatory control.
- *Conceptual Explanation and Learning Support*—Tasks in this category were for when AI tools are used as tutors for clarification, knowledge retrieval and new information over planning.
- *Information Seeking and Curiosity*—Tasks in this category were hypothetical questions and for use instead of as a quicker search engine.
- *Tailored Recommendations*—Tasks in this category highlighted the personalised suggestions that save an individual time from searching multiple sources of information.

Finally, there was a category for unspecified tasks where the task description was too vague, to be able to be categorised without knowing the underlying rationale of why the participant had mentioned it or how AI was being used to support that task (e.g. 'project' or 'to do tasks').

Discussion

Our study set out to explore how neurodivergent academics—staff and students in UK HE—use AI tools in both academic and everyday contexts and identify what can be learned for neurodivergent-specific use of AI Tools through an executive function lens. We found that participants were familiar with a range of AI tools, including some that they had not used. We suggest the familiarity-use gap may be due to the need for neurodivergent people, like others with non-visible disabilities, to self-source their own solutions in order to meet demand needs of neurotypical environments, which may prompt finding tools that are not subsequently used [31]. Alternatively, the breadth of familiarity may reflect a high interest in technology and in making processes more efficient, often a self-identified strength in neurodivergent individuals [11], [18].

GenAI chatbots, particularly ChatGPT, were the most commonly used AI tools in both academic and non-academic settings. This may reflect both their familiarity and versatility across multiple tasks. GenAI popularity is unsurprising given the natural language interface providing intuitive interaction, along with public attention at their launch.

Boubker [32] found that perceived usefulness, ease of use and social influence all predicted students' adaptation of ChatGPT. Furthermore, Labadze et al., [33]'s review of 67 studies found that AI chatbots can be particularly helpful for learners and educators since they offer immediate and personalised assistance based on individual's unique needs and preferences, which can also be reflected in the type of tasks reported by our participants.

By comparison, tools specifically developed for neurodivergent people (e.g. goblin.tools), which scaffold support according to level of need, were used less often by our participants in 2025, despite being the target users for these tools. One reason could be the open-source community-driven approach behind these tools, and having not been embedded in everyday technology, potentially diminishing word-of-mouth impact. Adoption may therefore depend on how long a tool has been available, and how widely it is promoted, as well as how much it meets users' needs.

Another reason GenAI chatbots may be used by neurodivergent academics could be that they accommodate unstructured input: thinking does not need to be organised or precisely worded, and instructions can be clarified iteratively based on outputs. This may support thinking-aloud processes, akin to the "rubber duck" effect, where the thinking and understanding happens as a result of the perceived interaction with another. Externalising thought in this way makes the content of one's thinking available for inspection, drawing on meta-cognition—the monitoring, regulation and organization of one's own thinking [34].

The BRIEF-framed analysis of our tasks also reflected that holding and manipulating information through working memory was one of the main uses of AI tools given the broad range of tasks mapped to it. GenAI might be experienced as particularly helpful for students with ADHD to support their metacognitive skills as it can provide them with an external thinking "sandpit" to output their internal information and they can examine it without relying on memory alone—a process supported by working memory [35]. The tasks participants most commonly used AI for were broadly similar to those reported in previous research of AI use in HE [36].

Table 2. Mapping neurodivergent ai task use to executive function

EF Factor & Process	Definition	Example Quotes
Behaviour Regulation Inhibit	Control impulses; stop a behaviour at the appropriate time. Filtering irrelevant information	Summarising notes/articles /papers/notes/book; condensing text; ask for summaries of advice (instead of going through blogs, reddit etc); collating information
Behaviour Regulation Self-Monitor	Awareness of how one's own communication affects others.	editing email communication (to tone down its abruptness); checking the tone of a text or email & how to response appropriately; professional sounding emails and letters; appropriate ways to relate with people
Emotional Regulation Shift and Flexibility	Move flexibly between situations, ideas, or frames of reference. Consider reader writer perspective	Advice on friendships and relationships; casual conversations; writing application letters, CVs, cover letters; translating text into English; coming with responses to people I am communicating with.
Emotional Regulation Emotional Control	Modulate emotional responses appropriately.	seeking moral support; as a friend; times when I needed support but it was 'quicker' to use ai; counselling
Meta-cognition Working Memory	Hold and manipulate information in mind to complete a task.	Research; writing essays/ reports/paragraphs; finding papers for research; drafting letters / proposals/ template; simplifying text I do not understand; note-taking; rewording: feedback, unclear sentences, things, sentences; rewriting things so they're simpler
Meta-cognition Initiate	Begin a task or generate output independently.	Drafting email / assisting me form emails; brainstorming; content generation or ideas; social media posts; blogging & article writing / travel blog writing / newsletter writing; help writing/responding texts; Using it more so as a different Google or baseline starting point; for generating text; writing sections of papers; ideas for writing CVs, job interviews, linkedin etc; give me some ideas on some of my project stuff; drafting possible research project ideas; ideas / idea generation / idea formation; ideas on how to make tables
Meta-cognition Plan	Anticipate future needs; set goals; sequence steps toward them.	time management; planning; scheduling meetings and classes; planning projects or essays; creating a plan for work I will complete; creating dissertation timeline planning long runs; breaking down to lists as small as possible & giving the expected time each task will take to do (time estimation); preparing research drafting proposals; dissertation help; possible essay structure for one assignment. Study support; help with paragraph structure; present idea; shopping list and recipe creating; writing code / coding / help with coding (R and Python); predict potential risks in my business
Meta-cognition Organisation	Keep tasks, records, or spaces structured and orderly.	daily diaries, creating routines, flow of my work, organising workload, helping me write to do lists, ask to break down large tasks, breaking things down so they're easy to understand, making budgets, making invoices, flashcard making, breaking down the points of a research paper to help processing; data analysis
Meta-cognition Task-Monitor	Check one's own work during or after completion to catch errors.	Proofreading; spellchecking; editing; checking documents; punctuation; editing work i have written myself to check for grammar mistakes; forming citations; help with r studio code and errors, programming problems, finding synonyms

However, the underlying reasons for using these tools may be unique to a neurodivergent population, relating to their particular strengths and challenges surrounding EF, a unique strength of using an EF framework to ground their AI use. Our BRIEF-framed analysis showed that AI tools were mostly used to scaffold metacognitive processes to support self-regulation of task execution: starting, holding, sequencing and revising work. Initiation, Planning and Organisation were some of the most frequently mentioned processes, suggesting that participants used AI to help close the gap between intention and action.

For example, participants described using AI tools to create to-do lists, prioritise tasks, or begin projects by asking general questions or finding references. Executive functions have been shown to moderate the intention behaviour link, with those higher in EF more likely to act on their intentions [40]. Generative AI tools may therefore support people in translating intentions into action by taking on some of the executive demands involved. We tentatively suggest that this represents a form of cognitive offloading of EF.

Tools that generate instructions, organise material and suggest decisions may therefore prompt movement from intention to action and may support the reduction of task inertia or choice-paralysis, a phenomenon experienced as an inability to initiate a task when multiple competing demands are associated with similar estimated priorities [37]. Neurodivergent staff and students reported GenAI use to support processes which they have mentioned can be one of the challenges of academia such as juggling multiple responsibilities [11] or estimating the appropriate time for a task. Similarly, tasks which AI tools were frequently used included proofreading (e.g. using Grammarly) and support with coding. These tasks require sustained high attention-to -detail, which may be effortful for some neurotypes, particularly for those with dyslexia or ADHD. Interestingly, some of the tasks that were not part of the BRIEF mapping, such as creative thinking, divergent-thinking or curiosity-driven questions reflect well some of the highlighted strengths that have been reported by neurodivergent individuals as well as by their peers [38].

Although fewer tasks mapped onto emotional regulation, this component likely interacts with cognitive and behaviour regulation, governing emotional control, motivation and perspective-taking. Monitoring tone in a message, and evaluating social situations to compose a "neutral" tone are tasks that can carry considerable emotional weight for neurodivergent people, particularly given heightened rejection sensitivity [23] along with struggles around double-empathy tension increasing the uncertainty of how message will be received [11]. We suggest that communication-oriented uses may involve offloading emotional risk: participants may use AI tools to seek clarification without the social cost or exposure to rejection, a risk that may feel greater given that neurodivergent people are more likely to have experienced social rejection in the past [23]. GenAI chatbots may thus offer a form of psychological safety, in which neurodivergent individuals need to self-monitor less, whether in their choice of language, the order and clarity of their ideas, or the emotional response associated with asking questions. In this sense, we suggest these "private" digital spaces could become safe havens for some neurodivergent individuals. We are currently exploring this possibility with follow-up research.

Framing EF in terms of the fit between a person's profile and the demands of their environment shifts focus towards what can be changed outside of the individual: how tasks are structured, how expectations are communicated, what "hidden" implicit steps are needed within a task, and how much unsupported executive load an environment assumes its users can carry. The adjustments process illustrates this Sisyphean task: securing support is itself an executively demanding, self-directed, multi-step task [39] that assumes a constant neurotypical functioning capacity. Thus, just like Sisyphus carrying his stone up the mountain for it to be sent down, the mechanism intended to reduce executive cognitive load relies on the capacity it is meant to accommodate.

Because the mapping is organised around the EF underlying the skills, rather than tied to a specific neurotype, our mapping has relevance beyond neurodivergence: it can be applied to any group encountering similar struggles with EF demands. This is in keeping with the principles of Universal Design for Learning [40], which focus on making learning and academic practice accessible to as many people as possible, reducing the need for individual disclosure or accommodation. First-generation students, for example, often have to learn the culture and "hidden curriculum" of academia, as do international students after moving to a new country. Similarly, assimilating and monitoring tone is something that has been observed during cultural adaptation, and may be a barrier for creating meaningful relationships. These individuals may experience similar benefits from changes in institutional practice, as well as from being provided resources that can support scaffold per task.

Our findings reflect some of the most common struggles neurodivergent academics are experiencing and highlight ways in which this population is finding "fast" ways to alleviate their experiences. However, they cannot speak to the long-term implications of this AI use, and the results can raise concerns about loneliness and isolation if individuals are turning to AI for socioemotional support rather than building connections with others. Related concerns by individuals and institutions have been raised in regards to the longer-term impact on self-efficacy; the belief in one's ability to carry out a task [41]. It has been noted that frequent use of AI may result in dependency, consequently reducing a student's ability to solve problems independently [42].

This introduces a potential tension between short-term and long-term outcomes. Nonetheless, we argue there is nuance to this critique, since research has shown that cognitive off-loading combined with cognitive vigilance can support a Human-AI interaction to foster transformative learning [43].

Critical, informed use of AI tools as assistive technology to supplement cognitive demands could also foster a sense of self-efficacy and self-achievement, particularly amongst individuals who are constantly challenged by the interaction of their struggles with the environmental demands of HE. Research has shown that when GenAI tools function as a scaffold, rather than a substitute for skill development, they can be great tools for learners. Zhu et al. [35] worked alongside students with ADHD and expert coaches of ADHD and co-developed an intervention of metacognitive use of GenAI to support their struggles. Their findings reflect the fine balance that is needed when adapting a new tool for accessibility, since when used consciously to build metacognitive scaffold (often a challenge with ADHD), then GenAI tools can strengthen these EF skills. However, when the design is not careful or the use is not mindful and reflective, there is a risk of fostering dependency on supplemented metacognition. We therefore acknowledge that caution is merited until long-term outcomes can be established.

Limitations

Although we explored the experiences of staff and students, our respondents were predominantly students. This could bias perspectives on academic tasks towards learners, and may not fully reflect how neurodivergent staff in HE might be using these tools. Existing evidence [44] suggests that AI use does change with increased academic experience, finding that graduate students with more mastery were more likely to be more systematic and methodical across four self-regulated learning strategies, compared to undergraduates who were more reactive and exploratory. Future studies should look into how neurodivergent staff and students are using these strategies. Our data captured only the tasks participants reported using AI for; we inferred how and why. Although these inferences were theoretically informed, they relied on our interpretation influenced by our lived experiences as staff and students within HE rather than participants' lived experience. Future research should ask neurodivergent staff and students directly how and why they use AI for each task, and whether their approach differs from their neurotypical peers. Furthermore, given that workload and environment demands are different for members of staff depending on their role and level of seniority, data is still lacking on neurodivergent members of staff outwith lecturers and researchers, for example focusing on professional services and technicians. An additional limitation is the rapid pace of technological development. We are aware of tools that have emerged since we collected our data that could serve academic (e.g. Claude, Notebook LM) and non-academic purposes (e.g. Canva AI), but we would expect the underlying EF processes that these tools support to be broadly consistent.

Conclusion

Our study highlights the EF processes underlying some of the tasks that neurodivergent staff and students use AI tools for in UK HE. This provides a psychological lens through which to

interpret the use of AI that our participants reported, and our analysis demonstrates how AI might be meaningfully scaffolding several of these processes. However, caution is needed in terms of establishing an evidence-base for the longer-term best practice use of AI and how it can be used to support neurodivergent academics.

Declarations

AI use disclosure: In keeping with the research theme, the authors engaged with AI tools during the research process. At key stages in the paper, we have highlighted how AI was used. Additionally, Claude was used in editing for cohesion, proofreading and consistency and to help reduce wordcount and for synthesis of participants' individual reflexivity statements to a single paragraph. Google Labs was used to search for key papers. Copilot was used for "rubber duck" R coding during analysis and in editing text.

References

[1] G. Carden and J. Freeman, 'AI and the Future of Universities', Higher Education Policy Institute, 193, Oct. 2025. doi: https://www.hepi.ac.uk/wp content/uploads/2025/10/AI-and-the-Future-of Universities.pdf.
[2] R. Stephenson and C. Armstrong, 'Student Generative AI Survey 2026', Higher Education Policy Institute, 199, 2026. doi: https://www.hepi.ac.uk/reports/student-generative-ai survey-2026/.
[3] N. Doyle, 'Neurodiversity at work: a biopsychosocial model and the impact on working adults', Br. Med. Bull., vol. 135, no. 1, pp. 108–125, Oct. 2020, doi: 10.1093/bmb/ldaa021.
[4] I. A. Apperly, R. Lee, S. W. van der Kleij, and R. T. Devine, 'A transdiagnostic approach to neurodiversity in a representative population sample: The N+ 4 model', JCPP Adv., vol. 4, no. 2, p. e12219, 2024, doi: 10.1002/jcv2.12219.
[5] J. Shaw and F. Selman, 'An asset, not a problem: meeting the needs of neurodivergent students', Unite Students, 2023.
[6] 'Table 5 - HE academic staff by disability and academic employment function 2014/15 to 2024/25 | HESA'. Accessed: Oct. 13, 2026. [Online]. Available: https://www.hesa.ac.uk/data-and-analysis/staff/table-5
[7] Y. Bolourian, S. M. Zeedyk, and J. Blacher, 'Autism and the university experience: Narratives from students with neurodevelopmental disorders', J. Autism Dev. Disord., vol. 48, no. 10, pp. 3330–3343, Oct. 2018, doi: 10.1007/s10803 018-3599-5.
[8] D. Mellifont, 'Ableist ivory towers: a narrative review informing about the lived experiences of neurodivergent staff in contemporary higher education', Disabil. Soc., vol. 38, no. 5, pp. 865–886, May 2023, doi: 10.1080/09687599.2021.1965547.
[9] S. Wray and G. Kinman, 'Supporting staff wellbeing in higher education', Education Support, 2021. [Online]. Available: https://www.educationsupport.org.uk/media/x4jdvxpl/es supporting-staff-wellbeing-in-he-report.pdf
[10] F. Wu, Y. Dang, and M. Li, 'A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education', Behav. Sci., vol. 15, no. 4, p. 467, Apr. 2025, doi: 10.3390/bs15040467.
[11] S. C. Jones, 'Autistics working in academia: What are the barriers and facilitators?', Autism, vol. 27, no. 3, pp. 822–831, Apr. 2023, doi: 10.1177/13623613221118158.
[12] A. Diamond, 'Executive Functions', Annu. Rev. Psychol., vol. 64, pp. 135–168, 2013, doi: 10.1146/annurev-psych-113011143750.
[13] A. Miyake, N. P. Friedman, M. J. Emerson, A. H. Witzki, A. Howerter, and T. D. Wager, 'The unity and diversity of Executive Functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis', Cognit. Psychol., vol. 41, no. 1, pp. 49–100, Aug. 2000, doi: 10.1006/cogp.1999.0734.
[14] A. Folker et al., 'A longitudinal study of adolescent-to-young adult Executive Function development in seven countries', Dev. Sci., vol. 28, no. 4, p. e70040, 2025, doi: 10.1111/desc.70040.
[15] L. E. Knouse, G. Feldman, and E. J. Blevins, 'Executive functioning difficulties as predictors of academic performance: Examining the role of grade goals', Learn. Individ. Differ., vol. 36, pp. 19–26, Dec. 2014, doi: 10.1016/j.lindif.2014.07.001.
[16] Y. Moriguchi, 'The early development of executive function and its relation to social interaction: a brief review', Front. Psychol., vol. 5, p. 388, Apr. 2014, doi: 10.3389/fpsyg.2014.00388.
[17] R. B. Rafael, H. Jia, M. Rouel, B. M. Wootton, and D. Mitchison, 'Attention deficit/hyperactivity disorder (ADHD) related strengths in adults: A scoping review', J. Atten. Disord., p. 10870547261425737, Mar. 2026, doi: 10.1177/10870547261425737.
[18] C. Chow and K. Cooper, 'What are the lived experiences of strengths in autistic individuals? A systematic review and thematic synthesis', Autism Adulthood, vol. 8, no. 1, pp. 14 26, Feb. 2026, doi: 10.1089/aut.2023.0172.
[19] M. Ronksley-Pavia et al., 'A scoping literature review of generative artificial intelligence for supporting neurodivergent school students', Comput. Educ. Artif. Intell., vol. 9, p. 100437, Dec. 2025, doi: 10.1016/j.caeai.2025.100437.
[20] M. F. Sulaimani and D. M. Gut, 'Hidden curriculum in a special education context: The case of individuals with autism', J. Educ. Res. Pract., vol. 9, no. 1, pp. 30–39, 2019.
[21] N. Kenny et al., '"Tell me what my job is": a qualitative exploration of the experiences of autistic academic staff working in higher education in Ireland', Autism, vol. 30, no. 7, pp. 1724–1737, Jul. 2026, doi: 10.1177/13623613261440799.
[22] N. R. Bressman et al., 'Strengths, challenges, and experiences of neurodivergent faculty in academia', Integr. Comp. Biol., vol. 66, p. icag019, Jan. 2026, doi: 10.1093/icb/icag019.
[23] A. Rowney-Smith, B. Sutton, L. Quadt, and J. A. Eccles, 'The lived experience of rejection sensitivity in ADHD - A qualitative exploration', PLOS One, vol. 21, no. 1, p. e0314669, Jan. 2026, doi: 10.1371/journal.pone.0314669.
[24] X. Zhao, A. Cox, and X. Chen, 'The use of generative AI by students with disabilities in higher education', Internet High. Educ., vol. 66, p. 101014, Jun. 2025, doi: 10.1016/j.iheduc.2025.101014.
[25] J. Liao and C. S. Lee, 'The potential of generative AI in supporting neurodiversity in higher education: A systematic review', Proc. Assoc. Inf. Sci. Technol., vol. 62, no. 1, pp. 1013–1018, 2025, doi: 10.1002/pra2.1331.
[26] R Core Team, R: A language and environment for statistical computing. (2026). doi: 10.32614/R.manuals. https://www.R project.org/.
[27] H. Wickham et al., 'Welcome to the Tidyverse', J. Open Source Softw., vol. 4, no. 43, p. 1686, Nov. 2019, doi: 10.21105/joss.01686.
[28] G. A. Gioia, P. K. Isquith, P. D. Retzlaff, and K. A. Espy, 'Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample', Child Neuropsychol., vol. 8, no. 4, pp. 249–257, Dec. 2002, doi: 10.1076/chin.8.4.249.13513.
[29] C. W. Little, J. Taylor, A. Moltisanti, C. Ennis, S. A. Hart, and C. Schatschneider, 'Factor structure and etiological architecture of the BRIEF: A twin study', J. Neuropsychol., vol. 11, no. 2, pp. 252–276, Jun. 2017, doi: 10.1111/jnp.12084.
[30] H. L. Bunt, A. Goddard, T. W. Reader, and A. Gillespie, 'Validating the use of large language models for psychological text classification', Front. Soc. Psychol., vol. 3, Feb. 2025, doi: 10.3389/frsps.2025.1460277.
[31] L. G. Hamilton and S. Petty, 'Compassionate pedagogy for neurodiversity in higher education: A conceptual analysis', Front. Psychol., vol. 14, Feb. 2023, doi: 10.3389/fpsyg.2023.1093290.
[32] O. Boubker, 'From chatting to self-educating: Can AI tools boost student learning outcomes?', Expert Syst. Appl., vol. 238, p. 121820, Mar. 2024, doi: 10.1016/j.eswa.2023.121820.
[33] L. Labadze, M. Grigolia, and L. Machaidze, 'Role of AI chatbots in education: systematic literature review', Int. J. Educ. Technol. High. Educ., vol. 20, no. 1, p. 56, Oct. 2023, doi: 10.1186/s41239-023-00426-1.
[34] L. Guo, 'Using metacognitive prompts to enhance self regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments', J. Comput. Assist. Learn., vol. 38, pp. 811 832, Feb. 2022, doi: 10.1111/jcal.12650.
[35] Z. Zhu, J. Yu, and Y. Luo, 'Scaffolding metacognition with genAI: exploring design opportunities to support task management for university students with ADHD', in Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, in CHI '26. New York, NY, USA: Association for Computing Machinery, Apr. 2026, pp. 1–24. doi: 10.1145/3772318.3790697.

[36] I. M. Castillo-Martínez, D. Flores-Bueno, S. M. Gómez Puente, and V. O. Vite-León, 'AI in higher education: a systematic literature review', Front. Educ., vol. 9, Aug. 2024, doi: 10.3389/feduc.2024.1391485.
[37] K. L. Buckle, K. Leadbitter, E. Poliakoff, and E. Gowen, '"No way out except from external intervention": first-hand accounts of autistic inertia', Front. Psychol., vol. 12, p. 631596, Jul. 2021, doi: 10.3389/fpsyg.2021.631596. A. Gogne, 'Strengths in neurodivergence', in Neurodevelopmental Disorders in Adult Women: Special Considerations in the Perinatal Period,
[38] A. Gogne, Ed., Cham: Springer Nature Switzerland, 2025, pp. 145–160. doi: 10.1007/978-3-031-86588-6_7.
[39] G. Brewer, E. Urwin, and B. Witham, 'Disabled student experiences of Higher Education', Disabil. Soc., vol. 40, no. 1, pp. 108–127, Jan. 2025, doi: 10.1080/09687599.2023.2263633.
[40] N. Martin, M. Wray, and J. Krupa, Eds, Universal Design for Learning: A Critical Approach. London: Routledge, 2025. doi: 10.4324/9781041057796.
[41] A. Bandura, 'Self-efficacy: Toward a unifying theory of behavioral change.', Psychol. Rev., vol. 84, no. 2, pp. 191 215, Mar. 1977, doi: 10.1037/0033-295X.84.2.191.
[42] S. Zhang, X. Zhao, T. Zhou, and J. H. Kim, 'Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior', Int. J. Educ. Technol. High. Educ., vol. 21, no. 1, p. 34, May 2024, doi: 10.1186/s41239-024-00467-0.
[43] S. Wang and H. Zhang, 'Pedagogical partnerships with generative AI in higher education: how dual cognitive pathways paradoxically enable transformative learning', Int. J. Educ. Technol. High. Educ., vol. 23, no. 1, p. 11, Mar. 2026, doi: 10.1186/s41239-026-00585-x.
[44] M. Perifanou and A. A. Economides, 'How undergraduate and graduate students self-regulate GenAI use in project-based learning: choosing, creating, filtering, and stopping strategies', J. Educ. Comput. Res., p. 07356331261469412, Jul. 2026, doi: 10.1177/07356331261469412.