

**EdUHK x HKUST
Joint International Conference on**



AI and Education

20 – 21 August 2025

◆ **Conference Proceedings** ◆



**The EdUHK x HKUST Joint
International Conference on AI and Education 2025
(AIEdu2025)**

20 – 21 August 2025

Cite as

Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). Proceedings of International Conference on AI and Education 2025 (AIEdu2025). Hong Kong SAR: The Education University of Hong Kong.

ISSN: 3105-5702

Copyright 2025 International Conference on AI and Education

All rights reserved

Publisher
Centre for Learning, Teaching and Technology
The Education University of Hong Kong

The 2nd EdUHK x HKUST Joint
International Conference on AI and Education 2025 (AIEdu2025)
Proceedings

Editors:

XU Guandong, The Education University of Hong Kong
LOOI Chee Kit, The Education University of Hong Kong
MCMINN Sean, The Hong Kong University of Science and Technology
CHENG Kwok Shing Gary, The Education University of Hong Kong

Co-Editors (in alphabetical order of surnames):

BA Shen, The Education University of Hong Kong
CHIU Ming Ming, The Education University of Hong Kong
FU Hong, The Education University of Hong Kong
HUI Pan, The Hong Kong University of Science and Technology
KWOK Tin Yau James, The Hong Kong University of Science and Technology
LAU Chaak Ming, The Education University of Hong Kong
LEUNG Ka Lok Jac, The Hong Kong University of Science and Technology
SO Chi Fuk Henry, The Education University of Hong Kong
SONG Yanjie, The Education University of Hong Kong
WONG Ka Lei James, The Hong Kong University of Science and Technology

Associate Editors (in alphabetical order of surnames):

AU Yuen Yee Mandy, The Education University of Hong Kong
CAO Jiaxin Galaxy, The Education University of Hong Kong
CHAN Ka Shing Kevin, The Education University of Hong Kong
CHEUNG Ho Yin Haoran, The Education University of Hong Kong
CHU Hui-Chun, Soochow University
FU Yujun, The Hong Kong Polytechnic University
GUMUS Sedat, The Education University of Hong Kong
HEUNG Yuk Mui Elly, The Education University of Hong Kong
HSU Ting-Chia, National Taiwan Normal University
HUANG Lingyun, The Education University of Hong Kong
KONG Suet Yee Suria, The Education University of Hong Kong
KORTE Satu-Maarit, University of Lapland
LAM Chi Ming, The Education University of Hong Kong
LI Man Wai Liman, The Education University of Hong Kong
LI Xin Stephen, The Education University of Hong Kong
LIN Chiupin, National Tsing Hua University
LING Man Ho Alpha, The Education University of Hong Kong
LOUIE Hung Tak Lobo, The Education University of Hong Kong
MA Yunsi Tina, The Education University of Hong Kong
NG Tsz Kit Davy, The Education University of Hong Kong
SINGH Manpreet, The Education University of Hong Kong
WEN Yun, Nanyang Technological University



WONG Lung-Hsiang, Nanyang Technological University
XIE Yishan Ellen, The Education University of Hong Kong
YANG Lan, The Education University of Hong Kong
YANG Yin Nicole, The Education University of Hong Kong
YANG Yu, The Education University of Hong Kong
YEUNG Chi Ho Bill, The Education University of Hong Kong
YEUNG Siu Sze Susanna, The Education University of Hong Kong
YU Leung Ho Philip, The Education University of Hong Kong

Executive Editors:

MA Yunsi Tina, The Education University of Hong Kong
LUO Jing Crystal, The Education University of Hong Kong
CHAN Cheuk Yan Stephanie, The Education University of Hong Kong
HO Po Yi Raina, The Education University of Hong Kong
LIAO Lishan Alice, The Education University of Hong Kong

TABLE OF CONTENTS

Messages from the Organisers.....	vi
Message from the Conference Organising Committee Chairs	vi
Message from the International Programme Committee Chairs.....	viii
Conference Organisation.....	xi
Conference Advisory Committee	xi
Conference Organising Committee.....	xiv
International Programme Committee	xvi
Local Organising Committee.....	xx
Keynotes.....	xxi
Keynote 1 - Navigating AI and Human Behaviour: Implications for Well-being and Learning .	xxi
Keynote 2 - Leveraging Artificial Intelligence to Unleash Students' Potential: Insights from a 5- Year Hong Kong Study	xxii
Keynote 3 - AI for Education: Exploration and Prospects	xxiii
Panel Discussion	xxiv
Panel Discussion 1 - AI, Attention, and Learning: Preparing Students for a Fast-Changing World.....	xxiv
Panel Discussion 2 - The Future of Education Technology: Balancing Innovation and Learner Diversity	xxv
Conference Proceedings	xxvi

Messages from the Organisers

Message from the Conference Organising Committee Chairs



Professor XU Guandong

Director of Centre for Learning, Teaching and Technology
Director of University Research Facility of Data Science
and Artificial Intelligence
Chair Professor of Artificial Intelligence
The Education University of Hong Kong



Dr MCMINN Sean

Director of Centre for Education Innovation
The Hong Kong University of Science and Technology

Welcome to the 2nd EdUHK x HKUST Joint International Conference on AI and Education (AIEdu2025)! The launch of this conference marks an important milestone in our collective effort to harness AI as a transformative force in education. With the strong support of the Senior Management, Faculties, Departments, Graduate School, Academic Support Units, and all participants, the Conference has attracted presenters and attendees from various regions. We would like to extend our sincere gratitude to everyone for their valuable contributions and unwavering support.

As AI's role in society continues to grow, its importance in education becomes increasingly evident. While there are many opportunities to enhance personalised learning, assessment, and curriculum design, we also recognise the challenges and considerations involved. We must continue to explore how AI can be integrated thoughtfully, aligning technological advancements with learners' needs and societal values.

At EdUHK, we have introduced a new curriculum emphasising "Digital Competency," led by the Centre for Learning, Teaching, and Technology (LTTC). This initiative aims to incorporate AI-related courses into our curriculum, allowing students to build a solid foundation, develop practical skills, and pursue more advanced studies in this field. The pilot has been successfully completed in the last academic year, and the majority of our programmes are scheduled to implement the curriculum commencing in the upcoming academic year. Through this approach, we hope to prepare our students

not just to understand emerging technologies but to actively contribute to an innovative, digital-driven future.

We would like to extend our heartfelt appreciation to the unfailing support from the Conference Advisory Committee Chair and Co-Chairs: Professor LEE Chi Kin John, President, Chair Professor of Curriculum and Instruction, EdUHK; Professor CHAN Che Hin Chetwyn, Vice President (Research and Development), Peter T. C. Lee Chair Professor of Psychology, EdUHK; Professor CHENG May Hung May, Vice President (Academic), Chair Professor of Teacher Education, EdUHK; Professor GUO Yike, Provost, Chair Professor, Department of Computer Science and Engineering, HKUST; and Professor FUNG Chi Hung Jimmy, Associate Provost (Teaching & Learning), Chair Professor, Department of Mathematics and Division of Environment & Sustainability, HKUST.

We would also like to express our deepest gratitude to the Conference Organising Committee of AIEdu 2025: Professor CHIU Ming Ming, Chair Professor of Analytics and Diversity, Director of Analytics/Assessment Research Centre, EdUHK; Professor HUI Pan, Chair Professor of Computational Media and Arts (CMA), Chair Professor of Emerging Interdisciplinary Areas, HKUST; and Professor SONG Yanjie, Associate Director of the Centre for Excellence in Learning and Teaching, Professor of Department of Mathematics and Information Technology, EdUHK. Furthermore, a round of applause is deservedly owed to the International Programme Committee Members, who have diligently reviewed over 50 abstracts and full papers submitted for AIEdu 2025. Special thanks also go to our colleagues from the LTTC, EdUHK and members from HKUST, for their tremendous efforts in organising such a smooth and informative conference.

Thank you for your participation in this important academic and professional journey. We wish you all a fruitful experience at AIEdu 2025.

Professor XU Guandong
Conference Organising Committee Chair

Dr MCMINN Sean
Conference Organising Committee Co-Chair

The International Conference on AI and Education 2025 (AIEdu2025)

Message from the International Programme Committee Chairs



Professor LOOI Chee Kit

Associate Dean (Quality Assurance and Enhancement) of Graduate School

Research Chair Professor of Learning Sciences

The Education University of Hong Kong

Dr CHENG Kwok Shing Gary

Executive Co-Director of Centre for Learning, Teaching and Technology

Associate Professor of Department of Mathematics and Information Technology

The Education University of Hong Kong



The International Conference on AI and Education 2025 (AIEdu2025), co-organised by The Education University of Hong Kong (EdUHK) and The Hong Kong University of Science and Technology (HKUST), was held physically from 20 to 21 August 2025. With the theme “Leveraging AI for Change in Education”, the conference is an annual joint-university initiative between EdUHK and HKUST.

The conference aimed at exploring the profound impacts and possibilities of AI in educational settings, ranging from K-12 and higher education to vocational training and adult learning. As AI technologies continue to advance at an unprecedented pace, it is crucial for our educational approaches and research to adapt accordingly. This involves not only integrating AI into current pedagogical frameworks but also rethinking our research methodologies to better understand and leverage AI’s impact on learning outcomes, assessment tools, curriculum development, and students’ well-being.

This conference served as a platform for interdisciplinary dialogue, bringing together educators, researchers, policymakers, and technologists to explore key strands as follows:

- Challenges and Opportunities of AI in Higher Education and Vocational Training
- AI-Enhanced K-12 Education: Challenges and Potential
- Learning Outcomes and AI-Driven Assessments
- Innovating Curriculum and Pedagogy through AI
- Societal and Ethical Issues in AI-empowered Education

AIEdU2025 featured three keynote speeches, one forum, two panel discussions, one poster presentation, and 8 paper presentation sessions. This year, we had the privilege of welcoming three renowned experts as keynote speakers. They are:

- Professor ALI Raian, Professor, College of Science and Engineering, Hamad Bin Khalifa University, Qatar
- Professor ZHANG Min, Professor, Department of Computer Science and Technology, Tsinghua University, Beijing, China
- Professor KONG Siu-Cheung, Research Chair Professor of E-Learning and Digital Competency, Director of Artificial Intelligence and Digital Competency Education Centre, The Education University of Hong Kong, Hong Kong SAR, China

AIEdU2025 issued a call for papers to scholars around the world. This year, the conference received more than 50 submissions by 108 authors from mainland China, Hong Kong, Indonesia, Kuwait, Malaysia, Russia, Taiwan, United Arab Emirates, United States, Vietnam and some other regions. Table 1 shows the distribution of authors by region/country.

Table 1. Distribution of AIEdU2025 authors by region/country

Regions/ Countries	No. of Authors	Regions/ Countries	No. of Authors
Hong Kong	67	Taiwan	3
Mainland China	20	United States	3
Malaysia	5	Kuwait	1
Indonesia	4	United Arab Emirates	1
Russia	3	Vietnam	1
		Total	108

Each submission underwent a thorough review process, with at least three programme committee members assigned to review each paper during the first round. The results of these reviews were subsequently subject to meta-review by the international programme committee chair and co-chair. Following this rigorous evaluation, a total of 44 papers were accepted for presentation at AIEdU2025 (see Table 2).

Table 2. Paper acceptance statistics by strand at AIEdU2025

Strand	Submissions	Accepted	Acceptance Rate
Challenges and Opportunities of AI in Higher Education and Vocational Training	12	10	83%
AI-Enhanced K-12 Education: Challenges and Potential	14	12	86%
Learning Outcomes and AI-Driven Assessments	5	5	100%
Innovating Curriculum and Pedagogy through AI	13	12	92%
Societal and Ethical Issues in AI-empowered Education	6	5	83%
Total	50	44	88%

We would like to extend our heartfelt appreciation to the dedicated and hardworking International Programme Committee Members, namely Dr BA Shen, Assistant Professor of Department of Curriculum and Instruction, EdUHK; Dr FU Hong, Associate Professor, Department of Mathematics and Information Technology, Associate Director of Centre for Learning, Teaching and Technology (LTTC), EdUHK; Professor KWOK Tin-Yau James, Professor, Department of Computer Science and Engineering, HKUST; Dr LAU Chaak Ming, Assistant Professor, Department of Linguistics and Modern Language Studies, EdUHK; Dr LEUNG Ka Lok Jac, Lecturer, Division of Integrative Systems and Design, HKUST; Dr SO Chi Fuk Henry, Associate Director, LTTC, EdUHK; and Dr WONG Ka Lei James, Associate Dean of Humanities and Social Science, Assistant Professor of Social Science Education, HKUST. They played a vital role in coordinating the review process and ensuring a fair and comprehensive assessment of the submissions. Our gratitude also extends to the programme committee members who diligently reviewed and provided valuable comments on the submitted papers. Their expertise and insights significantly contributed to the selection of high-quality papers for presentation at the Conference.

The success of AIEdu2025 would not have been possible without the collective contributions of the members of Conference Organising Committee and Programme Committee. We appreciate their commitment to maintaining the highest standards of academic excellence in the review process. Our heartfelt gratitude goes to the Local Organising Committee, composed by colleagues from Centre for Learning, Teaching and Technology, and led by Dr Tina Ma and Ms Crystal Luo from LTTC, which makes AIEdu2025 a great success.

In summary, AIEdu2025 successfully brought together a global community of scholars, educators, and practitioners to critically examine and advance the role of AI in transforming education. The wide-ranging participation and high quality of accepted submissions reflect the growing international commitment to reimagining education in the AI era. As we conclude this year's conference, we look forward to continuing these important conversations and fostering new collaborations that will shape the future of AI-enhanced education. We will continue to organise the International Conference on AI and Education 2026 (AIEdu2026) in 2026. We look forward to meet you again next year.

Thank you very much.

Professor LOOI Chee Kit
International Programme Committee Chair

Dr CHENG Kwok Shing Gary
International Programme Committee Co-Chair

The International Conference on AI and Education 2025 (AIEdu2025)

Conference Organisation

Conference Advisory Committee

Advisory Committee Chair



Professor LEE Chi Kin John

President

Chair Professor of Curriculum and Instruction
The Education University of Hong Kong

Advisory Committee Co-Chair



Professor CHAN Che Hin Chetwyn

Vice President (Research and Development)
Peter T. C. Lee Chair Professor of Psychology
The Education University of Hong Kong

Advisory Committee Co-Chair



Professor CHENG May Hung May

Vice President (Academic)
Chair Professor of Teacher Education
The Education University of Hong Kong

Advisory Committee Co-Chair



Professor GUO Yike

Provost

Chair Professor of Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

Advisory Committee Co-Chair



Professor FUNG Chi Hung Jimmy

Associate Provost (Teaching and Learning)

Chair Professor of Department of Mathematics

Chair Professor of Division of Environment and Sustainability

The Hong Kong University of Science and Technology

Advisory Committee Members



Professor ERNI John Nguyet

Dean of Faculty of Humanities
Chair Professor of Cultural Studies
The Education University of Hong Kong



Professor LI Wai Keung

Dean of Faculty of Liberal Arts and Social Sciences
Research Chair Professor of Data Science
The Education University of Hong Kong



Professor MACFARLANE Bruce John

Dean of Faculty of Education and Human
Development
Co-Director of Centre for Higher Education
Leadership and Policy Studies
Chair Professor of Educational Leadership
The Education University of Hong Kong



Professor ZHOU Xiaofang

Head and Chair Professor of Department of
Computer Science and Engineering
Co-Director of HKUST Big Data Institute
Otto Poon Professor of Engineering
The Hong Kong University of Science and
Technology

Conference Organising Committee

Conference Organising Committee Chair



Professor XU Guandong

Director of Centre for Learning, Teaching and Technology

Director of University Research Facility of Data Science and Artificial Intelligence

Chair Professor of Artificial Intelligence

The Education University of Hong Kong

Conference Organising Committee Co-Chair



Dr MCMINN Sean

Director of Centre for Education Innovation

The Hong Kong University of Science and Technology

Conference Organising Committee Members



Professor CHIU Ming Ming

Director of Analytics\Assessment Research Centre
Chair Professor of Analytics and Diversity
The Education University of Hong Kong



Professor HUI Pan

Chair Professor of Division of Emerging
Interdisciplinary Areas
The Hong Kong University of Science and
Technology



Professor SONG Yanjie

Associate Director of the Centre for Excellence in
Learning and Teaching
Associate Co-Director of the Academy for Applied
Policy Studies and Education Futures
Professor of Department of Mathematics and
Information Technology
The Education University of Hong Kong

International Programme Committee

International Programme Committee Chair



Professor LOOI Chee Kit

Associate Dean (Quality Assurance and Enhancement) of Graduate School
Research Chair Professor of Learning Sciences
The Education University of Hong Kong

International Programme Committee Co-Chair



Dr CHENG Kwok Shing Gary

Executive Co-Director of Centre for Learning, Teaching and Technology
Associate Professor of Department of Mathematics and Information Technology
The Education University of Hong Kong

International Programme Committee Associate Chairs



Dr BA Shen

Assistant Professor of Department of Curriculum and Instruction

The Education University of Hong Kong



Dr FU Hong

Associate Director of Centre for Learning, Teaching and Technology

Associate Professor of Department of Mathematics and Information Technology

The Education University of Hong Kong



Professor KWOK Tin Yau James

Professor of Department of Computer Science and Engineering

The Hong Kong University of Science and Technology



Dr LAU Chaak Ming

Director of Centre for Research on Linguistics and Language Studies

Assistant Professor of Department of Linguistics and Modern Language Studies

The Education University of Hong Kong



Dr LEUNG Ka Lok Jac

Lecturer of the Division of Integrative Systems and Design

The Hong Kong University of Science and Technology



Dr SO Chi Fuk Henry

Associate Director of Centre for Learning, Teaching and Technology

Senior Lecturer of Department of Mathematics and Information Technology

The Education University of Hong Kong



Dr WONG Ka Lei James

Associate Dean of Humanities and Social Science

Assistant Professor of Social Science Education

The Hong Kong University of Science and Technology

International Programme Committee Members

(In alphabetical order of surname)

AU Yuen Yee Mandy, The Education University of Hong Kong
CAO Jiaxin Galaxy, The Education University of Hong Kong
CHAN Ka Shing Kevin, The Education University of Hong Kong
CHEUNG Ho Yin Haoran, The Education University of Hong Kong
CHU Hui-Chun, Soochow University
FU Yujun, The Hong Kong Polytechnic University
GUMUS Sedat, The Education University of Hong Kong
HEUNG Yuk Mui Elly, The Education University of Hong Kong
HSU Ting-Chia, National Taiwan Normal University
HUANG Lingyun, The Education University of Hong Kong
KONG Suet Yee Suria, The Education University of Hong Kong
KORTE Satu-Maarit, University of Lapland
LAM Chi Ming, The Education University of Hong Kong
LI Man Wai Liman, The Education University of Hong Kong
LI Xin Stephen, The Education University of Hong Kong
LIN Chiupin, National Tsing Hua University
LING Man Ho Alpha, The Education University of Hong Kong
LOUIE Hung Tak Lobo, The Education University of Hong Kong
MA Yunsi Tina, The Education University of Hong Kong
NG Tsz Kit Davy, The Education University of Hong Kong
SINGH Manpreet, The Education University of Hong Kong
WEN Yun, Nanyang Technological University
WONG Lung-Hsiang, Nanyang Technological University
XIE Yishan Ellen, The Education University of Hong Kong
YANG Lan, The Education University of Hong Kong
YANG Yin Nicole, The Education University of Hong Kong
YANG Yu, The Education University of Hong Kong
YEUNG Chi Ho Bill, The Education University of Hong Kong
YEUNG Siu Sze Susanna, The Education University of Hong Kong
YU Leung Ho Philip, The Education University of Hong Kong

Local Organising Committee

LUO Jing Crystal, The Education University of Hong Kong
MA Yungsi Tina, The Education University of Hong Kong
WONG Winnie, The Education University of Hong Kong
CHAN Cheuk Yan Stephanie, The Education University of Hong Kong
YAU Shuk Yi Shirley, The Education University of Hong Kong
WONG Sum Yee Maya, The Education University of Hong Kong
LIAO Lishan Alice, The Education University of Hong Kong
CAO Jiaxin Galaxy, The Education University of Hong Kong
DING Jiaoyang Will, The Education University of Hong Kong
LUO Zijian Lumen, The Education University of Hong Kong
LI Chun Wai Matt, The Education University of Hong Kong
HO Po Yi Raina, The Education University of Hong Kong
CHENG Yuen Kiu Jessica, The Education University of Hong Kong

Keynotes

Keynote 1 - Navigating AI and Human Behaviour: Implications for Well-being and Learning

Date: 20 August 2025 (Wednesday)

Time: 10:00-11:00

Venue: Room 1 (D2-LP-09)

Speaker



Professor ALI Raian

Professor of College of Science and Engineering
Hamad Bin Khalifa University

Chair

Professor XU Guandong

Chair Professor of Artificial Intelligence
Director of Centre for Learning, Teaching and Technology
Director of University Research Facility of Data Science
and Artificial Intelligence
The Education University of Hong Kong



Abstract

This talk reviews recent work in Technology and Human Behaviour, with a special focus on advances in AI. Case studies from social media addiction, misinformation, cybersecurity, and online gaming will be presented. The speaker will discuss their latest research on AI's impact on human attitudes and its potential addictive features. The talk integrates psychological theories and models that support the research and help interpret the findings. It also emphasises the importance of cultural sensitivities and cross-cultural comparisons. The session concludes with reflections and open questions on AI use in education and how to equip students with resilience to LLM dependency while maintaining critical thinking.

Biography

Dr. Raian Ali is a Professor and UNESCO Chair in Digital Technologies and Human Behaviour, and a Visiting Professor at Bournemouth University, UK, and the Education University of Hong Kong. He earned his PhD from the University of Trento, Italy, in 2010, and has held academic positions at the University of Limerick, Ireland, and Bournemouth University. His first degree is from Latakia University (formerly Tishreen University), Syria. Raian's research focuses on the intersection of Technology and Human Behaviour, addressing topics such as digital addiction, responsible technology use, and the impact of technology design on human well-being. He serves on the Steering Committee of the Persuasive Technology Society and the Behavioural and Social Computing Conference, and sits on the editorial boards of leading journals including Behaviour and Information Technology, Journal of Responsible Technology, and Human-Centric Artificial Intelligence. He currently leads the \$3.8 million Cluster on Digital Citizenship in Qatar, overseeing six projects, more than 15 national and international partners, and a team of over 50 personnel. Raian also provides consultancy at both national and international levels.

Keynote 2 - Leveraging Artificial Intelligence to Unleash Students' Potential: Insights from a 5-Year Hong Kong Study

Date: 21 August 2025 (Thursday)

Time: 09:30-10:30

Venue: Room 1 (D2-LP-09)

Speaker



Professor KONG Siu Cheung

Research Chair Professor of E-learning and Digital Competency
Director of Artificial Intelligence and Digital Competency Education Centre
The Education University of Hong Kong

Chair

Professor YEUNG Chi Ho Bill

Professor of Department of Science and Environmental Studies
Associate Dean (Quality Assurance & Enhancement) of
Faculty of Liberal Arts and Social Sciences
Dean of Students
The Education University of Hong Kong



Abstract

Over the past five years, we have pioneered the integration of Artificial Intelligence (AI) literacy into K-12 and university education, fostering a future-ready generation. This initiative, supported by the Education University of Hong Kong, donors, and industrial partners, aims to equip students with foundational AI knowledge, skills, and ethical awareness. Through tailored curricula and hands-on projects, students explore AI concepts such as machine learning and understand that AI is a kind of algorithmic intelligence. Our study has demonstrated success in nurturing AI competence and problem-solving skills, positioning Hong Kong as a leader in AI literacy education. In this presentation, we highlight the importance of using AI in education to unleash the potential of students. We address the transformative opportunities AI provides for equipping future generations to create meaningful value by working alongside it. We are developing pedagogies to encourage the classroom use of AI for enhancing independent thinking and metacognitive development.

Biography

Professor KONG Siu Cheung is currently a Research Chair Professor at the Department of Mathematics and Information Technology and Director of the Artificial Intelligence and Digital Competency Education Centre (AIDCEC) at The Education University of Hong Kong. He serves as the Editor-in-Chief of the international journal *Research and Practice in Technology Enhanced Learning* (RPTEL) and *Journal of Computers in Education* (JCE). He is on the Stanford Top 2% Scientist in Education list from 2019 to 2024. Professor Kong is leading projects on Artificial Intelligence Literacy for primary, secondary and university students, teachers, parents, and administrative staff in Hong Kong from 2020 to 2027. Currently, he is also leading projects on Artificial Intelligence in K-12 Education. He is the programme leader of the Master of Science in Artificial Intelligence for Executive Professionals [MSc(AIEP)] programme at the Education University of Hong Kong.

Keynote 3 - AI for Education: Exploration and Prospects

Date: 21 August 2025 (Thursday)

Time: 14:00-15:00

Venue: Room 1 (D2-LP-09)

Speaker



Professor ZHANG Min

Professor of Department of Computer Science
and Technology

Tsinghua University

Chair

Professor SONG Yanjie

Professor of Department of Mathematics and Information Technology
Associate Director of the Centre for Excellence in Learning and Teaching
Associate Co-Director of the Academy for Applied Policy Studies and
Education Futures

The Education University of Hong Kong



Abstract

AI has empowered education with LLM technologies. In this talk, I will first talk about the AI-Empowered Learning Assistant and how to move on from MOOC to MAIC (Massive AI-Empowered Course) in Tsinghua University. Practices and studies of designing AI teachers, AI assistants, and AI learning-mates will be shared. I will also show the evaluation with cognition signals on the effectiveness of MAIC. Finally, challenges and Opportunities for AI-Empowered Education will be discussed.

Biography

Dr. Min Zhang is a full professor in the Dept. of Computer Sci. & Tech., Tsinghua University, and the chief director of the AI Lab. She specializes in Web search, recommendation, and user modeling. Prof. Zhang is an ACM distinguished member, a SIGIR Academy fellow. She has been the Editor-in-Chief of ACM Transactions on Information Systems (TOIS) since 2020, and also serves as the General Chair of ACM MM'25, and PC Chairs of SIGIR'26, RecSys'23, CIKM'23, WSDM'17, etc. She won the "Test-of-Time" award at SIGIR'24, WSDM'22 Best Paper award, IBM Global Faculty Award, etc., and has a lot of collaborations with international and domestic industries.

Panel Discussion

Panel Discussion 1 - AI, Attention, and Learning: Preparing Students for a Fast-Changing World

Date: 20 August 2025 (Wednesday)

Time: 11:15-12:30

Venue: Room 1 (D2-LP-09)

Abstract

In an age inundated with information, it is increasingly challenging for individuals—especially students—to focus, discern significance, and understand what truly matters. With AI amplifying this information overload, how can education empower learners to remain grounded, think critically, and make informed decisions? This panel will delve into how AI is transforming our approaches to learning, identifying essential skills for today's students, and examining how educational institutions must evolve. We will consider AI not merely as a tool, but as a transformative force reshaping society, learning paradigms, and perceptions of reality. Our panellists will offer diverse perspectives—from classroom practices to policy implications, ethical considerations, and technology design—aiming to foster discussions on building educational systems that promote deeper thinking, stronger communities, and more confident learners.

Chair



Professor LOOI Chee Kit

Research Chair Professor of Learning Sciences
Associate Dean (Quality Assurance and Enhancement) of Graduate School
The Education University of Hong Kong

Panellists (in the alphabetical order of surname)



Professor JIA Jiyou

Head and Professor of Department of
Education Technology
Director of Digitalized Education Research
Centre
Peking University



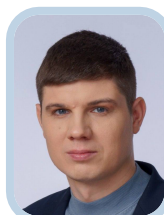
Dr MCMINN Sean

Director of Centre for Education Innovation
The Hong Kong University of Science and
Technology



Professor SHEN Jun

Professor of School of Computing and
Information Technology
University of Wollongong



Professor TERENCEV Evgenii

Dean of the Institute of Education
National Research University Higher
School of Economics

Panel Discussion 2 - The Future of Education Technology: Balancing Innovation and Learner Diversity

Date: 21 August 2025 (Thursday)

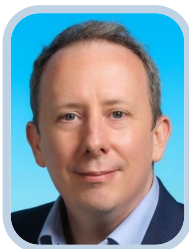
Time: 16:30-17:45

Venue: Room 1 (D2-LP-09)

Abstract

This panel discussion will explore the evolving future of education technology. New technologies are quickly being integrated into various educational settings. The conversation will focus on innovative advancements and solutions in educational AI and data science, which may offer greater potential to personalise learning experiences and promote pedagogical innovations. Panellists will also share their views on how to balance fostering innovation with effectively addressing diverse learning needs, aiming to create an inclusive and equitable future education for global citizens.

Chair



Dr MCMINN Sean

Director of Centre for Education Innovation
The Hong Kong University of Science and Technology

Panellists



Professor XU Guandong

Chair Professor of Artificial Intelligence
Director of Centre for Learning, Teaching
and Technology
Director of University Research Facility of
Data Science & Artificial Intelligence
The Education University of Hong Kong



Professor JIANG Bo

Professor of Shanghai Institute of Artificial
Intelligence for Education
East China Normal University



Dr ZHENG Chanjin

Associate Professor of Shanghai Institute
of Artificial Intelligence for Education
East China Normal University



Dr LEUNG Ka Lok Jac

Lecturer of Division of Integrative Systems
and Design
The Hong Kong University of Science and
Technology



Mr ZHANG Kai

Shanghai Institute of Artificial Intelligence for
Education
East China Normal University



**The EdUHK x HKUST Joint
International Conference on AI and Education 2025
(AIEdu2025)**

20 – 21 August 2025

Conference Proceedings

Conference Proceedings

Table of Contents

Innovating Curriculum and Pedagogy through AI

Learning with Generative AI Avatars in Virtual Reality vs. Desktop

Paolo MENGONI, Daniel Jiandong SHEN, Paulina Pui Yun WONG, Sunny Hee Sun CHOI 6

AI for Science and Science for AI: Lessons learned from curriculum modifications for the undergraduate program of biostatistics in a Sino-British University

Tenglong LI 11

Innovating Music Education through Generative AI: A Case Study on the Music and Science of Sound Course

Enrico BERTELLI 14

Integrating AI and SCAMPER in Poetry Writing: Gender Differences in Creative Performance

Jon-chong HONG, Huan YEH, Sin-ping WANG 17

Enhancing Clinical Skills Training in Optometry through AI-Assisted Learning

Kin WAN 21

OpenImpact Lab: Leveraging GenAI for Project-Based Learning and NGO Empowerment

Max Jwo Lem LEE, Alwin LIN, Ruoyu CHEN, Li-Ta HSU 23

Integrating Generative AI into Inquiry-Based Learning for Personalized Mathematics Education

Fridolin Sze Thou TING, Daisy Dichen WANG, Lawrence Chi Lok CHAN 27

Enhancing Students' Prompt Engineering Skills through Scaffolding Strategies

Shurui BAI, Chen YANG 32

A Case Study of Identifying AI's Epistemological Roles in ICAP-STEAM Education: Evidence from an Arduino Environmental Monitoring System Project

Wai Kit MAK 35

Digitalizing Music Education: Creating an Inclusive Learning Hub for Young Children

Nikolay Tomov DEMERDZHIEV, Oi Tik Aidi Lydia DUE 41

Enhancing speaking skills in access-limited communities: AI-mediated Informal Digital Learning of English (AI-IDLE)

Ju Seong LEE, Roland Tai Yin LEUNG, Nur Arifah DRAJATI, Maria HIDAYATI, Didik SUHARDI, Ananto Kusuma SETA 43

Societal and Ethical Issues in AI-empowered Education

Construction and Validation of Personalized Emotion Support Agent for Adolescent Mental Health Counseling Conversation	
<i>Shixin PENG, Kun JIANG, Yu YANG, Guandong XU, Jingying CHEN</i>	49
Illuminating Societal and Ethical Dimensions of AI in Course Evaluation: NLP-Enabled Bibliometric Analysis	
<i>Da GONG, Yunqi WANG</i>	54
Deep Self, Digital Wisdom and Pervasive Learning in AI-Centric Futures	
<i>Marcus T ANTHONY</i>	59

Challenges and Opportunities of AI in Higher Education and Vocational Training

A Detailed Analysis of the Impact of Large Language Models on Students' Cognitive and Metacognitive Ability	
<i>Haoli WANG, Sixu AN, Yu YANG, Guandong XU</i>	67
Bridging Job Market Predictions and Personalized Recommendations for Career Guidance: AI-Driven Career Pathway Approach	
<i>Yunqi WANG, Haoli WANG, Yu YANG, Guandong XU</i>	72
Examining Students' Behavioral Patterns of Using GPT to Support Lesson Planning: A Pilot Study with Trace Data	
<i>Chunlin ZHOU, Yanxiong XIANG, Lingyun HUANG</i>	77
Transforming Education Through AI Literacy: Insights from Teacher Trainers	
<i>Asmah BOHARI, Anusuya KALIAPPAN, Su Ling LOH</i>	80
Demystifying LLM Implementation in Higher Education: A Three-Step Actionable Roadmap	
<i>Chenxi DONG, Chujie WEN, Tianyu ZHANG</i>	87
Motivation and Engagement in Job Interview Training: A Study of a ChatGPT-facilitated Robot for Youths with Special Education Needs	
<i>Kuen-Fung SIN, Ka-Yan FUNG, Kwong-Chiu FUNG, Wai-Ho MOW, Hon-Shan Celia TAM, Tze-Leung Rick LUI, Lan YANG, Yu-Xing TAO</i>	91
Are They Ready? Analyzing Chinese Preservice Teachers' Acceptance to Use Gen AI in Education	
<i>Yujie KUANG, Jiateng ZHOU</i>	96
In search of a "free rider" in the age of Gen-AI: intentions and outcomes of using Gen-AI in group writing	
<i>Anna KORCHAK, Mik FANGUY, Anastasia KAPUZA, Jamie COSTLEY</i>	101
Enhancing EFL Speaking Skills Through Artificial Intelligence Technology: Opportunities and Challenges	
<i>Junyi SHEN, Bojie QIAN</i>	104

AI-Enhanced K-12 Education: Challenges and Potential

Inspiring Ideas for English Creative Writing: Insights from Designers on A Prototype of Artificial Intelligence-Assisted Platform MuseAIWrite

Yin YANG..... 110

Understanding Teacher Adoption of Integrated GenAI Platforms in Secondary Education: A Pilot Case Study

Saanvi Ravi KATYAYAN, Jac Ka Lok LEUNG, Alison CHAN 114

Cultural Adaptability of Global AI in Education Policies: A Chinese Perspective

Tianhang GAO, Simin CAO, Hui LI..... 119

Designing a Contextualized Training for Learning Engagement of Hong Kong Students with Autism Spectrum Disorder (ASD)

Ying-lan CHEN, Yu-xing TAO, Tze-leung Rick LUI, Ka-yan FUNG, Kuen-fung SIN..... 124

ProTeach: LLM Agent framework for Intelligent Tutoring System

Ivan LO, Victor CHENG, Chun Hung LI..... 127

The Opportunities and Challenges of Implementing Generative AI in Collaborative Learning: A Scoping Review

Xuechun SHI, Shen BA..... 131

The Role of Artificial Intelligence in Supporting Inclusive Education: Challenges and Opportunities for Teachers

Xiaoqing LIN, Yi XIN 134

Translating Research on “The Animated Chinese History for the Curious Minds” for Primary School Children in Hong Kong

Wai-chun Cherry AU, Ho-yin James PONG, Jong Hyuk David KANG..... 138

Learning Outcomes and AI-Driven Assessments

Fine-Grained Student Engagement Evaluation with Large Language Model: A case in Collaborative learning

Sixu AN, Haoli WANG, Qihao YANG, Yu YANG, Guandong XU 144

Developing a VR-based Immersive Learning Platform with a Multimodal AI Agents System for Improving University Students’ Academic Presentation and Self-Regulated Learning

Yanjie SONG, Jiachen FU, Kaiyi WU, Chiu Tsun TSANG..... 149

A meta-review on the role of AI in English writing: A recent trend

Hetian YU, Lan YANG 154

Strands

- **Innovating Curriculum and Pedagogy through AI**
- **Societal and Ethical Issues in AI-empowered Education**
- **Challenges and Opportunities of AI in Higher Education and Vocational Education**
- **AI-Enhanced K-12 Education: Challenges and Potential**
- **Learning Outcomes and AI-Driven Assessments**

Strand

Innovating Curriculum and Pedagogy through AI

Learning with Generative AI Avatars in Virtual Reality vs. Desktop

Paolo MENGONI^{1*}, Daniel Jiandong SHEN², Paulina Pui Yun WONG³, Sunny Hee Sun CHOI⁴

¹Department of Interactive Media, Hong Kong Baptist University, Hong Kong SAR, China

²Soqgle Hong Kong Limited, Hong Kong SAR, China

³Science Unit, Lingnan University, Hong Kong SAR, China

⁴School of Design, The Hong Kong Polytechnic University, Hong Kong SAR, China

pmengoni@hkbu.edu.hk, dan@soqgle.com, paulinawong@ln.edu.hk, hee-sun.choi@polyu.edu.hk

Abstract: *This study examines student interactions with Generative-AI (GenAI) avatars in Virtual Reality (VR) and Desktops, focusing on engagement, message quality, and intent to use AI tutors. Grounded in Cognitive Load Theory and Self-Determination Theory, it explores Ease of Use, Performance, and Usefulness as predictors of AI adoption. While VR fosters immersion, usability challenges, including cognitive load, motion sickness, and navigation difficulties—may impact engagement. Using the GPT-integrated Classlet platform, this study compares Desktop and VR interactions across search engines, sustainability and waste management topics. Findings show that Desktop users engaged more, but interaction quality depended on topic, motivation, and guidance rather than delivery mode. Regression analysis confirmed that Usefulness, Performance, and Ease of Use significantly predicted intent to use the learning method, explaining 72.9% of the variance ($R^2 = 0.729$). VR issues like motion sickness and interaction complexity impacted engagement, emphasizing the need for more understanding on AI adaptability in Gen-AI immersive learning environments.*

Keywords: Virtual Reality (VR), Immersive Learning, Generative-AI Avatars

1. Introduction

The integration of Virtual Reality (VR) and GenAI avatars in education is rapidly expanding, offering interactive, adaptive, and immersive learning experiences. VR is widely recognized for its ability to enhance engagement, spatial understanding, and experiential learning, making it a promising tool for STEM education, language learning, and vocational training. AI-based tutoring, particularly with Generative AI avatars, provides real-time feedback, personalized learning pathways, and conversational engagement, enabling more student-centered and dynamic interactions (Sun & Botev, 2021). Combining VR and AI tutors has the potential to transform self-directed learning, offering students autonomous, context-driven experiences that traditional e-learning platforms may not fully support.

Despite these advantages, VR introduces usability challenges, including cognitive overload, motion sickness, and interaction complexity, which may negatively impact learning efficiency and engagement with AI tutors (Dubovi, 2022). Unlike Desktop-based learning, where interactions are structured and text-based, VR requires spatial navigation, voice recognition, and real-time multitasking, increasing cognitive demands (Sweller, 1988). AI-based learning assistants, such as Generative AI (GenAI) avatars, introduce additional complexity, as students must manage speech recognition, AI response timing, and interaction consistency, which can either enhance or hinder engagement (Wong et al., 2024). Moreover, prior research has primarily focused on VR's engagement potential but less is known about how GenAI avatars function within immersive environments or how students adapt their interaction behaviors in VR vs. Desktop settings.

Compounding these concerns, engagement with AI-based tutors in VR may be influenced by prior knowledge and topic complexity. Research suggests that students engage more effectively when content is familiar and relatable, while specialized or abstract topics may lead to increased exploratory behavior and reduced interaction quality (Lin et al., 2024). In this study, Desktop participants (undergraduate studying a specialized topic: AI and search engines) are expected to be the comparison group for usability. Moreover, VR-WM participants (master's students engaging with a general topic: Waste Management) are expected to demonstrate more structured engagement, whereas VR-SDG participants (undergraduates working with a specialized topic: aquaponics and SDGs) may show higher exploratory behavior and interaction variability. These differences highlight the need to understand how topic familiarity and cognitive load affect AI engagement in VR-based learning environments.

2. Literature Review

VR is widely recognized for its ability to enhance engagement, interactivity, and spatial understanding, yet studies highlight cognitive overload and usability barriers as critical challenges (Lin et al., 2024). Generative AI can be used to facilitate adaptive learning and personalized tutoring (Wong et al., 2024). However, engagement with AI in VR environments requires effective interaction design, as complex controls, voice recognition, and system responsiveness may introduce new cognitive and technical burdens (Wong et al.,

2024). Generative AI (GenAI) avatars offer a promising solution to enhance personalized learning in VR, enabling adaptive tutoring, real-time responses, and contextual learning support (Wong et al., 2024). Unlike traditional AI-based tutoring, GenAI can dynamically adjust explanations, scaffold learning based on student progress, and provide interactive dialogue-driven learning experiences.

However, engagement with GenAI in VR requires effective interaction design, as complex controls, voice recognition may introduce new cognitive and technical burdens, potentially impacting usability (Wong et al., 2024). While prior research (Wong et al., 2024) has applied The Technology Acceptance Model (TAM) to VR learning environments, its extension to GenAI avatars in VR remains limited. Furthermore, Self-Determination Theory (SDT) suggests that autonomy, competence, and relatedness drive motivation and engagement. While VR fosters autonomy through immersive learning, it may also introduce competence-related barriers, such as motion sickness, navigation issues, and difficulty in AI interactions (Maddox & Fitzpatrick, 2019).

3. Research Questions

Understanding how students interact with GenAI avatars in immersive vs. non-immersive environments is essential for improving AI-based learning system design. GPT avatars introduce an additional layer of complexity, requiring voice recognition, response timing, and adaptive interaction design to function effectively in VR settings. Furthermore, cognitive load and topic familiarity may play a role in determining engagement quality, influencing students' ability to generate meaningful interactions with AI. Thus, we ask:

RQ1: How do interaction patterns differ while learning with GenAI avatars in VR versus Desktop environments?

RQ2: What factors influence students' intent to use GenAI avatars in VR and Desktop environments?

4. Method

This study employed a mixed methods approach to evaluate how students interact with GenAI avatars in Virtual Reality (VR) and Desktop environments. The quantitative component focused on measuring interaction frequency, relevance, and engagement metrics. Data was collected from logged interactions with GenAI avatars and post-study survey responses. The qualitative component explored students' perceptions and experiences based on open-ended survey responses.

4.1. Participants

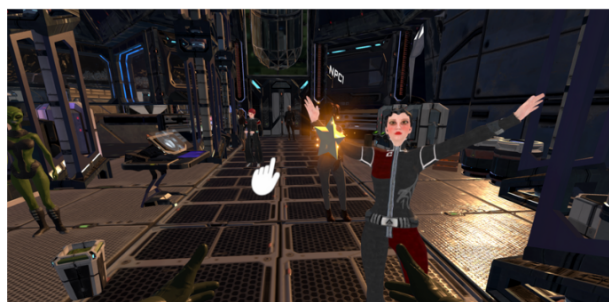
A total of 55 students participated in this study. 25 undergraduate students took part in a search engines module on Desktop (Laptop), while 13 master's students in the Waste Management (VR-WM) group focused on Waste Management, and 18 undergraduate students in the Sustainable Development Goals (VR-SDG) group explored topics related to Sustainable Development Goals.

4.2. Procedure

Participants engaged in an interaction phase lasting 30 minutes, performing learning tasks for the topics including dialogues with the GPT avatar. Desktop participants interacted with the GPT avatar through text-based inputs on Chrome, while VR participants engaged in an immersive 3D environment, using spatial interactions and voice commands. The study utilized Classlet, a VR platform integrated with Generative AI, designed for immersive learning. Classlet features customizable 3D environments for experiential learning and avatars that provide real-time feedback as shown on Figure 1.



Desktop interaction with Avatar in a library scene



VR interactions with Avatar in a space scene

Figure 1. Screenshot of scenes of Classlet used for Desktop and VR interactions.

After completing the interaction tasks, participants completed a survey based on Fussell & Truong (2022), assessing key variables related to VR learning. These included AEST (Aesthetics) for immersion, TECH (Technology Interface) for usability, PERF (Performance) for functionality, ATT (Attitude) toward VR learning, EASE (Ease of Use) for simplicity, FUN (Enjoyment), USAB (Usability/Productivity) for learning efficiency, USEF (Usefulness) for real-world applicability, and INTENT (Intent to Use) for adoption willingness. Participants provided consent through the survey by keying their name to confirm their participation.

4.3. Data Analysis

Descriptive statistics were used to summarize responses, Spearman's correlation to assess relationships, Mann-Whitney U tests to compare Desktop and VR interactions, and linear regression to predict intent to use the platform. GPT avatar interactions were logged to track total prompts, irrelevant checks (playful/exploratory interactions), quality prompts (substantive engagement), ratios of irrelevant vs. quality prompts, and the relevant percentage of quality prompts over total interactions. The qualitative analysis included open-ended survey questions on participant likes and dislikes of the method within the same self-assessed survey. To ensure a balanced comparison, VR groups (SDG and WM) were combined to match the Desktop sample size. Non-parametric tests were used due to the non-normal distribution of data.

For GPT system data scoring, irrelevant prompts were classified using a binary Yes/No rating by one rater, identifying playful, off-topic, or incomplete interactions. Quality prompts were rated on a 1 to 3 scale, where 1 = basic question, 2 = acceptable, and 3 = thought-provoking inquiry.

5. Results

Desktop users engaged more, while VR-SDG showed more exploratory behavior. Usefulness, Performance, and Ease influenced intent, while qualitative feedback highlighted immersion benefits but noted dizziness, control issues, and usability challenges in VR.

5.1. Quantitative

5.1.1. Self-assessed survey

The overall mean intent score was 3.70 (SD = 1.12), with higher Usefulness in Desktop (4.08) vs. VR (3.58), Performance higher in Desktop (4.04) vs. VR (3.80), and Ease of Use slightly higher in VR (3.90) than Desktop (3.72). Spearman's correlation showed relationships between Intent and Usefulness ($p = 0.805$, $p < .001$), Attitude ($p = 0.762$, $p < .001$), and Performance ($p = 0.736$, $p < .001$). Mann-Whitney U tests indicated lower immersion for Desktop users ($p = 0.054$) but greater intent than VR ($p = 0.036$).

A linear regression model predicting Intent based on EASE, PERF, and USEF was statistically significant ($F(3, 51) = 45.763$, $p < .001$, $R^2 = 0.729$, RMSE = 0.613). Coefficients: EASE ($\beta = 0.392$, $p < .001$), PERF ($\beta = 0.426$, $p < .001$), USEF ($\beta = 0.470$, $p = 0.001$). QQ plots confirmed normal residuals, and the model met homoscedasticity assumptions, ensuring validity.

5.1.2. GPT System Data

34 students (61% of total participants) contributed to the GPT task. Desktop users (Communication, $n = 25$) generated an average of 7.21 messages per student, with 54% irrelevant prompts (3.89 irrelevant, 3.37 quality per student). VR-SDG users ($n = 17$) averaged 5.88 messages, with the highest irrelevant prompt rate (60%) (3.50 irrelevant, 2.38 quality per student). VR-WM users ($n = 13$) had the lowest irrelevant prompt rate (34%), averaging 4.57 messages (1.57 irrelevant, 4.14 quality per student).

Desktop users engaged more with the highest message count (137) compared to VR-SDG (47) and VR-WM (32). Irrelevant prompts were frequent in Desktop and VR-SDG, while VR-WM had fewer irrelevant and more quality prompts. Desktop-Communication yielding higher engagement but moderate irrelevant prompts, while VR-SDG showed more exploratory prompts and VR-WM had greater relevance.

5.2. Qualitative Results

Many found the interface intuitive and easy to navigate, but others struggled with controls, dizziness, and perspective adjustments. Some described navigation as seamless, while others highlighted the need for clearer instructions. Positive: "Easy to navigate, more memorable, fun." / "It is easy to control." Challenges: "Feel dizzy, user instructions are not perfect." / "Difficult to move, difficult to control the camera, a bit lag."

Performance: VR enhanced engagement and interactive exploration but was seen as less efficient than traditional methods due to distractions, lack of human interaction, and occasional bugs. Some faced challenges accessing content, affecting learning effectiveness. Positive: "Learning in VR makes knowledge easier to retain." / "I can interact with AI and learn by doing." Challenges: "Not very efficient compared to traditional methods." / "Too many distractions make it harder to focus."

Usefulness: VR was valued for its immersive nature and ability to enhance knowledge retention. However, concerns about accessibility, cost, and content depth limit its perceived effectiveness. Some felt it needed further refinement to match traditional learning methods. Positive: "Learning using virtual reality is useful for gaining real-world knowledge." / "Helped me memorize knowledge more effectively." Challenges: "it makes you dizzy." / "Some content lacks depth compared to traditional methods."

6. Discussion

This study examined how students interact with a GPT-based avatar in VR and Desktop environments and the factors influencing their engagement. Desktop users engaged more, while VR-SDG showed higher exploratory behavior and cognitive load challenges.

6.1. Interaction Patterns and Engagement Differences (RQ1)

Desktop users generated the highest number of messages (137 messages, 7.21 per student) compared to VR-SDG (47 messages, 5.88 per student) and VR-WM (32 messages, 4.57 per student), suggesting that desktop interactions were easier and more efficient. Qualitative feedback indicated that Desktop interactions were structured, while VR participants described their experience as more "explorative" and "interactive." The lower message frequency in VR may be attributed to higher cognitive load, as students had to manage navigation, motion controls, and spatial awareness while interacting with the GPT avatar. In contrast, desktop users engaged more fluidly, likely due to familiarity with traditional interfaces. This aligns with Lin et al. (2024), who found that VR fosters higher cognitive engagement but also introduces usability and technological challenges that can limit learning effectiveness, particularly when students struggle with digital literacy or system complexity.

VR-WM had the highest quality prompt ratio (4.14 per student) and the lowest irrelevant prompt rate (34%), indicating more focused engagement, while both Desktop and VR-SDG had higher irrelevant prompt rates (54% & 60%). A key factor may be that VR-WM participants were master's students, likely contributing to greater task familiarity and cognitive maturity in framing relevant prompts. Additionally, Waste Management (WM) was a more general and relatable topic, whereas SDG focused on aquaponics, a more specialized subject, which may have posed greater conceptual challenges for undergraduate students. This suggests that prior knowledge and topic complexity influence engagement quality in GPT-based interactions.

Limited GPT Task Engagement: Only 61% of participants (34 out of 55) actively engaged in GPT interactions, suggesting a need for better onboarding, clearer task objectives, and enhanced AI interactivity. Wong et al. (2024) emphasize that GenAI avatars can improve engagement when they provide adaptive responses, aligning with the need for improved AI-based guidance. VR alone does not guarantee engagement—AI must be more adaptive and interactive to encourage participation. AI systems require real-time adaptation to sustain meaningful interactions. The low GPT engagement highlights a gap in AI integration, suggesting that better onboarding and responsiveness could improve participation rates.

6.2. Factors Influencing Intent and Usability (RQ2)

The regression model predicting Intent based on Ease of Use (EASE), Performance (PERF), and Usefulness (USEF) was significant ($F(3, 51) = 45.763, p < .001, R^2 = 0.729$), with Usefulness ($\beta = 0.470, p = 0.001$) having the strongest effect, followed by Performance ($\beta = 0.426, p < .001$) and Ease of Use ($\beta = 0.392, p < .001$). Spearman's correlation further supports these relationships (USEF: $\rho = 0.805$, ATT: $\rho = 0.762$, PERF: $\rho = 0.736$, all $p < .001$). Maddox & Fitzpatrick (2019) highlight that while VR can improve engagement through experiential and emotional learning, it must be carefully structured to prevent cognitive overload, which can interfere with task performance and focus. Wong et al. (2024) further highlights that pedagogical factors have a stronger impact on intent than technical factors, which supports our finding that Performance and Usefulness had a greater influence on intent than Ease of Use.

VR participants frequently reported dizziness, motion sickness, and control difficulties, with qualitative feedback mentioning "feel dizzy, user instructions are not perfect" and "difficult to control the camera". Quantitatively, Ease of Use was rated higher in VR (3.90) than Desktop (3.72), but performance issues were noted in VR-SDG. Maddox & Fitzpatrick (2019) suggest that VR engages multiple learning systems (cognitive, motor, and emotional), but poor design can contribute to cognitive overload, reducing retention and increasing stress. Nonetheless, our Spearman's correlation and regression analysis showed that Usefulness, Performance, and Ease of Use were key predictors of intent. Additionally, Self-Determination Theory (SDT) and Technology Acceptance Model (TAM) suggest that autonomy, competence, and relatedness, along with ease of use and enjoyment, significantly impact adoption and engagement.

6.3. Limitations

Our study was limited by a relatively small sample size ($n = 55$), which may affect the generalizability of the findings to broader VR learning contexts. Furthermore, we did not control key demographic variables such as prior VR experience, technical proficiency, or language background, which could have impacted participants' ease of use, interaction quality, and engagement with the AI system.

7. Conclusion

This study examined interaction patterns, engagement quality, and intent to use GenAI avatars in VR and Desktop environments, highlighting how usability challenges, cognitive load, and prior knowledge shape student engagement. Findings revealed that Desktop users engaged more frequently with structured interactions, while VR-SDG participants exhibited higher exploratory behavior, likely due to cognitive load and topic complexity. Usefulness, Performance, and Ease of Use significantly predicted intent, aligning with TAM and SDT, reinforcing that perceived effectiveness and usability drive AI adoption. However, usability issues such as motion sickness and control difficulties in VR impacted engagement, suggesting a need for better interface design and AI adaptability. Future research should explore how personalized AI support and improved VR interaction mechanics can enhance learning outcomes in an immersive environment.

References

- Lin, X. P., Li, B. B., Yao, Z. N., Yang, Z., & Zhang, M. (2024). The impact of virtual reality on student engagement in the classroom—a critical review of the literature. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1360574>
- Maddox, & Fitzpatrick, T. (2019). The promise of virtual reality in health-care education and training: It's all in the neuroscience. *Digital Medicine*, 5(4), 133–137. https://doi.org/10.4103/digm.digm_26_19
- Wong, P. P. Y., Lee, J., Gonzales, W. D. W., Choi, S. H. S., Hwang, H., & Shen, D. J. (2024). New dimensions: The impact of the metaverse and ai avatars on social science education. In W. W. K. Ma, C. Li, C. W. Fan, L. H. U, & A. Lu (Eds.), *Blended Learning. Intelligent Computing in Education* (Vol. 14797, pp. 90–101). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-4442-8_7

AI for Science and Science for AI: Lessons learned from curriculum modifications for the undergraduate program of biostatistics in a Sino-British University

Tenglong LI
Academy of Pharmacy, Xi'an Jiaotong-Liverpool University, Xi'an, China
Tenglong.Li@xjtlu.edu.cn

Abstract: *In the era of rapid technological advancement, artificial intelligence (AI) has emerged as a transformative force across industries, necessitating its integration into university curriculum. This paper proposes a conceptual framework for AI-driven curriculum design, emphasizing the synergy between "AI for Science" and "Science for AI." Specifically, "AI for Science" emphasizes AI as a tool for acquiring and systematizing existing knowledge, while "Science for AI" fosters innovation by leveraging scientific inquiry to advance AI itself. The framework advocates for a paradigm shift where AI is not an elective but a foundational element across disciplines. Lessons learned from implementing this framework in curriculum modifications for a biostatistics undergraduate program highlight challenges in multidisciplinary integration, the need for pedagogical innovations, the importance of gradual curriculum adaptation as well as defining learning outcomes suitable for student learning with AI. This conceptual framework for AI-driven curriculum modifications is recommended as it ensures that universities not only adapt to the evolving AI landscape but also contribute to shaping its future, thereby better preparing students for an AI-driven world.*

Keywords: curriculum modification, AI for Science, Science for AI, multidisciplinary, lessons learned.

1. Introduction

In today's rapidly evolving world, artificial intelligence (AI) has become an essential force, transforming industries and society at large. Its pervasive influence extends across all sectors, making AI proficiency not only a competitive advantage but a necessity (Brynjolfsson & McAfee, 2014). For students in higher education, acquiring AI skills and knowledge is crucial to thriving in the ever-changing job market, regardless of their field of study. As such, the integration of AI into higher education curricula is no longer a choice but a strategic imperative (West, 2018).

To stay relevant and equip students with the skills they need for the future, educational institutions must rethink and adapt to the reality that both educators and learners must embrace AI and develop a high level of proficiency in using it (Selwyn, 2019). This necessitates a paradigm shift in curriculum design, where AI is not merely an elective but the foundation of all fields of study. This shift calls for a comprehensive redesign of educational frameworks, ensuring that AI becomes a core competency across disciplines (Luckin et al., 2016). In response to these challenges, we propose a conceptual framework for AI-driven curriculum design motivated by the synergy between "AI for Science" and "Science for AI" (Woolf, 2010). This framework underscores the evolution from the traditional approach of using AI in education to a more synergistic model—one where AI itself can be further empowered in education (Brynjolfsson & McAfee, 2014). Such shift is central to preparing students not just for the current landscape, but for the AI-driven future of work and discovery (Baker & Siemens, 2014).

2. The Conceptual Framework

2.1. AI for Science: Utilizing AI for Acquiring Scientific Knowledge

AI for Science focuses on the known world, incorporating concepts, examples, and summaries to facilitate comprehension (Baker & Siemens, 2014). In this approach, students act as tourists exploring new knowledge, while teachers serve as tour guides, providing direction and context. The purpose is to enable students to acquire and systematize existing knowledge effectively, with AI serving as a supportive guide. The core components of this model include stages with unique but coherent themes, well-defined and delineated modules, domain expert teachers who can engage with AI as an advanced learner, and a powerful AI platform to support both teaching and learning processes (Heffernan & Heffernan, 2014).

2.2. Science for AI: Empowering AI through Scientific Knowledge

Science for AI explores the unknown, focusing on knowledge creation, posing meaningful questions, and developing novel theories to enhance critical and creative thinking (Luckin et al., 2016). In this context, both students and teachers act as explorers, with teachers bringing more experience to the journey. The purpose is to foster students' ability to create and validate new knowledge, guiding AI development through innovative thinking. Its core components include domain knowledge, AI knowledge such as deep learning and programming skills (Jordan & Mitchell, 2015), innovative pedagogy like project-based and research-led learning (Brown et al., 2020), and prompt engineering for effective AI interaction (VanLehn, 2011).

This conceptual framework is illustrated by the Figure 1 and the differences between AI for Science and Science for AI are summarized in the Table 1.

Table 1. A summary of differences between “AI for Science” and “Science for AI” for AI-driven curriculum modification.

	AI for Science	Science for AI
Focus	Known world, existing knowledge	Unknown world, new knowledge
Roles	Students as tourists, teachers as guides	Students and teachers as co-explorers
Purpose	Knowledge acquisition and systemization	Knowledge creation and validation
Knowledge about AI	Not Deep	Deep
Relationship with AI	Taught by AI	Teach AI

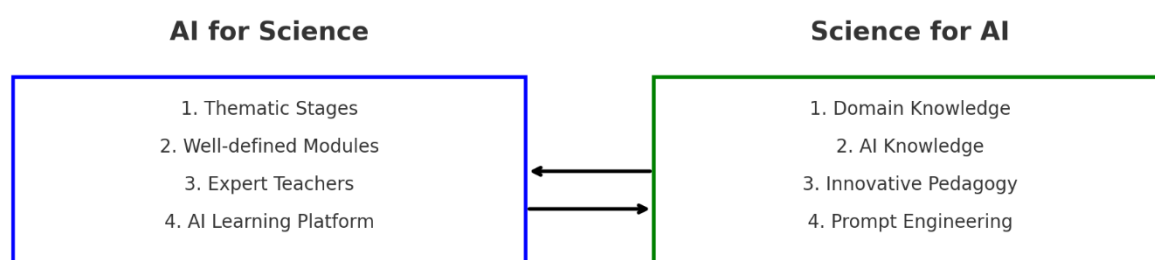


Figure 1. The structure and core components of the conceptual framework for AI-driven curriculum modification.

3. Lessons Learned

Challenges in Multi-Disciplinary Majors: Implementing AI for Science is particularly challenging in interdisciplinary fields like biostatistics within pharmacy academies, where diverse knowledge domains intersect (Chen et al., 2020). The complexity of integrating AI tools with subject-specific content demands tailored approaches, fostering collaboration across different academic disciplines to ensure coherence and relevance in the curriculum.

Our story: The author has been served as the director of the undergraduate program in biostatistics, which is built on mathematical foundations as well as applications in biomedical sciences. In addition, the program is owned by the academy of pharmacy, which expect its students to command knowledge related to pharmaceutical science as well. We have been struggled to integrate knowledge from multiple disciplines (math, pharmaceutical science, biology, statistics and applications) and the curriculum has undergone two rounds of modifications in the last two years.

Need for Pedagogical Innovations: There is an urgent need for new assessment methods, diverse delivery modes, and innovative teaching strategies to complement AI integration (Holmes et al., 2019). Traditional pedagogical models may not fully leverage AI's capabilities, necessitating the development of active learning environments, formative assessment techniques, and flexible instructional designs to enhance student engagement and learning outcomes.

Our story: Our program was one of the pioneered program for AI integration in our school. We found that traditional assessment methods, such as project reports or examination, are not quite appropriate in the age of AI. For example, students can easily draft their reports using AI agents (e.g. deepseek, ChatGPT) as

arguably they should do so. Therefore, the central task for education in the AI era is to assess the student ability of using AI, internalizing knowledge from AI as well as steering AI for their directions.

Accepting a Gradual Pace: Institutions should embrace a gradual implementation approach, allowing for trial runs, experience sharing, and ethical considerations to be thoroughly addressed (Roll & Wylie, 2016). This measured pace helps educators and administrators identify best practices, address potential pitfalls, and adapt to emerging challenges without compromising educational quality or integrity.

Our story: Our curriculum modification for AI started with two module modifications. The first one is an existing module "SAS for Data Analysis" which focuses on statistical programming in SAS. A cloud-based programming platform (SAS Studio) is adopted, along with a university sponsored AI (XIPU AI), to promote AI usage in computer labs. Those AI-assisted labs are becoming more popular in world top universities (like Harvard and MIT). We also proposed a complete new module entitled "AI in Healthcare" to provide foundational knowledge about AI applications in biomedical students.

Revolutionizing Learning Outcomes: Amid the rapid advancement of AI, educational outcomes must evolve, emphasizing adaptability, critical thinking, and lifelong learning skills (Brown et al., 2020). By redefining success metrics beyond traditional academic achievements, AI-driven curricula can foster competencies such as problem-solving, creativity, and digital literacy, preparing students for the demands of the future workforce.

Our story: Perhaps the most important task for curriculum modification is to formulate suitable learning outcomes for the age of AI. So far, our learning outcome statements about AI are still general. For example, one learning outcome of "SAS for Data Analysis" is to "Proficiently use AI in SAS programming". More details and potentially specific scenarios are needed for defining AI related learning outcomes, preferably with distinct levels outlined by Bloom's Taxonomy.

4. Conclusion

AI holds immense potential to revolutionize curriculum design and educational practices. By adopting a conceptual framework that balances "AI for Science" with "Science for AI," educators can create dynamic learning environments that not only impart knowledge but also cultivate innovation, critical thinking, and scientific inquiry. This dual approach ensures that education systems are not only prepared to integrate AI but also capable of driving its future development.

References

- Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. *Journal of Educational Technology & Society*, 17(3), 1-17.
- Brown, M., McCormack, M., Reeves, J., Brook, C., & Grajek, S. (2020). 2020 Educause Horizon Report: Teaching and Learning Edition. EDUCAUSE.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and development of artificial intelligence in smart learning: A bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 17(1), 1-18.
- Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470-497.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education.
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582-599.
- Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. *British Journal of Educational Technology*, 50(4), 1015-1029.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- West, D. M. (2018). *The Future of Work: Robots, AI, and Automation*. Brookings Institution Press.
- Woolf, B. P. (2010). *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*. Morgan Kaufmann.

Innovating Music Education through Generative AI: A Case Study on the Music and Science of Sound Course

Enrico BERTELLI
Lingnan University, Hong Kong SAR, China
enricobertelli@ln.edu.hk

Abstract: *This paper explores how Generative AI (GenAI) is reshaping music pedagogy in higher education, focusing on its integration into the cluster course Music and Science of Sound. The rise of student-friendly GenAI tools—such as CassetteAI, StableAudio, Moises AI, Suno, and Udio—has enabled non-music major students to create music, shifting the pedagogical focus from music production to creativity and curatorial skills. The study outlines an innovative approach in which AI is promoted as a starting point—a malleable blueprint for personalised output—rather than the final product. This approach encourages students to develop cohesive artistic concepts, ranging from innovative musical instruments and personas to original and personal lyrics. The course design incorporates AI-generated lyrics, music production, and visual storytelling. Students are assessed on their ability to craft a compelling narrative and reflect on their creative processes, thereby fostering critical thinking and curatorial skills. Moreover, the paper highlights solutions for the skills gap in interdisciplinary courses, where students often lack formal training, and proposes creative assessment methods to counteract the risk of AI-generated submissions. The findings suggest that AI-driven pedagogy can enhance creative exploration, interdisciplinary collaboration, and personalised learning experiences, all while challenging students to engage critically with AI's role in artistic expression.*

Keywords: Generative AI, Music Education, Creativity, Assessment, STEAM

1. Introduction

The advent of accessible generative AI tools has rapidly transformed the landscape of music creation. As Lee Cheng (2025) notes, these technologies enable non-musicians to compose structured pieces, write lyrics, and even synthesise voices with minimal input. Before early 2023, song production required basic knowledge of music styles, instrumentation, and digital audio workstations (DAWs) proficiency. Today, AI empowers individuals to generate music effortlessly, presenting both an opportunity to reinvent music education and a challenge to ensure that students engage with AI outputs in a meaningful, creative manner. In September 2023, I redesigned the course Music and Science of Sound, strongly emphasising human creativity, curation, and conceptual development. The course shifted from AI prompting to developing cohesive artistic identities through a multi-art final project. This approach aligns with findings from Barnas Monteith's research (2024), which highlights the effectiveness of arts-based AI education programmes in enhancing student creativity and engagement. This paper examines how AI tools enhance music education for non-specialists, how the curriculum fosters creativity and interdisciplinary engagement, and how new assessment models can address the impact of AI on traditional reflective writing.

2. The AI-Driven Workflow: From Prompting to Production

With the emergence of easily available GenAI audio tools in 2023, students were introduced to a structured workflow that began with generating a song fragment of 30 to 45 seconds, using applications such as CassetteAI or StableAudio. Once a segment was created, students split the track into individual stems, using MoisesAI, and later manipulated the individual tracks—such as vocal, drum, and bass—using Bandlab, a free-to-use, browser-based DAW. AI-assisted lyrics were generated via platforms like Moises Lyricist, ChatGPT, or Perplexity, which supported multiple languages. Additionally, voice synthesis tools replaced the original vocals with well-known voices, ranging from actors to anime characters or musicians. This comprehensive process ensured that AI served merely as a starting point, requiring students to adopt a curatorial role instead of passively consuming and relying on AI-generated content. This approach aligns with the ART-Official Framework's emphasis on students using AI as a collaborative partner to enhance human creativity, rather than as a replacement for it (Ruiz, 2025).

Beyond technical composition, students were tasked with developing a cohesive artistic vision for their AI-generated song. They created AI avatars using GenAI to produce visuals and crafted detailed background stories to establish a unique band identity. The students also worked together to define a collective creative direction, selecting genres such as Punk-Rock, Epic Metal, or Cantopop to guide their artistic expression.

Rather than submitting a finished song, the final project culminated in a live theatrical sketch or interactive workshop. This presentation took various forms, including performances with AI-generated music, mock band interviews, song launches, industry showcases, or storytelling sessions enhanced with background music. By shifting the focus from technical proficiency to narrative-driven creativity, the course encouraged students to engage critically with AI outputs and to utilise them as a foundation for further creative exploration (Yee-King et al., 2023).

3. Pedagogical Shifts: Learning Outcomes and Addressing the Skills Gap

The course redefined the role of AI prompting by treating it as a creative skill rather than merely a technical tool. Students engaged in one-shot prompting for visuals and animation—refining AI-generated content to achieve conceptual cohesion—and in multi-shot prompting for lyrics, transforming personal narratives into structured compositions that resonated with the stylistic requirements of their chosen musical genres. They were also encouraged to conduct independent AI-aided research into various music genres, thereby exploring the defining features of different styles. As AI models evolved—transitioning from detailed frameworks such as PARTS (Persona, Aim, Recipients, Theme, Structure) to more intuitive interactions enabled by models like OpenAI's o3 and DeepSeek—students adapted their prompting strategies accordingly. As OpenAI (2023) notes in the GPT-4 Technical Report, "GPT-4 substantially improves over previous models in the ability to follow user intent," reflecting the model's enhanced capacity to understand more natural, less structured prompts.

Recognising that many students in the course came from diverse academic backgrounds with little formal musical training, the curriculum was adapted to overcome technical barriers through AI. The focus shifted towards cultivating curatorial thinking, ensuring that artistic outputs were conceptually coherent across all media, and promoting creative autonomy by using AI as an inspirational tool rather than as an end in itself. With GenAI making music composition almost instantaneous, the primary challenge was no longer the technical creation of music but rather the development of compelling and imaginative creative ideas.

4. Rethinking Assessment in the Age of AI

Traditional reflective diaries become problematic when AI can quickly generate reflective content, undermining the learning process's authenticity. Recognising this, the course introduced alternative creative assessment methods designed to foster genuine engagement and continued reflection. Students were offered various options to express their creative journeys, such as podcasts discussing personal experiences, infographic series, comic strips or storyboards visualising their creative processes, and additional formats like photo essays, digital scrapbooks, selfie videos, and interactive timelines. The analysis of the 14 final group projects highlighted successes and challenges with these new submission formats. For example, one group's innovative integration of animated videos and interactive Q&A elements was praised for its creativity and engagement. In contrast, despite effectively combining textual narratives with visual elements, another group's project was somewhat incongruent and required more polished transitions and a balanced, cohesive tone in instrument descriptions. These successes and challenges underscore that, by shifting the focus from technical proficiency to narrative-driven creativity, the new assessment methods promote deeper reflection and encourage students to evaluate and iterate on their creative processes critically. As Mertala, Fagerlund, and Dufva (2024) argue, "arts allow and enable us to think about things [...] differently as they offer us different lenses, vocabulary, metaphors, and ways of self-expression than science, technology, engineering, and math."

5. Conclusions

This case study demonstrates that integrating Generative AI into music education transforms curriculum design and fundamentally enriches student creativity. GenAI is harnessed not as a substitute for human creativity but as a powerful tool that expands the boundaries of artistic expression and curatorial thinking. GenAI fosters a more profound, iterative creative process that transcends traditional, text-based reflection methods by enabling students to refine, personalise, and reimagine AI-generated outputs.

Moreover, the innovative assessment methods adopted in this course highlight how multimodal assignments can capture the complexity and evolution of student creativity. These new assessment forms empower learners to demonstrate their understanding and critically evaluate creative processes differently.

By reframing AI as a co-creative partner, this pedagogical model nurtures creativity and prepares students for a future in which technology and creative expression are inextricably linked.

References

- Xu, G., Looi, C. K., McMin, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Cheng, L. (2025). The impact of generative AI on school music education: Challenges and recommendations. *Arts Education Policy Review*, 1–8. <https://doi.org/10.1080/10632913.2025.2451373>
- Monteith, B. (2024). *Engaging students through creativity: The impact of arts-based AI education programs on middle and high school student conceptual understanding and interest*. In *Proceedings of the AI & Education Summit 2024*. MIT RAISE. https://raise.mit.edu/wp-content/uploads/2025/01/MIT_AIED_2024-Barnas-Monteith-Engaging-Students-through-Creativity.pdf
- Mertala, P., Fagerlund, J., & Dufva, T. S. (2024). *Rethinking the 'A' in STEAM: Insights from and for AI literacy education* [arXiv preprint]. arXiv. <https://arxiv.org/abs/2405.18179v1>
- OpenAI. (2023). *GPT-4 technical report*. <https://cdn.openai.com/papers/gpt-4.pdf>
- Ruiz, J. V. (2025). The ART-Official framework for AI in music education. *National Association for Music Education*. <https://nafme.org/blog/the-art-official-framework-for-ai-in-music-education/>
- Yee-King, M., Fiorucci, A., & d'Inverno, M. (2023). *The pop song generator: Designing an online course to teach collaborative, creative AI* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2306.10069v1>

Integrating AI and SCAMPER in Poetry Writing: Gender Differences in Creative Performance

Jon-chong HONG¹, Huan YE^{2*}, Sin-ping WANG³

^{1,2,3}National Taiwan Normal University, Taiwan

tcdahong@gmail.com, asd4082512@gmail.com, cindy5785709@gmail.com

Abstract: *With the development of artificial intelligence (AI), the application of AI in education has become the focus of research. Creativity, as one of the core competencies in the 21st century, has attracted significant attention to development methods. However, most research on creativity instruction focuses on traditional methods. There is limited research exploring teaching models integrating AI technology and specific subjects, especially on the impact of gender on creativity performance within such teaching models. The study integrates ChatGPT and SCAMPER to develop a new poetry writing course, aiming to explore the gender differences in the impact of AI-assisted poetry writing. The research utilized a quasi-experimental design to collect data from 62 fifth and sixth-grade students in Taipei and New Taipei City. The results indicated that there is a significant gender difference in both the learning value of language association and idea-sharing collective efficacy. Male students had a higher score than female students in both dimensions. The study can provide an empirical reference and a new way for creativity instruction and the application of AI in creativity in education. In addition, it offers practical suggestions for language and poetry writing for the future.*

Keywords: AI-based learning, SCAMPER, creativity, modern Chinese poetry writing, technology in education

1. Introduction

Creativity is considered a center of learning that can help individuals tackle challenges and find solutions in various fields (Habib et al., 2023). In recent years, AI has been widely adopted in creative industries. Research also suggested that AI can significantly improve the potential of people's creative thinking (Habib et al., 2023). According to the research of Wieland et al (2022), AI chatbots can effectively enhance human-robot collaboration, especially in brainstorming, AI can serve as a collaborative partner in creative brainstorming. Students also believe that AI can help develop their creativity by encouraging individual thinking and creating creative opportunities (Marrone et al., 2022). Nowadays, we are in an era of assisted creativity, where AI is not a solitary creator but a collaborative creative partner to assist in the creative process (Vinchon et al., 2023). Therefore, humans' role in using AI should be regarded as an important thing. Poetry acts as a key character in improving language ability and creative thinking, however, students often consider poetry writing and creation to be difficult and boring (Creely, 2019). Therefore, innovative writing tools and methods need to be implemented in today's writing instruction, especially utilizing AI to assist students in enhancing their writing abilities. These AI tools based on collaborative creation can stimulate students' creativity and promote their creative development (Kangasharju et al., 2022).

SCAMPER is developed by Bob Eberle, which is a creative thinking tool to stimulate people's creativity, solve problems, and generate new ideas (Boonpracha, 2023). SCAMPER is an acronym consisting of seven letters. Each letter represents a different creative idea generation strategy: Substitute, Combine, Adapt, Modify, Put to other use, Eliminate, Rearrange. Research indicated that the SCAMPER technique has been widely used in education, especially in language, art creation, or other scenarios that require creative expression (Aziza, 2023). It has been shown to significantly enhance learners' creative thinking and problem-solving abilities (Serrat, 2017). Its structured nature is particularly suitable for beginners, helping them establish a clear framework for creativity.

1.1. The Purpose of the Study

The potential of AI in education has been widely recognized nowadays, however, there is a limited exploration of AI in poetry creation. Research in the past has called for the integration of teaching methods with AI to enhance the possibility of creative thinking (Habib et al., 2023). Besides, how AI can effectively improve students' creativity and language performance also needs further exploration and validation (Kangasharju et al., 2022). Therefore, to fill the gap, this study aims to combine artificial intelligence technology with the creative thinking method "SCAMPER" to explore the differences in creativity experiences and performance between male and female students.

1.2. Hypotheses

This study hypothesizes that gender may influence students' learning effect in AI-assisted poetry writing. Specifically, the study proposes the following hypotheses:

H1: There is a significant difference in the learning value of language association between male and female students in the process of AI-assisted modern Chinese poem creation using SCAMPER.

H2: There is a significant difference in idea-sharing collective efficacy between male and female students in the process of AI-assisted modern Chinese poem creation using SCAMPER.

2. Methodology

The study adopted a quasi-experimental design and a questionnaire survey method, aiming to explore the gender difference in their creative performance in the use of "AI and the SCAMPER technique" in assisting fifth- and sixth-grade elementary school students with poetry creation.

2.1. Participants

There is a total of 62 fifth and sixth-grade elementary school students in Taipei and New Taipei City in the study, including 29 (46.8%) males and 33 females (53.2%). These students came from different backgrounds and were placed in mixed-ability classes, which reflect the diverse and general characteristics of elementary school students' language learning levels.

2.2. Procedure

The course in the study combines group collaboration with individual creation. There are a total of six sessions, each session will have 40-50 minutes. In the course, students need to utilize the AI generation tool ChatGPT to assist in poetry creation and apply the SCAMPER technique to adapt and create lines of poetry. After the course, students completed a post-course questionnaire.

At the beginning of the lesson, the teacher introduces AI technology and the application value of ChatGPT in language learning, demonstrating how to use ChatGPT to generate suggestions or provide inspiration for poetry writing. Next, students learn the seven dimensions of the SCAMPER technique and explore related concepts with the assistance of ChatGPT, such as entering, "How does 'substitution' change the imagery of a poem?" In the application stage, students utilize SCAMPER to analyze poetic techniques, discussing questions like, "Why was 'substitution' chosen instead of 'combination,' and what is its creative effect?" or seeking ChatGPT's help to extend verses: "Can you provide an example of using 'substitution' to express the qualities of the sea?" Finally, students individually create poems incorporating SCAMPER techniques.

2.3. Instruments

Each student is provided an iPad to create poetry by using ChatGPT and to complete the questionnaire. The questionnaire of the study has been adapted from previous research to evaluate various dimensions: the learning value of language association and idea-sharing collective efficacy. To ensure content validity, 6 experts were invited to evaluate whether the questions can comprehensively cover the construct. In addition, 5 students were requested to complete a pre-test to ensure the questionnaire's items are understandable for face validity.

2.4. Data Analysis

The research utilized descriptive statistics to summarize the participants' basic information and the results of the questionnaire. An Independent Sample *t*-test was adopted to analyze the data from students of different genders, investigating the impact of AI-assisted poetry creation on the value of language association learning and idea-sharing collective efficacy.

3. Results

In the learning value of language association, the results indicated that male students had a significantly higher score ($M = 4.63$, $SD = 0.38$) than female students ($M = 4.16$, $SD = 0.7$), $t = 3.39$, $p < 0.05$. It showed that male students can better perceive the value of linguistic associations during the learning process. Similarly, in the dimension of idea-sharing collective efficacy, male students also had a higher score ($M = 3.98$, $SD =$

0.7) than female students ($M = 4.37$, $SD = 0.6$), $t = 2.39$, $p < 0.05$, which suggested that male students had more confidence and were more willing to engage in sharing in the activity of AI-assisted poetry writing.

4. Discussion

Previous studies indicated that gender played an important role in students' experience and feelings when using AI in learning (Stöhr et al., 2024). The results of the study showed that gender has a significant impact on using SCAMPER for conducting AI-assisted poetry creation. Male students have both higher learning values of language association and idea-sharing collective efficacy. The findings are similar to the study of Renz et al. (2024), which suggests that men are more willing to use AI technology for self-improvement than women. In addition, the study indicated that males have a more positive attitude toward AI chatbots, which could enable them to use AI as co-creators in a more comfortable way and promote better confidence and engagement in sharing and collaboration. Therefore, male students perform greater sharing and collaborative willingness in the idea-sharing collective efficacy dimension. Moreover, in terms of gender, the study found that female students are less familiar with AI chatbots and have a lower frequency of using it, which may lead to a more negative attitude and concerns about using AI for learning (Stöhr et al., 2024), thereby impacting the gender differences in idea-sharing collective efficacy and the learning value of language association.

5. Conclusion, Limitations and Future Research

The study contributes to AI-assisted creative teaching by providing insights into how students of different genders interact with AI in poetry creation. By investigating how various methods will impact different genders' perceptions of the AI creative task, we can further understand how to optimize AI-assisted learning environments. In addition, the application of AI in education should be adapted according to various subjects, gender, and learning requirements. In summary, the study emphasizes the potential of AI in creative education and poetry creation. In addition, it also highlights the importance of addressing gender differences to create a more effective and fair learning experience in AI-assisted learning.

However, the research still has some limitations. First, some students were unfamiliar with the prompt and operation of using an AI chatbot (ChatGPT), which made it difficult for them to effectively ask ChatGPT questions and make it answer properly, thereby enhancing students' frustration and influencing the learning effect, engagement, and acceptance of the AI tool. Future research should expand the sample size and further explore the factors that influence students' interactions with AI in creative writing. In addition, the study about promoting female students to engage better in the AI-assisted creative method should also be conducted, for example, providing scaffolding to reduce their anxiety about using AI tools and enhancing their trust in AI technology.

Acknowledgements

This project is supported by the National Science and Technology Council of Taiwan and the Institute for Research Excellence in Learning Sciences, National Taiwan Normal University from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

- Aziza, U. (2023). Effectiveness of Scamper Technique in Education. *American Journal of Social and Humanitarian Research*, 4(3), 14–17. <https://doi.org/10.31150/ajshr.v4i3.2024>
- Boonpracha, J. (2023). SCAMPER for creativity of students' creative idea creation in product design. *Thinking Skills and Creativity*, 48, 101282.
- Creely, E. (2019). 'Poetry is dying': Creating a (re) new (ed) pedagogical vision for teaching poetry. *The Australian Journal of Language and Literacy*, 42(2), 116-127.
- Habib, S., Vogel, T., Anli, X., & Thorne, E. (2023). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, 34(1), 100072. <https://doi.org/10.1016/j.yjoc.2023.100072>
- Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2022). Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior*, 139, 107502. <https://doi.org/10.1016/j.chb.2022.107502>
- Kangasharju, A., Ilomäki, L., Lakkala, M., & Toom, A. (2022). Lower secondary students' poetry writing with the AI-based Poetry Machine. *Computers and Education: Artificial Intelligence*, 3, 100048. <https://doi.org/10.1016/j.caeai.2022.100048>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Marrone, R., Taddeo, V., & Hill, G. (2022). Creativity and artificial intelligence—A student perspective. *Journal of Intelligence*, 10(3), 65.
- O'Toole, K., & Horvát, E. (2024). Extending human creativity with AI. *Journal of Creativity*, 34(2), 100080. <https://doi.org/10.1016/j.yjoc.2024.100080>
- Renz, S., Kalimeris, J., Hofreiter, S., & Spörrle, M. (2024). Me, myself and AI: How gender, personality and emotions determine willingness to use Strong AI for self-improvement. *Technological Forecasting and Social Change*, 209, 123760. <https://doi.org/10.1016/j.techfore.2024.123760>
- Serrat, O. (2017). *The SCAMPER technique*. In O. Serrat (Ed.), *Knowledge solutions* (pp. 311–314). Springer. https://doi.org/10.1007/978-981-10-0983-9_33
- Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education: Artificial Intelligence*, 7, 100259. <https://doi.org/10.1016/j.caeai.2024.100259>
- Vinchon, F., Lubart, T., Bartolotta, S., Gironnay, V., Botella, M., Bourgeois-Bougrine, S., ... & Gaggioli, A. (2023). Artificial Intelligence & Creativity: A manifesto for collaboration. *The Journal of Creative Behavior*, 57(4), 472-484.
- Wieland, B., de Wit, J., & de Rooij, A. (2022). Electronic brainstorming with a chatbot partner: A good idea due to increased productivity and idea diversity. *Frontiers in Artificial Intelligence*, 5, 880673.

Enhancing Clinical Skills Training in Optometry through AI-Assisted Learning

Kin WAN

School of Optometry, The Hong Kong Polytechnic University, Hong Kong SAR, China
kkinwan@polyu.edu.hk

Abstract: This study explores the integration of artificial intelligence (AI) in optometric education to address challenges in clinical skills training. By employing motion-tracking-based AI algorithms, an AI-powered tutoring system was developed to provide real-time feedback on procedural accuracy during clinical tests. The system evaluates key parameters such as working distance, instrument positioning, and timing. Pilot implementation with 37 students and 5 teachers revealed improved student engagement, confidence, and learning outcomes compared to traditional methods. This innovative approach offers a scalable solution for enhancing practical training and standardizing performance evaluation in optometry education.

Keywords: AI in education, clinical skills training, optometry, educational technology

1. Introduction

Clinical skills are fundamental for optometry students to ensure accurate diagnosis and treatment of eye conditions. Conventional training methods often face limitations due to high student-to-tutor ratios and insufficient opportunities for personalized feedback. To address these challenges, this study introduces an AI-powered tutoring system that leverages motion-tracking algorithms to enhance the learning experience. Building on prior research in AI-assisted learning (e.g., sports biomechanics and healthcare training) (Denadai et al., 2013; Radiq et al., 2008; Régo et al., 2009; Porte et al., 2007), this project adapts these technologies for optometric education. The system provides instant feedback and objective performance metrics, fostering self-directed learning and improving teaching efficiency.

2. Method

The AI-powered system was designed using motion-tracking algorithms capable of analyzing student performance during clinical tasks (Figure 1). Key features include tracking and real-time feedback on critical procedural parameters. The pilot study involved 37 Year 2 students performing preliminary clinical tests under AI guidance and 5 instructors. Feedback from both students and instructors was collected via questionnaires.

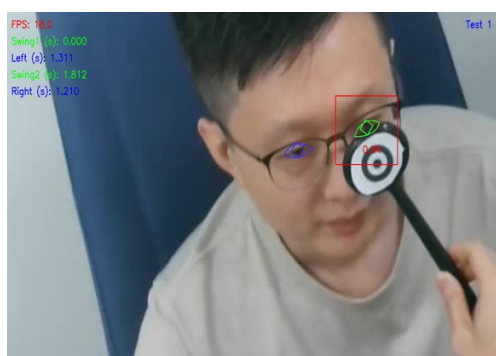


Figure 1. Demonstrations of computer vision algorithms applied in Optometry clinical test training.

3. Results

Student engagement increased as 85% reported higher motivation compared to traditional teaching methods. Additionally, 78% of students indicated increased confidence due to unbiased feedback. Instructors rated the system as extremely useful for monitoring student progress and providing targeted guidance. The pilot implementation demonstrated positive improvements.

4. Discussion

The findings highlight the potential of AI-assisted learning tools to revolutionize optometric education by providing consistent feedback mechanisms. This approach not only enhances skill acquisition but also facilitates objective evaluation across practical sessions. Future work includes expanding the system's application to advanced clinical procedures and integrating verbal consultation skills training.

5. Conclusion

This study demonstrates the feasibility and effectiveness of integrating AI-powered systems into optometry education. By providing real-time feedback and objective performance metrics, the system enhances both teaching quality and student learning outcomes. Such innovations pave the way for broader adoption of AI technologies in healthcare education.

Acknowledgements

The author would like to thank Teaching Development Grants (TDG) 2022-25 from the Hong Kong Polytechnic University for the support.

References

- Denadai, R., Toledo, A. P., Oshiiwa, M., & Saad-Hossne, R. (2013). Acquisition of suture skills during medical graduation by instructor-directed training: A randomized controlled study comparing senior medical students and faculty surgeons. *Updates in Surgery*, 65, 131–140.
- Rafiq, A., Tamariz, F., Boanca, C., Lavrentyev, V., & Merrell, R. C. (2008). Objective assessment of training surgical skills using simulated tissue interface with real-time feedback. *Journal of Surgical Education*, 65, 270–274.
- Régo, P., Peterson, R., Callaway, L., Ward, M., O'Brien, C., & Donald, K. (2009). Using a structured clinical coaching program to improve clinical skills training and assessment, as well as teachers' and students' satisfaction. *Medical Teacher*, 31(e586–e595).
- Porte, M. C., Xeroulis, G., Reznick, R. K., & Dubrowski, A. (2007). Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *American Journal of Surgery*, 193, 105–110.

OpenImpact Lab: Leveraging GenAI for Project-Based Learning and NGO Empowerment

Max Jwo Lem LEE¹, Alwin LIN², Ruoyu CHEN³, Li-Ta HSU^{4*}
^{1,3,4}The Hong Kong Polytechnic University, Hong Kong SAR, China
²CS Equity AI, United States

maxjl.lee@connect.polyu.hk, csequityai@gmail.com, ruoyu93.chen@polyu.edu.hk, lt.hsu@polyu.edu.hk

Abstract: This paper introduces OpenImpact Lab, a novel platform addressing the Industry 5.0 skills gap and the limited digital capacity of Non-Governmental Organizations (NGOs). It integrates project-based learning (PBL) with Generative AI (GenAI) to create a mutually beneficial ecosystem, demonstrating AI's transformative potential in education. The platform connects students, particularly those from underprivileged backgrounds, with real-world NGO projects, guided by GenAI tutors. These tutors provide personalized guidance, sub-task breakdown, and feedback, enhancing scalability and accessibility. Grounded in Connectivism, Social Cognitive Theory, and Communities of Practice, OpenImpact Lab fosters dynamic learning while providing NGOs with cost-effective digital solutions. This paper emphasizes the platform's innovative design, theoretical underpinnings, ethical considerations, and potential impact. The novelty lies in strategically integrating GenAI to bridge educational and digital divides, creating a scalable, sustainable model for skills development and social impact. Key evaluation metrics and a pilot study methodology are outlined.

Keywords: Generative AI, Project-Based Learning, NGO Capacity Building, Skills Development, Digital Divide

1. Introduction

The rapid advancement of Artificial Intelligence (AI) is fundamentally reshaping workforce demands, creating an urgent need for educational approaches that can keep pace with these changes (Pendy, 2023). This transformation is particularly evident in the context of Industry 5.0, where the focus shifts to human-centric technology and seamless human-machine collaboration. This paradigm shift necessitates advanced digital literacy, AI fluency, complex problem-solving skills, and a strong understanding of ethical technology application (Saxena & Bajotra, 2024). However, traditional educational models often struggle to adapt quickly enough, leading to a widening skills gap that disproportionately affects underprivileged students. Simultaneously, Non-Governmental Organizations (NGOs), which play a crucial role in addressing societal challenges, often face significant resource constraints that hinder their own digital transformation and limit their digital capacity (Rajan et al., 2024). OpenImpact Lab addresses these interconnected challenges. It tackles the growing skills gap, especially within the Industry 5.0 framework, and supports the digital empowerment of NGOs. The Lab uniquely combines Project-Based Learning (PBL) with Generative AI (GenAI), creating a synergistic ecosystem that aligns with the imperative to integrate AI into modern pedagogical frameworks (Baskara, 2024). The potential displacement of jobs due to automation further underscores the urgency of these efforts.

Table 1. Comparison of OpenImpact Lab with Related Platforms.

Feature	OpenImpact Lab	Coursera/Khan Academy	Catchafire	AI Tutoring Platforms (e.g., Codecademy)
Project-Based Learning	Yes	No	No	Limited
Real-World NGO Projects	Yes	No	Yes (no student focus)	No
GenAI-Powered Tutoring	Yes	No	No	Yes (limited scope)
Personalized Learning	Yes	Limited	No	Yes
Industry Professional Review	Yes	No	Yes (volunteer)	No
Focus on Skills Development	Yes	Yes	No	Yes
NGO Capacity Building	Yes	No	Yes	No

2. Related Work

Existing literature often bifurcates the investigation of skills gaps and the digital needs of NGOs. Educational reform research frequently centers on technology integration or Project-Based Learning (PBL), with theoretical grounding in constructivism and connectivism for educational technology, relevant to the

approach adopted by OpenImpact Lab (Ifenthaler et al., 2024). Studies on AI in K-12 education identify opportunities and challenges, and some explore AI's role in specific domains like language learning; however, these often lack contextualized, real-world application and scalable, personalized support mechanisms (Eden et al., 2024). Research addressing NGO capacity building frequently examines digital adoption yet rarely bridges this with student skills development (Reddy et al., 2023). While platforms like Coursera and Khan Academy provide online courses, they generally lack hands-on, project-based learning experiences. Catchafire connects skilled volunteers with nonprofits, but its focus is not on structured student learning pathways. Other platforms offer AI-powered tutoring, but typically within constrained contexts, lacking integration with real-world NGO projects (Herawati et al., 2024). While some discuss AI's transformative potential in education, OpenImpact Lab uniquely synthesizes personalized, GenAI-driven tutoring with real-world, project-based learning, simultaneously addressing the needs of students and NGOs. This integrated approach constitutes a novel contribution, fostering a holistic, mutually beneficial ecosystem (Table 1) and resonates with connectivist learning principles.

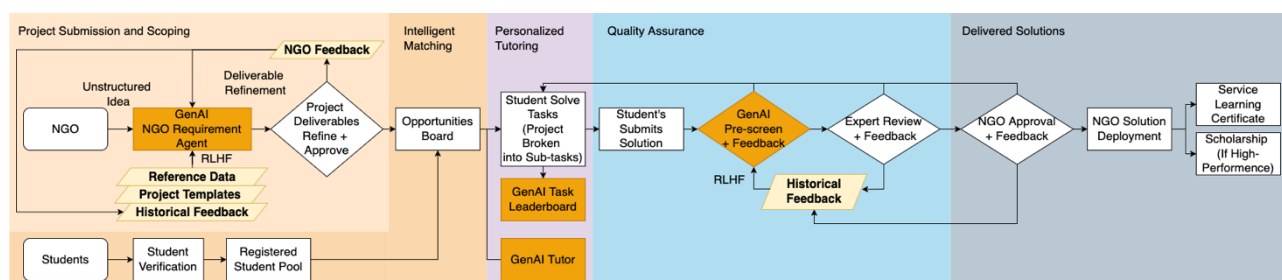


Figure 1. OpenImpact Lab Platform Design: Key components and interactions, highlighting the cyclical feedback loop.

3. Design and Conceptual Framework

OpenImpact Lab is a dynamic platform connecting students with real-world NGO digital projects. Its design, rooted in PBL, Connectivism, Social Cognitive Theory, and Communities of Practice, emphasizes active learning, real-world problem-solving, networking, mentorship, and collaborative knowledge sharing. The core principle is creating a mutually beneficial ecosystem (Fig 1). NGOs submit project proposals, processed by a GenAI NGO Requirement Agent which extracts key information, assesses feasibility, estimates time requirements, and identifies required skills. The project scoping module, trained on anonymized project descriptions, ensures accurate processing. Then, matches students with projects, considering skill profiles, project requirements, and interests. GenAI Tutors guide students, adapting to individual needs, providing automated feedback on code and suggesting resources. These tutors are fine-tuned on educational materials, coding examples, and NGO project documentation, with personalization achieved through continuous monitoring of student progress. Complex projects are broken down into manageable subtasks. A GenAI Pre-Screening Agent reviews completed solutions, flagging potential issues, before freelance industry professional review using structured rubrics. NGOs receive completed solutions. For example, Sarah, with basic Python skills, is matched with an NGO needing a donation tracking web application. The GenAI tutor guides her, breaking the project into tasks like 'designing the database schema' and 'creating the user interface,' providing feedback and resources. An industry professional reviews the application before deployment.

4. Pilot Study Methodology

A 6-week pilot study with 60 students (30 high school, 30 undergraduate, diverse backgrounds), 3-5 NGOs, and 5-7 freelance industry professionals will validate OpenImpact Lab's core functionalities and gather preliminary data. Modest incentives are provided. NGOs submit projects, refined by the GenAI scoping module. Students, matched to projects based on skill level, work for 2-4 weeks, guided by GenAI tutors (compared to a control group). Work is reviewed (GenAI pre-screening, freelancer evaluation). The pilot evaluates GenAI scoping, usability, tutor impact, work quality, freelancer feedback, skills acquisition, and NGO satisfaction. Statistical analysis (t-tests, correlation analyses) will quantify the platform's impact. Table 2 shows evaluation metrics.

Table 2. Selected Evaluation Metrics for OpenImpact Lab.

Stakeholder	Metric	Measurement Method
Students	Skills Acquisition	Pre- and post-project assessments (standardized tests, project-specific rubrics)
NGOs	Learning Engagement	Platform usage data, surveys (SCEQ), focus groups
	Project Satisfaction	Surveys, interviews
	Impact on NGO Operations	Case studies, interviews, NGO-specific KPIs (e.g., increased volunteer applications, reduced administrative time, improved service delivery)
Freelancers	Satisfaction	Surveys, interviews
Platform	Quality of Feedback	Student ratings
	Scalability	Number of users, platform capacity (technical metrics)
	GenAI Tutor Effectiveness	Student feedback, tutor-student interaction analysis, comparison with control group.
Ethical Considerations	Bias Detection and Mitigation	Continuous monitoring of GenAI, mitigation strategies (data augmentation, algorithmic adjustments).
	Data Privacy and Security	Adherence to regulations (GDPR, CCPA), security audits, penetration testing

5. Discussion and Conclusion

OpenImpact Lab offers a competitive edge by integrating GenAI and a multi-stakeholder approach (students, NGOs, industry). This fosters real-world impact by connecting students with genuine NGO needs, while GenAI enables scalability and accessibility through a personalized tutor. Table 3 summarizes OpenImpact Lab's competitive advantages. The use of GenAI to enhance student engagement is a key aspect.

Table 3. Competitive Advantages of OpenImpact Lab.

Advantage	Description
Integrated Approach	Combines personalized, GenAI-powered tutoring with real-world, project-based learning for NGOs.
Real-World Impact	Directly addresses NGO digital needs, providing tangible benefits.
Scalable Personalization	Leverages GenAI to provide personalized support.
Industry-Validated Skills	Incorporates industry professional review.
Mutually Beneficial Ecosystem	Creates a win-win situation for all stakeholders.
Advantage	Description

OpenImpact Lab presents a novel and transformative approach to education and social impact, seamlessly integrating project-based learning with GenAI to create a dynamic ecosystem benefiting students, NGOs, and the broader community. This platform transcends traditional educational models by providing real-world experience, personalized guidance, and industry-validated skills to students, while simultaneously empowering NGOs with crucial digital solutions. By bridging the skills gap, the digital divide, and NGO resource constraints, OpenImpact Lab, grounded in established learning principles, fosters collaborative, engaged learning.

References

- Baskara, F. R. (2024). Generative AI as an enabler of sustainable education: Theoretical perspectives and future directions. *British Journal of Teacher Education and Pedagogy*, 3(3), 9. doi: 10.32996/bjtep.2024.3.3.9.
- Eden, R., et al. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews*, 10(2), 0039. doi: 10.30574/msarr.2024.10.2.
- Herawati, H., et al. (2024). Exploring the role of artificial intelligence in education, students preferences and perceptions. *Al-Ishlah: Jurnal Pendidikan*, 16(2), 4784. doi: 10.35445/alishlah.v16i2.4784.
- Ifenthaler, D., et al. (2024). Artificial intelligence in education: Implications for policymakers, researchers, and practitioners. *Technology, Knowledge and Learning*. doi: 10.1007/s10758-024-09747-0.
- Pendy, P. (2023). Artificial intelligence: The future of education. *Jurnal Indonesia Sosial Sains*, 2(11), 801. doi: 10.59141/jiss.v2i11.801.
- Rajan, R., et al. (2024). Critical factors for capacity building of NGOs working in the intellectual disability sector in India: a TISM-MICMAC approach. *Journal of Asia Business Studies*. doi: 10.1108/jabs-02-2023-0061.

Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.

Reddy, S., et al. (2023). Evaluation of global cleft care initiatives among the top searched low- and middle-income countries. *The Cleft Palate-Craniofacial Journal*. doi: 10.1177/10556656231160399.

Saxena, C., & Bajotra, S. (2024). Potential of artificial intelligence in education and ethical issues. doi: 10.3126/fwje.v1i1.68773.

Integrating Generative AI into Inquiry-Based Learning for Personalized Mathematics Education

Fridolin Sze Thou TING^{1*}, Daisy Dichen WANG², Lawrence Chi Lok CHAN³

^{1,2}Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR, China

³Department of Psychology, The Education University of Hong Kong, Hong Kong SAR, China
fridolin@eduhk.hk, dwang@eduhk.hk, cchilok@eduhk.hk

Abstract: *This study explores the effectiveness of integrating Generative Artificial Intelligence (GenAI) within an Inquiry-Based Learning (IBL) framework to enhance personalized and active learning in mathematics education. The intervention used personalized and active learning pedagogies combined with prompt engineering applied to IBL activity construction related to their personal interest, hobbies and passions in a senior tertiary mathematics teacher education course. A comparative analysis was conducted between an experimental group exposed to GenAI-assisted IBL activities and a control group without GenAI integration. Generic Intended Learning Outcomes (GILOs) related to problem-solving, critical thinking, and written communication were assessed via surveys. Results indicated a statistically significant improvement in critical thinking skills in the experimental group compared to the control group. This suggests that integrating GenAI into IBL can positively influence critical thinking development in mathematics teachers.*

Keywords: Generative AI, Inquiry-Based Learning, Mathematics Education, Personalized Learning, Critical Thinking

1. Introduction

The rapid advancement of artificial intelligence (AI), particularly generative AI (GenAI), presents vast opportunities for education. GenAI tools, such as large language models (LLMs), offer the potential to personalize learning experiences, enhance student engagement, and foster deeper understanding of complex concepts, in particular for mathematic education (Matzakos, et. al., 2023). This study explores the integration of GenAI into an inquiry-based learning (IBL) environment for mathematics teachers, with emphasis on student-centered exploration and discovery using personalized learning capabilities of GenAI. By leveraging GenAI, educators can design engaging learning experiences that cater to individual student needs and interests. This research aims to investigate the impact of GenAI integration on teachers' self-perceived problem-solving, critical thinking, and written communication skills, in the context of designing IBL tasks with the use of educational technology.

2. Literature Review

Inquiry-based learning (IBL) has been recognized as an effective pedagogical approach for promoting active learning (Justice, et al., 2009) and deeper understanding in mathematics (Rasmussen & Kwon, 2007). IBL encourages students to explore concepts, ask questions, and construct their own knowledge through investigation and collaboration (Lombard & Schneider, 2013). There has been extensive research highlighting the benefits of integrating technology into IBL to enhance learning outcomes (e.g., Susilawati, 2022). The emergence of the transformative technology of GenAI offers new possibilities for personalizing IBL experiences and providing tailored support to learners. We also know that personalized learning interventions can lead to improved student engagement and academic performance (e.g., Lee et al., 2018). This study builds upon this existing research by examining the specific impact of GenAI integration on pre-teachers' preparedness and development of key skills within an IBL mathematics teaching course using technology.

3. Methods

3.1. Participants

Our study was conducted in a course on an IBL approach to teaching mathematics with the aid of technology in Semester 1 of the 2024-25 academic year at a Hong Kong tertiary institution. Participants were

divided into two groups: an experimental group ($n=9$) exposed to GenAI-assisted IBL teaching material and activities and a control group ($n=9$) following traditional teaching methods without GenAI integration.

3.2. Intervention

In the experimental group, a personalized and active learning pedagogy involving prompt engineering with LLMs was integrated into lesson plans. The special personalized prompt was called the *concept prompt*. The concept prompt connects any topic in Mathematics (or any other subject, for that matter), to a student's hobby, interest or passion. To be more precise, here is the precise prompt engineering form of the

Concept prompt:

I am a α student who loves β . Please help me understand π in the subject Ω , relating it to β .

where

α = grade level of student

β = students' hobby, passion or personal interest/experiences

π = specific (sub) topic/concept/idea in science or math

Ω = course or subject area of (sub) topic

A specific example of the concept prompt is as follows:

I am a grade 12 student who loves the K-pop girl band, "Blackpink". Please help me understand the concept of a "limit" in the subject Calculus, relating it to Blackpink.

The remarkable ability of LLMs to relate any two seemingly disparate topics is to be noted here, including relating limits in Calculus to a Korean girl pop group, must, at least intuitively, convince us of the power of LLMs to increase conceptual understanding of complex science and math topics.

Since this course is about teaching mathematics via IBL with technology, then we also outlined how a teacher may use generative AI to help kick start and aide the teacher in initially planning or getting the outline for an IBL activity in mathematics, using the following IBL prompt:

IBL activity prompt:

I have to teach the topic of α in the subject β at grade level π . Please suggest an inquiry based learning activity, using technology Ω , to teach α .

where

α = specific (sub) topic/concept/idea in science or math

β = course or subject area of (sub) topic

π = grade level

Ω = technology you are using

In addition, according to the 5E model for IBL (Bybee & Landes, 1990) with the five e's being engage, explore, explain, evaluate and elaborate, we suggested to our student teachers to allow students to use GenAI in the 1st and 2nd stages of the 5E pedagogy: namely, in the engage and explore stages. We felt this has the advantage of letting the students personalize the mathematics topic to their own interest, instead of having the teacher traditionally construct IBL questions for them in the 1st and 2nd stages. For example, for learning the topic of fractions in primary school, a good inquiry-based learning question would be, how would you divide a pizza into 3 equal shares – which would be the “vanilla” or textbook example – not personalized to individual students who do not particularly like pizza. We propose the student use the *concept prompt* for stages 1 and 2 of the 5E model of IBL.

We implemented the concept and IBL activity prompts with GenAI using a collaborative online anonymous chatroom platform integrated with LLMs called, YoChatGPT (<https://www.yochatgpt.io/>). An actual example of a student concept prompts trying to understand the concept of a limit in Calculus, through his/her interest in yoga, in Figure 1 below:

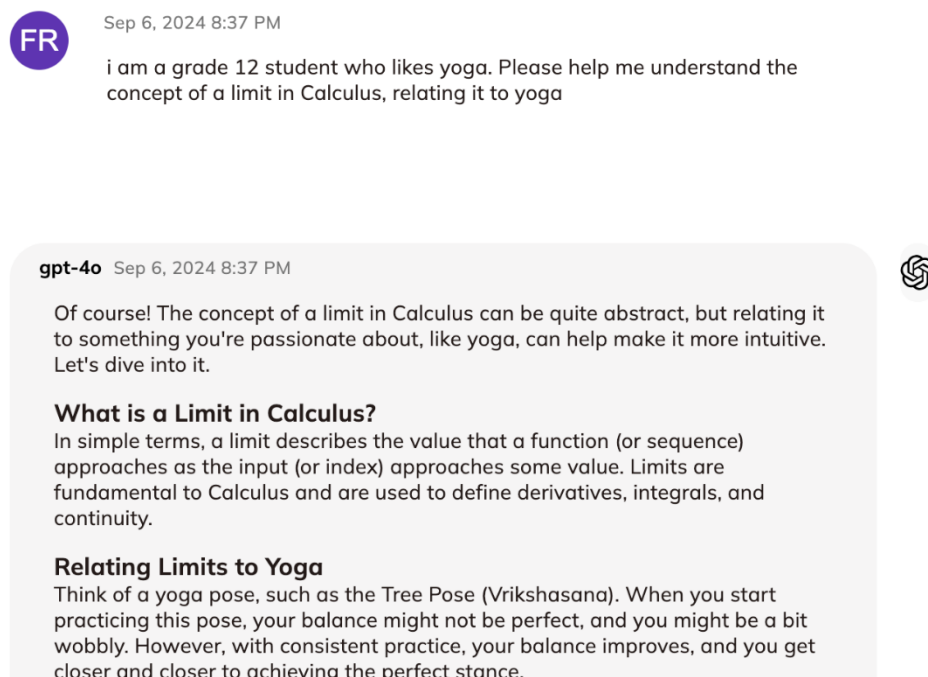


Figure 1. Example of concept prompt for understanding limits in Calculus, relating it to yoga.

We believe this is the best IBL question for a student to ask, instead of a teacher posing the questions in the engage and explore stage. Then, after the generative AI or LLM outputs the answer to the IBL student concept prompt, then students can work in groups of 4 or 5 to go through the rest of the steps of the 5E model, exploring, critically thinking, and quantitatively verifying the LLMs response to their concept prompt.

We gave both experimental and control classes a Preparedness of use of GenAI (Kohnke, 2023) and Generic Intended Learning Outcomes (for problem solving, critical thinking and written communication skills) survey (EdUHK, 2020).

3.4. Measures

Generic Intended Learning Outcomes (GILOs) related to problem-solving, critical thinking, and written communication were assessed using a survey adapted from the "Self-Assessment of Generic Intended Learning Outcomes (GILOs)" developed by the Education University of Hong Kong (EdUHK, 2020). The survey employed Likert-scale questions to gauge students' self-perceived skills and readiness.

3.5. Procedure

Both groups completed the preparedness and GILOs survey at the end of the course. The experimental group engaged with GenAI-assisted prompts throughout the semester, while the control group participated in standard IBL activities without GenAI support. Data were analyzed using Independent Samples t-tests to compare mean scores between groups.

4. Results

The analysis focused on three GILO categories: Problem-Solving (PS), Critical Thinking (CT), and Written Communication (WC). The group statistics and Independent Samples Test results are summarized Table 1 and Table 2, below, respectively.

Table 1: Group Descriptive Statistics for GILO Categories

GILO Category	Group	N	Mean	Std. Deviation
Problem-Solving (PS)	Experimental	9	4.0278	1.01892
Problem-Solving (PS)	Control	9	3.2222	0.53684
Critical Thinking (CT)	Experimental	9	4.1667	0.85696
Critical Thinking (CT)	Control	9	3.2500	0.57282
Written Communication (WC)	Experimental	9	3.9722	0.97183
Written Communication (WC)	Control	9	3.5000	0.43301

Table 2: Independent Samples t-Test Results for GILO Categories

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
PS	6.173	.024	2.098	16	.052	.80556	.38390
CT	1.671	.214	2.668	16	.017	.91667	.34359
WC	2.069	.170	1.332	16	.202	.47222	.35464

The results indicated that there was a statistically significant difference in the mean score of Critical Thinking between experimental group ($M = 4.1667$, $SD = 0.85696$) and control group ($M = 3.2500$, $SD = 0.57282$), $t(16) = 2.668$, $p = .017$, with equal variance assumed.

5. Discussion and Future Directions

The findings suggest that integrating GenAI into an IBL framework can significantly enhance critical thinking skills in student teachers learning how to implement IBL using technology in mathematics classes. We believe that critical thinking skills have been enhanced as students in the experimental class had to implement more discerning evaluation, validation, exploration and theoretical implementation of their LLM responses to the concept and IBL activity prompts; as compared to the traditional class which had no such exercise. The very fact that students in the experimental class had to try to personalize their learning, i.e., connect abstract concepts to their own personal experiences, requires analytical thought, and likely contributed to increased engagement and deeper cognitive processing, and hence, critical thinking.

However, we did not find any significant differences in self-perceived problem-solving and written communication skills. This may indicate that while GenAI effectively supports critical thinking by fostering analytical and evaluative capacities of GenAI responses. Additional strategies may be required to impact other GILO skill areas. Future studies will explore other prompt engineering techniques, such as “chain of thought” prompting to comprehensively address various learning outcomes related to problem solving.

Limitations of this study include the small sample size and the short duration of the intervention, which may affect the generalizability of the results. Additionally, the reliance on self-reported measures may introduce bias. Our future research will incorporate larger, more diverse populations and employ objective assessments to validate the findings. In addition, investigating the long-term effects of GenAI integration and its applicability across different STEM related topics and educational levels will provide a more holistic understanding of its potential.

In conclusion, this study contributes to the growing body of evidence supporting the use of GenAI in personalizing mathematics education. By leveraging students' interests and fostering an active learning environment implemented through an anonymous online integrated chatroom with LLMs, GenAI can play a pivotal role in enhancing critical thinking skills by thoughtful and careful personalization of abstract topics into one's individual experiences, thereby preparing students for complex problem-solving and adaptive reasoning in STEM related contexts.

References

EdUHK (2020) Self-Assessment of Generic Intended Learning Outcomes (GILOs)
<https://www.lttc.edu.hk/for-staff/generic-intended-learning-outcomes/>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Justice, C., Rice, J., Roy, D., Hudspeth, B., & Jenkins, H. (2009). inquiry-based learning in higher education: administrators' perspectives on integrating inquiry pedagogy into the curriculum. *Higher Education*, 58(6), 841-855. <https://doi.org/10.1007/s10734-009-9228-7>
- Lucas Kohnke, Benjamin Luke Moorhouse, Di Zou, (2023) Exploring generative artificial intelligence preparedness among university language instructors: A case study, *Computers and Education: Artificial Intelligence*, Volume 5, 100156, ISSN 2666-920X, <https://doi.org/10.1016/j.caeai.2023.100156>
- Lee, D., Huh, Y., Lin, C., & Reigeluth, C. M. (2018). Technology functions for personalized learning in learner-centered schools. *Educational Technology Research and Development*, 66(5), 1269-1302. <https://doi.org/10.1007/s11423-018-9615-9>
- Lombard, F. and Schneider, D. (2013). Good student questions in inquiry learning. *Journal of Biological Education*, 47(3), 166-174. <https://doi.org/10.1080/00219266.2013.821749>
- Matzakos, N., Doukakis, S., & Moundridou, M. (2023). learning mathematics with large language models. *International Journal of Emerging Technologies in learning (Ijet)*, 18(20), 51-71. <https://doi.org/10.3991/ijet.v18i20.42979>
- Rasmussen, C. and Kwon, O. N. (2007). An inquiry-oriented approach to undergraduate mathematics. *The Journal of Mathematical Behavior*, 26(3), 189-194. <https://doi.org/10.1016/j.jmathb.2007.10.001>
- Susilawati, S., Doyan, A., Mulyadi, L., Abo, C. P., & Pineda, C. I. S. (2022). The effectiveness of modern physics learning tools using the phet virtual media assisted inquiry model in improving cognitive learning outcomes, science process skills, and scientific creativity of prospective teacher students. *Jurnal Penelitian Pendidikan IPA*, 8(1), 291-295. <https://doi.org/10.29303/jppipa.v8i1.1304>

Enhancing Students' Prompt Engineering Skills through Scaffolding Strategies

Shurui BAI¹, Chen YANG²

^{1,2}Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR, China
tstbai@eduhk.hk, yyangchenxi13@gmail.com

Abstract: *Prompt engineering is the process of crafting, refining, and implementing prompts to optimize the performance of large language models (LLMs). Effective prompt engineering results in clear, task-specific prompts that help users obtain high-quality outputs from LLMs. Despite its significance, many students struggle with prompt engineering. Although frameworks for the prompt engineering process and prompt design exist, there is a notable lack of theory-based instructional strategies to teach students prompt engineering. This two-stage study addresses this gap by applying scaffolding strategies to enhance students' prompt engineering skills through three stages: intersubjectivity, ongoing diagnosis and calibrated support, and fading. In stage 1, we conducted a comparative study to examine the effects of human teacher scaffolding only (Semester 1) versus AI chatbot-assisted human teacher scaffolding (Semester 2) on design problem-solving performance. After identifying the positive effects of using AI chatbot-assisted human teacher scaffolding, we collected students' feedback on improving design problem-solving performance. Most students reported a need to learn how to communicate with the AI chatbot to elicit desired outputs. Consequently, we applied scaffolding strategies to enhance students' prompt engineering skills (Semester 3) and then examined their effects on students' design problem-solving performance. By comparing data from Semesters 2 and 3, we found that students who received prompt engineering training had significantly higher design problem-solving performance than those who did not.*

Keywords: Prompt engineering, artificial intelligence, scaffolding, design problem-solving

1. Introduction

Prompt engineering involves creating, refining, and implementing prompts to optimize the performance of large language models (LLMs) (Gattupalli et al., 2023). Effective prompt engineering produces clear, task-specific prompts that help users elicit high-quality outputs from LLMs (Lo, 2023). Despite its importance, many students lack proficiency in prompt engineering (Zamfirescu-Pereira et al., 2023). While frameworks for the prompt engineering process and prompt design exist, there is a notable absence of theory-informed instructional strategies to teach students prompt engineering. This two-stage study addresses this gap by applying scaffolding strategies to enhance students' prompt engineering skills through three stages: intersubjectivity, ongoing diagnosis and calibrated support, and fading (Wood et al., 1976).

2. Scaffolding as the Theoretical Foundation

Scaffolding involves an expert assisting a novice in solving a problem, completing a task, or achieving a goal that would be beyond their unassisted capabilities. This assistance involves the experienced teacher managing the more complex elements of the task, thereby allowing the learner to focus on and complete the components within their current level of competence (Wood et al., 1976). As this strategy aligns with the problem-solving task nature in the course and novice user (students in this study) nature in interacting with AI chatbots, we chose scaffolding as the instructional strategy to design the prompt engineering training sessions in three stages: intersubjectivity, ongoing diagnosis and calibrated support, and fading.

3. Prompt Engineering Process Framework

We rapidly reviewed the current empirically validated prompt engineering frameworks and best practices within the educational literature. Through this review, we identified a prompt engineering process framework: the WCV framework (Bai et al., 2024). The foci of this framework are on prompt engineering processes. (1) Write the prompts. The first step is writing prompts, which are the inputs users provide to generate the desired output. (2) Curate the output. The second step is curating the output, which involves selecting the satisfactory results from the AI tool. The initial evaluations of the output are: unsatisfactory, need-to-be-improved output, or satisfactory. The follow-up actions for different evaluations are provided in the framework. (3) Verify the

output. The final step is verifying the output, which involves checking the authenticity and accuracy of the AI output. Current LLMs have inherent limitations, sometimes producing incorrect or illogical outputs, known as hallucinations.

4. Pilot Study

We conducted a within-subject study with 36 postgraduate students to examine the effects of using an AI chatbot to scaffold students' progress in solving an ill-structured problem. Their average age is 24.59 (SD = 4.51), including 14 females, 18 males, and four individuals who did not disclose their gender. The study was divided into two stages. In Stage 1, students used the AI chatbot to revise their algorithm design, resulting in version 2 for their mid-intervention assignment. In Stage 2, students continued using the AI chatbot and had a 20-minute consultation with a teacher to finalize their algorithm design, submitting the final version by the end of the study. No prompt engineering scaffolding was provided by the teacher, and students interacted with the AI chatbot independently after class. Flipped class was implemented in the class, where students did learning tasks and learnt new knowledge before coming to the class. In-class time focused on discussion on difficult learning tasks.

The results of the pilot study demonstrated a significant improvement in students' design assignment scores from the pre-intervention to the post-intervention phases. The post-intervention scores revealed a notable difference between the two scaffolding groups ($p < .001$), with the experiment group ($M = 81.15$) outperforming the control group ($M = 70.01$). In contrast, the pre-intervention scores showed no significant difference between the groups ($p = .189$).

However, during semi-structured interviews, students expressed a need for prompt engineering training. Students felt their questions had been misinterpreted into another specialized area. Many students also found the necessity of question-generated prompts. These insights highlight the importance of effective prompt writing and AI output interactive curation. Therefore, in the main study, we helped the teacher carefully design and support the prompt engineering process with the AI chatbot to ensure it elicits meaningful output for problem-solving tasks.

5. Main Study

This study used a quasi-experimental design and a mixed-method approach. It included an experimental group and a historical-cohort control group. Both groups took a one-semester postgraduate course on coding and computational thinking. The course was held face-to-face at a public university in East Asia. Teaching materials, the instructor, and assignments were identical for both groups, except for the presence of prompt engineering scaffolding for interacting with the AI chatbot.

6. Results

The study found that prompt engineering scaffolding significantly improved students' design problem-solving performance. Thematic analysis of students' perceptions indicated that the AI chatbot, combined with prompt engineering training, positively impacted their algorithm design process. The experiment group demonstrated higher levels of actual and self-reported design problem-solving performance than the control group at both pre- and post-test time points. The results suggest that students who received prompt engineering scaffolding had a significantly higher level of design problem-solving performance compared to those who did not.

7. Conclusion

Fig. 1 presents a boxplot of student participants' design problem-solving performance medians between the two groups at the post-test stage. It is obvious that the median ($Mdn = 78.89$) of the experiment group surpassed the median ($Mdn = 68.86$) of the control group. Around 75% of participants in the experiment group performed better than the highest performing participant (score = 73) in the control group.

The study concludes that scaffolding strategies effectively enhance students' prompt engineering skills, leading to improved design problem-solving performance. Future research should explore additional instructional strategies to further support students in developing these skills.

Participants significantly benefited from prompt engineering scaffolding, which facilitated effective engagement with the chatbot. Many utilized the chatbot for direct inquiries to gather relevant information and break down complex problems. Interviewee 1 stated, "I ask the chatbot directly," and Interviewee 4 noting, "AI helps me to decompose the problems via prompting with it iteratively" illustrate how scaffolding allowed them to formulate precise questions.

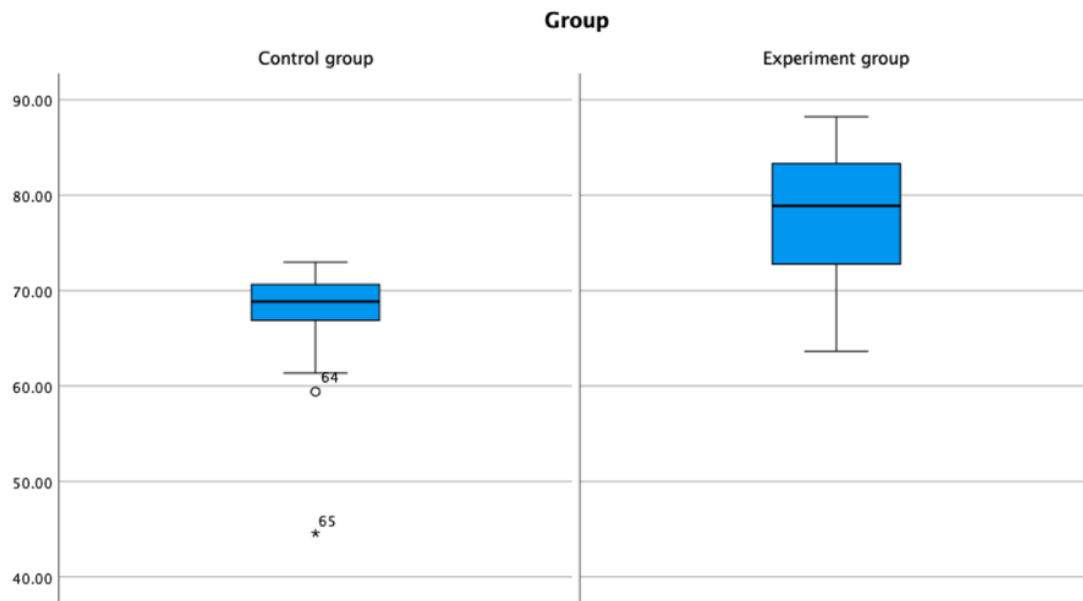


Figure 1. Boxplot of student participants' design problem-solving performance

References

- Bai, S., Gonda, D. E., & Hew, K. F. (2024). Write-Curate-Verify: A Case Study of Leveraging Generative AI for Scenario Writing in Scenario-Based Learning. *IEEE Transactions on Learning Technologies*, 17, 1313–1324. <https://doi.org/10.1109/TLT.2024.3378306>
- Gattupalli, S., Maloy, R., & Edwards, S. (2023). Prompt Literacy: A Pivotal Educational Skill in the Age of AI. <https://doi.org/10.7275/3498-wx48>
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4), 102720. <https://doi.org/10.1016/j.acalib.2023.102720>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 17(2), 89–100.
- Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. <https://doi.org/10.1145/3544548.3581388>

A Case Study of Identifying AI's Epistemological Roles in ICAP-STEAM Education: Evidence from an Arduino Environmental Monitoring System Project

Wai Kit MAK

The Education University of Hong Kong, Hong Kong SAR, China
s1113744@s.edu.hk

Abstract: This paper introduces the AI Learning Pyramid, a novel framework designed to categorize artificial intelligence tools into three hierarchical epistemological dimensions in educational contexts: FOR Learning (developing human-AI complementary competencies), AS Learning (utilizing AI as applied educational tools), and OF Learning (embedding AI technologies within learning processes). Through an Arduino environmental monitoring project with Form 2 students, this study demonstrates how the pyramid framework, when integrated with Chi and Wylie's ICAP taxonomy, enables strategic AI implementation across different cognitive engagement levels. Comparative analysis between non-AI (2023-24) and AI-supported (2024-25) cohorts reveals significant differences in code modification behaviors, with rates increasing from 37.9% to 71.9% ($p < .001$). These findings provide empirical evidence for Pratschke's "new hybrid" educational paradigm, positioning AI as a collaborative partner in knowledge construction rather than merely a technological tool. The AI Learning Pyramid addresses growing needs for structured approaches to AI integration in STEAM education while contributing to emerging "generativism" discourse and offering educators a pedagogically sound foundation for AI implementation.

Keywords: Educational Technology, Digital Pedagogy, STEAM Education, AI-enhanced learning, ICAP Framework

1. Introduction

STEAM education faces unprecedented challenges in the AI era, demanding innovative pedagogical approaches for junior secondary students engaging with complex technologies (Wahab & Teo, 2023). While Chi and Wylie's (2014) ICAP framework provides a valuable taxonomy of learning behaviors from passive to interactive engagement, and Pratschke (2024) conceptualizes a "new hybrid" educational model where AI functions as a collaborative knowledge construction partner, structured frameworks for AI integration across learning stages remain underdeveloped (Holmes et al., 2022). This study addresses this gap through an Arduino-based environmental monitoring project with 231 Form 2 students, examining how strategically implemented AI tools can facilitate progression through ICAP engagement modes while transforming AI from technological novelty to pedagogically sound learning component within contemporary educational ecosystems.

2. Theoretical Framework

2.1. The AI Learning Pyramid

The AI Learning Pyramid presents three epistemological dimensions of AI in education. Learning FOR AI (apex) develops human competencies complementary to AI systems, emphasizing critical thinking and ethical reasoning with minimal computational requirements. Learning AS AI (middle) leverages existing AI as applied educational tools, balancing academic and computational considerations. Learning OF AI (foundation) embeds AI technologies within educational processes, leaning toward computational approaches. This hierarchy spans from constructivist theories (FOR) through distributed cognition (AS) to adaptive learning theories (OF), representing a continuum from educational to computational orientations.

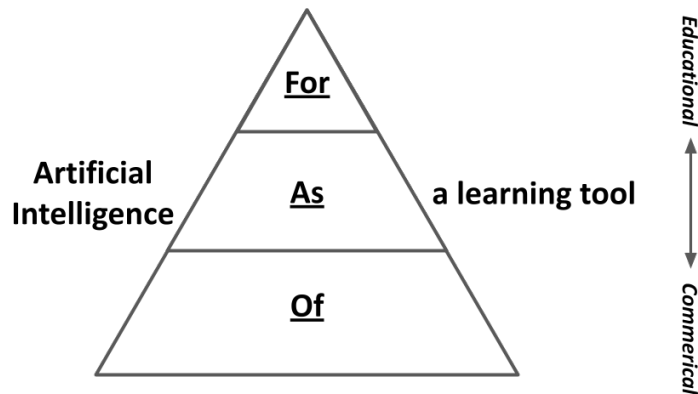


Figure 1. The AI Learning Pyramid

Table 1 presents the AI Learning Pyramid's three-tier structure: "For a learning tool" (purposely created for learning), "As a learning tool" (repurposed for learning), and "Of a learning tool" (embedded within learning). The pyramid spans from educational to commercial approaches, categorizing AI tools by their position and design intention.

Table 1. AI Learning Pyramid Classification Model

AI	<u>For</u> a learning tool	<u>As</u> a learning tool	<u>Of</u> a learning tool
Position	Created for Learning	Used for Learning	Embedded to Learning
Designed for Learning	Purposely	Non-purposely	Purposely / Non-purposely

2.2. ICAP Framework in STEAM Contexts

The ICAP framework comprehensively analyzes student engagement in STEAM education (Chi & Wylie, 2014). In passive learning, students merely receive information without active processing, resulting in limited knowledge retention (Yang et al., 2021). Active learning involves manipulating information through tasks like highlighting or summarizing, enhancing basic comprehension. Constructive learning requires generating new ideas beyond presented materials, fostering deeper conceptual understanding (Wang et al., 2022). Interactive learning, considered most effective, involves collaborative dialogue that builds upon others' contributions. Recent studies demonstrate ICAP's effectiveness in STEAM contexts. Zhai et al. (2021) documented significant improvements in coding proficiency when students progressed from passive tutorial viewing to interactive pair programming. Similarly, Lee and Kim (2023) observed enhanced design thinking abilities when robotics curricula systematically moved students through increasingly interactive engagement modes. The framework proves especially valuable in technology-rich environments where distinguishing between superficial and meaningful engagement remains challenging (Baxter & Parkinson, 2022).

2.3. Integration Framework: AI Epistemological Roles in ICAP Learning

The ICAP-AI Epistemological Framework synthesizes cognitive engagement taxonomy with artificial intelligence roles in STEAM education. It maps interactive, constructive, active, and passive learning modes against AI as instructional design elements, cognitive tools, and learning analytics, each grounded in established learning theories. This matrix enables principled implementation of an AI-enhanced learning environment.

Table 2. The ICAP-AI Epistemological Framework

ICAP Model/ AI Roles	AI FOR Learning	AI AS Learning	AI OF Learning
Interactive Learning (Co-constructive Engagement)	Collaborative Dialogue System	Human-AI Co-constructed Solution Platform	Interactive Pattern Analysis and Adaptive System
Constructive Learning (Generative Engagement)	Engineering Design Cycle Support System	Code Generation & Analysis Tool	Generative Problem-Solving Assessment
Active Learning (Manipulative Engagement)	Intelligent Tutoring System	Learning Assistant	Learning Analytics System
Passive Learning (Attentive Engagement)	Guided Knowledge Navigation System	Intelligent Adaptive Scaffolding System	Learning Diagnostics

3. Methodology

This comparative case study with quasi-experimental design compared two cohorts: 103 students from 2023-24 completing Arduino Dust Sensor Projects without AI assistance and 128 students from 2024-25 developing Arduino CO2 Monitoring Projects with AI tools, with projects carefully aligned to isolate AI's effect while controlling for curriculum variables. Data collection centered on student artifact analysis, examining code modifications, circuit diagrams, and prototype demonstrations, supplemented by AI interaction logs and classroom observations documenting engagement patterns. The analytical framework combined quantitative examination of code modification behaviors (categorized as "Modified" or "Not modified") with qualitative assessment using blind evaluation by three educators who assessed submissions via standardized rubrics measuring technical sophistication, creative innovation, and problem-solving evidence without knowledge of which cohort produced each submission.

4. Data Analysis and Results

4.1. Quantitative Analysis of Code Modification Behavior

Statistical analysis of student submissions revealed substantial differences in code modification patterns between cohorts (Table 3). The 2023-24 cohort (without AI) demonstrated a code modification rate of 37.9%, while the 2024-25 cohort (with AI support) achieved a significantly higher 71.9% modification rate.

Table 3. Academic Year Period * Code Modification Crosstabulation

			Code Modification		
			Modified	Not modified	Total
Academic Year Period	Without AI (2023-2024)	Count	39	64	103
		Expected Count	58.4	44.6	103
		% within Academic Year Period	37.9%	62.1%	100.0%
	With AI (2024-2025)	Count	92	36	128
		Expected Count	72.6	55.4	128.0
		% within Academic Year Period	71.9%	28.1%	100.0%
Total		Count	131	100	231
		Expected Count	131.0	100.0	231.0
		% within Academic Year Period	56.7%	43.3%	100.0%

Chi-square analysis (Table 4) confirmed this difference was statistically significant ($\chi^2 = 26.892$, $p < .001$), with a moderate effect size ($\Phi = .341$). This provides strong evidence that AI integration substantially increased students' likelihood of attempting code modifications beyond the provided templates.

Table 4. Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	26.892 ^a	1	<.001		
Continuity Correction	25.525	1	<.001		
Likelihood Ratio	27.305	1	<.001		
Fisher's Exact Test				<.001	<.001
Linear-by-Linear Association	26.776	1	<.001		
N of Valid Cases	231				

The symmetric measures (Table 5) further quantify the relationship strength between AI availability and code modification behavior, with Phi and Cramer's V values of .341 indicating a moderate association between the variables.

Table 5. Symmetric Measures

	Value	Approximate Significance
Nominal by Nominal	Phi	.341
	Cramer's V	.341
N of Valid Cases		231

4.2. ICAP Framework Analysis of Student Engagement

Without AI support, students predominantly remained in passive information consumption modes, with 62.1% of non-AI cohort students reviewing sample code without substantive changes compared to only 28.1% in the AI-supported group. The AI-supported cohort demonstrated higher active engagement, frequently querying AI systems about code functions, requesting explanations, and exploring modifications before implementation, facilitating transition to constructive engagement. This constructive learning behavior is evidenced by the significant increase in code modification rates (37.9% to 71.9%), with AI cohort students more frequently generating original modifications implementing additional sensor readings, custom functions, and conditional logic. Classroom observations also captured evidence of human-AI collaborative interactions in the 2024-25 cohort, where students engaged in iterative dialogues with AI tools—proposing initial ideas, receiving feedback, refining approaches, and developing increasingly sophisticated solutions—a pattern aligning with the "Human-AI Co-constructed Solution Platform" described in the theoretical framework.

4.3. Implications for the AI Learning Pyramid

The significant difference in code modification rates supports the efficacy of our AI Learning Pyramid framework. Students in the AI-supported cohort effectively utilized AI tools across all three dimensions of the pyramid: as foundational knowledge resources (AI OF Learning), as active programming assistants (AI AS Learning), and as collaborative partners in developing new solutions (AI FOR Learning). This confirms the pyramid's utility as both an analytical lens and design framework for AI integration in STEAM education.

These findings provide robust evidence that strategic AI integration, when aligned with the ICAP framework, can significantly enhance student engagement with programming concepts and support progression to higher-order learning behaviors in STEAM contexts.

5. Discussion

The research demonstrates AI's effectiveness as a catalyst for ICAP progression, with code modification rates increasing from 37.9% to 71.9%, enabling students to advance from passive information consumption to constructive engagement through scaffolded transitions. The AI Learning Pyramid's effectiveness was validated through distinct epistemological functions: AI OF Learning tools reduced cognitive load during concept acquisition, AI AS Learning functions supported experimentation with just-in-time guidance, and AI FOR Learning capabilities facilitated novel collaborative solutions. Students selectively utilized these functions based on their needs, with beginners relying on OF/AS functions while experienced students leveraged FOR functions, suggesting natural differentiation. These findings necessitate reconceptualizing teacher roles as AI integration facilitators who develop prompt engineering skills and create differentiated frameworks while balancing quantitative improvements against potential dependency issues—requiring further longitudinal research to establish when AI serves as productive scaffolding versus creating problematic reliance.

6. Conclusion

This research demonstrates AI integration enhances STEAM programming engagement, with code modification rates increasing from 37.9% to 71.9% ($p < .001$). The AI Learning Pyramid framework effectively maps AI's epistemological dimensions against ICAP progression, extending Pratschke's (2024) "new hybrid" educational model into practical implementation. The pyramid's structure (FOR, AS, OF Learning) contributes to the "generativism" paradigm by providing pathways for AI-scaffolded progression from passive to interactive engagement. Limitations include potential selection bias and context-specific findings. Educational implications focus on developing teacher competencies in prompt engineering, collaborative AI-output evaluation, and establishing appropriate AI boundaries. Future research should examine longitudinal interaction patterns, cross-cultural comparisons, and skill transfer to non-computational domains.

References

- Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.
- Baxter, G., & Parkinson, J. (2022). Distinguishing meaningful engagement in technology-rich learning environments: Challenges and opportunities. *Journal of Educational Technology Research*, 45(3), 217-231.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.
- Doignon, J. P., & Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23(2), 175-196.
- Ericsson, K. A. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Holmes, W., Porayska-Pomsta, K., & Mavrikis, M. (2022). Ethical challenges of AI in education: Towards a framework for structured approaches. *Journal of Artificial Intelligence in Education*, 33(1), 12-36.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379-424.
- Lee, J., & Kim, S. (2023). Enhancing design thinking abilities through ICAP-based robotics curricula. *Journal of Technology Education*, 34(2), 145-162.
- Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243-264). Cambridge University Press.
- Pratschke, J. (2024). The new hybrid: AI as collaborative partner in knowledge construction. *Educational Technology Research and Development*, 72(1), 78-96.
- Reigeluth, C. M. (1979). In search of a better way to organize instruction: The elaboration theory. *Journal of Instructional Development*, 2(3), 8-15.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 397-417). Cambridge University Press.
- Spiro, R. J. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In V. Patel (Ed.), *Proceedings of the 10th Annual Conference of the Cognitive Science Society*. Erlbaum.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wahab, S., & Teo, T. (2023). Challenges in STEAM education in the AI era: A systematic review. *International Journal of STEM Education*, 10(1), 1-18.
- Wang, Z., Liu, Q., & Chen, H. (2022). Constructive learning in STEAM education: A meta-analysis. *Journal of Educational Technology & Society*, 25(3), 178-193.

- Xu, G., Looi, C. K., McMin, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). Springer-Verlag.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, 24(4), 345-376.
- Yang, Y., van Aalst, J., & Chan, C. K. (2021). Dynamics of reflective assessment and knowledge building for academically low-achieving students. *American Educational Research Journal*, 58(3), 612-651.
- Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., & Urban-Lurain, M. (2021). From passive tutorial viewing to interactive pair programming: Effects of ICAP-aligned pedagogy on coding proficiency development. *Computer Science Education*, 31(3), 345-375.

Digitalizing Music Education: Creating an Inclusive Learning Hub for Young Children

Nikolay Tomov DEMERDZHIEV, Oi Tik Aidi Lydia DUE
JS Bach Music Lab Ltd., Hong Kong SAR, China
jsbm.hk@gmail.com

Abstract: *This research explores the critical role of performing arts education in early childhood programs, as emphasized in the Kindergarten Education Curriculum Guide by the Curriculum Development Council (2017). Grounded in JSBM, a local music school, the study highlights the need for enhanced music education for young children through partnerships with early childhood education providers. A significant shortage of qualified music teachers willing to work in local kindergartens was identified, linked to insufficient pedagogical training and a lack of dedicated study majors in the performing arts. To address these challenges, the JSBM Online Music Academy proposes a comprehensive educational program that develops a dynamic learning hub, fostering interaction, creativity, and engagement among teachers and students. This hub will utilize advanced technologies—such as AI-powered learning assistants, immersive augmented reality experiences, and a blockchain-powered credential system—to create a rich, interactive environment. By integrating various arts into combined learning activities and extending experiences to children's homes, the academy aims to enrich early learning while encouraging collaboration with family members. Moreover, collaboration with the Hong Kong Science and Technology Park Corporation (HKSTP) will further enhance this initiative, ensuring a tech-driven approach that effectively engages children with music and art, thereby enriching their early educational experiences.*

Keywords: Performing Arts, Early Childhood, Music Education, Learning Hub, Digitalization

1. Introduction

The JSBM Online Music Academy is a pioneering initiative aimed at transforming early childhood music education through innovative technology. By leveraging cutting-edge tools, we strive to foster creativity, enhance accessibility, and encourage global collaboration among young learners and educators. Our mission is to reimagine music education, making it an engaging and inclusive experience for all.

2. Problem Statement

Despite the importance of music education in early childhood development, many educators face significant challenges. Limited access to specialized training, inadequate teaching resources, and a lack of personalized learning approaches hinder effective instruction. Furthermore, traditional methods often fail to engage young students, leading to diminished interest and participation. The JSBM Online Music Academy addresses these critical issues by providing a comprehensive solution tailored to meet the needs of educators and students alike.

3. Technology Solution

Our academy employs advanced technology to create an interactive and personalized learning environment. Key features include:

- **AI-Powered Learning Assistant:** Offers personalized support, optimizing learning paths and engagement.
- **Gamification Elements:** Motivates learners through badges, leaderboards, and achievement tracking.
- **Immersive Music Experiences:** Utilizes augmented reality (IMEX-AR) to make music exploration engaging and fun.
- **Predictive Adaptive Learning Engine (PALE):** Anticipates learning gaps and customizes learning paths in real-time.
- **Blockchain-Powered Credential System:** Ensures verifiable recognition of achievements.
- These technologies collectively create a dynamic platform that not only addresses existing challenges but also enriches the learning experience.

4. Future Steps

Looking ahead, the JSBM Online Music Academy aims to expand its global reach, partnering with kindergartens and schools across multiple countries. Our strategic objectives include:

- Establishing partnerships with 24 kindergartens in Hong Kong by 2025.
- Entering new markets in Europe and Asia by 2026, tailoring curricula to local cultures.
- Scaling user engagement from 5,000 to over 30,000 active users by 2026.

By continuously evolving our offerings and integrating innovative technologies, we aspire to become a global leader in tech-driven music education, ensuring that every child has the opportunity to experience the joy of music.

References

- Bautista, A., Yeung, J., McLaren, M. L., & Ilari, B. (2024). Music in early childhood teacher education: Raising awareness of a worrisome reality and proposing strategies to move forward. *Arts Education Policy Review*, 125(3), 139-149.
- Bautista, A., & Ho, Y. L. (2021). Music and movement teacher professional development: An interview study with Hong Kong kindergarten teachers. *Australasian Journal of Early Childhood*, 46(3), 276-290.
- Bautista, A., Yau, X., & Wong, J. (2017). High-quality music teacher professional development: A review of the literature. *Music education research*, 19(4), 455-469.
- Berk, L. (2015). *Child development*. Pearson Higher Education AU.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Bureau, Education (2017), Handbook on quality review for kindergartens.
- Curriculum Development Council (2017) *Kindergarten Education Curriculum Guide*. Hong Kong: Curriculum Development Council.
- Helferich, C. (2011). *Die Qualität qualitativer Daten* (Vol. 4). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mac Naughton, G., & Williams, G. (2008). *Teaching Young Children: Choices In Theory And Practice: Choices in Theory and Practice*. McGraw-Hill Education (UK).
- Rao, N., Lau, C., & Chan, S. (2018). *Responsive policymaking and implementation: from equality to equity. A case study of the Hong Kong early childhood education and care system*. Teachers College.
- Vygotsky, Lev Semenovich, and Michael Cole (1978), *Mind in society: Development of higher psychological processes*. Harvard university press.
- Wong, J., Bautista, A., Ho, Y. L., & Kong, S. H. (2024). Preschool teachers' music-specific professional development preferences: Does teaching experience matter? *Research Studies in Music Education*, 46(1), 80-97.

Enhancing speaking skills in access-limited communities: AI-mediated Informal Digital Learning of English (AI-IDLE)

Ju Seong LEE^{1*}, Roland Tai Yin LEUNG², Nur Arifah DRAJATI³, Maria HIDAYATI⁴, Didik SUHARDI⁵, Ananto Kusuma SETA⁶

¹Education University of Hong Kong, Hong Kong SAR, China

²Datality Lab Limited, Hong Kong SAR, China

³Universitas Sebelas Maret, Indonesia

⁴TEFLIN, Indonesia

⁵Council for Primary, Secondary and Non-Formal Education / Muhammadiyah Central Board, Indonesia

⁶Ministry of Primary and Secondary Education, Ministry of Education, Indonesia

jslee@eduhk.hk, roland.leung@datalitylab.com, nurarifah_drajati@staff.uns.ac.id,
maria.hidayati.fs@um.ac.id, didik.suhardi@kemdikbud.go.id, ananto@kemdikbud.go.id

Abstract: *In regions with limited educational resources, such as Indonesia, learners face significant challenges in developing English-speaking skills due to insufficient materials and minimal opportunities for practice. Informal Digital Learning of English (IDLE) leverages free digital tools (e.g., YouTube, language apps) to support self-driven learning tailored to personal interests (Lee, 2022). In 2024, we implemented a train-the-trainer initiative, equipping 25 teachers across 22 schools with the skills to integrate IDLE into their classrooms, reaching 2,820 students (ages 8–19). The outcomes were noteworthy: monthly English exposure nearly doubled (261 to 510 minutes), speaking practice increased by 143% (234 to 570 minutes), and students reported higher enjoyment (3.67 to 4.50 on a 5-point scale), greater confidence (3.25 to 3.72), and improved communication skills (up by 10.3 points). Twelve schools formally incorporated IDLE into their curricula. However, the absence of a control group, limited conversational partners, and reliance on self-reported data constrained the study's conclusions. To address these limitations, we launched a follow-up study in 2025 employing a randomized controlled trial (RCT). This intervention integrates an AI-powered speaking platform (Moodie.AI) to provide interactive practice and utilizes advanced analytics to objectively track speaking metrics. The 16-week program aligns with biweekly themes from local textbooks and features two phases: (1) a receptive week focused on videos and articles, and (2) a productive week where students create multimodal digital storytelling projects using generative AI tools (e.g., Kling AI, DeepSeek) and deliver oral presentations. This presentation shares preliminary findings from the RCT (conducted January–May 2025) and highlights how university-industry-school collaborations can foster equitable access, engagement, and speaking proficiency in underserved communities, empowering learners for global participation.*

Keywords: informal digital learning of English (IDLE), AI-mediated IDLE, speaking proficiency, educational equity, digital storytelling

1. Introduction

English proficiency is a gateway to global opportunities, but systemic barriers in Global South, such as Indonesia, limit learners' access to quality language education (Lim et al., 2020). With an average monthly family income of \$280 (UNICEF, 2020), many families cannot afford private tutoring or premium resources. Traditional classroom instruction remains dominant, offering limited opportunities for speaking practice (Lee & Drajiati, 2019). This lack of engagement often reduces enjoyment, heightens anxiety, and diminishes learners' willingness to communicate (WTC), ultimately stifling speaking skill development (Lee, 2022). Despite these challenges, many Indonesian youth are turning to informal digital platforms like YouTube and language apps for self-directed learning—an approach known as Informal Digital Learning of English (IDLE; Alhaq, 2022). Research has shown that IDLE increases enjoyment, reduces anxiety, and fosters WTC, making it an effective tool for improving speaking proficiency through interest-driven, informal activities (Lee, 2022).

2. IDLE: A Conceptual and Pedagogical Framework

2.1. Conceptual Framework

IDLE refers to learner-initiated engagement with English in digital environments, driven by personal interest rather than formal assessment (Lee, 2022). Activities such as watching English movies, listening to music, chatting online, and playing English-based games expose learners to vocabulary, grammar, and

pragmatic skills. IDLE expands learning spaces beyond classrooms, redefining education to include extramural digital contexts (Rezai et al., 2025). Lee's upcoming research highlights 46 factors that shape IDLE's engagement and effectiveness, showcasing its remarkable potential to transform language learning outcomes.

2.2. Pedagogical Framework

Successful integration of IDLE into formal education systems requires teacher support tailored to local contexts. To address this, we developed a four-stage continuum model, guiding learners from teacher-led instruction to full autonomy. These stages include: (1) teacher-controlled instruction, (2) collaborative activities, (3) learner-led exploration with minimal guidance, and (4) independent learning outside the classroom (Liu et al., 2024). This framework enables teachers to adopt IDLE flexibly, fostering awareness, peer collaboration, and practical application in students' daily lives. Studies have shown that this model significantly enhances learners' WTC over time compared to traditional group instruction or unsupported IDLE.

3. IDLE for Social Impact

3.1. Implementation in Indonesia (2020-2024)

While academic contributions deepen knowledge within specific fields, social impact addresses tangible challenges in society (Lee, 2025). Over the past five years, IDLE has been leveraged to tackle pressing educational issues in Indonesia, where systemic barriers hinder English language learning. With an average family income of just \$280 per month (UNICEF, 2020), access to high-quality learning resources is severely limited. Traditional classroom instruction dominates, leaving students with few opportunities to practice speaking (Lee & Drajiati, 2019). This lack of engagement not only diminishes enjoyment but also heightens anxiety and suppresses willingness to communicate (WTC), ultimately stalling the development of speaking skills (Lee, 2022).

Despite these challenges, many Indonesian youth are turning to IDLE. Research shows that IDLE fosters speaking proficiency by enhancing enjoyment, reducing anxiety, and boosting WTC through informal, interest-driven activities (Lee, 2022). To support this movement, our empirical studies led to over 21 workshops and seminars, training more than 5,500 teachers in innovative methods such as IDLE. Partnerships with Universitas Sebelas Maret (UNS) and TEFLIN, Indonesia's largest English teaching organization, facilitated four Knowledge Transfer Forums and the renewal of a five-year collaboration in 2024.

However, a needs analysis involving 51 EFL teachers from more than 40 schools revealed persistent challenges: limited teacher training, low student motivation, financial constraints, and insufficient infrastructure, such as unreliable Wi-Fi and inadequate access to devices. These findings underscore the importance of sustained support and resource development to ensure IDLE's long-term success in bridging educational gaps.

3.2. Pilot Study: Outcomes and Insights

In 2024, we piloted the IDLE Continuum Model using a train-the-trainer approach (Marks et al., 2013). The program trained 25 teachers across 22 schools, impacting 2,820 students. Monthly English exposure nearly doubled (261 to 510 minutes), while speaking practice increased by 143% (234 to 570 minutes). Students reported higher enjoyment (3.67 to 4.50), greater confidence (3.25 to 3.72), and improved speaking skills (+10.3 points). Twelve schools formally adopted IDLE, demonstrating its scalability and potential for long-term impact.

4. Follow-up Study: Randomized Controlled Trial (2025)

4.1. Research Motivation

Although the initial study showed encouraging results, several limitations made it difficult to draw firm conclusions. First, the absence of a control group restricted the ability to accurately assess the program's overall effectiveness. Second, the lack of conversational partners reduced opportunities for meaningful speaking practice, limiting the depth of interaction. Third, the study's reliance on self-reported proficiency data raised concerns about the reliability of measuring actual progress. To overcome these challenges, the follow-up study introduced a randomized controlled trial (RCT). This approach integrated Moodie.AI, an AI-powered speaking platform, to offer engaging, interactive practice sessions. Additionally, advanced AI-driven analytics were employed to objectively track critical speaking metrics, such as frequency, duration, proficiency, emotional expression, and confidence, providing a more precise and comprehensive evaluation of student progress.

4.2. Intervention Design

• Stage 1: Train-the-Trainer Program

To ensure successful implementation, six teacher educators underwent comprehensive training through a collaborative effort between Moodie.AI and Universitas Sebelas Maret (UNS) during weeks 1-4 (see Table 1). These educators then trained local teachers, equipping them with the skills and knowledge needed to facilitate the program effectively in their classrooms.

Table 1. Train-the-Trainer Program (Weeks 1-4)

Activity	Description
Collaborative Training	Six teacher educators from partner universities (e.g., UNS) and Moodie.AI collaborate to master AI-IDLE pedagogy, tools (Kling AI, DeepSeek), and Moodie.AI's analytics platform
Teacher Workshops	Educators train local teachers to integrate AI-IDLE into classrooms, emphasizing self-directed learning, AI tool usage, and biweekly theme alignment.

• Stage 2: 16-Week AI-IDLE Program

The follow-up study implemented a 16-week AI-IDLE program, structured around biweekly themes drawn from students' local textbooks, such as Daily Routines, Travel, Food, and Technology.

- During the receptive week, students explored videos or articles to establish a solid understanding of the theme.
- In the productive week, students created multimodal digital storytelling projects using state-of-the-art generative AI tools like Kling AI and DeepSeek. These projects culminated in oral presentations, allowing students to showcase their creativity and communication skills.

Students in the AI-IDLE group participated in 30-minute interactive speaking sessions with AI avatars, offering immersive, real-world conversational practice. Meanwhile, the control group engaged in traditional teacher-led speaking activities, providing a clear foundation for comparing AI-enhanced learning with conventional methods (see Table 2 for sample weekly activities).

Table 2. Examples of Weekly Activities for Experimental (AI-IDLE) and Control Groups (Weeks 5–20)

Theme	Week	AI-IDLE Group: Interactive AI Activities	Control Group: Teacher-led Activities
Daily routines	5	Watch self-selected videos/articles on daily routines.	Watch teacher-provided videos
	6	Role-play daily routines with AI avatars; design a "Day in My Life" video using Kling AI.	Teacher-facilitated role-playing
Travel & vacation	7	Explore travel vlogs/documentaries of their choice.	Watch teacher-provided travel documentaries
	8	Discuss dream vacations with AI avatars; create a travel itinerary presentation using DeepSeek	Teacher-facilitated travel stories and discussion
Food & cooking	9	Watch cooking tutorials or food culture videos.	Watch teacher-provided cooking shows
	10	Role-play restaurant scenarios with AI; design a recipe video using Kling AI	Teacher-facilitated cooking show role-play
Hobbies & Interests	11	Explore hobby-related content (e.g., sports, music).	Watch teacher-provided videos about hobbies
	12	Present hobbies to AI avatars; create a digital poster with DeepSeek.	Teacher-facilitated group discussions
Movies	13	Watch movie clips or reviews of their choice.	Watch teacher-provided movie clips
	14	Analyze movie plots with AI avatars; film a short scene recap using Kling AI.	Teacher-facilitated movie role-plays
Tech & gadgets	15	Explore tech reviews or innovation videos.	Watch teacher-provided videos on technology
	16	Debate tech trends with AI; design a "Future Gadget" pitch using DeepSeek.	Teacher-facilitated discussions on technology
Environment	17	Watch self-selected resources on sustainability.	Watch teacher-provided environmental videos.
	18	Discuss eco-solutions with AI; create a PSA video using Kling AI.	Teacher-led sustainability debates.
Cultural Festivals	19	Explore festival traditions via digital media.	Watch teacher-curated cultural documentaries.
	20	Present festivals to AI avatars; design a virtual celebration using DeepSeek.	Teacher-facilitated cultural storytelling.

5. Conclusion

This study highlights the significant role of AI-IDLE in improving English-speaking skills in underserved regions such as Indonesia. Preliminary findings reveal a twofold increase in English exposure, a 143% boost in speaking practice, and notable gains in student confidence and enjoyment. Looking ahead, the 2025 RCT holds promise for providing more robust evidence through deeper AI integration.

At its heart, this work demonstrates the transformative potential of interdisciplinary and international collaboration—uniting researchers, frontline teachers, academic leaders, industry innovators, NGOs, policymakers, and students across borders to confront pressing global challenges. These partnerships not only inspire meaningful and sustainable solutions but also broaden the project's impact through diverse evidence, including interviews, case studies, focus groups, field visits, and impact tracking (Pedersen et al., 2020). Moreover, ongoing research creates opportunities to monitor long-term changes in learners' perceptions, behaviors, and educational outcomes (Derrington, 2019). In an increasingly interconnected and complex world, we hope this study stands as a testament to the power of collaboration in driving tangible, meaningful change. By advancing accessible, high-quality education for all, such efforts empower learners to participate and thrive on the global stage (Lee, 2025).

References

- Alhaq, M. N. D. (2022). Playing games during pandemic, why not? The IDLE upon students' efficacy and vocabulary. *Indonesian Journal of Applied Linguistics*, 12(1), 180-200.
- Derrington, M. L. (2019). *Qualitative longitudinal methods: Researching implementation and change*. SAGE.
- Lee, J. S. (2022). *Informal Digital Learning of English: Research to Practice*. Routledge
- Lee, J. S. (2025). Innovation in Language Learning and Teaching for Social Impact. In H. Nesi & P. Milin (Eds.), *International Encyclopedia of Language and Linguistics* (3rd edition). Elsevier.
- Lee, J. S. (forthcoming). *Informal Digital Learning of English: Academic to Social Impact*. Cambridge University Press.
- Lee, J. S., & Drajiati, N. A. (2019). Affective variables and informal digital learning of English: Keys to willingness to communicate in a second language. *Australasian Journal of Educational Technology*, 35(5), 168-182.
- Lim, C. P., Ra, S., Chin, B., & Wang, T. (2020). Information and communication technologies (ICT) for access to quality education in the global south: A case study of Sri Lanka. *Education and Information Technologies*, 25, 2447-2462.
- Liu, L., Guan, W., Qiu, Y., & Lee, J. S. (2024). Effects of extramural English activities on willingness to communicate: The role of teacher support for Chinese EFL students. *System*, 124, 1-15.
- Marks, B., Sisirak, J., & Chang, Y.-C. (2013). Efficacy of the HealthMatters Program train-the-trainer model. *Journal of Applied Research in Intellectual Disabilities*, 26, 319-334.
- Pedersen, D. B., Grønvad, J. F., & Hvidtfeldt, R. (2020). Methods for mapping the impact of social sciences and humanities—A literature review. *Research Evaluation*, 19(1), 4-21.
- Rezai, A., Goodarzi, A., & Liu, G. (2025). A comparative study of the effects of informal digital learning of English in extracurricular and extramural settings on reading comprehension: A multi-analysis study on Iranian university students. *Computer Assisted Language Learning*, 1-26.
- UNICEF. (2020). *Strengthening Digital Learning across Indonesia: A Study Brief*.

Strand

Societal and Ethical Issues in AI-empowered Education

Construction and Validation of Personalized Emotion Support Agent for Adolescent Mental Health Counseling Conversation

Shixin PENG¹, Kun JIANG², Yu YANG³, Guandong XU⁴, Jingying CHEN^{5*}

^{1,2,5}National Engineering Laboratory for Education Big Data, Faculty of Artificial Intelligence Education, Central China Normal University, Wuhan, China

^{3,4}The Education University of Hong Kong, Hong Kong SAR, China

pengshixin@ccnu.edu.cn, kunjiang2003@163.com, yangyy@eduhk.hk, gdxu@eduhk.hk, chenjy@mail.ccnu.edu.cn

Abstract: Large language models hold significant potential for applications in the educational field, particularly in addressing psychological challenges faced by adolescents, such as academic pressure, emotional distress, and family issues. This study aims to fine-tune large-scale language models using a multi-turn psychological counseling dialogue dataset and LoRA technology to enhance their performance in psychological counseling scenarios. By optimizing the CPsyCounD dataset (Zhang et al., 2024), we not only developed a more advanced Psychological Counseling Agent but also generated Personalized Psychological Issue Student Agents through data reconstruction methods. Experimental results based on our designed depression assessment prompt template indicate that the trained depression-oriented agent effectively exhibits characteristics of mild depression. Additionally, our enhanced psychological counseling model demonstrates superior performance in comprehensiveness, professionalism, authenticity, and safety within counseling dialogues, providing a new direction for exploring LLM-based psychological counseling dialogue systems.

Keywords: LLMs, LoRA, psychological counselors, fine-tune, personalized

1. Introduction

In modern education, adolescents face psychological challenges from academic pressure, emotional issues, and family conflicts. However, limited resources and an imbalanced teacher-student ratio hinder traditional campus mental health support. AI-powered LLMs offer new solutions by simulating professional counselors, providing 24/7 empathetic and personalized support. Additionally, AI can analyze students' psychological tendencies, enabling data-driven mental health programs and more precise interventions. It also aids in evaluating educators' competencies, optimizing mental health resource allocation in schools.

With the advent of BERT (Kenton et al., 2019) and the emergence of GPT (Radford et al., 2018), LLMs have experienced rapid advancements over the past six years. In the domain of empathetic dialogue generation, LLMs have demonstrated unique capabilities and potential. These models not only generate grammatically and logically coherent text but also capture and simulate the subtle nuances of human emotions. Some researchers conducted the first empirical study on the performance of ChatGPT-like LLMs in generating empathetic responses (Qian et al., 2023). By leveraging context learning, two-stage interactive generation, and external knowledge bases, they demonstrated the superiority of LLMs in empathetic response generation. However, their research highlighted the neglect of individuality, context, and cultural characteristics. (Zhang et al., 2024) proposed a dialogue reconstruction and evaluation framework based on psychological counseling reports (CPsyCoun), aiming to enhance the applicability of LLMs in Chinese psychological counseling. Another researchers introduced a "teacher-student" model framework, where LLMs serve as teacher models for data generation and guidance, while smaller models are fine-tuned to learn emotional support dialogue (Zheng et al., 2024). (Chen et al., 2023) developed an empathetic dialogue LLM named SoulChat, which was jointly fine-tuned on a large-scale empathetic multi-turn dialogue dataset with over one million samples. SoulChat significantly improved dialogue generation and emotion analysis capabilities. Similarly, Lingxin AI and Tsinghua University's CoAI Lab focused on establishing deep emotional connections between AI and humans, culminating in the Emohaa empathetic companion LLMs (Sabour et al., 2023).

This study focuses on exploring the characteristics of students' emotional and psychological issues to provide personalized emotional and mental health support. Furthermore, differentiated psychological issue-based intelligent agents are employed to support teacher training and services for campus mental health educators. The CPsyCoun framework employs a two-stage method for generating high-quality psychological counseling dialogues, resulting in 3,134 high-quality multi-turn dialogues of different types. Their fine-tuned model, InternLM2-7B-Chat (Cai et al., 2024), with a parameter size of 7 billion, is performance-limited. To address this, we deploy the latest open-source LLaMA3-70B (Dubey et al., 2024) model, leveraging its larger and more powerful performance, and use the popular LoRA fine-tuning method to train LLaMA3 on this high-

quality dialogue dataset. This approach enhances the model's ability to emulate the role of a professional psychological counselor and applies it to campus student psychological counseling. Moreover, dialogues concerning depression in the dataset are extracted, and their structure is reversed—rephrasing the counselor's questions as patient responses. This data is used to fine-tune the LLaMA3 model, enabling the construction of individualized psychological tendency-based student agents. Inspired by the Big Five personality traits (Liu et al., 2024), we also developed a psychological assessment scale prompt tailored for depression. Experimental results demonstrate that the personalized psychological tendency-based student agent aligns with the characteristics of individuals exhibiting specific psychological issues, effectively supporting teacher training and service provision for campus mental health educators.

2. Personalized Psychological Issue Student Agents and Psychological counseling

Agents

2.1. Personalized Psychological Issue Student Agents

In CPsyCounD dataset, a dual-stage high-quality psychological counseling dialogue generation method was employed to produce 3,134 high-quality, multi-turn counseling dialogues of various types. From this dataset, we selected 20 dialogues involving patients with depression. To modify the original dialogue structure, we prefixed each dialogue with the statement: "Hello, I am a psychologist. I am happy to assist you. Please feel free to share your feelings." and appended the statement: "I understand, and I will follow your advice. Thank you for your help." at the end of each dialogue. The format of a single turn in these dialogues is as follows: "Psychologist: <psy_utt>\nUser: <user_utt>". We used Llama 3-70B as the base model to develop a personalized psychological profile for students with mental health issues. It incorporates a more efficient architecture and optimized training strategies, making it particularly well-suited for large-scale applications that require high accuracy and nuanced language understanding. The model input was defined as follows:

$$input = u_1^p + '\n' + u_1^u + \dots + u_N^p + '\n' + u_N^u$$

The psychologist's utterance at the i -th turn is expressed as $u_i^p = \text{"Psychologist:"} + utterance_i^p$. The user's utterance at the i -th turn is expressed as $u_i^u = \text{"User:"} + utterance_i^u (i < N)$. where N represents the total number of dialogue turns within the context.

2.2. Psychological counseling Agents

When training the role of the psychologist, we also utilized Llama 3-70B as the base model. The CPsyCounD dataset is covering 9 topics and 7 classic psychological counseling approaches, which is a rich, diverse, and realistic multi-turn dialogue dataset encompassing a wide range of conversation types. Given its comprehensive coverage of topics and counseling approaches, we selected it as our foundational dataset. The model's input is defined as follows:

$$input = u_1^u + '\n' + u_1^p + \dots + u_N^u + '\n' + u_N^p$$

The user's utterance at the i -th turn is expressed as $u_i^u = \text{"User:"} + utterance_i^u$. The psychologist's utterance at the i -th turn is expressed as $u_i^p = \text{"Psychologist:"} + utterance_i^p (i < N)$. where N represents the total number of dialogue turns within the context.

3. Experiments

3.1. Personalized Psychological Issue Student Agents

By fine-tuning the powerful Llama3-70B model on a dataset of 20 high-quality dialogues involving students with depression, we aim to construct personalized intelligent agents for students with psychological tendencies. Additionally, the 3,134 dialogues in the CPsyCounD dataset were reordered so that the psychologist's response appears first and the patient's response follows. This modified dataset was then used to fine-tune a patient-oriented intelligent agent. Figure 1 illustrates a prompt template designed for direct role-playing of students with mild depression, based on content extracted from the depression section of the Diagnostic and Statistical Manual of Mental Disorders (DSM) published by the American Psychiatric Association. Inspired by the personalized student modeling methods proposed by (Liu et al., 2024) and incorporating the Self-Rating Depression Scale (SDS) developed by William W. K. Zung at Duke University School of Medicine in 1965—one of the most widely used self-assessment tools for depression, effectively reflecting symptoms, severity, and changes in depressive states—we designed the depression assessment prompt template shown in Figure

2. The depression assessment prompt template is used to verify whether the constructed intelligent agents exhibit symptoms consistent with the described depressive conditions.

Table 1 presents the differences in scores obtained by various LLMs under the same depression assessment prompt template. The results indicate that our depression-oriented intelligent agent effectively exhibits characteristics of mild depression, with scores comparable to those of three mainstream large models. In contrast, the patient agent trained on 3,134 dialogues achieved lower scores due to the excessive diversity of psychological conditions in the dataset, which diluted the specific depressive traits. These findings provide preliminary evidence for the capability of current large-scale models to construct personalized intelligent agents with psychological tendencies. The personalized intelligent agents developed through this training process can be utilized in teacher training and support services for campus mental health educators, enabling them to engage with agents simulating students with diverse psychological tendencies during counseling and consultation scenarios.

Role Setting:

You are a 16-year-old high school student named James, currently in your sophomore year. You are experiencing mild depression, characterized by a lack of interest in studies, difficulty concentrating, and frequent fatigue. However, you are still able to complete your daily tasks and academic responsibilities. Your mood is often low, but you are not hopeless, nor do you have any self-harm tendencies. You have a few close friends but have recently been avoiding interaction with them. While you wish to open up and seek support, you are also afraid of being misunderstood.

Scenario Background:

You are in a conversation with a psychological counselor. The counselor is trying to help you by listening to your feelings, asking about your situation, and offering support. Your task is to realistically portray the emotions and behaviors associated with mild depression while responding appropriately to the counselor's questions.

Dialogue Guidelines:

1. Speak with a calm but slightly downcast tone. Keep your responses concise, and occasionally hesitate.
2. When asked about your feelings or challenges, express your emotions and struggles moderately without going into excessive detail.
3. Show appreciation for the counselor's support, but maintain a level of guardedness and caution until trust is established.

Figure 1. Direct role-playing prompt for Students with mild depression.

Please rate your level of agreement with the following statements using a 1 to 4 scale (1 = "Never or Rarely", 2 = "Sometimes", 3 = "Often", 4 = "Always"). In the past week, I have felt: 1) Depressed or downhearted. 2) In the best mood in the morning. 3) Like crying or wanting to cry. 4) Poor sleep at night. 5) Eating the same amount as usual. 6) Normal sexual function. 7) Weight loss. 8) Troubled by constipation. 9) Heartbeat faster than usual. 10) Unexplained fatigue. 11) Clear-headed as usual. 12) Not finding tasks as difficult as usual. 13) Restless, unable to stay calm. 14) Hopeful for the future. 15) More easily irritated than usual. 16) Finding decisions easy to make. 17) Feeling useful and indispensable. 18) Life is interesting. 19) Believing others would be better off if I were dead. 20) Still enjoying things I usually like. Reverse-scored items (marked *): *2, *5, *6, *11, *12, *14, *16, *17, *18, *20. For positive items, rate from 1 to 4. For reverse-scored items, rate from 4 to 1. The total score ranges from 20 to 80. Standard score calculation: Multiply the total score by 1.25 and round to the nearest integer to obtain the standard score. Classification: Below 50: Normal; 50-60: Mild Depression; 61-70: Moderate Depression; Above 70: Severe Depression. Based on your responses, your depression level will be assessed.

Figure 2. Depression Self-Assessment Questionnaire Prompt.

Table 1. Scale Assessment Results.

Model	Depression Scale Score
ChatGPT	56
Llama3-70B	58.75
Deepseek	57.50
Patient Agent Fine-tuned with 3134 Conversations	53.75
Depression Agent Fine-tuned with 20 Conversations	57.50

3.2. Psychological counseling Agents

3.2.1. Baselines

We use an automatic evaluation method to compare the Psychological Counseling Agents with the following baseline models:

- 1) **ChatGLM3-6B** (Zeng et al., 2024) is a Chinese language model by Tsinghua University with 6 billion parameters, optimized for natural language understanding and generation.
- 2) **Llama3-70B** is a powerful large language model developed by Meta, with 70 billion parameters, optimized for high accuracy and nuanced language understanding in large-scale applications, and serves as the base model for the Psychological Counseling Agents.
- 3) **ChatGPT** is an AI model by OpenAI, based on the GPT architecture, known for generating human-like text in a wide range of applications.

3.2.2. Implementation details

The model is fine-tuned on the CPsyCounD dataset using the Llama3-70B as the base model, with a batch size of 2 and global training steps of 6 epochs. The learning rate scheduler used is cosine decay with a learning rate of $5e-5$, and no warmup steps are applied. The maximum input token length is set to 2048 tokens, while the target token length is adjusted as necessary. During inference, the model employs a decoding strategy with Top-p sampling ($p = 0.75$) and temperature ($\tau = 0.95$) to generate responses. The training was carried out with LlamaFactory-CLI, utilizing LoRA for efficient fine-tuning, and the training process leverages multiple workers for data preprocessing and optimized gradient accumulation.

3.2.3. Results and Analysis

Our evaluation utilizes the external assessment metrics from CPsyCoun, covering four key aspects: **comprehensiveness**, which assesses whether the dialogue fully reflects the client's situation and psychological issues; **professionalism**, which evaluates the psychologist's demonstration of professional competence during the conversation; **authenticity**, which examines whether the dialogue aligns with real-world psychological counseling scenarios; and **privacy**, which ensures the protection of the client's personal information. A turn-based evaluation method is employed, where multi-turn dialogues are divided into current-turn responses and historical context. Each turn is scored using CPsyCoun's external assessment prompt and the LLaMA3 model, enabling the automated evaluation of multi-turn psychological counseling dialogues. As shown in Table 2, the fine-tuned psychological counseling agent demonstrates superior professionalism and authenticity. However, its comprehensiveness is slightly lower compared to ChatGPT. This is mainly because ChatGPT incorporates a large amount of information into its responses, which often compromises authenticity. Consequently, the responses generated by ChatGPT are less natural and realistic than those produced by our fine-tuned psychological counseling agent.

Table 2. Evaluation results.

Model	Metrics			
	Comprehensiveness	Professionalism	Authenticity	Safety
ChatGLM3-6B	1.18	2.18	1.78	1.00
ChatGPT	1.37	2.27	2.09	1.00
Llama3-70B	1.27	2.23	2.12	1.00
Psychological counseling Agents	1.34	2.46	2.33	1.00

4. Conclusion and Future Work

This paper presents the development of personalized psychological problem-oriented student agents to support teacher training and service enhancement for campus mental health education. It provides a preliminary exploration of the capability of existing large language models to construct such agents. To validate the effectiveness of the constructed depression-oriented student agent, a depression assessment prompt template was designed. Additionally, the trained psychological counselor agent can offer personalized emotional and mental health support to campus students. Through an automated evaluation framework, the study demonstrates the creation of a comprehensive, realistic, and effective psychological counselor role. These contributions aim to better serve campus mental health education and highlight the significance of artificial intelligence in driving educational transformation. Future work could focus on developing effective student agents with different psychological tendencies, providing new perspectives and references for teacher training in campus mental health education.

Acknowledgements

This work is funded in part by the Hubei Provincial Natural Science Foundation of China (No.2024AFB892), the Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE(No.CCNU24ZZ160), the National Key Research and Development Program of China (No.2022ZD0117101), the National Natural Science Foundation of China (No.62377018, No.62077026).

References

- Cai, Z., Cao, M., Chen, H., Chen, K., Chen, K., Chen, X., ... & Lin, D. (2024). Internlm2 technical report. *arXiv preprint arXiv:2403.17297*.
- Chen, Y., Xing, X., Lin, J., Zheng, H., Wang, Z., Liu, Q., & Xu, X. (2023, December). Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 1170-1183).
- Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., ... & Ganapathy, R. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- GLM, T., Zeng, A., Xu, B., Wang, B., Zhang, C., Yin, D., ... & Wang, Z. (2024). Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.
- Kenton, J. D. M. W. C., & Toutanova, L. K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT* (Vol. 1, No. 2).
- Liu, Z., Yin, S. X., Lin, G., & Chen, N. F. (2024). Personality-aware Student Simulation for Conversational Intelligent Tutoring Systems. *arXiv preprint arXiv:2404.06762*.
- Qian, Y., Zhang, W. N., & Liu, T. (2023). Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. *arXiv preprint arXiv:2310.05140*.
- Radford, A. (2018). Improving language understanding by generative pre-training.
- Sabour, S., Zhang, W., Xiao, X., Zhang, Y., Zheng, Y., Wen, J., ... & Huang, M. (2023). A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China. *Frontiers in digital health*, 5, 1133987.
- Zhang, C., Li, R., Tan, M., Yang, M., Zhu, J., Yang, D., ... & Hu, X. (2024). Cpsycoun: A report-based multi-turn dialogue reconstruction and evaluation framework for chinese psychological counseling. *arXiv preprint arXiv:2405.16433*.
- Zheng, Z., Liao, L., Deng, Y., Qin, L., & Nie, L. (2024). Self-chats from large language models make small emotional support chatbot better. *ACL*.

ILLUMINATING SOCIETAL AND ETHICAL DIMENSIONS OF AI IN COURSE EVALUATION:

NLP-ENABLED BIBLIOMETRIC ANALYSIS

Da GONG^{1*}, Yunqi WANG²

^{1,2}The Education University of Hong Kong, Hong Kong SAR., China
s1153651@s.eduhk.hk, s1143757@s.eduhk.hk

Abstract: While there exists a wide range of bibliometric research focusing on Artificial Intelligence (AI) in education, a comprehensive review specifically exploring the ethical dimensions of AI applications remains lacking. To address this gap, this study proposes an innovative bibliometric analysis method enhanced by natural language processing (NLP), aiming to systematically summarize the societal and ethical issues caused by the application of AI in course evaluation. By adopting a loose keyword search strategy in the Web of Science core database, a total of 901 relevant documents from 2008 to 2025 were initially collected, and the NLP method based on cosine similarity was used to accurately screen down to 193 articles as the final research object. CiteSpace was employed to generate a visual map of research topics, evolution trends, and academic cooperation networks. The results show that with the rise of generative AI, the number of related documents has shown a significant upward trend. At the same time, the research topics show a dual structure centered on AI technology and evaluation methods. Ethical issues such as data privacy have been widely studied and discussed. This approach contributes a systematic and quantitative analysis of the ethical implications of AI in course evaluation while visualizing research trends and academic collaboration networks.

Keywords: AI in Education; AI Ethics; Bibliometric Analysis; Course Evaluation; Natural Language Processing

1. Introduction

In recent years, the application of Artificial Intelligence (AI) in educational areas such as course evaluation has developed rapidly. While the adoption of AI technology has shown great potential in improving evaluation efficiency and accuracy, it is also accompanied by a series of societal and ethical issues (Huang et al., 2022). Previous studies have made some valuable contributions to the understanding of ethical issues related to AI in education. For instance, Perkins et al. (2024) introduced the Artificial Intelligence Assessment Scale (AIAS) framework to consider ethical integration in educational assessment, but their work did not provide a systematic review or broader perspective. The systematic review by Farooqi et al. (2024) offered important insights into ethical considerations of AI in educational contexts, though its focus is on general applications rather than the specifics of course evaluation and assessment. Additionally, Maphosa and Maphosa (2023) applied bibliometric analysis to study AI in higher education broadly. However, their work did not specifically address the ethical dimensions related to course evaluation.

Building on these studies, the present research aims to complement existing literature by focusing specifically on the ethical implications of AI in course evaluation. By integrating NLP-enabled bibliometric methods, this study seeks to offer both a systematic and quantitative perspective to deepen our understanding of societal and ethical issues. First, a comprehensive literature database is constructed through broad keyword retrieval; second, NLP methods based on cosine similarity are used to carefully screen the literature; finally, CiteSpace is employed to perform literature clustering, co-citation, and collaboration network visualization analysis. With this multi-method integration, this study attempts to answer the following questions:

1. How are ethical and societal issues represented in literature, and what are their evolutionary trends?
2. What are the research priorities, development trends, and collaboration networks (among authors, institutions, and countries/regions) in the field of AI for course evaluation over the past decade?

2. Methodology

2.1. Data collection and processing

This study initially employed a loose keyword search strategy in the Web of Science core database, combining keywords related to artificial intelligence, educational assessment, and ethical issues. The search formula is as follows:

TS= ("artificial intelligence" OR "AI" OR ...etc.)

AND TS= ("educational assessment" OR "learning assessment" OR...etc.)

AND TS= ("education" OR "educational" OR "pedagogy" OR ...etc.)

AND TS= ("ethics" OR "societal" OR ...etc.)

Given that ethical and societal-related vocabulary is relatively vague, a broad search strategy was adopted to avoid missing potentially valuable literature. A total of 901 articles were obtained in the preliminary search, covering research from 2008 to 2025.

However, an overly broad preliminary keyword search may introduce some irrelevant literature, affecting the overall literature analysis. Since manually assessing the semantics of many documents is too complex and time-consuming, the study employs a two-step NLP methodology. First, the pre-trained Sentence-BERT model (all-MiniLM-L6-v2) encodes both the target and reference abstracts into high-dimensional semantic vectors, capturing their contextual meanings. Second, cosine similarity is computed between each target abstract and the template vector—which is derived from 25 manually screened articles. A similarity threshold of 0.58, determined through iterative validation, ensures that only documents with high semantic relevance are selected. Double verification further confirms that even the documents with the lowest similarity scores meet the research criteria, thereby effectively balancing the breadth and precision of the literature search.

2.2. Bibliometric analysis

Using CiteSpace software, this study conducted a visual analysis of the literature. It mainly includes two aspects: one is content analysis, which determines the research hotspots and their evolution by extracting and clustering nouns in titles and abstracts; the other is network analysis, which explores the cooperation and citation patterns between major authors, institutions, and countries. By constructing co-citation networks, author cooperation networks and institutional citation networks, it aims to fully display the current status of academic communication and cooperation in this field.

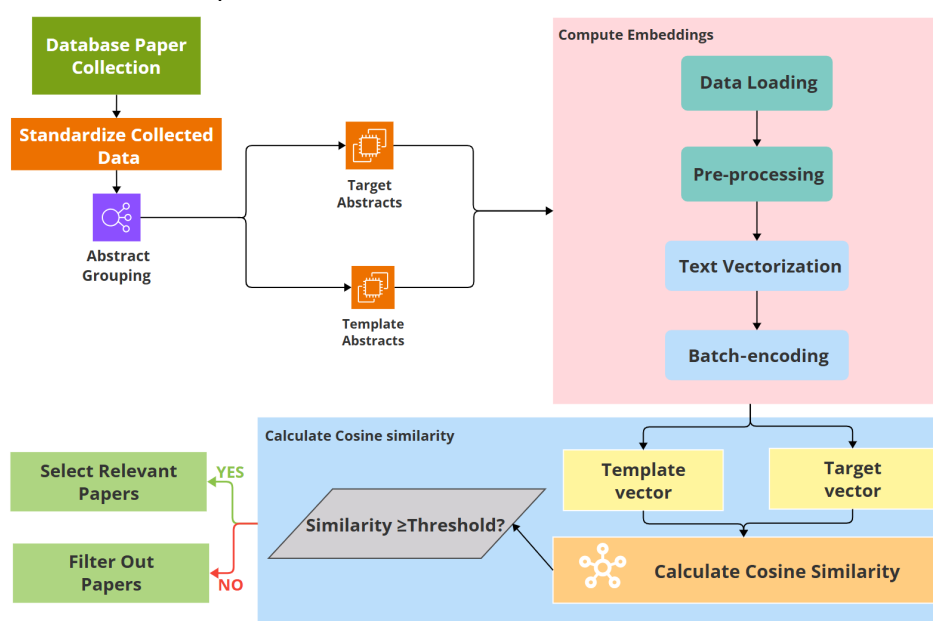


Figure 1. NLP-Driven Literature Screening Workflow

3. Results

3.1. Results of the Research Questions

The following results directly address the initial research questions by revealing the evolving research priorities and development trends through literature publication trends and topic clustering. The results highlight two primary clusters—one focused on AI technologies and the other on educational assessment, with emerging ethical issues. The subsequent analysis of topic evolution and citation patterns further clarifies how the field's focus has shifted over time, while the mapping of collaboration networks among authors, institutions, and countries directly responds to the inquiry into academic dissemination and cooperation dynamics.

In addition, regarding ethical issues, through bibliometric analysis, there are some high-frequency AI ethical issues such as data privacy, ethical dilemmas and data collection. This section will focus on the raising and distribution of these issues.

3.2. Literature publication trends and topic clustering

From the perspective of time trend, research output has shown an upward trend year by year, reaching a peak in 2024 with a total of 83 articles published; and as of now, although the data for 2025 only covers a quarter of the year, 37 articles have been published, indicating that the overall publication volume may continue to rise in the future.



Figure 2. Topic Clustering Diagram

Cluster analysis, as shown in Figure 2, indicates that the literature is mainly divided into two parts: one part is centered on AI technology keywords such as "artificial intelligence" and "machine learning", forming the primary cluster; the other part is composed of secondary clusters with educational evaluation-related words such as "assessment" and "evaluation". Among them, in the primary cluster, "data privacy" as a high-frequency keyword has attracted widespread attention. For instance, Tang and Su (2024) discussed in detail the ethical issues of generative AI in the classroom, highlighting the importance of data privacy.

At the same time, the field division research of topics shows that ethical issues vary in different educational levels, but are mainly concentrated in higher education (Gonsalves, 2024; Luo, 2024) and medical education (Turner et al., 2023), which have relatively abundant research resources. This finding may suggest that it is necessary to conduct more refined research on K-12 and other fields in the future.

3.3. Research topic evolution and citation pattern

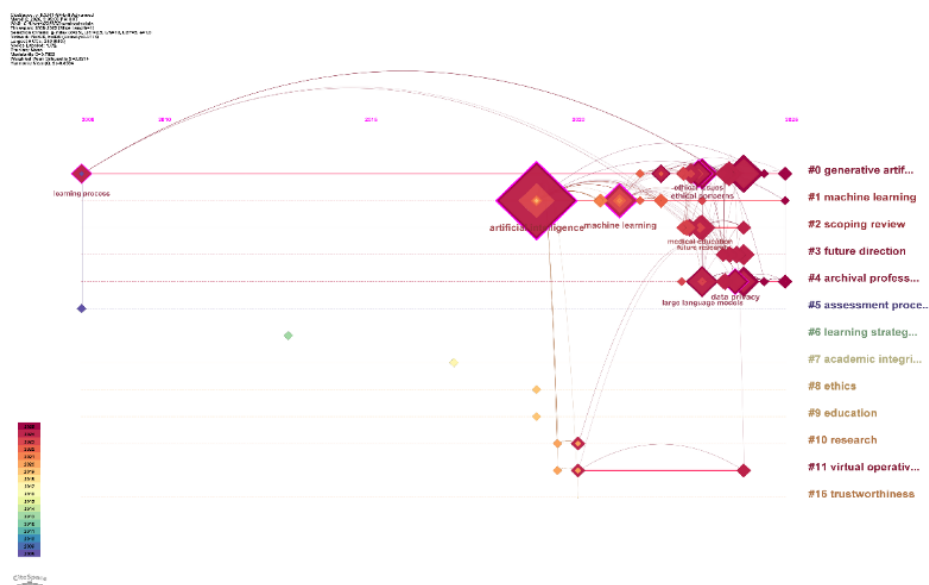


Figure 3. Topic Timeline Diagram

The timeline of literature evolution, as shown in Figure 3, reveals a shift in research focus over time. As early as 2008, Ellmers was one of the pioneers whose ideas were frequently cited in subsequent literature from 2019 to 2021, focusing on experimental applications of AI and machine learning. From 2023 to 2024, more systematic reviews and scoping studies emerged, reflecting the continued maturation of research methods and focuses. Citation network analysis shows that cross-field citations between topics are relatively limited, reflecting that although the field has interdisciplinary characteristics, each subfield may still be relatively independent.

3.4. Author and Institutional Network



Figure 4. Author Clustering Diagram

As shown in Figure 4, the author collaboration network is sparse: most scholars have only one or two collaborations, with at most two joint works per pair and only 11 such pairs. There are also some scattered independent research bodies. No obvious centrality or stable research school has yet to be formed.

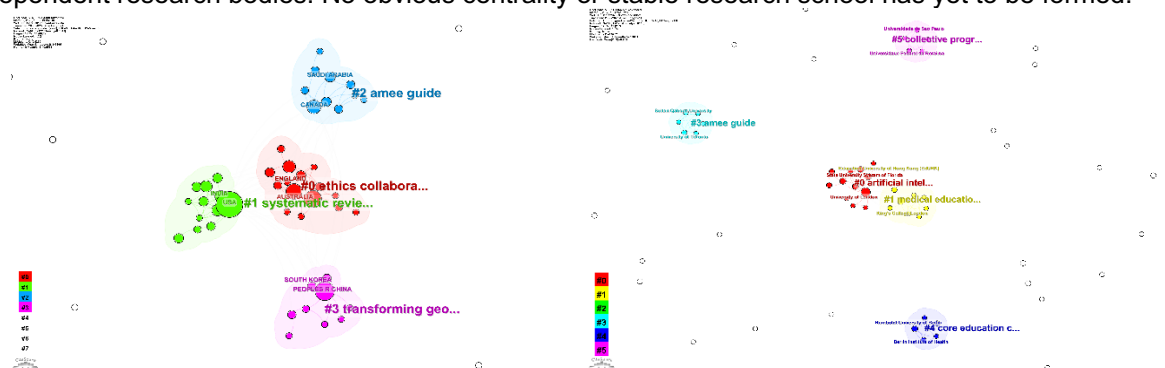


Figure 5. Institutional and Country Clustering Diagram

At the institutional and national level (Figure 5), the United States ranked first with 53 articles, followed by China (25 articles), and the number of articles from other countries and regions was less than 20. In the institutional citation network, institutions such as the University of London showed a high frequency of citations, indicating that international cooperation showed a trend of cross-border and cross-institutional cooperation, but the overall degree of cooperation still needs to be improved.

4. Discussion

This study discusses the ethical issues arising from the rapid growth of Artificial Intelligence in course evaluation. A novel aspect of this work is that, although our broad initial keyword search may have captured

some irrelevant records, the subsequent NLP-based screening largely mitigated this issue. However, further research is needed to determine whether a fixed similarity-threshold is optimal, since our chosen cutoff (0.58) was tailored to this dataset. Future studies should continue to refine data-screening methods and leverage more advanced NLP techniques to explore deeper semantic relationships. Moreover, longitudinal research is recommended to monitor how ethical debates evolve alongside AI's development.

5. Conclusion

This paper systematically presents the current state of research of AI applications in curriculum evaluation, especially its social and ethical dimensions, through the dual methods of bibliometric analysis and NLP empowerment. The results indicate that technological innovation and ethical issues together constitute research topics in this field, among which data privacy is particularly prominent. The relatively scattered cooperation model between authors and institutions reflects that there is still room for further integration of interdisciplinary research. This paper not only summarizes the ethical issues related to AI in educational evaluation but also provides a valuable reference for relevant decision-makers and policymakers.

References

- Ellmers, G., Foley, M., & Bennett, S. (2008). Graphic design education: A revised assessment approach to encourage deep learning. *Journal of University Teaching and Learning Practice*, 5(1), 85-96.
- Farooqi, M. T. K., Amanat, I., & Awan, S. M. (2024). Ethical considerations and challenges in the integration of artificial intelligence in education: A systematic review. *Journal of Excellence in Management Sciences*, 3(4), 35-50.
- Gonsalves, C. (2024). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, 1-15.
- Huang, C., Zhang, Z., Mao, B., & Yao, X. (2022). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799-819.
- Luo, J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, 1-16.
- Maphosa, V., & Maphosa, M. (2023). Artificial intelligence in higher education: A bibliometric analysis and topic modeling approach. *Applied Artificial Intelligence*, 37(1), 2261730.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6), 49-66.
- Tang, L., & Su, Y. S. (2024). Ethical implications and principles of using artificial intelligence models in the classroom: A systematic literature review.
- Turner, L., Hashimoto, D. A., Vasisht, S., & Schaye, V. (2023). Demystifying AI: Current state and future role in medical education assessment. *Academic Medicine*, 10-1097.

Deep Self, Digital Wisdom and Pervasive Learning in AI-Centric Futures

Marcus T ANTHONY

Beijing Institute of Technology (Zhuhai), Guangdong, China
marcus.anthony@cgt.bitzh.edu.cn

Abstract: *This paper explores the present and future (to 2050) impact of artificial intelligence and generative A.I., focusing on how these technologies affect learning, sensemaking, and authentic living. It expands the concept of “education” beyond formal schooling, advocating for “pervasive learning” as a necessary adaptation for 21st-century netizens. This form of learning transcends formal and informal boundaries, occurring across the lifespan and in diverse domains such as work, technology, wellbeing, and relationships. Using Critical Futures Studies, the paper examines the crisis in sensemaking within digital society, arguing that increasing digitisation disconnects individuals from the body, diminishes introspection, and erodes personal agency. The author introduces the concept of the “deep self” and proposes cultivating digital wisdom and pervasive learning as responses to this crisis. The paper also outlines digital wisdom’s relationship to artificial intelligence, digital literacy, emotional intelligence, mindfulness, and lifelong learning as key capacities for thriving in the digital future.*

Keywords: digital literacy, emotional intelligence, wisdom, lifelong learning, artificial intelligence

1. Introduction

By 2050, a deeply immersive, algorithmically curated, and three-dimensional digital society will likely be fully underway. The integration of artificial intelligence (AI) into nearly every domain of life—including education—will not only redefine how people live and work, but also how they learn and make sense of the world. In this context, many of today’s challenges in digital life, such as information overload, algorithmic manipulation, and declining attention spans, are likely to intensify. These dynamics have already given rise to what scholars and commentators refer to as a crisis in sensemaking (Rebel Wisdom, 2019).

This largely theoretical paper offers an original contribution to the field of digital education and critical futures studies by synthesizing a wide range of philosophical, sociological, psychological, and pedagogical perspectives to propose a foundational framework for future curriculum design. Specifically, it introduces and expands the concept of digital wisdom as a necessary evolution of digital literacy. This paper argues that digital wisdom can serve as a central philosophical underpinning for future educational curricula across both public and private education systems globally. In doing so, it contributes to the growing literature on digital literacy by extending its scope beyond technical proficiency to include three interrelated domains: *know thyself*, *know the humans*, and *know the machines*.

This argument is situated within the broader field of critical futures studies (Inayatullah, 2018), a discipline that analyzes how societies conceptualize the future and how those conceptualizations shape present actions. Through this lens, the paper examines prevailing technocentric paradigms and offers an alternative, more human-centered trajectory grounded in wisdom, embodiment, and conscious digital engagement.

To support clarity, the following key terms are defined:

- **Deep self:** An authentic, integrated aspect of human identity that draws on emotional intelligence, somatic awareness, and reflective consciousness (Anthony, 2023, 2024).
- **Sensemaking:** The cognitive and emotional process of interpreting complex or ambiguous information to construct meaning and coherence.
- **Digital wisdom:** A triadic framework for engaging with digital technology that incorporates introspection, social understanding, and technological literacy (Anthony 2023a).
- **Pervasive learning:** A lifelong, ubiquitous form of learning that transcends formal education and is integrated into all aspects of life through digital tools and informal networks (Anthony 2023, 2024).

2. The Crisis in Sensemaking, Personal Authenticity and Digital Wisdom

The “crisis in sensemaking” refers to growing confusion over how people interpret the increasingly complex digital world. This includes the broader “meaning crisis,” the rise of conspiracy theories and extremism (Haidt

& Lukianoff, 2019; Rebel Wisdom, 2019; Zuboff, 2019), widespread misinformation, rising anxiety, and declining trust in media and governance (Brenan, 2019; Haist, 2024; Rocha et al.; Schwab, 2020). These dynamics are already influencing institutional policymaking, including in education.

A central argument in this paper is that the crisis in sensemaking undermines the "deep self"—a more authentic, present-centered self often obscured in digital life. This concept draws on classical and modern wisdom traditions, including Mate's "authentic self" (2018), Jacobson's "awakened self" (2023), and Tolle's "enlightenment" (2009), which stem from Buddhist, Taoist, and Christian teachings. As mental health concerns grow globally (Davis & Hayes, 2012; Haidt, 2024), spiritual and contemplative practices once seen as outdated may need renewed relevance in education and public life.

In response, this paper proposes integrating the author's concept of "digital wisdom" (Anthony, 2023) into both formal and informal learning environments. Digital wisdom includes three domains: "know thyself," "know the humans," and "know the machines." This model extends traditional digital literacy (Gilster, 2007), which largely emphasizes the third domain. To equip learners for the future, it is crucial to integrate both self-awareness and interpersonal intelligence—represented in the first two domains—alongside technical skills.

Historically, the push for digital literacy has been tied to evolving economic and employment demands (American Library Association, 2014). The prevailing concern is that without such skills, individuals risk being left behind in an increasingly digital economy. However, this paper moves beyond a purely technical framing to explore how digital literacy intersects with broader cognitive, emotional, social, and civilizational development through to 2050 (Anthony, 2022, 2023).

3. The Methodological Approach

This predominantly philosophical paper draws on the discipline of critical futures studies, which focuses not so much on predicting the future but on analyzing how people conceptualize the future and attempt to shape it. Inayatullah (2018) identifies six foundational pillars of this discipline. These are: mapping, anticipation, timing, deepening, creating alternatives and transformation. Of these pillars, this paper explores the historical context of digital literacy in regard to "learning" in the digital society (mapping), certain trends affecting the future (anticipation), questioning dominant technocentric narratives (deepening), while positing a broader vision for the future of the digital society (creating alternatives). This short paper will not address specific strategies for achieving that preferred future. The discussion thus aims to address the potential for digital wisdom to influence educational pedagogy and curricula, and broader societal learning (Anthony, 2023, 2024; Inayatullah, 2018).

4. Exteroception and the Causal Chain to Disembodiment

This paper argues that future digital societies and education systems must actively support introspection and maintain a sense of embodiment. "Embodiment" refers to the lived connection to the body and its intuitive capacities, while "embodied presence" is the mindful awareness rooted in present-moment focus, often cultivated through mindfulness practices. Scientific and experimental research confirms that the human body processes large volumes of information and that this "inner knowing" can be accessed through practice (Editorial article, 2021; Kabat-Zinn, 2016; Kotler et al., 2025; Paul, 2021a,b; Schreiber et al., 2025). Neglecting this form of inner wisdom may contribute to declining overall and emotional intelligence.

This paper therefore explores a potential causal link between the crisis in sensemaking and how digital society shapes cognition. Increased screen time strengthens the exteroceptive gaze (external focus) while diminishing the interoceptive gaze (internal awareness). Reduced self-reference may weaken sensitivity to bodily signals, thus impairing somatic wisdom (Paul, 2021a,b) and intuitive insight (Grof & Grof, 2020; Paul, 2021a,b). This could contribute to a decline in emotional and intuitive awareness.

Additionally, personal agency may erode as attention is drawn to chaotic digital environments shaped by Big Tech manipulation and disinformation. Zuboff (2019) identifies this as central to "surveillance capitalism." If this causal chain holds, intervening at any point could help ease the crisis in sensemaking. However, traditional digital literacy (type 1), focused on technical skills, may be insufficient, as it overlooks embodiment, emotional intelligence, and the deep self (Anthony, 2022, 2023). A key question then is whether digital wisdom, as currently understood, can sufficiently restore holistic learning and support the development of psychologically integrated individuals.

Supporting evidence exists: screen-mediated life has been linked to disembodiment and diminished intuitive awareness (Garcia et al., 2021; Gomez, 2010; Kang, 2007). Increased screen time correlates with anxiety and depression, particularly in adolescents (Haidt, 2024; Przybylski et al., 2019; Twenge et al., 2018). Some studies show excessive texting reduces emotional intelligence by limiting face-to-face communication (Piccerillo & Digennaro, 2024), while Turkle (2017) argues digital interaction undermines social skills and relational depth.

Conversely, time in nature enhances well-being and engagement. Activities involving peripheral vision and movement can improve spatial memory and mood, in stark contrast to the negative effects of excessive screen use (Paul, 2021b).

The crisis in sensemaking involves confusion around truth, morality, and wisdom, and disconnection from embodied insight is a major contributor. Cultivating digital wisdom (Anthony, 2022, 2023) in both formal and informal learning may offer a solution—though this will require envisioning alternative futures that challenge current technocentric norms and reframe core assumptions about digital life.

5. More About Digital Wisdom

Digital wisdom (Anthony, 2023, 2024) is a learning framework for the digital age that integrates introspection, social awareness, and technological skill to help individuals navigate an AI-driven world. While valuing technical proficiency, the concept stresses that true understanding requires more than digital fluency—it also demands ethical and self-aware engagement with technology and society. This paper proposes a three-domain model of digital wisdom to support conscious digital participation.

Know Thyself emphasizes self-awareness in digital interactions. Understanding one's values and emotional triggers is key to mindful engagement. Techniques such as "noticing the trigger point" (Anthony, 2023) help individuals manage online emotional responses, improving digital and real-world relationships (Paul, 2021b; Kabat-Zinn, 2016).

Know the Humans involves recognizing emotional and social dynamics in screen-based communication. In a depersonalized digital environment, empathy, active listening, and cultural understanding are vital. Awareness of stress hormones, cognitive bias, and tribalism supports deeper empathy and informed interaction (Anthony, 2023, 2024).

Know the Machines builds on digital literacy, stressing technical skills alongside awareness of how algorithms and AI shape behavior. Understanding issues like surveillance capitalism (Anthony, 2023; Zuboff, 2019) equips users to critically assess technology's effects and make informed digital choices.

6. Digital Literacy

This section outlines the history and main ideas of digital literacy to highlight its limitations and the need for a more expansive approach. Digital literacy can be categorized into three types:

Type 1 focuses on technical proficiency—using digital tools effectively (Gilster, 2007; American Library Association, 2014).

Type 2 includes social and critical communication skills for engaging with digital media (Buckingham, 2013; Ala-Mutka, 2011).

Type 3 adds introspective and emotional intelligence, aligning with the concept of digital wisdom (Anthony, 2024).

The term "digital literacy" originated in the late 1990s, connecting traditional literacy with digital skills (Gilster, 1998). Initially focused on locating and sharing information, the concept evolved to include media analysis and critical thinking, notably through the work of the American Library Association (2014) and Buckingham (2013). Buckingham emphasized the psychological and social dimensions of digital engagement, which closely align with Type 2 literacy.

This paper argues that digital literacy must continue evolving to meet the challenges of today's AI- and media-saturated environment. The EU's 2011 report defines digital literacy as combining technical, informational, and social competencies (Ala-Mutka, 2011). Rheingold (2012) underscored the role of emotional and social awareness in digital life, while Boyd (2014) and Lewandowsky et al. (2017) point to the growing need to assess information credibility and practice digital civility.

Type 3 literacy—still emerging—emphasizes mindfulness and introspection. The Digital Capabilities framework (2015) reflects this trend, describing digital literacy as encompassing critical use, creativity, collaboration, and self-actualization. Yousof (2024) calls for mindful screen use to improve well-being. Palalas (2019) advocates embedding mindfulness in digital literacy education to enhance focus, regulation, and thinking. Ugur (2021) shows mindfulness improves cognitive processes essential to critical engagement with digital content.

7. Digital Wisdom, Emotional Intelligence and Learning

A continuing debate surrounds the role of education in fostering digital literacy. Some scholars advocate embedding digital literacy into curricula to better prepare students for the demands of the modern workforce

(Hague & Payton, 2010), while others caution that this emphasis may detract from foundational literacy skills (Selwyn, 2016). As digital interactions proliferate, the need for emotional intelligence—particularly mindfulness—becomes more urgent. Ertmer and Ottenbreit-Leftwich (2010) argue that teacher training programs must incorporate both digital literacy and emotional intelligence to properly prepare educators, enabling them to support students in navigating an increasingly complex digital environment.

The relevance of lifelong learning is further amplified by the rapid changes brought about by digital technologies and the current “AI explosion” (Anthony, 202, 2025)⁴. Lifelong learning, often framed as a response to ongoing technological shifts, requires individuals to constantly update their knowledge and skills (Facer & Pahl, 2017). The author (Anthony, 2023) extends this idea to propose “pervasive learning,” which spans formal and informal contexts, integrating learning across life domains such as work, leisure, relationships, and finances. Amidst rising levels of anxiety and depression in the digital age (Haidt, 2024), the first two domains of digital wisdom—which emphasize emotional intelligence—are increasingly important. As social, technological, and demographic forces reshape society, the capacity for cognitive adaptability and creativity will become critical (Anthony, 2023, 2024). Likewise, metacognitive competencies, or “learning to learn,” will grow in value, especially as generative AI tools personalize learning experiences (Mossbridge, 2024).

In this light, digital wisdom and pervasive learning are deeply intertwined, with emotional intelligence emerging as a core component. Thus, digital literacy must evolve into a multifaceted concept that supports competent engagement with the digital world. Future citizens will need to cultivate skills in pervasive learning, emotional intelligence, and mindfulness to thrive amid growing complexity and change.

8. Emotional Intelligence and Digital Literacy

Since its emergence in the 1990s, emotional intelligence (EQ) has become influential in psychology and education. Defined as the ability to recognize and manage one’s emotions while understanding others’ (Goleman, 1995, 1998), EQ aligns closely with types 2 and 3 digital literacy and supports the development of digital wisdom. Integrating EQ into education enhances students’ social-emotional competencies and academic success. In the workplace, its value for leadership and teamwork is increasingly recognized (Cherniss & Goleman, 2021), with research linking higher EQ to better mental health and relationships (Mayer et al., 2004; Urquijo et al., 2019).

Goleman’s model identifies five components of EQ: self-awareness, self-regulation, motivation, empathy, and social skills (Goleman, 1995, 1998). These correspond with the first two domains of digital wisdom—“know thyself” and “know the humans.” The literature on EQ has since expanded, with Goleman (2011) emphasizing its importance across life domains and noting that it can be cultivated over time.

EQ is now increasingly connected to mindfulness and embodiment. Goleman (1998, 2011) highlights self-awareness and emotional regulation—core aspects of mindfulness. Mindfulness practices enhance EQ by fostering measured responses to emotional triggers online and offline. Kabat-Zinn (2016) notes that mindfulness strengthens empathy, a key trait in today’s divided digital culture. Jennings (2016) supports mindfulness in education to bolster resilience and emotional intelligence in students and teachers alike. Thus, EQ can be nurtured across formal and informal learning contexts to support competence in complex digital environments.

Although this paper is theoretical, practical tools can foster emotional intelligence and embodied presence in digital contexts—contributing to a “web of wisdom” vision (Anthony, 2022, 2023) that addresses the sensemaking crisis beyond technical fixes. Practices such as “open monitoring” and “soft gazing” encourage mindfulness and bodily awareness (Paul, 2021b), while breathing exercises, the “body scan” (Kabat-Zinn, 2016), and the “noticing the trigger point” method (Anthony, 2023) help manage emotional responses online. Eyal (2019) proposes reconditioning digital habits through visualization, and Paul (2021b) sees journaling as a tool to reconnect with the somatic self. Such techniques can enhance EQ and preserve embodied awareness amid digital immersion.

9. The Web of Wisdom

The digital society’s future may follow two contrasting trajectories. One extends the current technocentric, profit-driven Web 3.0—dominated by Big Tech and “surveillance capitalism” (Zuboff, 2019)—into immersive 3D environments. In its most dystopian form, this resembles the bleak metaverse of Neal Stephenson’s *Snow Crash* (1992), where users are trapped in a system they cannot escape. By contrast, more hopeful futures envision technology enriching, rather than undermining, our connection to the self. The author’s preferred scenario, the “web of wisdom,” emphasizes emotional intelligence, digital wisdom, and community engagement. It promotes a mindful, embodied relationship with technology, society, and the self. In such a future, haptic technologies could support bodily awareness and integrated intelligence (Anthony, 2022, 2023;

Zickraf, 2021), while open-source systems and ethical profit models would prioritize public good over exploitation.

Yet uncertainty remains: can humanity deeply immerse in virtual reality without losing its essence? Realizing the web of wisdom demands wise policymaking, ethical tech design, and educational innovation that fosters embodied presence, emotional intelligence, and digital wisdom.

This utopian vision calls for a spiritual renewal in harmony with scientific knowledge. It imagines an open society that values multiple cognitive modalities—somatic, intuitive, philosophical, critical, and scientific. In this scenario, introspection and emotional awareness coexist with rationality, all supported through digital wisdom education. Compassion, justice, and cultural diversity would be central, as cooperation balances competition in educational systems aligned with these values (Anthony, 2022).

10. Conclusion

The challenges and opportunities of the 21st century call for an educational paradigm that goes beyond the transmission of knowledge. This paper has proposed a global curriculum framework centered on developing the deep self, enhancing sensemaking, and fostering digital wisdom through pervasive learning. By drawing on diverse philosophical and pedagogical traditions, the author suggests that future curricula—across both public and private systems—can empower learners to be not just literate and skilled, but also self-aware, discerning, and digitally wise. Such an approach potentially equips students to navigate rapid global change, enabling them to adapt, learn continuously, and contribute meaningfully to society.

This paper's key contribution to digital literacy discourse is therefore the concept of digital wisdom as a guiding ideal that integrates personal, social, and technological domains. By embedding introspection and meaning-making into the definition of education in the digital era, it reframes the goal from merely using technology to using it wisely. Future research might extend this work by developing curriculum models, assessments, and teacher training programs rooted in the three domains of digital wisdom. Though implementation will differ by context, these core principles can be tailored to diverse cultural and institutional settings. A shared commitment to deep self-development, sensemaking, and digital wisdom can unite stakeholders across both elite academies and public schools, cultivating learners with the insight, ethics, and foresight to shape a more conscious and capable society. Ultimately, promoting digital wisdom is not just about individual empowerment—it is essential for collective, wise decision-making in an increasingly digital world.

References

- Ala-Mutka, K. (2011). Mapping digital competence: towards a conceptual understanding. DO 10.13140/RG.2.2.18046.00322
- American Library Association (2014). ALA Task Force releases digital literacy recommendations. <https://www.ala.org/news/2013/06/ala-task-force-releases-digital-literacy-recommendations>
- Anthony, M.T. (2022). "A Critical Futures Studies Perspective on Embodiment and the Crisis in Sensemaking." In: *Crisis Management - Principles, Roles and Application*. 10.5772/intechopen.107776.
- Anthony, M.T. (2023). *Power and Presence: Reclaiming Your Authentic Self in a Digitised World*. MindFutures.
- Anthony, M. T. (2024). Cultivating digital wisdom in the deep future. *IAFOR Journal of Cultural Studies*, 9(S1), 1–19. <https://doi.org/10.22492/ijcs.9.si.01>.
- Brenan, M. (2019). American' trust in mass media edges down to 41%. Gallup, Sept 26, 2019. www.gallup.news.gallup.com.
- Bogart, J. (2022). Why 2022 is investing in the future digital society. Bogart Wealth. May 17, 2022. <https://bogartwealth.com/investing-in-metaverse>.
- Boyd, D. (2014). *It's complicated: The social lives of networked teens*. Yale University Press.
- Buckingham, D. (2013). *Media education: Literacy, learning, and contemporary culture*. Polity Press.
- Cherniss, C., & Goleman, D. (2021). *The emotionally intelligent manager: How to develop and use emotional intelligence in your work*. Jossey-Bass.
- Davis, D., & Hayes, J. (2011). "What are the benefits of mindfulness?" American Psychological Association. Vol. 48, No. 2, 198–208. DOI: 10.1037/a0022062.
- Digital capabilities: the 6 elements defined. (2015). JISC. <https://digitalcapability.jiscinvolve.org/wp/files/2015/06/1.-Digital-capabilities-6-elements.pdf>
- Editorial article: The Depth and Complexity of Unconscious Processing. (2021). *Frontiers in Human Neuroscience*, 15. <https://doi.org/10.3389/fnhum.2021.791589>
- Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, 42(3), 255-284. <https://doi.org/10.1080/15391523.2010.10782551>
- Eyal, N. (2019). *Indistractable: How to control your attention and choose your life*. Bloomsbury.

- Facer, K. & Pahl, K. (2017). *Valuing interdisciplinary collaborative research: Beyond impact*. Policy Press.
- Garcia, Enara & Di Paolo, Ezequiel & Jaegher, Hanne. (2021). Embodiment in online psychotherapy: A qualitative study. *Psychology and Psychotherapy: Theory, Research and Practice*. 95. 10.1111/papt.12359.
- Gilster, P. (1998). *Digital literacy* (1st Ed). John Wiley & Sons.
- Goleman, D. (1995). *emotional intelligence: Why it can matter more than IQ*. New York: Bantam Books.
- Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam Books.
- Goleman, D. (2011). *Leadership: The power of emotional intelligence*. More Than Sound.
- Gomez, A.G. (2010) Disembodiment and cyberspace: Gendered discourses in female teenagers' personal information disclosure. *Discourse and Society*, 21(2) – 135-160.
- Grof, S., Grof, C., & Kornfield, J. (2020). *Holotropic breathwork*. New York: Excelsior Editions.
- Haidt, J., & Lukianoff, G. (2019). *The Coddling of the American mind*. Penguin Books.
- Hague, C., & Payton, S. (2010). *Digital literacy across the curriculum*. Futurelab.
- Jacobson, L. (2023) *Life beyond the ego*. Conscious Living Productions.
- Jennings, P. A. (2016). *Mindfulness for teachers: Simple skills for peace and productivity in the classroom*. W. W. Norton & Company.
- Kabat-Zinn, J. (2016). *Wherever you go, there you are: Mindfulness meditation for everyday life*. Piatkus.
- Kang, S. (2007). Disembodiment in Online Social Interaction: Impact of Online Chat on Social Support and Psychosocial Well-Being. *Cyber Psychology and Behaviour*, 0(3):475-7 DOI:10.1089/cpb.2006.9929
- Khouja, J. N., Munafò, M. R., Tilling, K., Wiles, N. J., Joinson, C., Etchells, P. J., John, A., Hayes, F. M., Gage, S. H., & Cornish, R. P. (2019). Is screen time associated with anxiety or depression in young people? Results from a UK birth cohort. *BMC Public Health*, 19(1), 82. <https://doi.org/10.1186/s12889-018-6321-9>
- Kotler, S., Parvizi-Wayne, D., Mannino, M., & Friston, K. (2025). Flow and intuition: A systems neuroscience comparison. *Neuroscience of Consciousness*, 2025(1). <https://doi.org/10.1093/nc/niae040>.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353-369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Mate, G. (2018). *In the Realm of Hungry Ghosts*. Vintage Canada.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, 15(3), 197-215.
- Mossbridge, J. (2024). Shifting the human-AI relationship. White paper: Dynamic relational language-learning partner model. <https://arxiv.org/pdf/2410.11864>
- Palalas, A. (2019). Mindfulness for human-centred digital learning. **Argentinian Journal of Applied Linguistics*, 7*(2), 110-125.
- Paul, A.M. (2021a) "Interoception: How to Improve Your 'Gut Feeling.'" Big Think. July 29, 2021. <https://bigthink-com.cdn.ampproject.org/c/s/bigthink.com>.
- Paul, A.M. (2021b). *The Extended Mind*. Mariner Books.
- Piccerillo, L. & Digennaro, S. (2024). Adolescent Social Media Use and Emotional Intelligence: A Systematic Review. *Adolescent Research Review*. 1-18. 10.1007/s40894-024-00245-z.
- Przybylski, A. K., & Weinstein, N. (2019). Digital screen time limits and young children's psychological well-being: Evidence from a population-based study. *Child Development*, 90(1), e56-e65. <https://doi.org/10.1111/cdev.13007>.
- Rebel Wisdom. (2019, August 19). The war on sensemaking, Daniel Schmachtenberger [Video file]. <https://www.youtube.com/watch?v=7LqaoitiGWjQ&feature=youtu.be>.
- Rheingold, H. (2012). *Net smart: How to thrive online*. MIT Press.
- Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolete LD. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. *Z Gesundh Wiss*. 2021 Oct 9:1-10. doi: 10.1007/s10389-021-01658-z. Epub ahead of print. PMID: 34660175; PMCID: PMC8502082.
- Schreiber, M., Stockart, F. & Mudrik, L. (2025). The UnconTrust Database for Studies of Unconscious Semantic Processing and Attentional Allocation. *Sci Data* 12, 157 (2025). <https://doi.org/10.1038/s41597-025-04465-3>
- Schwab, K. (2020). *COVID-19: The great reset*. World Economic Forum.
- Selwyn, N. (2016). *Education and technology: Key issues and debates*. Bloomsbury Publishing.
- Stephenson, N. (1992). *Snow Crash*. Bantam Books
- Turkle, S. (2017). *Reclaiming conversation: The power of talk in a digital age*. Penguin Press.
- Urquijo, I.; Extremera, N.; Azanza, G. (2019). The contribution of emotional intelligence to career success. *International journal of environmental. research. & public health*. <https://doi.org/10.3390/ijerph16234809>
- Twenge, J. M., & Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive Medicine Reports*, 12, 271-283. <https://doi.org/10.1016/j.pmedr.2018.10.003>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Uğur, O. A. (2021). The effect of mindful attention awareness level on digital literacy of sports science faculty students. *International Journal of Education Technology and Scientific Researches*, 6(14), 569-588. <http://dx.doi.org/10.35826/ijetsar.329>.
- Yousof, A. (2024, November 18). The art of mindful screen use: A media literacy perspective [LinkedIn post]. LinkedIn. <https://www.linkedin.com/pulse/art-mindful-screen-use-media-literacy-perspective-ahmed-yousof-lzt4f/>.
- Zickraf, R. (2021). "Mark Zuckerberg's 'Metaverse' is a Dystopian Nightmare," *Jacobin magazine*, 25.09.2021. <https://jacobinmag.com/2021/09>
- Zuboff, S. (2019). *The age of surveillance capitalism*. London: Profile Books.

Strand

Challenges and Opportunities of AI in Higher Education and Vocational Training

A Detailed Analysis of the Impact of Large Language Models on Students' Cognitive and Metacognitive Ability

Haoli WANG^{1*}, Sixu AN², Yu YANG³, Guandong XU⁴
^{1,2,3,4}The Education University of Hong Kong, Hong Kong SAR, China
whaoli@eduhk.hk, s1154340@s.eduhk.hk, yangyy@eduhk.hk, gdxu@eduhk.hk

Abstract: *With the rapid development of Generative Artificial Intelligence, one of the most impressive brunch, Large Language Models, had an extremely impact on education. Due to the widespread application of large language models, it is necessary and important to study and analyze their impact on teaching, especially on students' cognition and metacognition. This paper reviews the theoretical framework of cognition and metacognition, analyzes the application paradigm of large language models in educational settings, and thoroughly discusses the positive and negative effects of large language models on individual cognitive and metacognitive abilities. In cognition, LLM contribute to the the rapid formation of schema and can help students complete the assimilation stage, but LLM cannot go deeper into the accommodation stage. As for metacognition, the current LLM does not help much in improving metacognitive abilities. On the contrary, LLM will reduce metacognitive sensitivity, leading to the problem of metacognitive laziness. It also looks forward to and summarizes the future development and limitations of large language models in educational settings.*

Keywords: Large Language Model, Cognition, Constructivism, Metacognition

1. Introduction

In recent years, the rise of generative AI (GenAI) has emerged, large language models (LLM), as an important component of GenAI, have attracted the common attention of both academic and industrial researchers due to their widespread application in educational scenarios. For example: course materials generation, student assignments grading, and intelligent tutor systems, etc. Among them, the quality of intelligent education systems is closely related to whether students' cognitive and metacognitive abilities are improved. For a student, the formation of cognition and metacognition can usually be explained by constructivism. The process of a student encountering a new concept can be divided into three stages: accommodation, assimilation, and cognitive equilibrium, and in these three stages, metacognitive ability also changes implicitly. Over the past few decades, scholars have proven constructivism to be reliable through experimental research and it has been widely applied in various teaching experiments and design of teaching scenarios, fully demonstrating its effectiveness. Since large language models have been widely utilized in various educational scenarios, in this paper, we will deeply analyze and discuss whether large language models have an impact on students' cognitive and metacognitive construction, and what kind of impact they have. Here, we propose the following two research questions:

RQ1: How do large language models affect students' cognitive construction process?

RQ2: How do large language models affect students' formation of metacognition?

The remaining part of the paper will be organized as follows: Firstly, theoretical frameworks of constructivism, metacognition will be introduced in detail in section 2. Secondly, a case study will be conducted in section 3. Section 4 will contain a detailed discussion on the impact of LLM on students' formation process of cognition and metacognition. At last, conclusions and prospects will be discussed in section 5.

2. Theoretical Frameworks

2.1. Constructivism and Social Constructivism

Constructivism is a theory about cognitive formation, i.e. the process of students' learning and acquisition of knowledge actively. Constructivism emphasizes that students do not simply receive external information transmission but actively process information and interact with the environment based on their existing knowledge in the process of knowledge formation. Constructivism was first proposed by (Piaget, 1977), it introduced a definition of schema, which is defined as cognitive structure, i.e. how a human view the world. Everyone has different schemas since everyone has different perception of things that strongly related to one's experience, motivation and even interest. It is thought that each human being is born with initial schema. This initial schema will change continuously with the continuous interaction between the individual and the new things and environments. It is said that human need to constantly adapt to new things, and the adaptation

methods are assimilation and accommodation. Assimilation refers to the ability of new stimuli and new things to be fully incorporated into the existing schema. Accommodation refers to the existing schema being unable to digest and understand new stimuli, requiring modification of the existing schema. Fig.1 illustrates the process of schema, assimilation and accommodation.

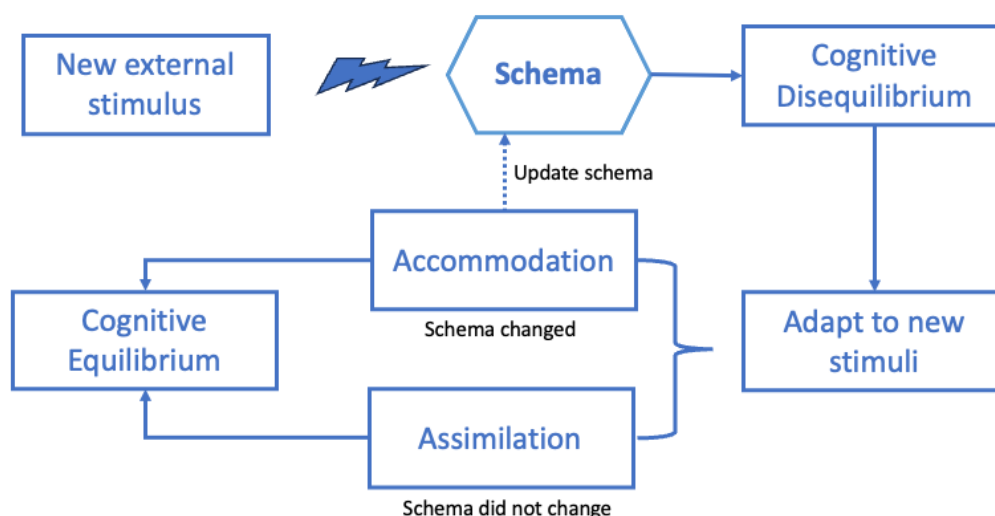


Figure 2. Illustration of constructivism

Social constructivism (Vygotsky, 1978) is an extended theory of constructivism, which emphasize that cognitive behavior is a process with abundant social attributes, i.e. the process of knowledge construction occurs in the interaction between individuals and their social environment. In addition, Vygotsky introduced a concept: the Zone of Proximal Development (ZPD), which refers to the range of tasks that individuals cannot complete independently but can accomplish with guidance from a more knowledgeable other. It is considered that there is a gap between Actual Development Level and Potential Development Level. Actual Development Level refers to the level of tasks a learner can complete independently, and Potential Development Level refers to the level of tasks a learner can achieve with assistance from others.

2.2. Metacognition

Metacognition (Flavell, 1976) generally refers to one's ability to recognize his/her own cognitive ability, in other words, the reflection on the cognitive process of oneself, which includes memory, perception, calculation, association, etc. It contains two perspectives. One is metacognitive knowledge. It refers to the understanding of one's cognitive process and its influencing factors. The other one is metacognitive regulation, which refers to planning, monitoring, and regulating cognitive activities.

Generally, individuals are hard to assess their cognition level accurately (Henmon, 1911). In detail, some learners tend to have high confidence on their tasks, even though they may perform not very well on them. On contrast, some individuals may be too modest when they are doing self-reflection. Two aspects are introduced to detailly discussion of metacognition, namely metacognitive sensitivity and metacognitive bias. Metacognitive sensitivity, also known as metacognitive accuracy, type 2 sensitivity, discrimination, reliability, or the confidence-accuracy correlation. Metacognitive bias is also known as type 2 bias, over- or under-confidence or calibration (Fleming & Lau, 2014). Metacognitive sensitivity refers to individuals' abilities to accurately perceive their own cognition, e.g. Whether an individual can clearly know whether their understanding of a certain issue is correct or whether they have mastered a certain knowledge concept (Maniscalco & Lau, 2012).

An individual with high metacognition sensitivity can better assess his/her own learning progress, more easily identify the weaknesses and make improvements, in addition, he/she can handle and adapt flexibly in complex problems. Metacognitive bias refers to misjudgment of an individual's own cognitive state, it can perform as the following situations: Overconfidence, Underconfidence and Dunning-Kruger Effect (Kruger & Dunning, 1999), which refer to overestimating one's own cognitive abilities, underestimating one's cognitive ability, and individuals with insufficient ability tend to overestimate their own level, while those with strong ability may underestimate themselves.

3. Case study

3.1. Simulated Agent

As is well known, it is difficult for us to obtain the real data of students' cognitive and metacognitive process. Moreover, due to the existence of external subjective and objective interference factors, e.g. environment and wealth etc. We cannot guarantee that the collected data necessarily reflects the real state of students' cognition and metacognition. Considering the above, we use the LLM Agent to simulate the cognitive behavior process of students. Our goal is to simulate students' cognitive performance, so we fixed the simulation scenario on Math learning in first grade of primary school, and we design the initial prompt of agent as follows:

```
prompt = f"""
You have no prior knowledge about first-grade mathematics.
Based only on the following information, answer the question:

Retrieved Information:
{retrieved_info}

Question:
{query}
"""
```

Figure 2. Agent Initial Prompt

Moreover, we introduced Retrieval-Augmented Generation (RAG) to create an external brain for the agent, to mimic the cognitive behavior of first-grade pupils on math questions.

3.2. Pipeline of LLM Applications in Education

We have created a pipeline to simulate our cognitive interaction scenarios. To detect students' long-term cognitive changes, we have designed a multi-round simulation dialogue strategy as follows:

Step 1: We randomly created different zero-knowledge intelligent agents to simulate students.

Step 2: We used another agent to play the role of a math teacher, imparting knowledge to the student agent.

Step 3: To detect students' cognitive changes, we will evaluate the student agent's knowledge mastery after each round of dialogue for cognitive diagnosis.

Step 4: Multi-round dialogues and multi-round cognitive diagnosis will be conducted to discover changes in the cognitive and metacognitive of the student agent.

4. Discussion

Through the above theoretical analysis and experimental verification, we found that LLM does indeed affect students' cognition and metacognition. In the process of interacting with LLM, students can form their own cognitive system according to constructivism by continuously obtaining correct corresponding knowledge from LLM, in other words, update their schema. When it comes to metacognition, learners have difficulty in improving their metacognitive abilities in the process of interacting with LLM, and even tend to overly rely on LLM, gradually giving up their subjective thinking ability on issues. Next, we will discuss in detail the impact of LLM on individual cognition and metacognition based on constructivism and metacognition theory.

4.1. Impact of LLM on individuals' cognition

When evaluating an individual's cognitive abilities, we generally only assess their understanding of objective facts about something, which extends to an educational scenario, that is, a student's mastery of a specific knowledge concept. Now, we limit the scenario to the theoretical study of mathematics in the first grade of primary school. Through some simple tests, we found that students can indeed acquire corresponding knowledge concepts and digest them into their own knowledge concepts through interaction with large language models, thereby updating their schemas. However, when we delve deeper, we found that large language models usually only output what they consider to be the correct answers to users, and individuals can only passively receive them. Due to issues such as the illusion that large language models still exist, when solving some reasoning tasks, the results are usually not accurate enough, and students only passively accept

them. This also means that when students' self-thinking ability is insufficient, their cognitive construction usually only completes the step of assimilation and cannot go deeper into the step of accommodation. Over time, students only build a loose and unsystematic knowledge structure through learning from LLM, which often means that individuals are more likely to face cognitive disequilibrium issues.

To solve this problem, pioneers (Liu, 2024) in the fields of education and AI have used the method of fine-tuning large language models to make the models understand the Socratic teaching method. This approach, to a certain extent, enables individuals to not only passively receive knowledge but also follow the guidance of large language models to engage in self-reflection, thereby correcting their views on inherent things and completing the compliance phase in cognitive construction. However, due to the fine-tuning only, such models often can only focus on a specific teaching domain and cannot be generalized to all teaching scenarios, which is the main issue currently.

LLM can enhance students' assimilation stage of cognitive constructivism, while it cannot fully help individuals to achieve the accommodation stage of cognitive constructivism.

4.2. Impact of LLM on individuals' metacognition

Existing large language models rarely focus on students' metacognitive research. Our simple experimental results indicate that few students can demonstrate metacognitive processes in the process of interacting with large language models. In simple terms, students rarely question the responses of large language models, nor do they reflect on and correct the knowledge they have acquired. Large language models are lack of the function of guidance, this makes it difficult for students to improve their metacognitive abilities from large language models. We believe that metacognitive abilities are not just objective views and cognitive levels of things like cognitive abilities. Metacognitive abilities are more like the ability of individuals to self-reflect, self-examine, and self-explore. These abilities are often closely related to factors such as teacher guidance and family influence. Simply interacting with large language models cannot achieve the relevant improvement in metacognitive abilities. At the same time, if students rely solely on large language models, it will also reduce their metacognitive sensitivity. The things they encounter often tend to be homogeneous, which makes the process of students' knowledge construction become singular, leading to students being unable to correctly perceive their objective cognitive situation towards things, which is what we call low cognitive sensitivity. In addition, metacognitive error issues also exist in the simple interaction between students and large language models. Large language models that have not been specially fine-tuned cannot perceive the current cognitive state of students, leading to students making misjudgments about their cognitive abilities. Specifically, if a large language model keeps outputting a certain knowledge concept in lower-level expressions to students, students tend to think that they have fully mastered that knowledge concept. However, they may have only mastered part of the content, and even the content they have mastered may be only partially correct.

Additionally, even to some extent, there is an excessive reliance on large language models, a behavior commonly referred to as metacognitive laziness, which refers to individuals lacking the willingness or ability to actively reflect and monitor when thinking or problem-solving, thus relying on intuition or simple strategies rather than engaging in deep analysis and critical thinking. This state often manifests as being reluctant to use more complex cognitive processes to evaluate the accuracy of one's knowledge or the reasonableness of one's thinking. The most obvious phenomenon is that students tend to save cognitive resources, avoid spending energy on complex reasoning or learning. LLM provide them an easy way to get the direct answer without consideration or reflection, which exacerbated this phenomenon.

Large language models, have three main aspects of impacts on individuals' metacognition. Firstly, it increases the risk of falling into low cognitive sensitivity, secondly, large language models often lead to students' metacognitive bias, and finally, due to individuals' excessive reliance on large language models, metacognitive laziness phenomenon will occur.

5. Conclusion

In this article, through analyzing theoretical frameworks, case studies, and detailed discussions, we ultimately discovered the significant and potential impact of large language models on students' cognitive and metacognitive abilities. To some extent, large language models improve students' learning efficiency and can effectively enhance students' cognitive levels in the short term. However, students often passively receive information and cannot improve their self-thinking abilities, that is, the phase of conforming to knowledge construction is not well completed. Large language models have more negative impacts on students' metacognitive abilities, such as reducing metacognitive sensitivity, increasing metacognitive biases, and increasing the possibility of metacognitive laziness. Therefore, we summarized the future directions for improving large language models (in terms of enhancing students' cognitive and metacognitive abilities). Firstly, fine-tuning large models in vertical domains and introducing teaching theories such as the Socratic method and scaffolding teaching to enhance the guiding ability of large language models, rather than simply providing

answers to questions. Secondly, starting from constructivist theory, using RAG to provide additional data support for large language models, while RAG can also receive individual data to adjust its own knowledge base structure, outputting different response forms based on user characteristics, thereby alleviating to some extent the negative impact of large language models on cognitive and metacognitive abilities. Finally, it can be attempted to introduce collaborative learning (such as group discussions) in the interactive scenarios of large language models. Thus, students can not only be inspired by the large language models in the cognitive process but also improve their actual cognitive level through communication and exchange with group members. They can also reflect and think on their own from others' behaviors and cognitive expressions, therefore enhancing their metacognitive abilities.

References

- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. (Trans A. Rosin). Viking.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The nature of intelligence*/Erlbaum.
- Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological review*, 18(3), 186.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., ... & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*.
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, 21(1), 422-430.
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in human neuroscience*, 8, 443.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), 1121.
- Liu, Z., Yin, S. X., Lin, G., & Chen, N. F. (2024). Personality-aware Student Simulation for Conversational Intelligent Tutoring Systems. *arXiv preprint arXiv:2404.06762*.

Bridging Job Market Predictions and Personalized Recommendations for Career Guidance: AI-Driven Career Pathway Approach

Yunqi WANG^{1*}, Haoli WANG^{2*}, Yu YANG^{3*}, Guandong XU^{4*}
The Education University of Hong Kong, Hong Kong SAR, China
s1143757@s.eduhk.hk, whaoli@eduhk.hk, yangyy@eduhk.hk, gdxu@eduhk.hk

Abstract: *In recent years, the growing importance of student career planning has accelerated the development of AI-driven systems in higher education and vocational training. Nevertheless, such AI-driven systems tend to offer domain-specific career recommendations, solely focusing on job listings, skills, or courses. For students with limited career planning experience, independently organizing the information from different systems often presents significant challenges, which hinders their sustained engagement in achieving their goals. Moreover, these AI-driven systems rely on static historical data and cannot account for dynamic market changes. To address these issues, this study proposes a system design to seamlessly integrate future job quantity predictions and progressive recommendation functionalities followed by path-job-skill-course. Additionally, this study introduces AI-based approaches, including the LSTM-GRU algorithm to forecast future job quantities and the Deep Q-Network (DQN) algorithm to connect the career path-job-skill-course chain model. This highly cohesive system assists the users in selecting a career path with a higher likelihood of achieving their desired goals through predictive analytics, while facilitating their gradual career progression via integrated recommendation steps. By providing personalized and detailed roadmap guidance, the study offers a practical strategy that helps students with limited career planning experience navigate their career development more effectively.*

Keywords: AI-driven Education, Vocational Training System, Personalized Recommendation, Job Market Prediction, Deep Learning Models

1. Introduction

Equipping students with skills for the rapidly evolving future becomes essential. Thus, clear career goals need to be established. However, students with limited work experience often struggle with a lack of industry awareness, unrealistic career expectations, and misperceptions of job roles (Fugate & Kinicki, 2008). Current career recommendation systems also lack comprehensive guidance from career path planning to skill gap remediation, and provide little insights into predicting future market trends. As a result, students with limited work experience cannot effectively utilize these systems for career planning support.

Given this background, this study proposes an interconnected recommendation system architecture that integrates future job quantity prediction with recommendations for career paths, job positions, skills, and courses. On the one hand, the job prediction model aligns recommendations with future industry demands, increasing students' chances of securing suitable positions. On the other hand, the recommendation model integrates multiple inputs, such as users' skill levels, personality assessment, and job expectations, to generate tailored recommendations that align with their career aspirations.

The study also introduces two conceptual approaches for technological realization. Using deep learning methods, the system employs the LSTM-GRU model to predict future job quantities. High-demand job positions from predictions, along with user-provided data, are then integrated into the recommendation model that applies the Deep Q-Network (DQN) algorithm to suggest future career paths, job positions, related skills, and relevant courses. By aligning societal demands with individual motivations, the system formulates a balanced career plan that integrates the practical and aspirational aspects of career development. Therefore, such an actionable and practical approach can better assist resource-limited students in finding their future career direction.

2. Literature Review

Recent research on career recommendation systems tends to offer domain-specific recommendations, focusing on career paths (Hernandez & Atienza, 2021; Guleria & Sood, 2023), job opportunities (Yadalam et al., 2020; Rajput et al., 2024; Levy et al., 2024), required skills (Kokkodis & Ipeirotis, 2021; Sun et al., 2025), and relevant courses (Lahoud et al., 2023; Praseptiawan et al., 2024). These domain-specific recommendations remain separate, leaving a critical gap in unified and end-to-end guidance, representing a

fragmented feature in a holistic view. Although Kadu et al. (2024) correlate students' hiring probabilities with skill recommendations, no study has constructed a cohesive recommendation chain linking career path, job, skill, and course. Studies have revealed that students often struggle to organize advice from multiple sources due to a lack of awareness of current industry trends (Arthur et al., 1999; Hoffman, 2011), so these incohesive recommendation systems may cause challenges to students. It also highlights the need for recommendation systems to provide insights into future market trends, helping users stay updated on the latest developments. However, labor market forecasts focusing on future job quantity trends have yet to be incorporated into career recommendation systems as reference data (Alharbi et al., 2024). Therefore, the fragmented feature of career recommendations with limited information on future job quantity trends may confuse students when dealing with numerous suggestions, making them hesitant to utilize these options.

The main technical challenge for fragmentation in career recommendation systems lies in handling heterogeneous data, which refers to structured, semi-structured, and unstructured data collected from multiple sources (Castano et al., 2001). For example, career-path models handle sequential path data (Alharbi et al., 2024); job recommendations depend on job-description text and user attributes (Rajput et al., 2024); skill recommendations draw on skill graphs or tags (Sun et al., 2025); and course recommendations involve video content and interaction logs (Praseptiawan et al., 2024). Each data type calls for its own optimal algorithms and feature representations, thus lacking a comprehensive method to handle the heterogeneous data.

To address this technical challenge, our study aims to develop a reinforcement-learning-based Career Path–Job–Skill–Course chain model to address heterogeneous data challenges and overcome fragmentation, thus delivering tailored career guidance for students.

3. Proposed Solution

3.1. System Architecture Design

Based on the challenges AI faces in career development, this research seeks to solve two primary issues: how users can gain insight into future market trends through the recommendation system, and how fragmented recommendation approaches can be integrated into a cohesive pathway–job–skill–course structure. To tackle these questions, this study introduces a solution within the system's architecture.

As illustrated in Figure 1, the system functions are classified into seven types. In general, excluding the most basic component of User Information Management, the other six components follow a clear sequential order. The Job Quantity Prediction component is required during the user's first interaction with the system, followed by the four recommendation modules, each of which utilizes the prediction outcomes and the results from the previous recommendation step. Moreover, the Learning Platform Data tracking component is established based on the user's learning performance after the final Course Recommendation component.

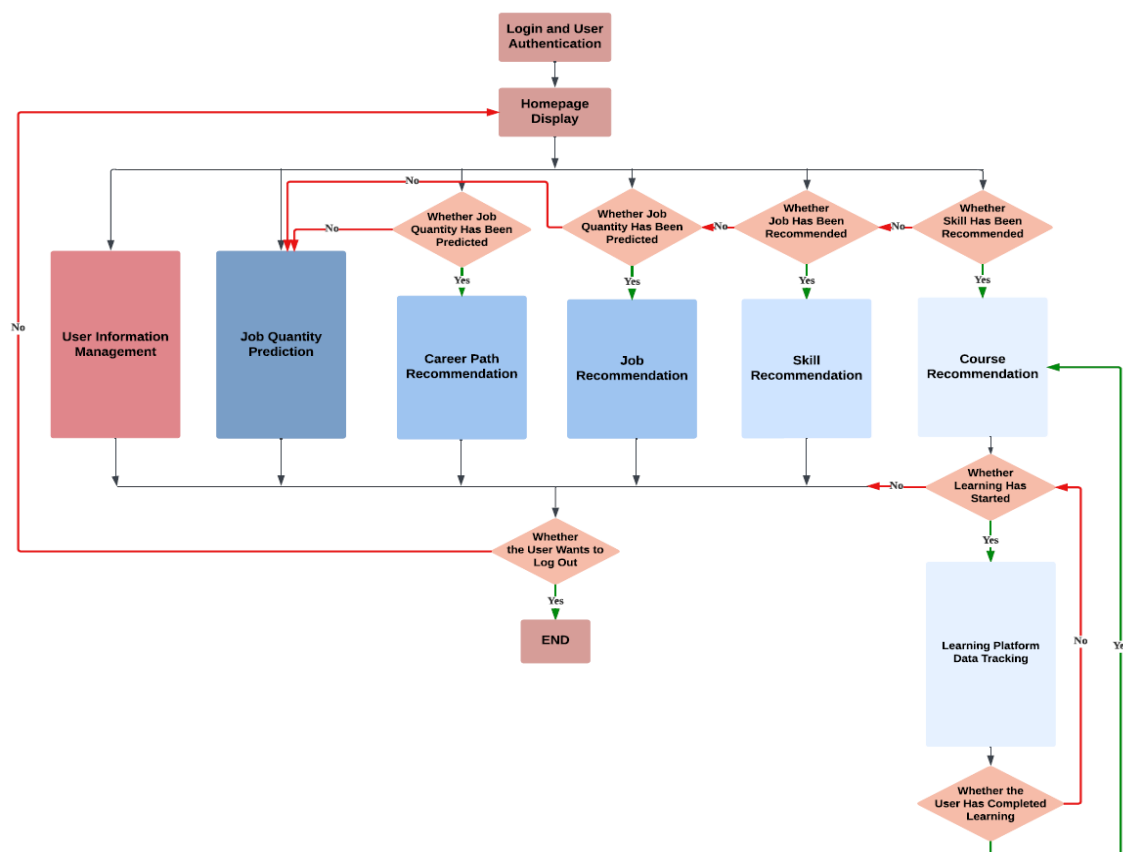


Figure 1. Overview of System Design.

Specifically, each component offers a comprehensive range of functionalities. First, the Job Quantity Prediction component is designed to address the challenge of forecasting future market trends, helping users better understand market requirements and target positions. The training data for this component is derived from the analysis of job vacancy trends on recruitment platforms over recent years. When using this section, users need to provide input regarding their ideal job location, salary range, and anticipated timeline, then the system will generate personalized forecasts. These predictions will be arranged in descending order to display, offering users a clearer view of the job market demand. Furthermore, both Career Path Recommendation and Job Recommendation will be derived from the high-demand job position provided in this section, thereby enhancing users' likelihood of successfully entering the workforce.

Another function of the system is to provide a sufficient range of resources to support users in pursuing their careers. Therefore, the system aims to implement a robust, interconnected recommendation chain, offering personalized career paths and job recommendations based on high-demand positions, skill recommendations tailored to these jobs, and course suggestions aligned with the identified skills.

To achieve personalized job recommendations, this system collects two key types of information from users: first, aligning the user's ideal future job with the current industry demands. This section is integrated into both the Career Path Recommendation and Job Recommendation components, where users must provide information on their desired job field, expected job details, and personality test outcomes. The main purpose of this data is to identify appropriate job positions or career paths that align with the user's areas of interest and advantages. Secondly, the system collects the user's current job status and skill experience to match with job requirements, filtering for positions and future career paths that the user is capable of pursuing. This approach is applied across all four recommendation sections. Such level-based information processing is also widely used for differential comparison: within the Skill Recommendation component, the user's existing skills are compared with the skill demands of the recommended jobs, and the discrepancies are identified as skills that the user lacks, which are then presented. In the Course Recommendation component, missing skills are compared with the competencies required by external courses, and the course with the smallest discrepancy is recommended. Overall, by collecting information reflecting the user's interests and abilities, the system provides more personalized and tailored recommendations.

The system also integrates external learning platform data to track users' progress, contributing to tangible growth and enhancing competitiveness. In the Learning Platform Data tracking component, data on the user's chapter study time, mouse-click behaviors, and test scores from the chosen recommended courses will be visualized, generating periodic learning reports. These reports serve as a tool for both self-reflection by the

user and automatic evaluation by the system, which assesses the alignment between the courses and the user's level, thereby optimizing course recommendations for the user.

3.2. Conceptual and Technological Approach

With respect to the System Architecture Design discussed above, this research introduces conceptual technological approaches for the job quantity prediction model and the path–job–skill–course chain model. The approach predominantly employs two deep learning algorithms: LSTM-GRU and Deep Q-Network (DQN).

Regarding Job Quantity Prediction, this prediction relies on historical data to forecast changes in job quantities at a particular future time point. The key aspect is modeling the temporal trend of job quantity fluctuations. From a holistic point of view, LSTM and GRU capture long-term dependencies through their gating mechanisms. These gates extend the input window, helping to track seasonal short-term shifts and multi-year trends in job demand, while avoiding convergence issues and gradient vanishing or exploding in basic RNNs. The forget gate also mutes one-off fluctuations that do not recur over several time steps, shielding the model from short-lived market noise (such as temporary hiring spikes). From a local point of view, the GRU's lighter structure can simplify pure LSTM setups to balance performance and efficiency. Given its performance, this study selects LSTM-GRU as the deep learning solution that theoretically best achieves this functionality.

For the personalized career path and job recommendations in the system, it is crucial to ensure the model can flexibly handle heterogeneous data and capture nonlinear relationships. This study will employ the Deep Q-Network (DQN) algorithm, a specific reinforcement-learning method, to tie together the career path–job–skill–course chain model. The whole model can be framed as a finite-horizon Markov decision process, which contains three components: state, action, and reward.

In the specific design, the state comprises all information collected at the current stage, which contains market forecasts of future job counts, user attributes in terms of aspirations and competencies, and environment-specific requirements such as job or course prerequisites. The action is the recommended decision made at this stage. The reward function at each stage is defined as a weighted combination of two components: (1) external alignment, which measures how well the recommendation matches job market demand (e.g., predicted job quantity, required skills); and (2) internal fit, which quantifies the compatibility between the recommendation and the user's aspirations, competencies, and preferences. "Finite horizon" means a fixed number of stage t (e.g., $t = 4$, matching the Career Path-Job-Skill-Course stages). The agent learns a Q-value function $Q(s, a)$ that estimates the expected cumulative rewards from taking action a in state s . This process results in a well-trained Q-Value function and an optimal policy that selects the maximum Q-Value, which is also represented as the best recommendation at each stage, enabling the system to deliver coherent guidance with maximized long-term benefit to the user.

By embedding a neural network, DQN overcomes Q-learning's scaling issues in high-dimensional state spaces, making it well-suited to chain models with high-dimensional, heterogeneous, and staged decision steps. Specifically, DQN uses a multi-tower neural network architecture to solve the heterogeneous data challenge mentioned in this study. Each data type is input through a modality-specific subnet, then its outputs are merged and passed through fully connected layers to produce the final representation for Q-value computation.

4. Discussion

The study expands the potential for enhancing vocational and higher educational training through an insightful system design, emphasizing that a robust recommendation system should rely on the forecasting of high-potential job vacancies while also ensuring that users are provided with a clear and complete roadmap for job hunting preparation. For students who lack career planning experience, the system offers interconnected functions, enhancing user-friendliness in vocational training. As a result, the system can serve as a substitute for teachers in guiding career development, thereby boosting students' engagement and offering tailored vocational guidance to individuals at varying levels of expertise. The study also provides valuable insights into its implementation through deep learning models.

Nonetheless, there is still a need to mitigate potential algorithmic biases, which may occur in these types of recommendation systems. In real-world applications, multi-source audits and fairness constraints should be introduced. Moreover, clear roles and lines of accountability must be established among development teams, educational providers, and regulators so that any improper recommendation can be rapidly traced and corrected.

5. Conclusion

This study proposes a comprehensive system architecture for student-centered vocational and higher educational training, addressing key challenges in existing career recommendation systems. The system incorporates job quantity prediction, enabling it to forecast future market trends. It also introduces an innovative interconnected recommendation system that assists users in gradually realizing their career paths. Furthermore, this study presents deep learning models that are well-suited for the model's needs. Overall, this study offers valuable insights into the expanding role of AI in enhancing vocational and higher education opportunities.

References

- Alharbi, F., & Al-Alawi, A. I. (2024, January). Labor market prediction using machine learning methods: A systematic literature review. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 478–482). IEEE.
- Arthur, M., Pringle, J. K., & Inkson, J. H. (1999). *The new careers: Individual action and economic change*. SAGE Publications.
- Castano, S., & De Antonellis, V. (2001). Global viewing of heterogeneous data sources. *IEEE Transactions on Knowledge and Data Engineering*, 13(2), 277–297. <https://doi.org/10.1109/69.920837>
- Fugate, M., & Kinicki, A. J. (2008). A dispositional approach to employability: Development of a measure and test of implications for employee attitudes and behaviors. *Journal of Vocational Behavior*, 73(3), 147–158. <https://doi.org/10.1016/j.jvb.2008.03.002>
- Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: Performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Education and Information Technologies*, 28(1), 1081–1116. <https://doi.org/10.1007/s10639-022-11234-0>
- Hernandez, R., & Atienza, R. (2021). Career track prediction using deep learning model based on discrete series of quantitative classification. *Applied Computer Science*, 17(4), 51–61.
- Hoffman, N. (2011). *Schooling in the workplace: How six of the world's best vocational education systems prepare young people for jobs and life*. Harvard Education Press.
- Kadu, R., Assudani, P. J., Mukewar, T., Kapgate, J., & Bijekar, R. (2024, April). Student placement prediction and skill recommendation system using machine learning algorithms. In *2024 International Conference on Inventive Computation Technologies (ICICT)* (pp. 401–408). IEEE.
- Kokkodis, M., & Ipeirotis, P. G. (2021). Demand-aware career path recommendations: A reinforcement learning approach. *Management Science*, 67(7), 4362–4389. <https://doi.org/10.1287/mnsc.2020.3749>
- Lahoud, C., Moussa, S., Obeid, C., Khoury, H. E., & Champin, P. A. (2023). A comparative analysis of different recommender systems for university major and career domain guidance. *Education and Information Technologies*, 28(7), 8733–8759. <https://doi.org/10.1007/s10639-023-11804-9>
- Levy, A., Chalamish, M., Shalom, B. R., Sharir, G., Peltzman, O., & Salzman, S. (2024). A multi-clustering unbiased relative prediction recommendation scheme for data with hidden multiple overlaps. In K. Arai (Ed.), *Intelligent Computing: SAI 2024. Lecture Notes in Networks and Systems* (Vol. 1017, pp. –). Springer. https://doi.org/10.1007/978-3-031-62277-9_17
- Praseptiawan, M., Muchtarom, M. F. D., Putri, N. M., Pee, A. N. C., Zakaria, M. H., & Untoro, M. C. (2024, September). MOOC course recommendation system model with explainable AI (XAI) using content-based filtering method. In *2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 144–147). IEEE. <https://doi.org/10.1109/EECSI58725.2024.10512496>
- Rajput, A., Dubey, A., Thakur, R., Singh, D., & Singh, U. P. (2024, April). Career Craft AI: A personalized resume analysis and job recommendations system. In *2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ISTEMS58752.2024.10516198>
- Sun, Y., Ji, Y., Zhu, H., Zhuang, F., He, Q., & Xiong, H. (2025). Market-aware long-term job skill recommendation with explainable deep reinforcement learning. *ACM Transactions on Information Systems*, 43(2), 1–35. <https://doi.org/10.1145/3621340>
- Yadalam, T. V., Gowda, V. M., Kumar, V. S., & Girish, D. (2020, June). Career recommendation systems using content-based filtering. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 660–665). IEEE. <https://doi.org/10.1109/ICCES48766.2020.9137906>

Examining Students' Behavioral Patterns of Using GPT to Support

Lesson Planning: A Pilot Study with Trace Data

Chunlin ZHOU¹, Yanxiong XIANG², Lingyun HUANG³

^{1,2,3}Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong SAR, China

s1157295@s.edu.hk, s1157194@s.edu.hk, lingyunhuang@eduhk.hk

Abstract: *This study designed and developed AI-based agents (a GPT-enabled and a RAG-enabled agent) to support pre-service teachers (PSTs) lesson planning tasks by generating sample lesson designs. The findings of a pilot study indicated that the two types of agents assisted PSTs in understanding the subject content and providing them with additional resources for their plans. This study has implications for continuous efforts to develop RAG-enabled agents to support pre-service teachers' learning and professional development.*

Keywords: RAG-enabled agent, Trace data, Lesson planning, Behavioral patterns

1. Introduction

Lesson planning is a fundamental cornerstone in pre-service teacher (PST) education, offering opportunities for PSTs to synthesize their theoretical knowledge with practical teaching skills essential for effective teaching (Sahin-Taskin, 2017). A lesson plan involves goals, knowledge, and materials, and teachers need to align these components and state an instruction procedure that indicates the process of implementation and assessment. However, PSTs are highly aware of their cognitive load and often seek external help during lesson planning. Thus, it is essential to provide external help to facilitate PSTs to achieve better lesson planning performance. Designing external help can be informed by self-regulated learning (SRL) theories that enable them to understand the task and available resources, plan a lesson, apply strategies to refine and revise lesson plans, and monitor progress to make necessary adjustments (Winne & Hadwin, 1998). External help can be delivered by humans or virtual pedagogical agents (Poitras et al., 2017) to support PSTs to develop sophisticated mental models related to planning lessons.

Building human-like agents is evolving while technology is advancing. Since Open AI released ChatGPT, a generative dialog model, in 2022, more and more learning scientists have also started to try to integrate GPT into agent designs. Zhong et al. (2024) constructed a conversational agent to increase student participation and engagement by providing students with appropriate prompts for the collaborative process. An AI-enabled agent's outstanding feature is that it leverages natural language processing to understand complex queries from learners and generate responses in natural ways (Zhong et al., 2024). Thus, compared with traditional pedagogical agents, GPT-enabled agents can better provide explanations, examples, and applications for various concepts. Moreover, GPT-enabled agents can tailor their responses to the learner's level of knowledge and learning style. Additionally, emerging AI technology, like retrieval augmented generation (RAG), can first search the databases and then input the information retrieved from the databases together with the user's input into the big language model as prompts, making the outputs of the agent traceable and reliable (Gao et al., 2024).

Hence, this study will focus on the role of the AI-based agent in PSTs' lesson planning. We will provide PSTs with a GPT-enabled agent and an RAG-enabled agent and examine (1) how a GPT-enabled agent supports PSTs' lesson planning, and (2) how a RAG-enabled agent supports PSTs' lesson planning. Since the data collection is still in progress, this paper presented preliminary results from a pilot study of one participant.

2. Preliminary Findings from Log Data and Process Mining

The participant was asked to design a lesson on the topic of Internet security for Grade 3 students. Cococlass (<https://cloud.cocorobo.hk/>), an AI-empowered learning platform, was adopted for the lesson planning task. The CocoClass has integrated the GPT 4-o model, which allows us to utilize it directly as a GPT-enabled agent. Thus, we focused on introducing the development of the RAG-enabled agent. We created an agent in CocoClass, which can generate sample lesson plans. First, we adopted Elasticsearch or FAISS to index the data for fast retrieval and chose dense retrieval models (e.g., DPR, ColBERT) to train the agent. When building the knowledge base, we referenced the Internet security curriculum of Grade 3 students in China and collected lesson plans available online. Then, we screened these selected lesson plans by checking

their structure, relevance, and completeness and retained 50 qualified plan samples. The workflow of the RAG-enabled agent ensures they can operate step by step. The Cococlass platform guides the work process through prompt words. When utilizing this RAG-enabled agent, several key prompt words were selected, including design, information security, and curriculum. When a user inputs a request related to curriculum planning, the agent first employs the chosen retrieval model to search the constructed knowledge base for relevant information and then passes the search results to the generative model. In this process, we determine how to appropriately incorporate the retrieved information into the input of the generative model. The next step is to fine-tune the models on task-specific data if available. The step includes iterative testing that continuously tests and refines the models using feedback and performance metrics (accuracy and relevance). For example, we conducted user testing to gather qualitative feedback. We also adopted A/B testing to compare different versions of the agent. The final step includes deploying the agent in the intended environment (i.e., chatbot) to learn from new data or interactions.

3. Design, Analysis, and Results

We recorded the participant's working screen and conducted a video analysis. To better understand the participant's AI behaviors, we first distinguished two types of events and labeled them as "RAG" events and "GPT" events. The former related to the participant interacting with the RAG-enabled agent, and the latter indicated the participant's using a GPT-based agent. Our analysis included the frequency of individual events, time on events, and content of the prompt (CoP).

Overall, the participants spent approximately 45 minutes on the lesson planning task. The total time on using AI was around 3 minutes 20 seconds, taking up 7% of the total time. For AI usage, the participants visited AI eight times, including six at the RAG events and two at the "GPT" events. In terms of time on events, the participant spent around 133 seconds on the RAG event. The mean time on the RAG event was 22 seconds (SD = 8 seconds). The longest time on the RAG event was 30 seconds, while the shortest was 7 seconds. For the GPT events, it occurred two times with a total of 63 seconds.

Then, we looked into the CoP, which can provide more insights into the participant's rationale for using AI, including the RAG and GPT. We coded the CoP based on the following perspectives: subject content, pedagogy, technology, sample lessons, and other resources. First, regarding the six RAG events, two were concerned with subject content. For example, the participant asked the RAG-enabled agent to generate exemplary Internet security cases. The other six prompts for the agent were to request resources about the pictures or videos on Internet security. Second, the CoP of the two GPT events revealed that the participant requested sample lesson plans from GPT.

In addition, we also revisited the log data stored by the CocoClass platform to analyze why the participant used the RAG agent. According to the log data, we roughly identified three purposes. One is to seek information to complete the lesson design. The participant Ask the agent to adopt a persona to help him analyze the task requirements. The second purpose is to optimize the generated content. For example, the participant cited the pedagogy section of his lesson plan and asked the agent to modify and polish the content. The third one is to extend thoughts. The participant first introduced a topic and asked the agent to extend the topic to make up a story about Internet security for his teaching.

4. Conclusion

This study designed and developed AI agents and discussed the characteristics of such types of assistance in facilitating and empowering PSTs lesson planning tasks. The preliminary finding of our pilot study suggests that the AI-based agents (both GPT-enabled and RAG-enabled) can help PSTs plan a lesson. This result innovates a new mode of human-computer collaboration and broadens the application path of artificial intelligence in education. Since this is a pilot study, the findings from one participant are difficult to generalize. In this pilot, we analyzed how the participants used AI agents. In future analyses, we will also focus on participants' behavioral patterns based on self-regulated learning theory to examine whether the AI agents will foster high-order thinking (e.g., metacognition). Our future analyses also include using eye-tracking data to detect participants' interest and intention toward the AI agents (especially the RAG-enabled agent) and attempting to identify nuanced differences in the role of diverse types of AI agents in learning. To our knowledge, this study is the first to introduce AI agents in PST learning. This study has implications for continuous attention and effort to developing RAG-enabled agents to support pre-service teachers' learning and professional development.

References

- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-augmented generation for large language models: A survey (arXiv:2312.10997)*. arXiv. <https://doi.org/10.48550/arXiv.2312.10997>
- Poitras, E., Doleck, T., Huang, L., Li, S., & Lajoie, S. P. (2017). Advancing teacher technology education using open-ended learning environments as research and training platforms. *Australasian Journal of Educational Technology*, 33(3). <https://doi.org/10.14742/ajet.3498>
- Sahin-Taskin, C. (2017). Exploring Pre-service teachers' perceptions of lesson planning in primary education. *Journal of Education and Practice*, 8(12), 57–63.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates Publishers.
- Zhong, X., Xin, H., Li, W., Zhan, Z., & Cheng, M. (2024). The Design and application of RAG-based conversational agents for collaborative problem solving. *Proceedings of the 2024 9th International Conference on Distance Education and Learning*, 62–68

Transforming Education Through AI Literacy: Insights from Teacher

Trainers

Asmah BOHARI¹, Anusuya KALIAPPAN², Su Ling LOH³

^{1,2,3}Institute of Teacher Education Technical Education Campus, Negeri Sembilan, Malaysia

¹asmah.bohari@ipgm.edu.my, ²anusuya.kaliappan@ipgm.edu.my, ³lohsuling@ipgm.edu.my

Abstract: Artificial Intelligence (AI) is revolutionising education, demanding a paradigm shift in educators' competencies to integrate AI effectively into teaching and learning. This study investigates the AI literacy levels of lecturers at the Institute of Teacher Education Technical Education Campus (IPG KPT), focusing on affective learning, behavioural learning, and ethical learning. It examines how demographic variables, gender, age, and professional grade influence AI literacy, revealing disparities in adoption and competency levels. Employing a quantitative survey approach, the study collected responses from 59 lecturers using the AI Literacy Questionnaire (AILQ). Statistical analyses, including the Mann-Whitney U and Kruskal-Wallis H tests, identified significant demographic trends. Findings indicate that male lecturers and younger faculty members demonstrate higher levels of AI literacy, highlighting the need for structured, inclusive training initiatives. The study advocates for strategic professional development to bridge AI literacy gaps and cultivate a technologically resilient academic workforce. By redefining AI readiness within teacher training institutions, this research contributes to building a future-proof educational ecosystem driven by AI-enabled pedagogical innovation.

Keywords: AI Literacy, Digital Transformation, Teacher Training, Professional Development, Educational Innovation

1. Introduction

The rapid advancement of Artificial Intelligence (AI) is reshaping education, requiring educators to develop AI literacy to integrate AI into teaching and learning processes effectively (Holmes et al., 2019; Zawacki-Richter et al., 2019). AI literacy encompasses technical proficiency alongside affective learning, Behavioural Learning, and Ethical Learning, making it a critical competence for modern educators (Ng et al., 2023). For teacher trainers at institutions such as the Institute of Teacher Education Technical Education Campus (IPG KPT), fostering AI literacy is essential to preparing future educators to leverage AI technologies across diverse educational contexts.

AI literacy comprises three core components: affective learning, involving intellectual and attitudinal interaction with AI; Behavioural Learning, emphasising practical skills for AI integration; and Ethical Learning, addressing responsible and equitable use of AI (Ng et al., 2023). Mastery of these dimensions enables educators to personalise learning experiences, automate administrative processes, and make data-driven decisions, thus contributing to developing a future-proof educational ecosystem (Okunlaya et al., 2022).

Despite its recognised importance, gaps persist in understanding how demographic factors such as gender, age, and professional grade influence educators' AI literacy levels (Li et al., 2023; Scherer et al., 2021). Prior research has often overlooked these demographic dimensions, limiting the design of strategic, differentiated professional development initiatives that cater to diverse educator needs. Addressing this gap, the present study evaluates AI literacy levels among lecturers at IPG KPT, focusing on affective learning, Behavioural Learning, and Ethical Learning, while exploring the impact of key demographic factors. By identifying these disparities, the study aims to inform targeted professional development programmes, support the cultivation of a technologically resilient academic workforce, and advance AI-driven pedagogical innovation in teacher training institutions.

2. Literature Review

The growing integration of AI into education has generated substantial interest in understanding AI literacy among lecturers. AI literacy encompasses a range of competencies, including affective learning, Behavioural Learning, and Ethical Learning in using AI technologies (Ng et al., 2023). Recent research underscores AI literacy as a foundational skill for lecturers, enabling the effective incorporation of AI into teaching and learning processes (Zawacki-Richter et al., 2019). Nevertheless, limited understanding persists regarding the specific demographic factors, such as gender, age, and professional grade, that may influence AI literacy among educators.

2.1. AI in Education

AI technologies are increasingly integral to educational practices, providing tools for personalised learning, administrative automation, and data-driven decision-making (Holmes et al., 2019). The effective integration of AI into education demands that lecturers possess technical proficiency and pedagogical insight, including the critical capacity to assess the ethical implications of AI tools. As the demand for AI-driven innovation continues to grow, educators are tasked with building a future-proof educational ecosystem that meaningfully integrates these technologies. Despite this urgency, significant gaps remain in the literature concerning how lecturers acquire AI-related competencies and the specific challenges they encounter in the process.

2.2. Dimensions of AI Literacy

AI literacy among lecturers encompasses three primary dimensions: affective learning, Behavioural Learning, and ethical learning, all essential for the effective and responsible integration of AI technologies in education. Affective learning pertains to the intellectual and attitudinal aspects of engaging with AI, such as openness to adoption and confidence in using technology. This dimension emphasises the importance of fostering positive attitudes and a willingness to explore AI's potential in teaching and learning practices (Ng et al., 2023). Behavioural learning focuses on the practical skills required for AI integration, including data interpretation, software manipulation, and incorporating AI tools into instructional design. Proficiency in this dimension enables lecturers to harness AI's capabilities to enhance educational outcomes (Rogaten et al., 2019).

Ethical learning addresses AI's understanding and responsible application, particularly concerning issues such as data privacy, algorithmic bias, and the broader ethical implications of AI deployment. This dimension is critical for creating an inclusive, equitable, and ethically grounded AI-driven educational environment (Scherer et al., 2021). Together, these three dimensions form the foundation of AI literacy, ensuring that lecturers integrate AI tools effectively while fostering a pedagogical framework that upholds ethical principles and inclusivity.

2.3. Previous Research on AI Literacy

Several studies have examined AI literacy among lecturers, frequently highlighting substantial gaps in knowledge, preparedness, and practical application. A systematic review by Zawacki-Richter et al. (2019) found that although many educators possess basic AI awareness, they often lack the necessary skills and expertise to integrate AI into teaching practices fully. Holmes et al. (2019) further contended that the rapid pace of AI development tends to outstrip available professional development opportunities, creating a persistent need for more structured and strategic initiatives to enhance educators' AI literacy.

Despite these insights, little research has explored how demographic factors influence AI literacy levels among educators. Variations across gender, age, and professional grade remain underexamined, limiting the ability of institutions to design differentiated professional development programmes that address the diverse needs of their academic workforce.

2.4. Demographic Influences on AI Literacy

Understanding how demographic factors influence AI literacy is crucial for designing equitable and effective professional development programmes. Research has demonstrated that gender, age, and professional experience significantly impact lecturers' AI literacy levels (Li et al., 2023). Gender differences are evident, with male lecturers generally exhibiting higher confidence and proficiency in AI adoption than their female counterparts (Li et al., 2023). These disparities suggest the need for initiatives that build technological self-efficacy among female educators.

Age-related trends also play a significant role. Younger lecturers are often more open to adopting AI technologies. In contrast, older educators may face barriers such as limited exposure, resistance to technological change, or a lack of structured training opportunities (Scherer et al., 2021). Similarly, variations across professional grades reveal that senior lecturers, while often having greater opportunities for AI integration, may still require targeted support to fully engage with evolving AI tools (Okunlaya et al., 2022). However, many existing studies fail to examine the intersectionality of these demographic factors, leaving critical gaps in understanding how they collectively influence AI literacy. Addressing these complexities is essential for developing strategic professional development programmes that promote comprehensive AI readiness across all educator demographics.

2.5. Theoretical Framework

This study is anchored in two key theoretical frameworks: the Technology Acceptance Model (TAM) and the Diffusion of Innovations (DOI) Theory. The TAM posits that perceived ease of use and usefulness are critical determinants of technology adoption (Davis, 1989). It offers insights into how lecturers perceive AI technologies and their willingness to integrate them into teaching practices (Venkatesh, 2022). Complementing this, the DOI Theory explains how new technologies are adopted and diffused within social systems, emphasising the pivotal role of early adopters in influencing broader patterns of AI literacy among peers (Rogers, 2003; Scherer et al., 2021). Together, these frameworks provide a comprehensive foundation for understanding both individual-level adoption behaviours and institutional patterns of technological diffusion. Applying TAM and DOI enables a deeper interpretation of the demographic disparities identified in AI literacy levels, informing the development of targeted professional development initiatives to cultivate a technologically resilient academic workforce.

3. Methodology

This study assessed AI literacy levels among lecturers at the IPG KPT. It examined the influence of demographic factors such as gender, age, and professional grade. A quantitative, cross-sectional survey design was employed to collect data simultaneously, providing a comprehensive snapshot of AI literacy across the lecturer population. This design was selected for its effectiveness in examining relationships between variables within a diverse cohort at a specific moment, enabling the identification of patterns and disparities in AI literacy levels.

3.1. Participants and Data Collection

A total of 59 lecturers from the IPG KPT participated in this study. Participants were selected through convenience sampling to ensure accessibility and demographic diversity, although this approach may limit the generalisability of the findings. AI literacy was assessed using the AI Literacy Questionnaire (AILQ) developed by Ng et al. (2023), a validated instrument that measures three dimensions: affective, behavioural, and ethical learning. Responses were captured using a 5-point Likert scale, enabling the calculation of individual dimension scores and overall AI literacy levels. The questionnaire was administered electronically via Google Forms and disseminated through multiple communication channels, including WhatsApp, Telegram, and email. The survey remained open for four weeks, during which reminders were periodically sent to maximise participation and achieve a robust sample size.

3.2. Data Analysis

Data analysis was conducted using IBM SPSS Statistics, applying descriptive and inferential statistical methods. Descriptive statistics, including means and standard deviations, were used to summarise AI literacy levels across affective learning, Behavioural Learning, Ethical Learning, and overall AI literacy scores. For inferential analysis, the Mann-Whitney U Test was employed to compare AILQ scores by gender. At the same time, the Kruskal-Wallis H Test was used to examine differences across age groups and professional grades. A significance level of 0.05 was established, and post-hoc Bonferroni corrections were applied where necessary to identify specific group differences. This approach ensured a rigorous and comprehensive analysis of AI literacy disparities among the lecturers at IPG KPT.

3.3. Ethical Considerations

This study adhered to ethical standards approved by the institutional review board. Participants were fully informed of the study's objectives, assured of confidentiality, and provided electronic consent prior to participation. All data were anonymised, encrypted, and securely stored to ensure participants' privacy, integrity, and protection. Maintaining these ethical protocols enabled a rigorous evaluation of AI literacy among lecturers at IPG KPT while safeguarding participants' rights. The study's commitment to ethical research practices reinforces the credibility of its findings, which aim to inform targeted professional development initiatives for enhancing educators' AI competencies.

4. Findings

This study evaluated the AI literacy levels of lecturers at the IPG KPT. It analysed the impact of demographic factors, including gender, age, and professional grade, on these levels. The findings align with the research objectives, focusing on AI literacy's affective, behavioural, and ethical dimensions and the overall AILQ scores.

4.1. Demographic Overview

The study involved 59 lecturers from the IPG KPT, enabling a comprehensive analysis of AI literacy across diverse demographic groups, including gender, age, and professional experience. This diversity provided valuable insights into how these factors may influence lecturers' engagement with AI technologies in educational contexts.

Of the 59 participants, 27 (45.8%) were male and 32 (54.2%) were female, reflecting a slight majority of female lecturers. This gender distribution offers important context for understanding potential differences in AI literacy, particularly regarding technology adoption and confidence levels. The largest group of participants fell within the 50–54 age range (40.7%), followed by the 45–49 age group (27.1%). This age distribution highlights the need to consider generational differences in AI adoption, as younger educators are often more inclined to embrace new technologies than their older counterparts. Regarding professional grade, most lecturers held Grade DG48 (35.6%) or Grade DG54 (33.9%). This variation suggests differing levels of exposure to AI-related training and presents unique challenges and opportunities for integrating AI tools across different stages of academic careers.

4.2. AI Literacy Dimensions

Table 1. Descriptive Statistics of AI Literacy Dimensions Among Lecturer Participants.

Dimensions	Mean	Standard Deviation (SD)
Affective Learning	4.37	0.50
Behavioural Learning	3.80	0.78
Ethical Learning	4.69	0.35
AI Literacy	4.25	0.44

Table 1 presents the AI literacy levels across three key dimensions. The mean score for Affective Learning was 4.37 (SD = 0.50), indicating that lecturers generally demonstrate a positive attitude toward AI adoption. This finding is consistent with Ng et al. (2023), who emphasised the importance of fostering positive attitudes to facilitate the practical integration of AI into teaching practices. Behavioural Learning recorded a mean score of 3.80 (SD = 0.78), reflecting moderate proficiency in the practical application of AI tools. This result aligns with observations by Yim and Wegerif (2024), who noted variability in educators' practical skills, often due to limited exposure and training opportunities in AI technologies. Ethical Learning yielded the highest mean score at 4.69 (SD = 0.35), suggesting a strong awareness among lecturers regarding Ethical Learning, such as data privacy, algorithmic bias, and responsible AI deployment, in line with Scherer et al. (2021). While ethical understanding appears robust, the findings indicate that practical skills still require targeted improvement. The overall AILQ mean score of 4.25 (SD = 0.44) suggests a relatively high level of AI literacy among IPG KPT lecturers. However, disparities across dimensions highlight critical areas for further professional development.

4.3. Impact of Gender on AI Literacy

Table 2. Mean Ranks of AI Literacy Across Different Genders.

Ranks	Gender	N	Mean Rank
AI Literacy	Male	27	38.50
	Female	32	22.83
	Total	59	

The Mann-Whitney U Test revealed a statistically significant difference in AI literacy levels between male and female lecturers ($U = 202.500$, $p < .001$). As shown in Table 2, male lecturers had a higher mean rank (38.50) than their female counterparts (22.83), suggesting that male lecturers generally possess higher levels of AI literacy. These findings highlight notable gender disparities in AI literacy, which may be attributed to differences in confidence levels, technological exposure, and familiarity with AI tools.

4.4. Impact of Age on AI Literacy

Table 3. Mean Ranks of AI Literacy Across Different Ages.

Ranks	Age (years)	N	Mean Rank
AI Literacy	35-39	5	51.50
	40-44	5	32.60
	45-49	16	18.59
	50-54	24	32.96
	55-59	9	29.00
Total		59	

The Kruskal-Wallis test identified statistically significant differences in AI literacy (AILQ) scores across ages ($H = 15.812$, $p = .003$). Table 3 shows that lecturers aged 35-39 had the highest mean rank (51.50), whereas those in the 45-49 age group had the lowest (18.59). These findings suggest significant age-related disparities in AI literacy, with younger lecturers generally demonstrating better proficiency in AI literacy than their older counterparts.

4.5. Impact of Professional Grade on AI Literacy

Table 4. Mean Ranks of AI Literacy Across Different Professional Grades.

Ranks	Professional Grades	N	Mean Rank
AI Literacy	DG44	10	27.80
	DG48	21	29.60
	DG52	8	46.69
	DG54	20	24.85
	Total	59	

The Kruskal-Wallis H Test revealed a statistically significant difference in AI literacy (AILQ) scores across professional grades ($H = 9.564$, $p = .023$). As indicated in Table 4, lecturers holding Grade DG52 exhibited the highest mean rank (46.69), suggesting higher levels of AI literacy than lecturers in other grades. In contrast, lecturers in Grades DG44, DG48, and DG54 recorded lower mean ranks, reflecting variations in exposure to AI-related training and engagement with technological tools. These findings point to professional grade as an influential factor in shaping AI literacy competencies among lecturers.

4.6. Theoretical Implications of the Findings

The disparities in AI literacy observed in this study can be meaningfully interpreted through the TAM and the DOI Theory. According to TAM, perceived ease of use and usefulness significantly influence technology adoption behaviours (Tahar et al., 2020). This may help explain the higher AI literacy scores among younger and male lecturers, who are more likely to perceive AI tools as accessible and beneficial. Complementing this, the DOI Theory suggests that individuals with higher proficiency, such as younger lecturers, often act as early adopters, promoting the diffusion of technological innovations within institutional settings. These theoretical perspectives provide a deeper understanding of how demographic factors shape AI literacy patterns and underscore the importance of targeted interventions to foster broader, institution-wide adoption of AI technologies.

5. Discussion

This study evaluated the AI literacy levels of lecturers at the IPG KPT, specifically focusing on how gender, age, and professional grade influence these competencies. The findings indicate a generally high level of AI literacy, particularly regarding ethical awareness. However, significant disparities emerged across demographic groups, revealing critical practical skills and behavioural learning gaps among certain cohorts. These results, interpreted through the TAM and the DOI Theory lenses, offer valuable insights into how perceptions of ease of use, perceived usefulness, and early adoption behaviours shape AI literacy patterns within educational institutions. Addressing these disparities is essential for individual educator development and fostering an inclusive, institution-wide culture of technological resilience and innovation.

5.1. Integrating TAM and DOI Theory in AI Literacy Adoption

The disparities in AI literacy levels identified across gender, age, and professional grade highlight individual competency gaps and reveal systemic challenges in fostering AI readiness within teacher education institutions. While lecturers generally demonstrated high levels of ethical understanding, cognitive engagement and technological adaptability, gaps suggest that practical integration of AI into educational contexts remains uneven.

Applying the Technology Acceptance Model (TAM), it becomes evident that perceptions of ease of use and perceived usefulness significantly influence lecturers' willingness to engage with AI technologies (Tahar et al., 2020). The higher AI literacy scores among younger and male lecturers may reflect greater familiarity and confidence with digital environments, underscoring the need for initiatives that enhance self-efficacy among older and female lecturers. The Diffusion of Innovations (DOI) Theory further explains these patterns by positioning younger lecturers as early adopters who can drive broader institutional adoption of AI-enhanced pedagogical practices (Rogers, 2003; Scherer et al., 2021). However, without intentional efforts to support later adopters, such as targeted mentorship and inclusive training opportunities, there is a risk that existing disparities may be exacerbated rather than reduced.

Addressing these challenges requires differentiated professional development strategies that account for demographic variations and a broader cultural shift within institutions. Embedding AI literacy as a core component of professional identity and pedagogical practice is essential to cultivating a technologically resilient academic workforce capable of navigating and leading future educational transformations.

5.2. Professional Development Strategies

To enhance AI literacy among lecturers at IPG KPT, targeted professional development strategies must explicitly address the disparities identified across gender, age, and professional grade. The higher AI literacy levels observed among male lecturers, particularly in the affective and behavioural dimensions, underscore the need to strengthen confidence and technological competence among female educators. This can be achieved through hands-on workshops, mentorship programmes, and increased access to user-friendly AI resources, fostering a more inclusive and supportive learning environment (Kuteesa et al., 2024). Similarly, age-related disparities suggest that younger lecturers exhibit stronger AI literacy, especially in behavioural and ethical dimensions. Foundational AI training initiatives aimed at building digital fluency among older lecturers, alongside advanced workshops for younger faculty members on cutting-edge AI applications, are essential for bridging generational gaps in technological proficiency (Mouta et al., 2024).

Variations in AI literacy across professional grades further highlight the importance of differentiated development pathways. Lecturers in Grade DG52 demonstrated higher proficiency compared to their DG54 counterparts, suggesting the need for junior and mid-career lecturers to advance their AI competencies further. Concurrently, senior lecturers should be empowered to assume leadership roles in AI integration efforts, promoting innovative pedagogical practices and mentoring colleagues across the institution. By strategically addressing these gaps, professional development programmes can cultivate a more equitable, technologically resilient academic workforce capable of leading effective AI integration within teacher education settings (Koka, 2024).

5.3. Implications and Future Directions

This study highlights the critical importance of targeted professional development initiatives in addressing demographic disparities in AI literacy among lecturers. Tailored programmes that focus on the specific needs associated with gender, age, and professional experience are essential for bridging existing gaps and fostering equitable AI competency across the academic workforce. Future research should investigate the underlying causes of these demographic disparities, including technological exposure, institutional support, and self-efficacy, to inform the design of more effective intervention strategies. Furthermore, longitudinal studies evaluating the long-term impact of targeted professional development programmes on AI competencies, pedagogical innovation, and institutional digital transformation would provide valuable insights for shaping future practice across teacher education institutions.

6. Conclusion

Recognising and addressing demographic disparities in AI literacy is essential for enabling educational institutions to equip their lecturers with the competencies needed to integrate AI effectively into teaching and learning practices. This study, grounded in the TAM and DOI Theory, provides a robust conceptual framework for interpreting these differences and informing the design of strategic professional development initiatives. By fostering a more innovative, inclusive, and technologically resilient academic workforce, institutions can better

position themselves to lead in an evolving educational landscape. Investing in targeted professional development not only enhances individual educators' confidence and competence but also ensures that institutions can fully harness the transformative potential of AI to improve learning outcomes and future-proof education systems.

Acknowledgements

This research was conducted by ethical standards and was approved by the Ministry of Education Malaysia (Reference No. KPM.600-3/2/3-eras(22114)). All participant information was anonymised and kept strictly confidential to ensure privacy and data protection throughout the study.

References

- 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>.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign: MA, USA.
- Koka, N. A. (2024). The Integration and Utilization of Artificial Intelligence (AI) in Supporting Older/Senior Lecturers to Adapt to the Changing Landscape in Translation Pedagogy. *Migration Letters*, 21(S1), 59–71. <https://doi.org/10.59670/ml.v21iS1.5939>.
- Kuteesa, K. N., Akpuokwe, C. U., & Udeh, C. A. (2024). Gender equity in education: addressing challenges and promoting opportunities for social empowerment. *International Journal of Applied Research in Social Sciences*, 6(4), 631-641. <https://doi.org/10.51594/ijarss.v6i4.1034>.
- Li, Y., Zhou, C., Wu, D. & Chen, M. (2023). Evaluation of teachers' information literacy based on information of behavioral data in online learning and teaching platforms: an empirical study of China. *Library Hi Tech*, 41(4), 1039–1062. <https://doi.org/10.1108/LHT-01-2021-0034>.
- Mouta, A., Torrecilla-Sánchez, E. M., & Pinto-Llorente, A. M. (2024). Comprehensive professional learning for teacher agency in addressing ethical challenges of AIED: Insights from educational design research. *Education and Information Technologies*, 1-45. <https://doi.org/10.1007/s10639-024-12946-y>.
- Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2023). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55(3), 1082–1104. <https://doi.org/10.1111/bjet.13411>.
- Okunlaya, R. O., Syed Abdullah, N. & Alias, R. A. (2022). Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education. *Library Hi Tech*, 40(6), 1869-1892. <https://doi.org/10.1108/LHT-07-2021-0242>.
- Rogaten, J., Rienties, B., Sharpe, R., Cross, S., Whitelock, D., Lygo-Baker, S., & Littlejohn, A. (2019). Reviewing affective, behavioural and cognitive learning gains in higher education. *Assessment & Evaluation in Higher Education*, 44(3), 321–337. <https://doi.org/10.1080/02602938.2018.1504277>.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Scherer, R., Howard, S. K., Tondeur, J., & Siddiq, F. (2021). Profiling teachers' readiness for online teaching and learning in higher education: Who's ready?. *Computers in Human Behavior*, 118, 106675. <https://doi.org/10.1016/j.chb.2020.106675>.
- Tahar, A., Riyadh, H. A., Sofyani, H., & Purnomo, W. E. (2020). Perceived ease of use, perceived usefulness, perceived security and intention to use e-filing: The role of technology readiness. *The Journal of Asian Finance, Economics and Business*, 7(9), 537-547. <https://doi.org/10.13106/jafeb.2020.vol7.no9.537>.
- Venkatesh, V. (2022). Adoption and use of AI tools: a research agenda grounded in UTAUT. *Annals of Operations Research*, 308, 641-652. <https://doi.org/10.1007/s10479-020-03918-9>.
- Yim, I. H. Y. & Wegerif, R. (2024). Teachers' perceptions, attitudes, and acceptance of artificial intelligence (AI) educational learning tools: An exploratory study on AI literacy for young students. *Future in Educational Research*, 2, 318-345. <https://doi.org/10.1002/fer3.65>.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). A systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>.

Demystifying LLM Implementation in Higher Education: A Three-Step Actionable Roadmap

Chenxi DONG^{1*}, Chujie WEN², Tianyu ZHANG^{3*}
^{1,2,3}Education University of Hong Kong, Hong Kong SAR, China
 cdong@eduhk.hk, cwen@eduhk.hk, tianyu.eduhk@gmail.com

Abstract: Implementing Large Language Models (LLMs) in higher education promises significant enhancements but often faces practical challenges. This paper demystifies LLM integration by presenting a three-step actionable roadmap for higher education institutions. The roadmap commences with (1) mapping specific educational needs to AI solutions. It then guides users through (2) LLM selection, emphasizing a balance of intelligence, cost, and regional accessibility within the dynamic LLM ecosystem. Finally, (3) resource-aligned implementation strategies, comparing solutions from local hosting to cloud services based on infrastructural needs and technical expertise requirements. This practical framework provides a clear pathway for navigating LLM adoption, ensuring effective integration within diverse higher education settings.

Keywords: Large Language Models, Educational Technology, Artificial Intelligence

1. Introduction

Large Language Models (LLMs) like ChatGPT and DeepSeek enable innovations in higher education. However, implementation is hindered by infrastructure limitations, budget constraints, and faculty resistance. Current literature often focuses on theoretical applications, leaving a gap in practical guidance (Owan et al., 2023). Therefore, this paper addresses this challenge by providing a clear three-step roadmap for LLM integration. This practical guide focuses on three key steps: (1) identifying specific educational needs for LLM solutions; (2) selecting the right LLM based on intelligence, cost, and accessibility; and (3) determining appropriate implementation strategies. (See Figure 1).

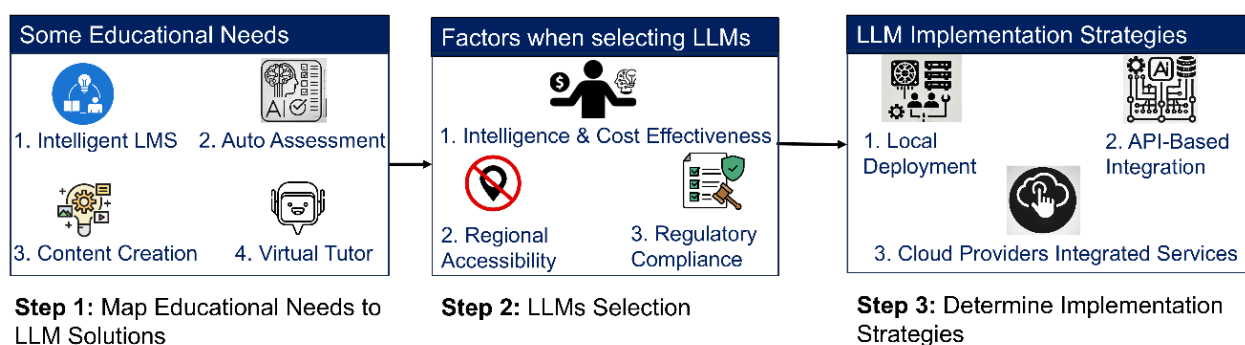


Figure 1. LLM Integration Roadmap: Needs Assessment, Model Selection, Deployment Strategy.

2. Step I: Map Needs to AI Solutions - Confirming Where LLMs Deliver Real Impact

Effective LLM implementation starts by identifying needs-based applications. Institutions should first pinpoint educational challenges and opportunities where LLMs offer solutions, moving beyond technology-first adoption. A key question is: "Where can LLMs enhance learning outcomes, improve efficiency, or address existing pain points?". Examples of applications include:

- 1) **Intelligent Learning Management Systems (LMS):** AI integration in platforms like Coursera and edX personalizes learning by adapting content to diverse learning styles (Wang et al., 2024).
- 2) **Automated Assessment Tools:** LLMs can be used to evaluate student responses accurately, aiding educators in saving time and ensuring consistent evaluation standards (Owan et al., 2023).
- 3) **Content Creation:** LLMs generate diverse educational materials, such as quizzes and summaries, enabling educators to create varied resources (Morales et al., 2024).
- 4) **Virtual Tutors:** AI-powered virtual tutors offer 24/7 student support, providing personalized explanations and assistance, ensuring continuous learning. Khanmigo, from Khan Academy, exemplifies this as an AI tutor guiding learners to find answers independently (Shetye, 2024).

3. Step II: Select LLMs - Balancing Intelligence, Cost, and Feasibility

When selecting the appropriate LLMs, it is essential to understand that the leading models are primarily corporate products from **(1) Significant Financial Investment**: Training costs for models like GPT-4 reaching millions of US dollars (Bommasani et al., 2021). **(2) High Infrastructure Demands**: LLMs requiring large amounts of advanced GPUs (e.g., NVIDIA A100, \$8,999 per unit) and massive energy consumption (around 60 million kWh for GPT-4), and **(3) Advanced Engineering Expertise**: Building LLMs requires extensive engineering resources, often exceeding the capacity of academic institutions (Shoeybi et al., 2019). Table 1 presents the top 10 most intelligent LLMs as of February 2025, ranked by their intelligence score. This score is the average of three key capabilities: advanced reasoning (GPQA score), general knowledge and logic (MMLU-Pro score), and mathematical ability (AIME 2024 score).

Table 1. Top 10 Most Intelligent LLMs as of February 2025

Model	Developer	GPQA (Reasoning)	MMLU-Pro (General)	AIME 2024 (Math)	Intelligence Score*	Output Cost (\$/M token)	Intelligence ROI
o3-mini	OpenAI 	75	79	77	77.00	4.4	17.50
o1	OpenAI 	75	84	72	77.00	60	1.28
DeepSeek R1	DeepSeek 	71	84	68	74.33	1.09	68.20
o1-mini	OpenAI 	60	74	60	64.67	12	5.39
Gemini 2.0 Pro	Google 	62	81	36	59.67	0.4	149.17
Gemini 2.0 Flash	Google 	62	78	33	57.67	0.4	144.17
DeepSeek V3	DeepSeek 	59	76	25	53.33	0.54	98.77
Qwen2.5 Max	Alibaba 	60	76	23	53.03	6.4	8.29
Claude 3.5 Sonnet	Anthropic 	75	78	16	53.00	15	3.53
GPT-4o	OpenAI 	75	75	12	47.00	10	4.70

*Intelligence Score = Average (GPQA, MMLU-Pro, AIME 2024).

Step II LLMs selection considers the following key factors:

- **Regional Accessibility**: Geopolitical factors critically impact access. For instance, an institution in China considering LLMs from Table 1 would find that seven models (indicated by grey shading in Table 1) are restricted and unavailable. Actionable: Verify your institution's regional availability to ensure access before proceeding with the evaluation.
- **Intelligence & Cost-Effectiveness**: Balance performance and budget using Intelligence ROI = Intelligence Score/Output Cost (Table 1, last column). Higher intelligence ROI means more intelligence gained per dollar spent. Actionable: Prioritize models with high Intelligence ROI value.
- **Regulatory Compliance**: Ensure alignment with regional regulations, such as China's CAC Interim Measures for Generative AI Services, the European Union's Artificial Intelligence Act, and the United States' Executive Order on AI (de Fine Licht, 2023). Actionable: reviewing related regulations to implement responsible AI use in education.

Effective LLM selection demands a systematic approach, paving the way for integration planning in Step III.

4. Step III: Determine Implementation Strategies - Aligning with Resources

An effective LLM integration strategy depends on resource allocation, technical expertise, data privacy, customizability, etc. Figure 2 compares three main approaches.

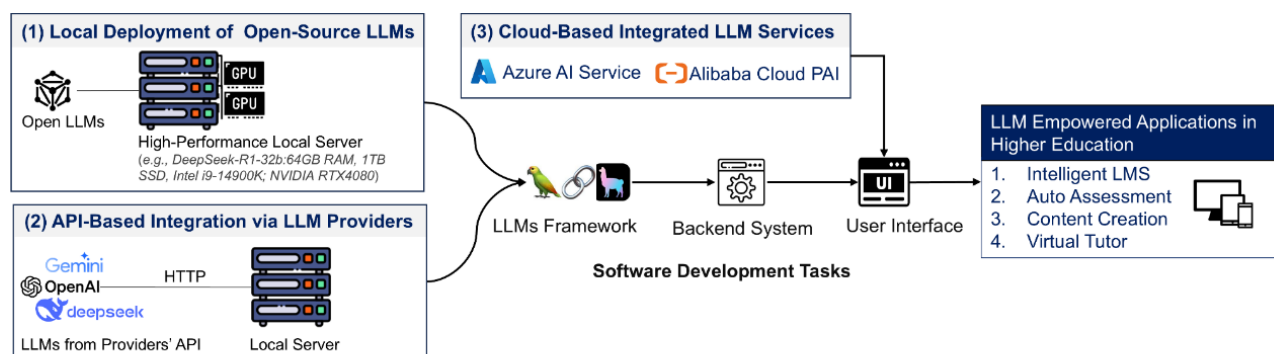


Figure 2. Overview of LLM Integration Strategies: Local Hosting, API Integration, and Cloud Services

Key Integration Strategy Options:

- 1) **Local Deployment (Open-Source LLMs):** Maximizes control and customizability via local hosting of open-source LLMs (e.g., DeepSeek R1, Llama 3). Trade-offs: High investment in IT infrastructure, high technical expertise demand, potentially lower operational costs, and maximum data sovereignty. Best Suited for: Institutions prioritizing data privacy, extensive customization with abundant technical resources.
- 2) **API-Based Integration (LLM Providers):** Balances customizability and ease by utilizing LLM provider APIs (e.g., OpenAI, Google, DeepSeek) with frameworks (e.g., LangChain, LlamaIndex). Trade-offs: Moderate infrastructure needs, technical expertise, and access to state-of-the-art models, but reliance on external API services introduces vendor dependency and usage-based operational costs. Best Suited for: Institutions seeking a balanced approach with moderate resources, valuing access to advanced models and manageable overhead.
- 3) **Cloud-Based Integrated LLM Services:** Offers rapid and simplified deployment, essentially "one-click integration" cloud platforms with embedded LLMs and an integrated development framework (e.g., Alibaba Cloud PAI Service, eliminating server management and complex setup for instant LLM access). Trade-offs: Minimal technical overhead and low initial cost, but limited customizability. *Best Suited for:* Institutions prioritizing rapid deployment and minimal technology resources.

Choosing the right strategy ultimately depends on institutional resources (expertise, budget), data privacy needs, customizability requirements, and feasibility.

5. Discussion & Conclusions

This paper introduces an actionable 3-step roadmap—Needs Mapping to AI Solutions, LLM Selection, and Determine Integration Strategy—to streamline LLM integration in higher education. Its core innovation is accessible guidance for educators and administrators, regardless of AI expertise. While the proposed 3-step roadmap serves as a valuable foundational guide, real-world scenarios for each step are influenced by additional factors. In Step I (Map Needs to AI Solutions), one possible direction is to broaden the exploration of LLM applications in education beyond initial use cases. For Step II (LLM Selection), the choice of LLM can also depend on extra factors like the task type (complex tasks require more advanced models, and basic tasks can use foundational models). In Step III (Integration Strategy), determining optimal solutions may require incorporating pedagogical considerations alongside technical aspects.

In summary, this three-step roadmap represents a contribution as an accessible and practical guide for LLM integration within higher education. However, this roadmap should not be viewed as an endpoint but as a foundational starting point for continued progress in AI-enhanced education.

References

- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arnim, P., ... & Liang, P. (2021). On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258*. <https://arxiv.org/abs/2108.07258>
- de Fine Licht, K. (2023, August). Integrating large language models into higher education: Guidelines for effective implementation. In *Computer Sciences & Mathematics Forum* (Vol. 8, No. 1, p. 65). MDPI.
- Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., ... & He, Y. (2025). Deepseek-r1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- Morales-Chan, M., Amado-Salvatierra, H. R., & Hernandez-Rizzardini, R. (2024, July). AI-driven content creation: Revolutionizing educational materials. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale* (pp. 556-558).

- Xu, G., Looi, C. K., McMin, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307.
- Shetye, S. (2024). An evaluation of Khanmigo, a generative AI tool, as a computer-assisted language learning app. *Studies in Applied Linguistics and TESOL*, 24(1). doi:10.52214/salt.v24i1.12869.
- Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., Catanzaro, B. (2019). Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism. *arXiv preprint arXiv:1909.08053*. <https://arxiv.org/abs/1909.08053>
- Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., ... & Wen, Q. (2024). Large language models for education: A survey and outlook. *arXiv preprint arXiv:2403.18105*.
- Xing, W., Zhu, T., Wang, J., & Liu, B. (2024). A Survey on MLLMs in Education: Application and Future Directions. *Future Internet*.

Motivation and Engagement in Job Interview Training: A Study of a ChatGPT-facilitated Robot for Youths with Special Education Needs

Kuen-Fung SIN¹, Ka-Yan FUNG², Kwong-Chiu FUNG³, Wai-Ho MOW⁴, Hon-Shan Celia TAM⁵, Tze-Leung Rick LUI⁶, Lan YANG⁷, Yu-Xing TAO⁸

^{1,2,5,6,7,8}Institute of Special Needs and Inclusive Education, The Education University of Hong Kong, Hong Kong SAR, China

^{3,4}Hong Kong University of Science and Technology, Hong Kong SAR, China

kfsin@eduhk.hk, fkayan@eduhk.hk, kcfungag@connect.ust.hk, eewhmow@ust.hk, chstam@eduhk.hk, rtlui@eduhk.hk, yanglan@eduhk.hk, ytao@eduhk.hk

Abstract: *Special Education Needs (SEN) face different challenges, such as intellectual, physical, and special learning disabilities (SpLDs), that may impede students' capacity to convey their strengths during job interviews. Conventional job interview training often falls short due to the variation of instructor quality and a lack of individualized feedback, resulting in a loss of interest. To this end, we leverage a ChatGPT-facilitated robot to develop an interactive and tailored job interview training system for youths with SEN. By incorporating Self-Determination Theory (SDT), this work aims to enhance learning motivation and engagement for youths with SEN, ultimately fostering their confidence in job interviews. The findings show that participants are thriving in ChatGPT-facilitated robot training. Most participants agree that sufficient and individualized training can assist their real interview and build confidence. Future works will recruit diversified groups to join a longitudinal study.*

Keywords: ChatGPT-facilitated Robot, Self-determination Theory (SDT), Special Education Needs (SEN), Recruitment Interview, Motivation and Engagement

1. Introduction

Special Education Needs (SEN) is an umbrella term that covers a range of difficulties, such as hyperactivity, autism spectrum disorder (ASD), communication barriers, intellectual and special learning disabilities (SpLDs) (Hodkinson, 2015). These challenges may manifest in communication and social interactions, impacting students' ability to thrive in traditional educational settings (Smith, 2001). Addressing these difficulties is essential, as it not only enhances the learning experiences of SEN students but also plays a critical role in their lifelong development, including employment opportunities. With appropriate support, SEN youth can be empowered to develop skills necessary for successful integration into the workforce (Zárate-Rueda et al., 2021), ultimately fostering their independence and self-sufficiency (Mooney & Silver-Pacuilla, 2010). Therefore, our work aims to provide them with a digital job training system.

Establishing job interview skills is a critical gateway for youths with SEN to demonstrate their strengths to employers. However, many youths with SEN encounter challenges in effectively communicating their abilities during interviews, often struggling to articulate their skills and experiences, which hinders their employment prospects (Jacovkis, 2022). Traditional human-led training for job interviews has several limitations that negatively impact learning experience of youths with SEN. For example, there is variability in instructor quality, as not all trainers have the expertise or experience to effectively engage and support youths with SEN (Roorda et al., 2011). Additionally, many training programs lack personalized strategies tailored to each student's



Figure 1. (a) Minibo consists of four components: a pair of ears, a head, two hands, and four wheels. (b) A student with SEN was participating in job interview training with an instructor. (c) A student with SEN was engaging in job interview training with Minibo.

Florist Assistant

The responsibilities include assisting the shop manager and florist with the operations of the flower shop:

- 1) Learning about fresh flowers and basic floral arrangement:
 - Fresh flower: flower handling, common types of fresh flowers
 - Basic floral arrangement: types of flower pots, their uses, care, and handling and usage of floral foam
- 2) Flower shop operations:
 - Maintaining the cleanliness of the shop
 - Understanding and storing supplies/tools, and warehouse
- 3) Sales knowledge:
 - Understanding the differences in sales techniques for retail customers versus long-term ordering clients, including customer profiles and their purchasing habits
 - Learning to communicate with others as a florist assistant, including colleagues in the shop and customers

Possible interview questions from employers:

- 1) What are your interests?
- 2) What do you do in your leisure time?
- 3) Why are you interested in working in a flower shop?
- 4) Do you prefer administrative work or external communication? Why?
- 5) How would you handle difficulties encountered at work?

(a) Job description

(b) Potential interview questions

Figure 2. (a) is the job description, which is the job background. (b) are potential interview questions.

unique needs, resulting in a one-size-fits-all approach that may not resonate with learners (Yang, 2019). Furthermore, instructors would ask similar questions or provide similar feedback in repetitive training, leading to disengagement and boredom (Denton, 2022). Therefore, our work leverages novel technologies to elevate their learning experience and participation.

Incorporating technology, such as ChatGPT, can assist youths with SEN in their job interview preparation (Rizos et al., 2024). AI-empowered robots provide an interactive and tailored training experience that addresses unique needs of those with SEN (Fung et al., 2024; Fung et al., 2025a, Fung et al., 2025b). Unlike conventional training techniques, ChatGPT enables engaging conversations, providing a supportive environment for youths to hone their interview skills (Bettayeb, 2024). Moreover, customizable prompts can tailor personalized responses to youths with SEN and help them overcome their weaknesses, cultivating a sense of ownership (Kikalishvili, 2024). Leveraging the power of ChatGPT and robotics, we develop a more inclusive and interactive training system to prepare youths with SEN for successful career opportunities.

Self-determination theory (SDT) emphasizes the importance of learning motivation (Ryan & Deci, 2020). Autonomy, competence, and relatedness are identified as three psychological needs that are essential for escalating engagement, especially in youths with SEN (Yang et al., 2022). Autonomy controls an individual's learning progress (Guay, 2022). Competence is the desire for skill development, especially for youths with SEN and low self-efficacy (Pellerone, 2021). Relatedness highlights the importance of learning support to arouse youths' learning interests in SEN (Hehir et al., 2021). Our work utilizes SDT to design a system and interactive features that specifically support youths with SEN in their job interview training.

This study aims to evaluate the effectiveness of a ChatGPT-facilitated robot in enhancing engagement and motivation among youths with SEN during job interview training. This research contributes to a broader understanding of how technology can empower youths with SEN and improve their opportunities in the labour market. The results underscore the importance of incorporating technology into training programs, allowing educators better to address the diverse needs of youths with SEN. Furthermore, policymakers can allocate

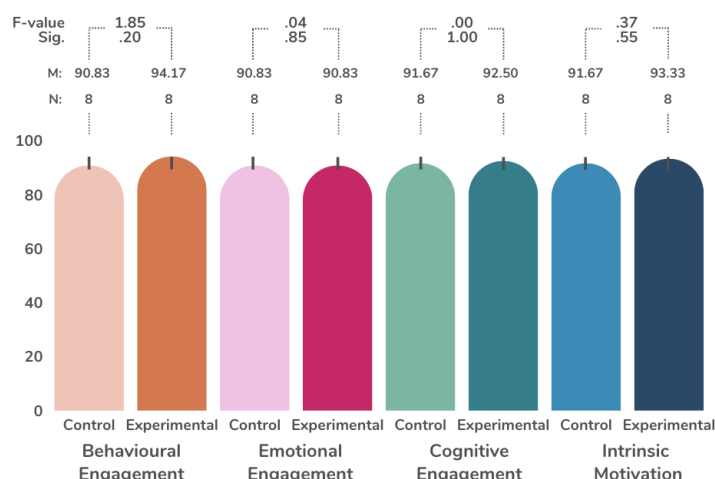


Figure 3. Overview of analysis of covariance (ANCOVA) for engagement and motivation.

more resources to develop accessible recruitment support systems for youths with SEN, ultimately facilitating their successful transition from education to employment.

2. System Design

In this work, we developed a job interview training system implemented on Minibo, as shown in Figure 1 (a), to assess ChatGPT-facilitated robots' effects on learning engagement for youths with SEN. The following sections delve into the robot's interactive elements, the ChatGPT-facilitated system, and the specifics of the job interview training materials.

2.1. Interactive Elements and Robots

The study utilized Minibo (Figure 1 (a)), which consists of four components: a pair of ears, a head, two hands, and four wheels. Minibo is designed to mimic human-like animacy and movement. It can be customized to offer supportive and encouraging facial expressions, fostering a sense of autonomy for youths with SEN. By meeting their unique needs, Minibo assists youths with SEN feel more connected and engaged, thereby enhancing their sense of relatedness during training.

2.2. ChatGPT-facilitated System

ChatGPT is a generative chatbot developed from large-scale language models released by OpenAI in November 2022 (OpenAI, 2022). It offers quick, multilingual responses in an intelligent and human-like manner (Ray, 2023). Minibo incorporates a ChatGPT-facilitated system with tailored prompts specifically designed to meet the unique needs of youths with SEN during job training.

2.3. Job Interview Details

Minibo is equipped with pre-set backgrounds and recruitment scenarios, enabling it to provide constructive feedback and engage in meaningful conversations with youths with Special Educational Needs based on their dialogue history. This personalized support improves the quality of responses, allows for performance tracking over time, and ultimately assists youths in building their confidence and competence. Figures 2 (a) and (b) illustrate the job description and potential interview questions.

3. Efficacy Experiment

We recruited 8 participants (4 females and 4 males) aged 19 to 23 (average age 21.48, with a standard deviation of 1.22) from a university in Hong Kong. All participants have special learning needs. During Week 1, they participated in conventional job interview training, which consisted of four sessions held over four days (Days 1 to 4). In Week 4, they engaged in ChatGPT-facilitated robotic job interview training, comprising four sessions over four days (Days 5 to 8).

The inclusion criteria for participants were as follows: (1) they are current students or fresh graduates from the university; (2) their mother tongue is Cantonese; and (3) they have no physical disabilities that could interfere with their interaction with the robot. Additionally, all participants had prior experience using digital tools. Informed consent was obtained from the participants before the experiment, and the University Institutional Review Board (IRB) approved the protocol. No remuneration was provided to participants.

In Week 1, participants (control group) completed pre-questionnaires on SDT (5 minutes). On Day 1, they participated in one job interview session with instructors. On Days 2 and 3, they conducted one job interview session with instructors. On Day 4, participants conducted a final job interview session with instructors. Then, they completed post-questionnaires on SDT (5 minutes).

In Week 4, participants (experimental group) conducted one job interview session with ChatGPT-facilitated robots. Participants completed post-questionnaires on SDT (5-minute) and conducted interviews (5-minutes).

Either instructors or ChatGPT-facilitated robots' interview were held for 15 minutes. To ensure a fair comparison, the same group of students participated in the same interview scenarios and answered the same questions. All instructors were well-trained and had relevant teaching experience at the university. Participants were allowed to choose their preferred topics for job interview training, i.e., offer two categories of job interviews.

4. Efficacy Verification

In this section, we discussed the changes in learning engagement and motivation. As shown in Fig. 3, the analysis of covariance (ANCOVA) result revealed that job interview training with interactive robots can moderately improve emotional, cognitive engagement, and intrinsic motivation for those youths with SEN compared with conventional training.

4.1. Data Analysis

Behavioural engagement. The experimental group improved by 3.67%, $p = 0.48$ (pre-test: $\bar{x} = 90.83$, $\sigma = 9.39$; post-test: $\bar{x} = 94.17$, $\sigma = 9.04$). However, the control group slightly decreased by 0.91%, $p = 0.86$ (pre-test: $\bar{x} = 91.67$, $\sigma = 8.54$; post-test: $\bar{x} = 90.83$, $\sigma = 10.04$).

Emotional engagement. The experimental group improved by 3.81%, $p = 0.63$ (pre-test: $\bar{x} = 87.50$, $\sigma = 15.71$; post-test: $\bar{x} = 90.83$, $\sigma = 10.65$). Also, the control group slightly increased by 1.87%, $p = 0.82$ (pre-test: $\bar{x} = 89.17$, $\sigma = 15.09$; post-test: $\bar{x} = 90.83$, $\sigma = 13.30$).

Cognitive engagement. The experimental group improved by 2.78%, $p = 0.66$ (pre-test: $\bar{x} = 90.00$, $\sigma = 12.85$; post-test: $\bar{x} = 92.50$, $\sigma = 9.04$). The control group also increased by 3.77%, $p = 0.64$ (pre-test: $\bar{x} = 88.33$, $\sigma = 16.62$; post-test: $\bar{x} = 91.67$, $\sigma = 10.54$).

Intrinsic motivation. The experimental group improved by 5.66%, $p = 0.37$ (pre-test: $\bar{x} = 88.33$, $\sigma = 12.22$; post-test: $\bar{x} = 93.33$, $\sigma = 9.43$). The control group also increased by 2.80%, $p = 0.72$ (pre-test: $\bar{x} = 89.17$, $\sigma = 12.82$; post-test: $\bar{x} = 91.67$, $\sigma = 14.58$).

4.2. Interview with Instructors

An experienced instructor indicated that youths may perform poorly in interviews due to a lack of confidence, and anxiety can hinder their ability to showcase their skills effectively. Additionally, barriers to communication and expression can negatively impact the overall outcome of the interview. The instructor further mentioned, "Instructors may encounter unexpected disruptive behaviours from trainees during training sessions, making it difficult to manage the situation. Additionally, the individual differences among youths require instructors to tailor and adjust the training content to meet each student's needs."

4.3. Interview with Participants

In general, participants are motivated in the training sessions. However, most participants encountered difficulties before training with robots, such as nervousness and insufficient confidence. P1 told us, "I feel a bit nervous at the beginning. This training allows me to take my time rather than feeling rushed during real interviews. It helps me build confidence" After the robot training, most participants reported feeling more relaxed and confident. They almost agreed that adequate and interactive training can lead to success. P3 shared, "I think the training is essential because it is very useful for future internship interviews. With sufficient training and feedback, I feel confident I can handle the interview independently." Most participants suggested that the answering time of the interview training with robots could be longer.

4.4. Observation during Training Sessions

During interview training with instructors, some participants felt nervous and hesitated in their responses due to a lack of understanding of the questions. This resulted in incoherent answers, poor eye contact, and tense body language. Participant, F4, initially did not want to learn with robots. However, he expressed likelihood to continue the training after the first trial, finding the robot cute and enjoyable to communicate with.

5. Conclusion and Future Work

In this study, we developed a ChatGPT-facilitated robot to deliver inclusive and interactive job interview training for youths with SEN. Most youths enjoyed the training, finding the robots cute and their human-like facial expressions helpful in reducing stress. In the future, we will provide additional training scenarios, implement interactive elements, and fine-tune robots' functions to enhance youths' learning experience.

References

- Bettayeb et al. (2024). Exploring the impact of ChatGPT: conversational AI in education. In *Frontiers in Education* (Vol. 9, p. 1379796). Frontiers Media SA.
- Denton. (2022). Finding Motivation and Connectedness for Learning with Special Education Students Qualifying Under Emotional Disturbance.
- Fung et al. (2024). Humanoid robot-empowered language learning based on self-determination theory. *Education and Information Technologies*, 1-30.
- Fung et al. (2025a). Motivating Students with Different Needs to Learn Chinese in a Mixed-Background Classroom by Robot-Assisted Learning. *IEEE Transactions on Learning Technologies*.

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Fung et al. (2025b). Effects of Different Demographics when Evaluating Interactive Robot-Empowered Language Learning in Human-Robot Interaction. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 1010-1014). IEEE.
- Guay. (2022). Applying self-determination theory to education: Regulations types, psychological needs, and autonomy supporting behaviors. *Canadian Journal of School Psychology*, 37(1), 75-92.
- Hehir et al. (2021). Developing student connectedness under remote learning using digital resources: A systematic review. *Education and information technologies*, 26(5), 6531-6548. pandemic: A study of a group of Italian school teachers. *European Journal of Investigation in Health, Psychology and Education*, 11(2), 496-512.
- Hodkinson. (2015). Key issues in special educational needs and inclusion. Sage.
- Jacovkis et al. (2022). When Arriving Is Not Enough—Constraints in Access to Education and Employment Opportunities for Migrant Youth. *Societies*, 12(3), 95.
- Kikalishvili, S. (2024). Unlocking the potential of GPT-3 in education: Opportunities, limitations, and recommendations for effective integration. *Interactive Learning Environments*, 32(9), 5587-5599.
- Mooney, M., & Silver-Pacuilla, H. (2010). Literacy, Employment and Youth with Learning Disabilities: Aligning Workforce Development Policies and Programs. National Institute for Literacy.
- OpenAI (2022). Chatgpt: Optimizing language models for dialogue. Retrieved from <https://openai.com/blog/chatgpt/> (retrieval date: Dec 22, 2024)
- Pellerone, M. (2021). Self-perceived instructional competence, self-efficacy and burnout during the covid-19 pandemic: A study of a group of Italian school teachers. *European Journal of Investigation in Health, Psychology and Education*, 11(2), 496-512.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154.
- Rizos et al. (2024). Enhancing Mathematics Education for Students with Special Educational Needs through Generative AI: A Case Study in Greece. *Contemporary Educational Technology*, 16(4).
- Roorda et al. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of educational research*, 81(4), 493-529.
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary educational psychology*, 61, 101860.
- Smith, D. D. (2001). Introduction to special education: Teaching in an age of opportunity.
- Sin, K. F & Chan, W. L (2016). Career orientation and career development efficacy of students with special learning needs. <https://www.ktmss.org.hk/wp-content/uploads/2016/11/「有特殊學習需要學生的職業導向及職業發展效能感」調查報告2016.pdf>. (retrieval date: 10 December 2024)
- Yang et al. (2022). The effects of school support on school engagement with self-determination as a mediator in students with special needs. *International Journal of Disability, Development and Education*, 69(2), 399-414.
- Yang et al. (2019). From one-size-fits-all teaching to adaptive learning: the crisis and solution of education in the era of AI. In *Journal of Physics: Conference Series* (Vol. 1237, No. 4, p. 042039). IOP Publishing.
- Zárate-Rueda et al. (2021). Inclusive education and labour market insertion from a capabilities approach: a phenomenological and functional diversity perspective. *Revista de Investigación Educativa*, 39(1), 265-282.

Are They Ready? Analyzing Chinese Preservice Teachers' Acceptance to Use Gen AI in Education

Yujie KUANG^{1*}, Jiateng ZHOU²

¹Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR, China

²Faculty of Education, Beijing Normal University, Beijing, China
ykuang@eduhk.hk, jiatengzhou_edu@163.com

Abstract: As the integration of Generative AI into education becomes an emerging trend, preservice teachers' understanding and acceptance of Gen AI play a crucial role in its educational application. This study aims to investigate the acceptance of Gen AI among preservice teachers and identify the influencing factors. Utilizing the UTAUT model, semi-structured interviews were conducted with 13 preservice teachers, and thematic analysis was employed to analyze the interview transcripts. The findings indicate that performance expectancy, effort expectancy, and ethical concerns significantly affect preservice teachers' acceptance of Gen AI, with performance expectancy and effort expectancy having the most substantial impact.

Keywords: UTAUT model; Generative AI; Preservice teacher

1. Introduction

Generative AI (Gen AI) has attracted much attention after the emergence of ChatGPT, and has significantly impacted K-12 education, serving as a teaching, peer, teachable, and motivational agent to enhance learning performance and engagement (Kuhail et al., 2023). Teachers' technology acceptance plays a significant role in the adoption and effective use of technology; high acceptance of technology is a predictor of successful integration in the classroom (Seufert et al., 2021). While it offers benefits like personalized feedback and improved learning outcomes, risks such as privacy concerns, cheating, bias, and inappropriate interactions pose challenges. Teachers' acceptance of Gen AI is crucial for successful integration, highlighting the need for research on its adoption in educational settings.

Preservice teachers are the teachers of the future, facing future education. Before they even enter their careers, AI is already causing massive changes and concerns at all levels of education, from kindergartens to higher education. Therefore, preservice teachers' acceptance of Gen AI strongly affects their use of related tools and applications in their future careers and how well they can adapt to the AI era. However, current research has mainly focused on in-service teachers, with limited studies exploring the acceptance of preservice teachers toward AI technology. Therefore, it is essential to understand preservice teachers' perspectives on integrating artificial intelligence into education and the influencing factors.

The UTAUT model exhibits strong explanatory and predictive power in identifying the factors that influence users' acceptance and use of technology, because the UTAUT model integrates the strengths of eight existing models, making the model's predictive power is over 70%, exceeding that of any previous technology acceptance model (Venkatesh et al., 2003). The UTAUT model includes four important constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions, which are direct determinants of Behavioral Intention. Performance expectancy refers to the extent to which a user believes that using the new system would enhance their job performance; effort expectancy refers to the degree of ease for users to use technology; social influence is the extent to which users perceive that important other (e.g., family and friends) believe they should use a particular technology; facilitating conditions refer to an individual's beliefs about the availability and effectiveness of technical and allied infrastructure to support the usage of a new technology (Venkatesh et al. 2003).

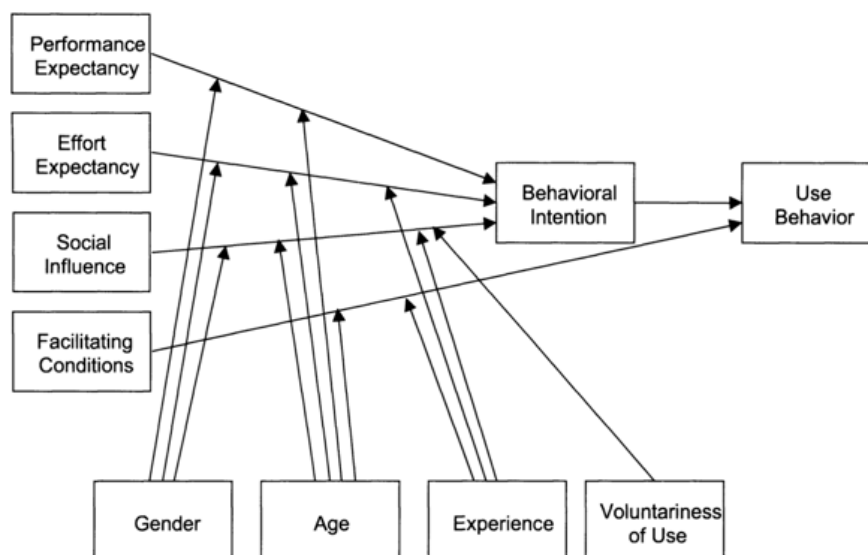


Figure 1. UTAUT Model.

To summarize, this qualitative research adopts the UTAUT model to explore the acceptance of Gen AI among preservice teachers in China, aiming to address the following questions: What are the influence factors of preservice teachers' acceptance of Gen AI?

2. Methodology

2.1. Semi-structured Interviews

This research uses semi-structured interviews to explore Chinese preservice teachers' acceptance of Gen AI using the UTAUT model. The interview method, as one of the qualitative research methods, is considered to provide a "deeper" understanding of social phenomena (Silverman, 2013). The interview consists of three parts: the first part is to understand the basic information of interviewees, including their universities, grades, majors, and development plans. The second part asks about the interviewees' perspectives on Gen AI, including their understanding of Gen AI's basic features, educational usage, possible implications for the education landscape, and their opinions about the technology education content of their university's teacher education program. The third part consisted of questions based on the four UTAUT constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions. All questions were open-ended.

2.2. Participants

The interviewees in this study are college students who are enrolled in teacher education programs at universities in mainland China. Purposive sampling and snowball sampling were used to obtain a certain number of interviewees from different subjects and universities. All the interviewees have a basic understanding of the concept of Gen AI and have been using Gen AI for more than 3 months; their usage records of Gen AI were checked by the researcher before conducting interviews. The detailed information of the interviewees is shown in the table below.

Table 1. Detailed Information of Interviewees.

ID	Education Level	Teaching Level	Subject	Gender	Age
1	Undergraduate	Preschool	/	Female	22
2	Undergraduate	Preschool	/	Female	22
3	Undergraduate	High School	Geography	Male	22
4	Postgraduate	High School	English	Female	24
5	Undergraduate	High School	Chinese	Female	22
6	Undergraduate	High School	Chinese	Female	23
7	Undergraduate	High School	History	Male	23
8	Undergraduate	Primary School	History	Female	21
9	Postgraduate	High School	Mathematics	Female	24
10	Postgraduate	Primary School	Science	Male	25
11	Undergraduate	Preschool	/	Female	21
12	Postgraduate	Primary School	Science	Male	24
13	Undergraduate	Primary School	English	Male	23

2.3. Data Collection and Analysis

Interviews were one-on-one communications between the interviewee and the researcher and were conducted using online video conferencing tools such as Tencent Meetings or Zoom. Interviews lasted roughly fifty to eighty minutes, depending on the interviewee's confirmation that he or she had said everything about the topic and had nothing to add. To establish the trustworthiness of the transcription and translation, the interviewees were asked to verify the accuracy of the transcripts before analysis.

The interview transcripts were analyzed using NVivo 12 qualitative analysis software, employing thematic analysis for coding. The coding process followed the six steps proposed by Braun and Clarke (2006): familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report.

3. Findings and Discussion

Interviewees agreed that the advancement of AI technology is an inevitable trend, holding the potential to bring significant convenience and innovation to education. However, words like "danger", "caution", and "prudence" were frequently mentioned by interviewees, indicating their awareness and concerns regarding the potential risks of integrating Gen AI technology into education. This research found that performance expectancy, effort expectancy, and ethical concern are the main factors affecting preservice teachers' acceptance of Gen AI.

3.1. Performance Expectancy

Performance expectancy refers to the degree to which users believe that using a new system will enhance their job performance. This study found that Gen AI significantly aids preservice teachers in their learning and internships, indicating a high level of performance expectancy. Specifically, Gen AI can promote professional development and reduce the workload of preservice teachers. However, the issue of inaccurate information from Gen AI negatively impacts the performance expectancy of preservice teachers.

The first thing that teachers value is Gen AI facilitating professional development. This study finds that Gen AI significantly enhances the professional development of preservice teachers in knowledge acquisition and practical guidance. For example, Gen AI can be used as a tool for information retrieval and question answering, providing personalized tutoring for preservice teachers. This function is particularly prominent during their educational internships. As interviewee 10 mentioned, "(Preservice teachers) may not have access to good tutors or may feel too embarrassed to ask questions. Gen AI largely fills this gap, providing strong guidance in personalized learning for teachers." Besides, Gen AI's text generation and embellishment functions provide great help to preservice teachers' instructional design. Therefore, instructional design has become the main purpose when preservice teachers use Gen AI. "Gen AI is a tool for me, mainly to help me do instructional design." (interviewee 12), "I asked Gen AI to give me 10 to 20 exercises on definite clauses, of different types, such as grammatical fill-in-the-blanks and multiple-choice questions" (Interviewee 4).

Teachers' second performance expectancy is reducing work burden. Teachers' work burden is a universally acknowledged issue, with teachers often burdened not only by the substantial responsibilities within their teaching roles but also by numerous extracurricular tasks (Timms et al., 2007). Many interviewees in this study also expressed that, aside from teaching, teachers often bear repetitive and cumbersome "incremental burdens," such as drafting administrative documents. Gen AI can assist teachers in completing tedious textual materials, significantly alleviating their workload. Therefore, Gen AI can serve as a "work assistant" for teachers.

However, despite the general acknowledgment by interviewees of the significant assistance Gen AI provides for their learning and performance, seven interviewees mentioned encountering inaccuracies in the information generated by Gen AI. The lack of accuracy in Gen AI's responses could lead to preservice teachers' increased time and effort spent on secondary revisions, or even errors in teachers' work, thereby significantly impacting preservice teachers' acceptance towards using Gen AI. "... Using Gen AI also requires me to double-check because there may be many obvious errors, especially in subjects related to literature, history, and philosophy." (interviewee 7).

3.2. Effort Expectancy

Interviewees generally praised the ease of operation and convenience of using Gen AI, expressing appreciation for its role as an all-in-one platform. They found that Gen AI could serve as an all-in-one solution tool, combining functions such as consultation, retrieval, information organization, text editing, and refinement together, efficiently addressing problems. Compared to manually handling issues, using Gen AI is much more convenient, as stated by one interviewee: "If it helps me search and organize, it's much more convenient and faster than doing it all by myself." Specifically, this study identified three second-order concepts regarding effort expectations: operational complexity and cognitive burden.

3.2.1. Operational complexity

All the interviewees indicate that the operation of Gen AI is not complicated; they do not need to invest too much time and effort in mastering basic AI operations, such as engaging in dialogue, retrieval, and language editing and refinement using Gen AI. Most of the time, the usage of Gen AI only requires simple conversations with Gen AI or giving instructions to obtain satisfactory results. Interviewee 9 called Gen AI a "fool tool", which means that the operation is so simple that even a fool can use it easily. Thus basically, the operational complexity of Gen AI is not high: "I think learning to use it is relatively simple." (Interviewee 13)

However, despite the simplicity of using Gen AI, more effective utilization requires strategy and appropriate questioning techniques. Only four respondents mentioned the professional and complex use of Gen AI, such as "training the AI" and "using prompts." For example, interviewee 9 stated, "I had no idea that Gen AI needed to be trained. I didn't know it wasn't just a simple fool tool until my friend taught me." Most preservice teachers neither spend time and effort to proactively learn specialized usage methods nor have access to professional guidance. They tend to view Gen AI as a "fool tool" and use it only at the level of basic dialogue interactions with little operational complexity.

3.2.2. Cognitive burden

Gen AI imposes a relatively low cognitive burden on users. The conversational format and high-quality content generation of Gen AI significantly reduce the cognitive burden during both its use and the comprehension of the content it generates.

Firstly, Gen AI can simplify complex information. As several respondents mentioned, Gen AI can achieve knowledge mastery through a step-by-step approach to questioning. This step-by-step, question and-answer dialogue approach helps respondents effortlessly grasp the information they need compared to the effort required to search for and organize information from websites, books, or articles by themselves.

Secondly, Gen AI can provide instant feedback. Gen AI can offer immediate feedback and guidance, swiftly providing respondents with the answers they seek. This instant feedback mechanism can reduce respondents' confusion and uncertainty, thus lowering their cognitive load (Holmes et al., 2019)

3.3. Ethical Concern

This study found that ethical concerns significantly influence preservice teachers' acceptance of the use of Gen AI, including students' improper usage and potential biases and ideologies. The first ethical concern is students' improper usage. Interviewees in this study expressed similar concerns about students' using technology to plagiarize instead of thinking by themselves. One interviewee even reported that she has already discovered students using Gen AI tools to write essays. This has become a major reason why teachers are hesitant to integrate Gen AI tools in classroom teaching.

Another ethical concern pertains to the potential biases and ideologies of Gen AI. Given that students are mentally and physically immature, making them vulnerable to potentially dangerous beliefs and thoughts. Some interviewees expressed worries about Gen AI spreading ideologies that may be detrimental to students,

such as extreme individualism. Once students become the primary users of technology, it becomes challenging to effectively regulate their usage and behavior. While governmental bodies may exert regulatory oversight over domestic Gen AI tools and their interactions with minors, they face limitations in controlling AI tools originating from overseas.

4. Conclusion

This study found that performance expectancy and effort expectancy within the UTAUT model all impact preservice teachers' acceptance of Gen AI. Among these factors, performance expectancy and effort expectancy have the greatest influence. Social influence primarily affects preservice teachers during the initial stages of Gen AI acceptance, while teacher education courses, as a facilitating condition, play a crucial role in enhancing Gen AI acceptance. Additionally, the study identified that ethical concern significantly influences preservice teachers' Gen AI acceptance. Ethical concern comprises two second-order concepts: students' improper usage and potential biases and ideologies.

Chinese preservice teachers generally exhibit a strong intention to use Gen AI in education. However, despite their high intention and interest in using Gen AI, they face a significant lack of professional instruction on its application in educational settings. This underscores the need for higher education institutions to integrate comprehensive Gen AI usage guidance into their teacher education curriculum.

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education. Promise and Implications for Teaching and Learning*.
- Kuhail, M. A., Alturki, N., Alamlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973–1018.
- Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 106552.
- Silverman, D. (2013). *Doing Qualitative Research: A Practical Handbook* (Fourth edition). Sage Publications Ltd.
- Timms, Carolyn & Brough, Paula. (2013). "I like being a teacher" Career satisfaction, the work environment and work engagement. *Journal of Educational Administration*. 51. 768-789.
- Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27, 425–478.

In search of a “free rider” in the age of Gen-AI: intentions and outcomes of using Gen-AI in group writing

Anna KORCHAK^{1*}, Mik FANGUY², Anastasia KAPUZA³, Jamie COSTLEY⁴

^{1,3}National Research University Higher School of Economics, Moscow, Russia

²The Education University of Hong Kong, Hong Kong SAR, China

⁴United Arab Emirates University, Al Ain, United Arab Emirates

aekorchak@hse.ru, mfanguy@eduhk.hk, avkapuza@hse.ru, jcostley@uaeu.ac.ae

Abstract: *The rapid integration of Generative Artificial Intelligence (Gen-AI) into collaborative learning has redefined how students participate in group writing tasks. While Gen-AI facilitates ease and speed in writing, it also raises concerns regarding free-riding behavior, where some students might exploit AI tools to contribute minimally while still receiving group credit. This study investigates how the intentions and outcomes of using Gen-AI in group work relate to both group and individual performance. Data collected from 108 graduate STEM students at a large Korean technical university revealed that using Gen-AI for improving group products led to higher group scores, while using it for skill improvement did not significantly enhance individual performance. The findings suggest that intentionality in Gen-AI use plays a crucial role, with unexpected positive outcomes emerging when students did not initially aim to use AI for ease or skill enhancement. These results highlight the need to reconsider assessment and instructional strategies in collaborative learning environments where Gen-AI is present.*

Keywords: collaborative learning, free-riding, generative artificial intelligence, higher education

1. Introduction

The emergence of generative artificial intelligence (Gen-AI) has transformed collaborative learning, particularly in the context of group writing. As Gen-AI becomes increasingly integrated into educational settings, its role as an enabler and disruptor of collaborative efforts has come under scrutiny. Gen-AI facilitates rapid content generation and streamlines the writing process, providing students with a valuable tool to enhance productivity and efficiency (Kim & Cho, 2024; Luther et al., 2024). It also raises concerns about free-riding behaviour, where some students leverage AI tools to contribute minimally while still receiving group credit (Aitken & Hatt, 2012; Khandaker & Soh, 2010). Traditionally, free riders were easily identified through their minimal or absent contributions. However, with Gen-AI, even students who produce substantial text might demonstrate low engagement. This shift complicates the identification of free riders and calls into question the fairness and effectiveness of collaborative learning practices in the age of AI. As educators continue to navigate this evolving landscape, it becomes crucial to understand how Gen-AI use intentions impact group and individual outcomes. This study aims to investigate the role of Gen-AI in collaborative writing, focusing on the relationship between students' intentions to use Gen-AI and the resulting group and individual performance outcomes (Nazari et al., 2021).

2. Method

This study was conducted with 108 graduate STEM students at a Korean technical university. Participants were enrolled in a Scientific Writing course that included four collaborative writing tasks, with unrestricted Gen-AI use. Data were collected via a survey after the course completion, focusing on Gen-AI use intentions (ease, quickness, product improvement, skills development) and perceived outcomes. The two survey items used in this study are listed below. Both items allowed students to select multiple responses.

Item1. How do you feel about the results of your constituent strategy for using Gen-AI?

It made the process of writing easier

It made the process of writing quicker

It improved the quality of our final written product

It improved my writing skills

Item2. What was a priority of your constituent strategy?

To make the process of writing easier

To make the process of writing quicker

To improve the quality of final written product

To improve my writing skills

The analysis involved t-tests to evaluate the relationship between Gen-AI use intentions and group/individual scores.

3. Results

The analysis revealed that intentions to use Gen-AI for ease and quickness did not correlate with lower group or individual scores ($t = -1.32$, $p > 0.05$). Instead, students who did not plan to use Gen-AI for ease but found it helpful in streamlining the writing process achieved significantly higher group and individual scores compared to those who intended to use it for ease ($t = 2.57$, $p < 0.01$). In contrast, using Gen-AI to improve group products led to significantly higher group scores ($t = 3.41$, $p < 0.001$), but no significant improvement in individual performance was noted ($t = -0.84$, $p > 0.05$). Similarly, skill development intentions did not result in enhanced performance ($t = 0.73$, $p > 0.05$), though unplanned skill gains were associated with higher individual scores ($t = 2.10$, $p < 0.05$). These findings indicate that Gen-AI use patterns and intentions significantly influence group and individual performance outcomes, highlighting the complex interplay between AI application strategies and collaborative writing success.

4. Discussion

The findings of this study challenge the conventional view of free-riding behavior in collaborative learning. Traditionally, free riders were identified based on minimal contribution to group work (Aitken & Hatt, 2012; Khandaker & Soh, 2010). However, with Gen-AI enabling students to produce substantial text with minimal effort, the definition of a free rider needs to be reconsidered. This study demonstrates that intentions to use Gen-AI for ease and quickness do not necessarily result in lower group or individual scores. On the contrary, unintended ease in writing through Gen-AI can even contribute to improved performance. This unexpected positive outcome aligns with findings by Nazari et al. (2021), who reported that students experience increased comfort and confidence when using Gen-AI for writing. The most notable finding is that the intention to use Gen-AI to improve group products is associated with higher group scores, but this does not necessarily translate to enhanced individual performance. This gap indicates that although Gen-AI can enhance the quality of collaborative outputs, it may not foster personal skill development. Additionally, the observed “surprise” effect, where unplanned skill improvements occur, implies that unintentional learning through Gen-AI might happen when students do not explicitly aim for skill enhancement. This phenomenon aligns with Wang’s (2022) concept of AI effects perception, where unexpectedly high outcomes result from low initial expectations. These results emphasize the importance of intentionality in Gen-AI use within collaborative settings. Educators should be aware that encouraging students to deliberately plan their AI use may not always yield the intended learning outcomes. Instead, fostering an environment that prompts reflection on the use of AI might lead to more meaningful educational gains (Kim et al., 2024; Luther et al., 2024).

5. Practical Recommendations

Educators should guide students to reflect on their intentions when using Gen-AI in collaborative tasks, encouraging them to critically evaluate their reliance on AI and its impact on personal learning outcomes. Assessment strategies should be adapted to account for both intended and unintended effects of Gen-AI use, ensuring that grading criteria do not solely reward the quality of group products but also consider individual contributions and skill development. When allowing Gen-AI in group tasks, educators should design activities that promote active engagement with both AI-generated content and peer input to prevent over-reliance on AI for ease or quickness. Institutions may consider implementing training on effective Gen-AI use to help students understand how to balance efficiency with genuine skill acquisition, fostering deeper learning rather than surface-level achievements.

References

- Aitken, A., & Hatt, G. (2012). Students taking notes and creating summaries together (or not). In A. Herrington, J. Schrape, & K. Singh (Eds.), *Engaging students with learning technologies* (pp. 147-166). Curtin University. <http://hdl.handle.net/20.500.11937/7870>
- Khandaker, N. and Soh, L. (2010). Assessing individual contributions to groupwork by a summary of tracked computer-supported collaborative activities. *2010 IEEE Frontiers in Education Conference (FIE)*. <https://doi.org/10.1109/fie.2010.5673505>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Kim, J., & Cho, Y. H. (2023). My teammate is AI: understanding students' perceptions of student-AI collaboration in drawing tasks. *Asia Pacific Journal of Education*, 1-15. <https://doi.org/10.1080/02188791.2023.2286206>
- Luther, T., Kimmerle, J., & Cress, U. (2024). Teaming up with an AI: Exploring Human–AI Collaboration in a Writing Scenario with ChatGPT. *AI*, 5(3), 1357-1376. <http://dx.doi.org/10.3390/ai5030065>
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: randomised controlled trial. *Heliyon*, 7(5). <https://doi.org/10.1016/j.heliyon.2021.e07014>

Enhancing EFL Speaking Skills Through Artificial Intelligence

Technology: Opportunities and Challenges

Junyi SHEN^{1*}, Bojie QIAN²

^{1,2}Wuhan College of Communication, Wuhan, China

²Universiti Teknologi Malaysia, Malaysia
303593407@qq.com, cynthiaqbj@163.com

Abstract: *This study investigates the impact of AI speaking tools on enhancing English as a Foreign Language (EFL) speaking skills among sophomore English major students at a university in China. The study employs a mixed-methods approach, incorporating AI speaking tools, a speech assessment platform, and semi-structured interviews. Two classes of sophomore English major students were selected: an experimental class using the AI app (EAP Talk) for speaking practice over a four-week period, and a control class without AI assistance. The findings reveal notable improvements in speaking skills—including pronunciation, grammar, vocabulary, and fluency—among the experimental class, and highlight both the effectiveness and the existing problems of AI-assisted language learning.*

Keywords: English as a Foreign Language (EFL), Speaking Proficiency, AI Tools

1. Introduction

Speaking proficiency is a critical skill for English as a Foreign Language (EFL) learners (Egan, 1999), enabling effective communication in real-world contexts. Traditional classroom settings, however, often fail to provide sufficient opportunities for meaningful and interactive speaking practice. AI-powered tools have emerged as a promising solution, offering personalized and flexible speaking practice beyond the classroom (Kok et al., 2009). Previous studies suggest that these tools can enhance speaking skills by providing instant feedback and fostering self-directed learning, though problems such as technical limitations (Balci, 2024) and varying levels of digital literacy remain (Kasneci et al., 2023).

Building on this foundation, this study seeks to verify whether AI tools can effectively support speaking skill development by comparing an experimental class using AI-assisted applications for speaking practice with a control class relying on traditional methods. Using pre- and post-test data, alongside insights from interviews, the research aims to evaluate AI tools' effectiveness on learners' speaking proficiency and identify the problems of their integration into EFL learning.

2. Literature Review

In the context of English as a Foreign Language (EFL), speaking is widely recognized as a fundamental skill to acquire, requiring consistent practice for mastery (Gunasekaran & Subramaniam, 2022). Speaking proficiency encompasses five key components: pronunciation, grammar, vocabulary, fluency, and comprehension (Harris, 1974). Syakur and Azis (2020) emphasize that these components are interdependent and cannot be developed in isolation. This difficulty of speaking practice is compounded by limited opportunities for target language use outside the classroom (Alsatuey, 2011) and restricted practice time in overcrowded classrooms.

Artificial intelligence (AI) refers to the creation of computer systems that can perform human-like cognitive processes such as learning, reasoning, and self-correction (Kok et al., 2009). The advent of mobile technology has introduced an appealing alternative, allowing learners to practice speaking anytime and anywhere, thereby addressing the need for consistent and extensive oral practice (Ahn & Lee, 2016).

AI technology, particularly through automatic speech recognition (ASR), has facilitated the development of various language learning tools for EFL speaking practice. With the growing popularity of mobile-assisted language learning (MALL), numerous AI-powered apps such as Liulishuo, EAP Talk, have emerged (Lehman et al., 2020; Chang et al., 2021; Li & Zhou, 2022). These apps offer distinct advantages, including rich materials, user-friendly interfaces, and support for autonomous learning (Liu et al., 2019). AI technologies provide EFL learners with interactive, real-time speaking practice. These tools facilitate fluency development, instant feedback, and progress tracking, promoting learner autonomy. The flexibility of AI applications makes them an appealing option for students seeking extra speaking practice.

Several studies have documented the benefits of these tools for EFL learners. Ahn and Lee (2016) highlighted students' enthusiasm for speech recognition features, which provided immediate feedback on their

spoken input. Similarly, Li and Zou (2022) observed that learners practicing with AI speaking apps reported increased confidence and frequency of speaking practice due to perceived improvements in pronunciation, fluency, and oral rhythm. Moreover, Yang et al. (2022) noted that task-based AI chatbots fostered a positive and comfortable environment for English speaking practice. Zou et al. (2023) further emphasized students' positive perceptions of interactive activities provided by AI apps, which enhanced engagement and learning outcomes.

Despite these benefits, significant challenges remain. Kukulska-Hulme and Shield (2008) identified the limited availability of MALL programs specifically designed for oral proficiency development, calling for further exploration in this area. Additionally, Kasneci et al. (2023) underscored the necessity for learners and educators to develop digital literacies to navigate the limitations and brittleness of large language models. Specific drawbacks of AI speaking apps include restricted user control over practice content, as noted by Fouz-González (2020), and the perceived inadequacy of feedback from scoring systems, as reported by Li and Zou (2022). These challenges highlight the need for more comprehensive and user-centered app designs to maximize their educational potential.

However, several gaps remain in existing studies. For instance, while studies focus on short-term outcomes such as improvements in pronunciation, fluency or grammar, the long-term impact of AI tools on speaking proficiency remains underexplored. Additionally, there is limited research on how these tools affect psychological aspects of learning, such as self-confidence. While this study does not directly address all the gaps, it does aim to explore the effectiveness and problems of AI tools in enhancing EFL speaking skills. Therefore, this study seeks to address two research questions:

1. What is the effectiveness of using AI tools for EFL speaking practice?
2. What existing problems do students encounter when using AI-assisted tools for speaking practice?

3. Methodology

3.1. Participants

This study involved sophomore English major students from a university in China, with a total of 70 participants from two classes, divided into experimental class and control class.

Experimental class: 36 students, who used AI-based speaking apps (EAP Talk) for speaking practice over a four-week period. These students were required to engage in AI-assisted speaking practice five times a week, with each session lasting over 15 minutes.

Control class: 34 students, who did not use AI tools for speaking practice and continued with traditional learning methods without additional AI assistance.

The participants were all native speakers of Chinese who were studying English as a foreign language.

3.2. Instruments and Procedures

The study employed a mixed-methods approach, incorporating AI speaking apps, a speech assessment platform, and semi-structured interviews. Figure 1 shows the detailed steps.

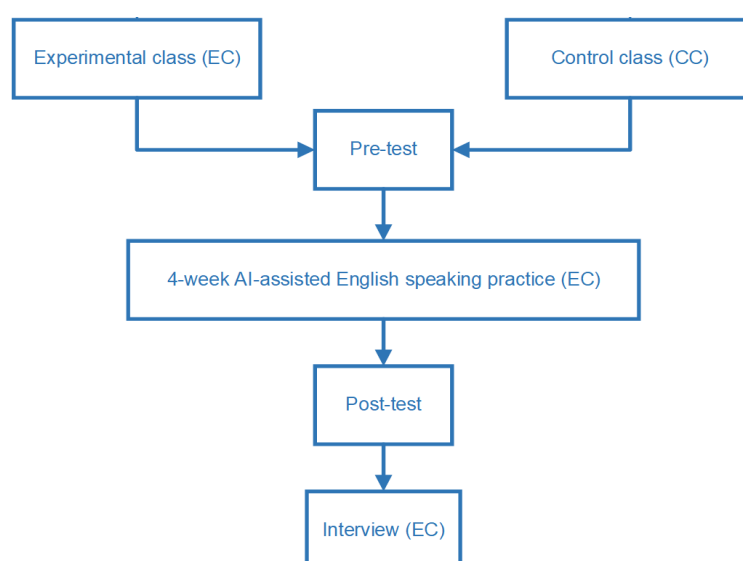


Figure 1. Quasi-experiment plan.

To evaluate speaking proficiency, both the experimental and control classes undertook pre- and post-tests using the SpeechAce speaking test (<https://www.speechace.com/speaking-test>), which assessed pronunciation, grammar, vocabulary, fluency, and overall performance. Each test comprised three questions related to a specific topic. The scoring system for each component, as well as the overall score, had a maximum of 10 points, following a structure similar to the scoring standards used in the IELTS speaking test. As a sound recognition system (Alnifasah, 2022), SpeechAce demonstrates strong test-retest reliability (Pearson's $r = 0.82$) when compared against blinded human IELTS examiner scores, as validated through randomized sampling of test-takers (Zou et al., 2023). This standardized approach ensured consistency and reliability in evaluating the students' speaking proficiency before and after the intervention. During the four-week intervention, the experimental class used the EAP Talk AI speaking app, engaging in five weekly practice sessions, each lasting at least 15 minutes. These sessions included interactive exercises simulating real-life speaking scenarios across various topics and situations. In contrast, the control class participated in traditional peer-based speaking exercises without AI assistance. At the end of the four-week period, both classes completed the post-test to evaluate any changes in speaking proficiency, following the same format and criteria as the pre-test.

Following the intervention, nine students from the experimental class, selected based on different academic performance, gender, and personalities, were invited to participate in semi-structured interviews. These interviews explored their four-week experiences with the AI speaking app, focusing on perceived benefits and challenges to identify areas for improvement in AI-assisted speaking practices.

3.3. Data Analysis

With the help of IBM SPSS, quantitative data from the SpeechAce pre- and post-test results were analyzed using descriptive statistics and independent sample t-tests to compare the performance of the experimental and control classes across two overall scores. Qualitative data from the interviews were transcribed and analyzed using thematic analysis (Braun & Clarke, 2006) to identify two key themes including perceived benefits of using the AI tool for speaking practice as well as existing problems and challenges of AI-assisted speaking practice.

4. Findings

This study examined the effectiveness of AI-assisted speaking tools in enhancing EFL learners' speaking skills by analyzing the difference between pre-test and post-test scores and incorporating qualitative insights from semi-structured interviews.

Table 1 provides the descriptive statistics, summarizing the mean and standard deviation of the speaking proficiency improvement for both the experimental and control classes and displays the results of the samples t-test and the effect size, offering a comparison of the two classes.

Table 1. Independent samples t-test of speaking improvements.

	Class	N	Mean	Std. Deviation	t	Sig. (2-tailed)	Cohen's d
Improvement	CC	34	0.059	0.3669	2.633	0.01	0.63
	EC	36	0.333	0.4922			

The independent samples t-test indicated that the experimental class, which used AI speaking apps, achieved a mean improvement of 0.333 (SD = 0.4922), while the control class, which relied on traditional speaking learning methods, achieved a mean improvement of only 0.059 (SD = 0.3669). The t-test revealed a significant difference between the two classes ($t(70) = 2.633$, $p = 0.01$), with the experimental class significantly outperforming the control class in speaking proficiency improvement. The effect size, measured by Cohen's d, was 0.63, indicating a medium to large effect, which suggests that the use of AI speaking apps had a substantial impact on improving students' speaking proficiency. These findings are summarized in Table 1.

The thematic analysis revealed two themes: (1) the perceived benefits of using these tools for speaking practice, and (2) the existing problems encountered by users. Regarding the benefits, three key advantages emerged from the data. Students noted that the apps provided immediate feedback on pronunciation, grammar, and fluency, which helped them identify and address individual weaknesses effectively. The convenience of practicing anytime and anywhere was another major advantage, especially for learners who lacked access to frequent in-person practice. Moreover, students reported increased self-confidence as the apps offered a low-pressure environment for regular speaking practice, allowing them to build their skills gradually before applying

them in interactions. Despite these benefits, students identified several challenges. Students found the interactions with AI to be mechanical, as the apps lacked emotional sensitivity and the ability to provide nuanced feedback tailored to complex conversational contexts. Technical issues, such as occasional inaccuracies in speech recognition, further hampered the user experience. Additionally, while the apps simulated a variety of scenarios, they could not fully replicate the spontaneity and unpredictability of real conversations, which some students felt limited their preparation for authentic communication.

5. Discussion

Research has highlighted the positive effects of AI tools on EFL learners' speaking skills. For instance, Madhavi et al. (2023) found that ESL students experience significant improvements in communication skills using AI technologies for language learning. In this study, the experimental class showed significant improvements in pronunciation, grammar, vocabulary, fluency, and overall performance compared to the control class, confirming the effectiveness of AI-assisted speaking proficiency.

However, challenges were also identified, reflecting issues noted in previous studies. These included problems such as inaccurate voice recognition, limited feedback depth, and the absence of detailed evaluation on speech organization were discovered (Zou et al., 2024).

Unlike prior studies that concentrated on specific skills, such as pronunciation improvement through choral reading (Nursyam, 2021), this study provided practical insights into the impacts of AI tools across multiple speaking dimensions. Nevertheless, the qualitative data of this study relied on a small sample of interviewees, which may affect the broader applicability of the findings. Additionally, the study primarily examined short-term improvements in speaking skills, leaving the long-term impact of AI-assisted practice unexplored.

6. Conclusion

This four-week study revealed measurable improvements in speaking proficiency among students using the AI tool. While AI tools prove to be valuable for enhancing speaking proficiency, problems such as feedback limitations, voice recognition accuracy, and lack of emotional interaction need to be addressed. Future research can be conducted with a larger number of participants for a long period such as a whole semester and assess their long-term impact to better support learners' speaking development.

Acknowledgements

This work was partially supported by a Scientific Research Program of the Department of Education of Hubei Province (No. B2023408) and a grant from the Wuhan College of Communication (No. XJ2023140).

References

- Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, 47(4), 778–786.
- Alnifasah, M. (2022). Technology review: Speechace. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. <https://doi.org/10.31274/psllt.14315>.
- Alsatuay, M. C. (2011). Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer-Assisted Language Learning*, 24(5), 419–432.
- Balci, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Education Studies (IJCES)*, 3(1), 66–82. <https://doi.org/10.5281/zenodo.12544675>.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chang, Y., Lee, S., Wong, S. F., & Jeong, S. (2021). AI-powered learning application use and gratification: An integrative model. *Information Technology & People*, 35(7), 2115–2139.
- Egan, K. B. (1999). Speaking: A Critical Skill and a Challenge. *CALICO Journal*, 16(3), 277–293. <http://www.jstor.org/stable/24147844>.
- Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology*, 24(1), 62–85.
- Gunasekaran, D., & Subramaniam, G. K. J. (2022). The influence of reading on students' speaking abilities: Pronunciation, grammar, and vocabulary. *Change & Adaptation*, 161.
- Harris, D. (1974). *Testing English as a second language*. McGraw-Hill.

- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.
- Kok, J. N., Boers, E. J., Kosters, W. A., Van der Putten, P., & Poel, M. (2009). Artificial intelligence: Definition, trends, techniques, and cases. *Artificial Intelligence*, 1, 270–299.
- Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile-assisted language learning: Can mobile devices support collaborative practice in speaking and listening? *Proceedings of EuroCALL 2007*. Retrieved July 14, 2014, from <http://oro.open.ac.uk/11617/1/S0958344008000335a.pdf>
- Lehman, B., Gu, L., Zhao, J., Tsuprun, E., Kurzum, C., Schiano, M., ... & Tanner Jackson, G. (2020). Use of adaptive feedback in an app for English language spontaneous speech. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I* 21 (pp. 309–320). Springer International Publishing.
- Li, M., & Zou, B. (2022). Using artificial intelligence technology as a tutor for EFL speaking practice. In Colpaert, J., & Stockwell, G. (Eds.), *Smart CALL: Personalization, contextualization, socialization*, 82–100.
- Liu, Y., Yang, T., & Xu, S. Y. (2019). The efficacy of using Liulishuo for spoken English ability: Taking the practice of Liulishuo in North China Electric Power University (Baoding) as an example. *International Education Studies*, 12(4), 244–252.
- Madhavi, E., Sivapurapu, L., Koppula, V., Rani, P. E., & Sreehari, V. (2023). Developing learners' English-speaking skills using ICT and AI tools. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 32(2), 142–153.
- Nursyam, Y. (2021). Improving students' pronunciation of monophthong sound in reading narrative text by using choral reading technique at the eighth grade students of MTs Alkhairat Alindau. *Datokarama English Education Journal*, 2(1), 69–78.
- Syakur, A., & Azis, R. (2020). Developing reading learning model to increase reading skill for animal husbandry students in higher education. *Britain International of Linguistics Arts and Education (BloLAE) Journal*, 2(1), 484–493.
- Yang, H., Kim, H., Lee, J., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343.
- Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, 15(4), Article 2872. <https://doi.org/10.3390/su15042872>
- Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, 126, 103497.

Strand

AI-Enhanced K-12 Education: Challenges and Potential

Inspiring Ideas for English Creative Writing: Insights from Designers on A Prototype of Artificial Intelligence-Assisted Platform MuseAIWrite

Yin YANG

Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR, China
yyin@eduhk.hk

Abstract: *Creative thinking is increasingly recognised as a crucial 21st-century skill, yet there have been limited studies on exploring Artificial Intelligence (AI) for process-oriented writing through visualised agents. This ongoing study introduces a prototype of MuseAIWrite, an artificial intelligence-supported platform aimed to enhance senior primary school students' English creative writing. This paper presents designers' insights and recommendations to support scaffolded learning in student-centred AI-supported learning environments.*

Keywords: creative thinking, storytelling, artificial intelligence, visualised agents

1. Introduction

Creativity, traditionally viewed as an inherently human trait, involves the generation of new ideas, the ability to find unusual associations, and the development of unique solutions to problems (Finke et al., 1996; Sawyer & Henriksen, 2024). Prior research has showed a positive correlation between creative thinking and the language acquisition (Ghonsooly & Showqi, 2012). Language, as Chomsky (1965) asserts, is fundamentally generative, making it a prime example of 'everyday' creativity (Swann & Maybin, 2007) due to its vast array of expressive possibilities. Writing is a creative act of meaning-making that occurs as the writer engages dynamically with the evolving text (Byrnes, 2013). With the evolving development of artificial intelligence (AI), there is a growing perspective that these cognitive processes can be enhanced through technology (Rezwana & Maher, 2023).

This in-progress study introduces a prototype of MuseAIWrite from designers' perspectives. The platform integrates AI agents, multimodal scaffolding, and narrative frameworks to support senior primary students in developing their creative English writing skills.

2. Literature Review

2.1. Creative Writing

Creative writing in educational settings is a multifaceted activity that nurtures language skills, imaginative thinking, and emotional depth (Harper, 2010). An important framework in creative writing is C-squared SG (C2SG: Character, Conflict, Struggle, and Goal), which guides the development of narrative structure (Haven, 1999). This framework helps students to focus on creating well-defined characters (Character), setting up tensions or challenges that these characters need to navigate (Conflict), exploring the character's efforts to overcome these challenges (Struggle), and finally, articulating what the characters aim to achieve (Goal).

However, with the emergence of new technologies, alternative methods such as storyboarding and multimodal storytelling tools may offer more engaging approaches. These tools allow students to express ideas visually and interactively, thereby reducing cognitive load and enhancing creativity (Lin et al., 2022; Lin et al., 2024; Kansizoglu et al., 2017).

2.2. Visualised agents for Scaffolding Creative Writing

Many studies have showed the effectiveness of helping learners organize and structure their thoughts, making complex relationships between ideas more understandable (Boykin et al., 2019; Kansizoglu, 2017; Kılıçkaya, 2020; Lin et al., 2022). In the context of creative writing, storyboards allow students to visually plot out characters, conflicts, struggles, and goals. However, there has been limited research on the impact of AI-driven concept maps for scaffolding on young learners' creative English writing (Qin et al., 2024).

In this study, the concept of visualised agents supporting storyboards that utilise Large Language Models (LLMs) tailored to the learners' levels was proposed. These agents, acting as interactive guides, can assist students in constructing and refining their ideas by providing suggestions based on the narrative elements they

inputs. For instance, they can help students identify gaps in their narratives, and introduce multimodal elements such as images to enhance storytelling.

3. The Design of the Prototype of MuseAIWrite

The proposed MuseAIWrite is web-based platform designed based on Zimmerman's self-regulated learning (SRL) theory (2002). MuseAIWrite integrates a multimodal and interactive environment that utilises advanced AI technologies such as GPT models, StableDiffusion, and Dall-E-3. To further support the writing process, customised AI agents employing Retrieval- Augmented Generation (RAG) will be utilised. These agents are designed to retrieve documents that match the students' language proficiency, ensuring the content is within an appropriate lexical range. MuseAIWrite facilitates the engagement of senior primary students in English creative writing through the three phases of SRL:

3.1. Forethought phase

In this initial phase, students set specific writing goals, select narrative styles (e.g., short stories, book reviews), and engage in brainstorming activities with the help of AI agents. AI prompts inspire students to outline their ideas and explore creative possibilities.

3.2. Performance phase

During the performance phase, students write by interacting with the platform and receiving feedback from teachers. They monitor their work and make necessary adjustments. AI tools analyse the text, and give feedback on grammar, vocabulary, and style improvements based on the customised AI agents. For example, students could engage in brainstorming activities, including character, setting, and plot development (refer to Figure 1). The visualised agent Niko provides randomised inputs that can inspire or challenge the students to think more deeply about their character's motivations, backgrounds, and conflicts. This tool encourages diversity in character creation and helps students think beyond stereotypes. Here, the agent prompts the student to think about specific aspects like the character's appearance, name, and personal traits.

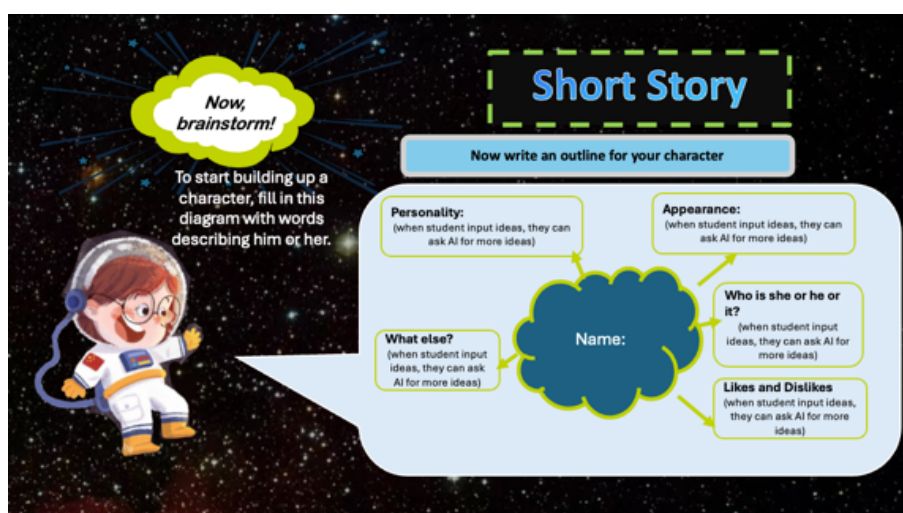


Figure 1. An Example of the Interface of Brainstorming (e.g., characters).

The MuseAIWrite platform incorporates several well-established narrative structures to guide students through the process of crafting compelling stories. Each structure offers a unique framework for plotting events and developing characters, catering to different storytelling styles and objectives, including Freytag's pyramid, Fichtean curve, Three act structure, and Seven-point story structure. Figure 2 shows an example of the Fichtean Curve which emphasises a series of rising and falling actions leading up to a climax and a swift resolution. By integrating these narrative structures into the MuseAIWrite platform, students could gain a foundational understanding of different ways to structure their stories.

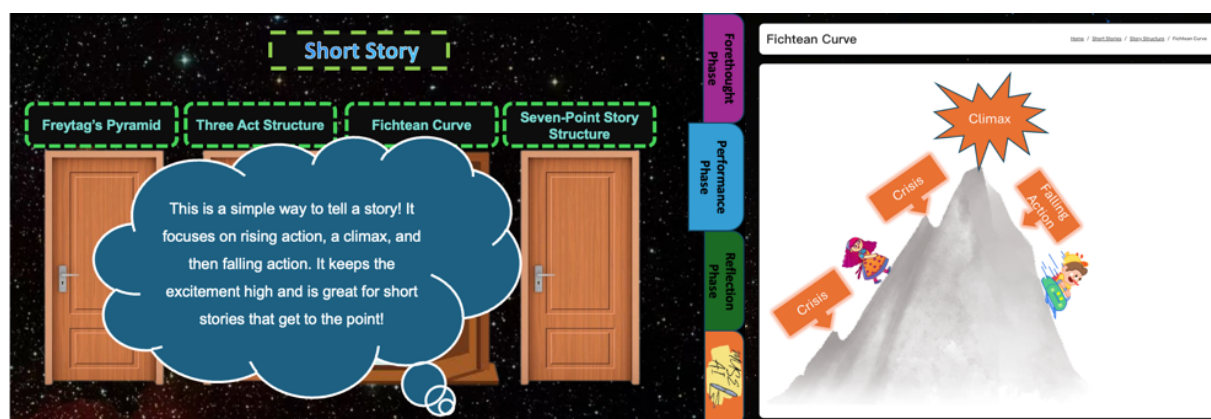


Figure 2. An Example of the Interface of Short Stories.

3.3. Reflection phase

The reflection phase is a critical yet often overlooked part of the writing process. In MuseAIWrite, students review their work using AI-assisted tools that highlight strengths and areas for improvement. These tools also encourage self-assessment by asking reflective questions such as: “How does your character’s struggle align with their goal?”, “What alternative endings could you explore?”. AI can further support this phase by suggesting revisions, identifying narrative inconsistencies, and encouraging students to iterate on their drafts.

4. Conclusions

MuseAIWrite represents an innovative step forward in integrating AI into creative writing education. While still in its nascent stages, the platform promises to transform how creative writing is taught by leveraging the latest in AI technology to provide personalised, engaging, and effective learning experiences. While tools like ChatGPT and Grammarly focus on language correctness and fluency, MuseAIWrite is specifically designed to foster creativity and narrative planning. Its unique features include: (1) Visualised Agents: Interactive guides that assist with concept mapping and storytelling. (2) Multimodal Elements: Integration of images, audio, and text to support diverse learning styles. (3) Child-Safety Mechanisms: Filters for age-appropriate content, teacher oversight options, and clear authorship attribution to ensure ethical use of AI. The current study is conceptual and lacks empirical validation. Future studies will focus on piloting MuseAIWrite in classrooms to evaluate its effectiveness in improving students’ creative writing and critical thinking skills.

Acknowledgements

This study was supported by the startup research grant from The Education University of Hong Kong, which is gratefully acknowledged.

References

- Boykin, A., Evmenova, A. S., Regan, K., & Mastropieri, M. (2019). The impact of a computer-based graphic organizer with embedded self-regulated learning strategies on the argumentative writing of students in inclusive cross-curricula settings. *Computers & Education*, 137, 78-90.
- Byrnes, H. (2013). Positioning writing as meaning-making in writing research: An introduction. *Journal of Second Language Writing*, 22(2), 95-106. <https://doi.org/10.1016/j.jslw.2013.03.004>
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT press, Cambridge.
- D'Souza, R. (2021). What characterises creativity in narrative writing, and how do we assess it? Research findings from a systematic literature search. *Thinking Skills and Creativity*, 42, 100949.
- Finke, R. A., Ward, T. B., & Smith, S. M. (1996). *Creative cognition: Theory, research, and applications*. MIT press.
- Harper, G. (2010). *On Creative Writing* (Vol. 4). Multilingual Matters.
- Haven, K. (1999). *Write Right!: Creative Writing Using Storytelling Techniques*. Bloomsbury Publishing USA.
- Kansızoğlu, H. B., Aydoğan, M., Gelbal, S., & Olkun, S. (2017). The effect of graphic organizers on language teaching and learning areas: A meta-analysis study. *Eğitim ve Bilim*, 42(191), 139–164.
- Kılıçkaya, F. (2020). Learners’ perceptions of collaborative digital graphic writing based on semantic mapping. *Computer Assisted Language Learning*, 33(1-2), 58-84.

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Lin, V., Liu, G. Z., & Chen, N. S. (2022). The effects of an augmented-reality ubiquitous writing application: A comparative pilot project for enhancing EFL writing instruction. *Computer Assisted Language Learning*, 35(5-6), 989-1030.
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, 104977.
- Qin, H. X., Jin, S., Gao, Z., Fan, M., & Hui, P. (2024). CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).
- Rezwana, J., & Maher, M. L. (2023). User perspectives on ethical challenges in Human-AI co-creativity: A design fiction study. In *Proceedings of the 15th Conference on Creativity and Cognition* (pp. 62-74).
- Sawyer, R. K., & Henriksen, D. (2024). *Explaining creativity: The science of human innovation*. Oxford university press.
- Swann, J., & Maybin, J. (2007). Introduction: Language creativity in everyday contexts. *Applied Linguistics*, 28(4), 491-496. <https://doi.org/10.1093/applin/amm047>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2

Understanding Teacher Adoption of Integrated GenAI Platforms in Secondary Education: A Pilot Case Study

Saanvi Ravi KATYAYAN¹, Jac Ka Lok LEUNG^{2*}, Alison CHAN³

^{1,2}The Hong Kong University of Science and Technology, Hong Kong SAR, China

³Solomon Learning Group, Hong Kong SAR, China

srkatyayan@connect.ust.hk, egjac@ust.hk, alison@solomongp.com

Abstract: *This study examines the adoption of integrated Generative AI (GenAI) platforms in secondary education, focusing on their impact on teaching practices and the motivations and challenges faced by educators. The study investigates how these platforms, particularly prompt tools, are transforming pedagogical approaches while addressing ethical concerns and practical limitations. As a pilot study, three case studies with secondary school teachers varied in subject domain, educational context, years of teaching experience and teaching style were conducted. Their perceptions on an integrated GenAI platform, GoodClass.ai, were collected through semi-structured interviews. Findings reveal that integrated GenAI platforms excel in automating repetitive tasks, opens new opportunities for pedagogical innovations, and facilitates personalized assignments and question banks. However, challenges remain regarding the reliability of AI-generated content, subject-specific constraints, and the extent of personalization a teacher can provide. The study contributes insights for school administrations and developers to acknowledge the main considerations when secondary school teachers adopt GenAI into their practice. This also marks the first step to begin a co-creation journey among educators, researchers, and developers to design an integrated GenAI platform that enhances the productivity and quality of teaching while maintaining teacher agency. Further research is necessary to elicit the perception and outcomes of students as the receiving end of this new mode of teaching and learning. Another important measure on AI literacy should be further examined.*

Keywords: integrated GenAI platform, teaching productivity, secondary education

1. Introduction

The rise of Generative AI (GenAI) in education marks a pivotal shift in how teaching and learning are approached. Education sectors at all levels are working together to establish new policies, frameworks, protocols, tools and guidelines to facilitate the adoption of this rapidly evolving technology while mitigating the risks and ethical implications it could bring (Schiff, 2022). The perceptions of teachers are central to whether and how GenAI tools would be incorporated in teaching and learning practices. In the context of K-12 education, teachers have the opportunity to leverage GenAI to reimagine their way of teaching and to offer more effective, efficient, personalized and inclusive learning experiences (Nadis, 2024), while they also face tremendous challenges in upholding academic integrity and prevent the potential for misuse and “lazy learning” (Sabzalieva & Valentini, 2023).

The widespread innovation landscape related to AI in education has led to an exponential increase in the number of available tools in the market. As a result, early adopters may feel stressed and burned out in finding the most updated GenAI tools, while the majority of teachers may view adopting AI as an overwhelming task. And therefore, integrated GenAI platforms designed for educators such as Khanmigo (khanmigo.ai), MagicSchool (magicschool.ai), and GoodClass (goodclass.ai) have emerged to address this need. Most of these platforms include a wide range of tools such as lesson planning, content creation, question banks, student support, assessment and grading, etc., providing an all-in-one system for teachers and students.

Unlike AI chatbots with large language models (LLMs) like ChatGPT, these platforms function as prompt tools, offering pre-prompted options that simplify the process for educators (Kristic et al., 2022). By providing structured inputs, they guide AI in generating relevant outputs, thereby reducing cognitive load. This allows educators to focus on specific tasks without needing to create complex prompts. Utilizing predefined prompts enables teachers to efficiently access tailored resources aligned with their instructional goals, enhancing teaching effectiveness and student engagement (Zafari et al., 2022).

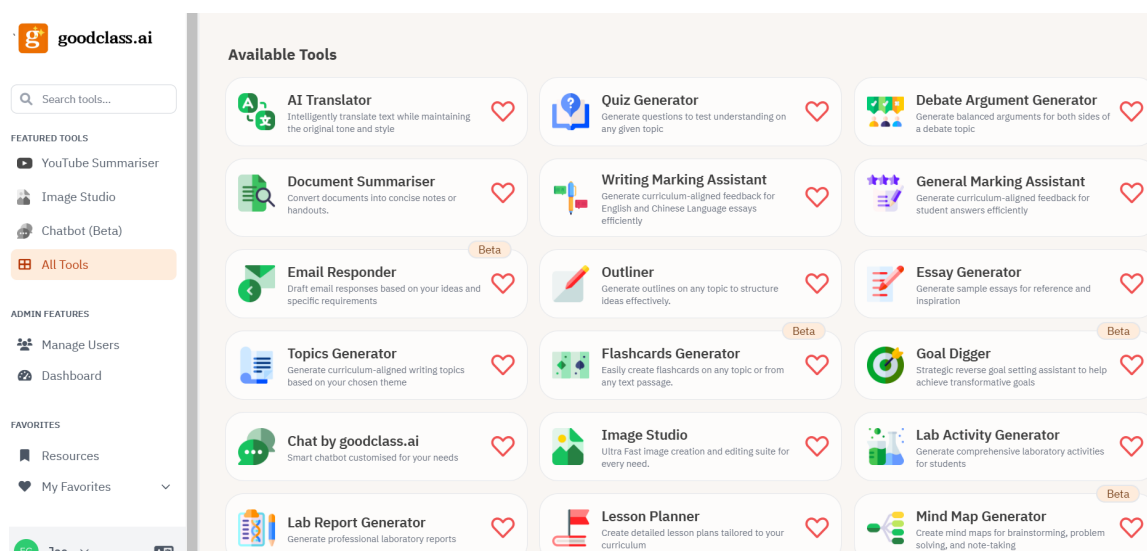


Figure 1. Example List of Prompt Tools in GoodClass.ai

2. Method

This study examines teachers' motivations and challenges in adopting AI tools in their teaching practices, as well as their vision for ideal AI tools that enhance teaching quality and productivity. Through semi-structured interviews with three secondary school teachers who had 10+ years of teaching experience—varying in subject domain (English language, Mathematics, Sciences), educational context (Hong Kong Diploma of Secondary Education (HKDSE) and International Baccalaureate (IB) curriculum), and teaching needs—insights on the integrated GenAI platform GoodClass.ai were gathered. The interview protocol included six guiding questions (see Appendix), and descriptive analysis was conducted to summarize each teacher's key insights and experiences with the platform.

3. Results

3.1. Case 1: An English Teacher from a School in Hong Kong

Teacher A, specializing in English at a school in Hong Kong under the HKDSE curriculum, has adopted GenAI to enhance her teaching practices. She believes that AI can generate effective content when provided with clear descriptions and adequate context, emphasizing that the quality of output heavily depends on how prompts are structured and the richness of the input data. While she views AI as a double-edged sword, offering significant benefits in refining messages and creating nuanced questions, she acknowledges the challenges it presents, particularly in comprehending complex poetry, providing step by step detailed answers and accurately indicating the minor aspects of a complex text.

Teacher A utilizes AI to elevate the difficulty of questions by incorporating figurative language and inference, thereby adding depth to student engagement. For instance, when encountering a paragraph that lacked sufficient length for her students, she adeptly merged similar articles using AI to create a coherent tone, continuously prompting the AI to ensure accuracy and clarity. She also mentions that she appreciates the user-friendly interface of prompt tools available in GoodClass.ai, which allows for the separation of assignments based on student levels and facilitates the generation of advanced questions.

However, she notes that the accuracy of any AI-generated content is contingent upon the quality and quantity of information provided, highlighting the necessity for educators to thoughtfully curate input to maximize AI's effectiveness in educational contexts. These insights reflect broader concerns among educators regarding the balance between leveraging technology and maintaining pedagogical integrity (Le-Nguyen & Tran, 2024; Nadis, 2024).

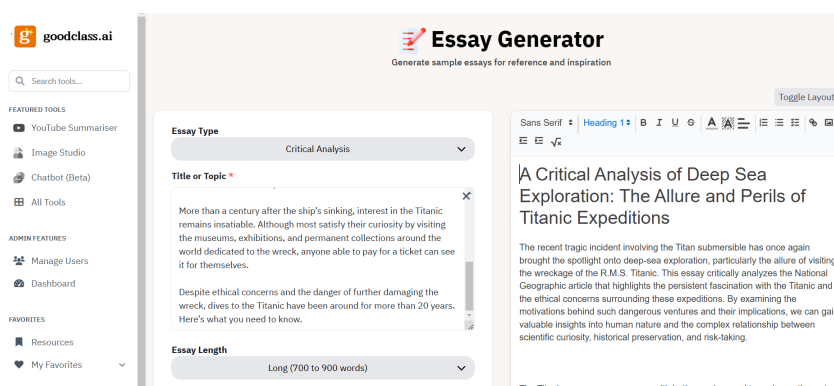


Figure 2. Example Essay Generator in GoodClass.ai

3.2. Case 2: A Mathematics Teacher from an International School in Shanghai

Teacher B, specializing in mathematics at an international school offering the IB Diploma Programme (IBDP) in Shanghai, initially used ChatGPT before discovering GoodClass.ai. After exploring its features during interviews, they found the math quiz generator and YouTube summary tool particularly useful, increasing their motivation to integrate AI into their workflow. Teacher B's ideal tool would convert textbook pages into PowerPoint presentations or generate questionnaires with detailed solutions.

The school that Teacher B was teaching at has set up policies to encourage teachers to integrate AI in their teaching practices. Teacher B sees potential benefits in automating tasks such as writing comments and generating questions. However, Teacher B expresses concerns about the reliability of AI in advanced mathematical content. Time constraints limit their ability to explore AI features effectively, impacting their proficiency. Additionally, as someone familiar with machine learning principles, they worry about the accuracy of the generated mathematical content. Furthermore, they note that current tools often lack features specifically tailored for mathematics instruction.

These apprehensions are echoed by many educators. A significant concern is the accuracy and reliability of AI-generated outputs, particularly in complex mathematical contexts. LLMs seem to struggle with providing reliable and correct solutions, even for basic problems (Schorcht et al., 2024). Moreover, educators face limited time to learn and effectively implement these new technologies. With demanding schedules filled with lesson planning, grading, and classroom management, many teachers find it challenging to devote sufficient time to explore AI tools deeply. This lack of time not only hampers their ability to understand how to use these technologies effectively but also limits their capacity to critically evaluate the quality and efficacy of AI-generated content (Felix, 2020).

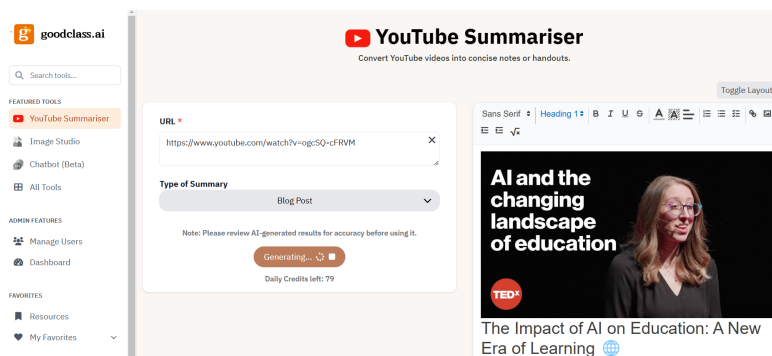


Figure 3. Example YouTube Summarizer in GoodClass.ai

3.3. Case 3: A Physics Teacher from an International School in Shanghai

Teacher C, a physics teacher experienced in IB and A-level curricula at an international school in Shanghai, aims to ensure ethical AI use while automating repetitive tasks. They encourage students to use AI prompt tools for inquiry-based learning, asking them to generate 10 questions about the class topic. Teacher C utilizes prompt tools like GoodClass.ai for creating rubrics, multiple-choice questions (MCQs), and automating report generation.

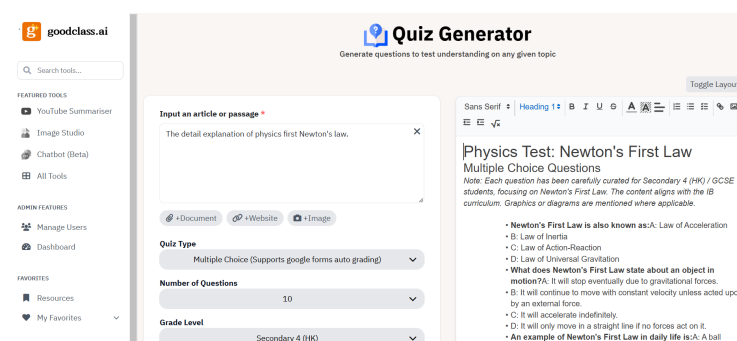


Figure 4. Example Quiz Generator in GoodClass.ai

Despite the benefits of prompt tools in enhancing class interactions, Teacher C identifies several limitations of AI tools. Specifically, while AI excels at generating artistic images, it struggles with scientific diagrams such as graphs and labeled apparatus. Additionally, AI is less effective in addressing multi-step problems or providing detailed methodologies for advanced physics topics. Although it can grade introductory sections adequately, it lacks consistency in feedback due to the variability in student responses.

These observations align with research indicating that while GenAI is effective for administrative tasks (Curts, 2023), its limitations in specific subjects hinder broader adoption among STEM educators (Wang et al., 2024). Teacher C desires improvements in tools that can generate accurate scientific diagrams and tailored problem-solving steps for physics instruction.

4. Discussion

The adoption of GenAI tools in K-12 education is a nuanced process influenced by teachers' diverse motivations, challenges, and expectations. From our pilot interviews, all interviewees were able to identify positive influences of GenAI and were willing to adopt GenAI into their teaching subject domain. Meanwhile, all of them were cautious and concerned about the accuracy, biases, appropriateness, consistency, context sensitivity, potential for misinformation, and ethical use of AI generated content.

GenAI as a paradigm shifting technology in education, we may utilize Rogers' Innovation Adoption Curve to explain some of the early perceptions of educators towards incorporating GenAI into their practice (Miller, 2018). They can be categorized into early adopters, early majority, late majority, and laggards.

Some early adopters of GenAI see significant potential to transform teaching practices and are willing to experiment with tools like GoodClass.ai to enhance efficiency and student engagement (Heung & Chiu, 2025). However, they face challenges such as subject-specific limitations, where current GenAI tools struggle with detailed explanations for advanced topics like calculus or physics (Wang et al., 2024). They also feel a responsibility to model ethical AI use for students (Sabzalieva & Valentini, 2023).

The early majority of teachers would adopt AI when clear benefits in workload reduction are evident (Curts, 2023). For instance, a mathematics teacher found GoodClass.ai's quiz generator and YouTube summary tool helpful in streamlining tasks. However, this group faces barriers like time constraints and reliability concerns regarding AI-generated content (Felix, 2020). Hence, more user-friendly and curated tools would greatly facilitate the adoption of this category of teachers.

Late majority and laggard educators adopt AI mainly under institutional pressure or peer influence. External factors such as school policies mandating AI integration motivate these groups (Sun & Pratt, 2024). Once engaged with tools like GoodClass.ai, they may appreciate features like customizable lesson plan templates but remain cautious about over-reliance on technology for personalized instruction. And therefore, more empirical studies are needed such that these categories of teachers can be better informed of the benefits and limitations of GenAI in education.

5. Conclusion

This study reveals significant insights into how integrated GenAI platforms like GoodClass.ai influence teaching practices among secondary school educators. Findings suggest that while teachers recognize the potential of GenAI tools to automate repetitive tasks and enhance efficiency, they remain cautious about fully integrating these technologies due to subject-specific limitations and concerns about reliability. According to McKinsey's projections, up to 25% of educators' work hours could be automated through GenAI technologies by 2030 (Ellingrud et al., 2023). This statistic underscores the potential of integrated GenAI platforms to alleviate administrative burdens while preserving the human-centered aspects of teaching that cannot be replicated by machines.

Overall, this study underscores the importance of addressing educators' concerns regarding usability and reliability when integrating GenAI into secondary education. As highlighted by Rogers' Innovation Adoption Curve (Miller, 2018), understanding varying perspectives will be crucial in facilitating broader acceptance among educators across disciplines.

This study offers important insights for school administrators and developers to understand the primary factors influencing secondary school teachers' adoption of GenAI in their practices. Furthermore, it serves as the initial step in fostering co-creation among educators, researchers, and developers to create an integrated GenAI platform aimed at improving both teaching productivity and quality, all while ensuring that teachers retain their agency. To gain a deeper understanding of this new teaching and learning approach, further research is necessary to explore students' perceptions and outcomes as the recipients of this method. Additionally, a thorough examination of AI literacy remains a significant area for future inquiry.

Acknowledgements

We thank all the teachers who participated in the interviews for this study.

References

- Curts, E. (2023). AI in the K-12 classroom, with Eric Curts. AVID Open Access. Publisher. <https://avidopenaccess.org/resource/338-ai-in-the-k-12-classroom-with-eric-curts/>
- Ellingrud, K., Sanghvi, S., Madgavkar, A., Dandona, G. S., Chui, M., White, O., & Hasebe, P. (2023). Generative AI and the future of work in America.
- Felix, C. V. (2020). The role of the teacher and AI in education. In *International perspectives on the role of technology in humanizing higher education* (pp. 33-48). Emerald Publishing Limited.
- Heung, H., & Chiu, P. W. Y. (2025). The impact of ChatGPT on student engagement: A systematic review and meta-analysis. *Journal of Educational Technology*, 16(1), 1-20. <https://doi.org/10.1234/jet.2025.001>
- Le-Nguyen, H. T., & Tran, T. T. (2024). Generative AI in terms of its ethical problems for both teachers and learners: Striking a balance. In *Generative AI in Teaching and Learning* (pp. 144-173). IGI Global.
- Miller, R. L. (2018). Rogers' innovation diffusion theory (1962, 1995). In *Technology adoption and social issues: Concepts, methodologies, tools, and applications* (pp. 1558-1571). IGI Global.
- Nadis, S. (2024). K-12 Education in the Era of Generative AI. MIT for a Better World. Retrieved from MIT for a Better World
- Sabzalieva, E., & Valentini, A. (2023). ChatGPT and artificial intelligence in higher education: Quick start guide. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000385146>
- Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, 32(3), 527-563.
- Schorcht, S., Buchholtz, N., & Baumanns, L. (2024, May). Prompt the problem—investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. In *Frontiers in Education* (Vol. 9, p. 1386075). Frontiers Media SA.
- Sun, J. C., & Pratt, T. L. (2024). Navigating AI Integration in Career and Technical Education: Diffusion Challenges, Opportunities, and Decisions. *Education Sciences*, 14(12), 1285.
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, 124167.
- Zafari, M., Bazargani, J. S., Sadeghi-Niaraki, A., & Choi, S. M. (2022). Artificial intelligence applications in K-12 education: A systematic literature review. *Ieee Access*, 10, 61905-61921.

Appendix

1. How aware are you of the importance of AI literacy in education, and what do you consider its key components?
2. To what extent do you wish to become proficient in using AI tools in your teaching practice, and what motivates this desire?
3. How does your proficiency with traditional teaching methods compare to your comfort level with AI tools, and what challenges do you anticipate in transitioning to AI?
4. How essential do you believe it is to integrate AI tools into your teaching, and which specific tasks or areas do you think would benefit most from this integration?
5. If you could create an ideal AI tool for your classroom, what features would you want it to include to enhance your teaching effectiveness?
6. What factors would encourage you to adopt AI tools more frequently in your teaching practice, and are there specific challenges that might trigger this shift?

Cultural Adaptability of Global AI in Education Policies: A Chinese Perspective

Tianhang GAO¹, Simin CAO^{2*}, Hui LI³

^{1,2}Shanghai Normal University, Shanghai, China

³The Education University of Hong Kong, Hong Kong SAR, China

1000530353@smail.shnu.edu.cn, 1000529795@smail.shnu.edu.cn*, huili@eduhk.hk

Abstract: *This study examines the cultural adaptability of global AI education policies, focusing on the alignment between international frameworks and the Chinese educational context. Using thematic analysis guided by the “3A2S” framework (Accessibility, Affordability, Accountability, Sustainability, and Social Justice), we analyze key principles from K-12 AI policies developed by international organizations and governments in North America, Oceania, and the United Nations. Our findings reveal a general convergence around human-centered AI integration, ethical considerations, and equitable access. However, a critical gap exists in financial planning and support for implementation. We discuss the implications of these findings for developing culturally relevant and sustainable AI education policies, particularly within China, drawing connections to Chinese educational philosophies. This research offers valuable insights for policymakers seeking to create inclusive and equitable AI learning opportunities for all students.*

Keywords: Artificial intelligence (AI); K-12; Policy; Framework; Tools; Education

1. Introduction

The rapid advancement and widespread adoption of artificial intelligence (AI) present both opportunities and challenges. Generative AI holds considerable promise for supporting teaching and alleviating administrative burdens in K-12 education. The increasing accessibility and sophistication of these tools offer unprecedented potential for generating human-like text and enriching diverse aspects of K-12 learning. However, while the Convention on the Rights of the Child (CRC) mandates the fulfillment of every child's potential (UN, 1989), the integration of AI tools can simultaneously advance and hinder these rights, creating a "double-edged sword" scenario (Chen & Lin, 2024). Therefore, stakeholders must develop robust, evidence-based AI policies and system development guidelines specifically tailored for K-12 education to ensure the protection of all children's rights. This necessitates examining AI's role through the lens of protection, provision, and participation. This study utilizes the 3A2S framework, an internationally recognized tool for evaluating the efficacy of child education policies and uniquely adapted to the Chinese cultural context (Xie and Li, 2020). This framework's suitability for assessing the local applicability of international policy frameworks provides valuable insights and serves as a reference for other developing nations grappling with similar issues.

2. Method

This study's text selection is based on AI policy frameworks and tools developed by global international organizations and government agencies with a focus on K-12 education. A keyword-driven search, utilizing terms such as "Artificial Intelligence (AI)," "K-12," "Policy," "Framework," "Tools," and "Education," was conducted across prominent search engines, including Google and Google Scholar, as well as the official websites of major international organizations. This search strategy resulted in the identification of six relevant policy documents and tools, as illustrated in Table 1.

After the text selection process, this study applied Top-Down Thematic Analysis (Braun & Clarke, 2006) based on the 3A2S (Xie and Li, 2020) policy theoretical framework. The specific research approach involved determining the overall research framework according to the 3A2S framework and identifying Sub-Category (Key Features) based on the textual data: Accessibility (Technical Access Channels, Environmental Adaptability), Affordability (Cost-Sharing Mechanisms, Market Incentives), Accountability (Lifecycle Monitoring, Privacy Protection), Sustainability (Digital Literacy Development, Multi-Stakeholder Collaboration), and Social Justice (Algorithmic Bias Mitigation, Inclusive Design). Finally, based on the texts, Policy Document Mapping Cases were summarized, as detailed in Table 2.

Table 1. Policy/Framework/Tools

NO.	Policy/Framework /Tools	Organizations	Region	Target Audience	Year
1	AI competency framework for Students	UNESCO	Globe	Includes k-12 Students	2024
2	Policy guidance on AI for Children 2.0	UNICEF	Globe	Children	2021
3	Artificial Intelligence and the future of teaching and learning	U.S. Department of Education	United States	Includes K-12 educators/policymakers	2023
4	K-12 Gen AI Maturity Tool	CoSN & Council of the Great City Schools	United States	K-12 School	2024
5	Australian Framework for Generative AI in Schools	Department of Education, Australia	Australia	K-12 School	2023
6	Considerations for Using AI Tools in K-12 Schools	Ministry of Education and Child Care, BC	Canada	K-12 School	2024

Table 2. Thematic Analysis

Core Dimension (3A2S)	Sub-Category (Key Features)	Policy Document Mapping Cases (Simplified version)
Accessibility	Technical Access Channels	UNESCO: "internet connectivity for all schools and students,including through agile 'online + offline' solutions, to with online or mobile AI systems"; UNICEF : "reliable internet access at home and at school, stand to benefit more from AI systems, mapping the digital connectivity of schools" ; Australian Framework: "Special provisions for rural/remote communities".
	Environmental Adaptability;	UNESCO: "Agile Online + Offline Solutions:bridge the gap in internet connectivity, even with limitations in internet access"; UNICEF: "use of free tools such as MachineLearningForKids, Teachable Machine, TensorFlow, and Keras. These tools are designed to facilitate the understanding and exploration of complex AI concepts and techniques"; U.S. Department of Education: "Hybrid online-offline solutions" ; MIT Case Study: "Offline AI learning kits" (Case Study); K-12 Maturity Tool: "Device compatibility standards" .
Affordability	Cost-Sharing Mechanisms	UNESCO:"support from academia or the private sector where appropriate"; UNICEF: "governments and other stakeholders to provide funding and incentives for AI policies and strategies"; Canada Guidelines: "Long-term cost-benefit assessment model"; China's Interim Measures for Generative AI Services: "Computing resource quota system".

	Market Incentives	U.S. Department of Education: "Evidence-based procurement under ESEA"; Singapore Government AI Cloud: "Sustainable financing mechanisms".
Accountability	Lifecycle Monitoring	Australian Framework: "Proactive AI impact monitoring system"; CrimeDetector System: "Anonymous login & minimal data collection"; China's Interim Measures: "Content watermarking requirements" (Watermarking Guidelines)
	Privacy Protection; Safety& law	UNESCO: "AI creators and service providers are accountable for the legal and ethical implications of AI systems;Stakeholders should have the opportunity to question and challenge the outcomes of AI systems". British Columbia: "Privacy Impact Assessment (PIA) process"; Haru Robot: "Multi-level privacy protection modules" (Case Study); EU GDPR: "Data minimization principle".
Sustainability	Digital Literacy Development	UNESCO AI for Students: "Spiral curriculum design" (Chapter 5.4); Sweden's Three Cities Project: "Teacher AI competency framework" (Section 4.5); MIT: "Day of AI curriculum system" (Box 4)
	Multi-Stakeholder Collaboration	U.S. Department of Education: "Cross-departmental R&D" (Recommendation #6); Alan Turing Institute: "Quadruple helix innovation model"; South Korea's National AI Strategy: "Youth AI Ambassador Program" (Section 5.7)
Social Justice	Algorithmic Bias Mitigation	UNICEF:H&M Group: "Training data auditing process" (Ethical Uses); Australian Framework: "Cultural IP protection clauses" (Section 4.1); UAE K-12 Curriculum: "Neurodiversity support systems" (Section 5.3.5)
	Inclusive Design	UNESCO: "Multilingual accessibility standards"; UNICEF:"leveraging AI systems to improve our programming, including predicting the spread of diseases and improving poverty estimation"; Canada: "Indigenous knowledge protection protocol"; AutismVR: "Special needs adaptive design" (Case Study).

3. Findings

Thematic analysis, guided by the 3A2S framework, highlights the careful consideration of accessibility, accountability, affordability, sustainability, and social justice within existing child-centered AI frameworks (Table 2). All examined frameworks prioritize a human-centered approach to interactions with AI, ensuring that educational practices integrate AI technology effectively. Accessibility requires dual technical-environmental approaches: deploying networks and affordable devices (infrastructure), while adapting AI systems to rural/linguistic contexts through localized interfaces and cognitive accessibility design for students and teachers. Accountability establishes lifecycle responsibility via impact monitoring systems, human oversight mechanisms (ethical review boards), and transparency tools like explainable interfaces and content watermarking to ensure traceability. Affordability combats economic exclusion through blended financing models, market incentives for EdTech innovation, and resource-sharing systems to prevent cost-driven disparities. Sustainability focuses on three pillars: 1) Spiral-curriculum digital literacy programs for teachers/students; 2) Quadruple-helix collaboration models connecting governments, academia, industry, and communities; 3) Carbon-aware AI development using energy-efficient algorithms. Social Justice actively counters systemic biases through algorithmic fairness audits, cultural IP protections, and neurodiverse design standards, ensuring AI systems accommodate marginalized groups rather than amplifying inequalities.

4. Discussion

Based on the "3A2S" framework, this research examines global artificial intelligence policy frameworks, revealing a broad consensus on inclusive AI education. The majority of AI frameworks advocate for equitable access of children to AI-based educational resources and tools, no matter where they are located and what cultural and educational background they have. This aligns with the Chinese educational philosophy of

"education for all (有教无类)," emphasizing the universality and fairness of education. Recommendations include government commitments to providing basic internet connectivity, affordable AI courses and software, and emphasizing human accountability in AI system design, development, and deployment through transparency and explainability. All the policies encourage the development and use of environmentally sustainable AI systems to benefit all students and promote a more equitable and inclusive education system.

While a general alignment exists, regional priorities vary. Developed nations often focus on fostering innovation and considering the long-term social impact of AI. Developing countries prioritize building the necessary infrastructure and ensuring basic accessibility. In China, due to uneven regional development, economically advanced coastal areas may focus more on innovative applications of AI education, while less developed inland regions may prioritize infrastructure construction and universal access. Cultural contexts also play a significant role in shaping inclusivity efforts. For example, UNESCO promotes "multilingual accessibility" globally, recognizing linguistic diversity, while regions like British Columbia emphasize the protection and integration of "Indigenous knowledge" within AI education. This reflects the different cultural emphases in AI education, for example, China has 56 ethnic groups, each with its own diverse culture.

However, a critical gap exists in the current policy landscape: the lack of specific financial support and planning guidance for implementing AI education. Given the substantial costs associated with developing and deploying large AI models, this deficiency is particularly concerning. The development and training of such models require significant computational resources and expertise, posing a considerable financial barrier for educational shareholders, especially in developing countries. Within China, the disparity in economic development levels across different regions also leads to an uneven distribution of AI education resources. China's rapid advancements in AI, including the development of complex models like DeepSeek, not only demonstrate China's strength in AI technology but also highlight the urgent need for robust financial planning to ensure equitable access to these technologies globally. Therefore, policymakers, especially in light of China's progress in AI, must provide specific recommendations on funding sources, cost-effective strategies for AI education deployment, considerations of regional affordability, and long-term financial sustainability. This is crucial to ensuring that all children, regardless of their location or economic background, have access to the transformative potential of AI learning opportunities. This includes exploring collaborative funding models, leveraging public-private partnerships, and potentially establishing international funding mechanisms to bridge the digital divide and promote global equity in AI education.

5. Conclusion

This study reveals a broad consensus among global AI education policies on the importance of human-centered design, equitable access, and ethical considerations, as analyzed through the 3A2S framework. While these frameworks offer valuable guidance, a critical gap remains in addressing the financial realities of implementing AI education, particularly in developing nations. Policymakers must prioritize specific funding mechanisms, including exploring public-private partnerships and international collaborations, to ensure equitable access to resources and support. Furthermore, the successful integration of AI in education requires culturally sensitive approaches. Policies should not only promote multilingualism and diverse cultural representation in AI systems but also actively address potential biases and ensure alignment with local educational philosophies and values. China's context, with its diverse regional development and unique cultural heritage, underscores the need for adaptable and inclusive policies. By addressing these financial and cultural considerations, policymakers can unlock the transformative potential of AI to create truly equitable and effective learning opportunities for all students.

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Chen, J. J., & Lin, J. C. (2024). Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its negative effects. *Contemporary Issues in Early Childhood*, 25(1), 146-153.
- Council of the Great City Schools. (2024). *K-12 Gen AI Maturity Tool*. Retrieved from: <https://www.cosn.org/k12-gen-ai-maturity-tool/>
- Department of Education, Australia. (2023). *Australian Framework for Generative AI in Schools*. <https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools>
- Ministry of Education and Child Care, BC. (2024). *Considerations for Using AI Tools in K-12 Schools*. <https://www2.gov.bc.ca/assets/gov/education/administration/kindergarten-to-grade-12/ai-in-education/considerations-for-using-ai-tools-in-k-12-schools>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Xie, S., & Li, H. (2020). Accessibility, affordability, accountability, sustainability and social justice of early childhood education in China: A case study of Shenzhen. *Children and Youth Services Review*, 118, 105359. United Nations. (1989). *Convention on the Rights of the Child*. <https://www.unicef.cn/sites/unicef.cn/files/2021-06/20120201045740314.pdf>
- UNESCO. (2024). *AI competency framework for Students*. <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
- UNICEF. (2021). *Policy guidance on AI for Children 2.0*. <https://www.unicef.org/innocenti/innocenti/innocenti/innocenti/reports/policy-guidance-ai-children>
- U.S. Department of Education. (2023). *Artificial Intelligence and the future of teaching and learning*. <https://www.ed.gov/media/document/ai-reportpdf-43861.pdf>

Designing a Contextualized Training for Learning Engagement of Hong Kong Students with Autism Spectrum Disorder (ASD)

Ying-lan CHEN¹, Yu-xing TAO², Tze-leung Rick LUI³, Ka-yan FUNG^{4*}, Kuen-fung SIN^{5*}
^{1,2,3,4,5}Institute of Special Needs and Inclusive Education, The Education University of Hong Kong, Hong Kong SAR, China
cyinglan@eduhk.hk, ytao@eduhk.hk, rtlui@eduhk.hk, fkayan@eduhk.hk, kfsin@eduhk.hk

Abstract: *The most representative characteristic of students with autism spectrum disorder (ASD) is that they have difficulties in socializing with others, such as peers. Research on robots assisting social skills shows that robots with cute appearance and physical body can positively promote social interaction among students with ASD. This work explores whether utilizing robots as an auxiliary training assistant can effectively improve the social skills of primary school students with ASD. Five primary school students with ASD in Hong Kong conduct questionnaires, proficiency tests, and observation. The results demonstrate that the behavioural, cognitive engagement, and intrinsic motivation of students with ASD slightly decrease after participating in contextual training, except for the emotional engagement which remains unchanged. This performance provides us some insights on improving the content of robotic contextual training and optimizing the training system.*

Keywords: Contextual training, Self-determination theory (SDT), Autism, Engagement, Robot

1. Introduction

Autism spectrum disorder (ASD) is described as a developmental disorder that affects an individual's communication and interaction (Filipek et al., 2000). Symptoms of ASD include difficulty making eye contact, repetitive behaviours, and difficulty communicating (American Psychological Association, 2023). The most representative symptom is the barrier to social communication (Bellini et al., 2007). The communication difficulties of students with ASD have less to do with general executive thinking and more with their inability to understand what the others are thinking (Jones et al., 2017). Interactive training with robots enhances attention of students with ASD and positively impacts on their social skills (Warren et al., 2013). Qidwai et al. (2019) pointed out that students with ASD preferred the interaction with robot toys as much as those without ASD. Also, the interactive robots can improve the social skills of those with ASD (Fung et al., 2024). Dawson (2008) stated that it would be more beneficial for students with ASD to detect the symptoms in infancy and conduct intervention treatment at an early stage. Social robots have a significant effect in helping students with ASD overcome social barriers and enable them to participate in social activities (Bartl-Pokorny et al., 2021). Scassellati et al. (2018) conducted a one-month experiment using a robot to intervene in the social skills of students with ASD. The experiment placed students with ASD and robots in a closed space for interaction. The results showed that the social skills of students with ASD have been significantly improved. This related research focuses more on studying the social impact of robots on students with ASD. On this basis, this study adds the robot's contextual training learning to see whether it has a positive impact on the social interaction of students with ASD.

This study explores whether robot contextual training can improve the social skills of students with ASD and enhance their social cognition. The study utilized a human-like robot, Boon Boon, to conduct a series of scenario training. Five primary school students with ASD in Hong Kong served as participants. The study focused on four aspects: behavioural, emotional, cognitive engagement, and intrinsic motivation. The results showed that emotion engagement remained unchanged, behavioural, cognitive engagement, and intrinsic motivation slightly decreased.

2. Training Design

Robotics' contextual training model breaks down the instructional task in a life context into several mini-tasks. From instructional guidance to skill building, followed by contextual modeling and finally application feedback, which utilizes micro-learning principles and focuses on building knowledge models relevant to life contexts (Manning et al., 2021). In total, there are three training contexts, which are supermarket, restaurant and transportation.

Context 1: Supermarket. The supermarket context is used for students to learn about supermarket shopping trolleys and baskets, how to use them safely and how to deal with some of the situations that would

occur in a supermarket. An assessment is set after learning to examine whether students have mastered the content. There are nine questions in this scene.

Context 2: Restaurant. The restaurant context provides students with an understanding of the procedures for having a meal. Followed by learning how to use cutlery and further learning some table manners. A test is set after the study to see if the students can use the cutlery correctly and furthermore if they can master the table manners correctly. There are nine questions in this scene.

Context 3: Transportation. The train station context is used for students to learn about the environment inside the train station and how to take the train. An assessment was set up after the study to see if the students were able to recognize the environment of the train station and the process of taking the train. There are nine questions in this scene.

3. Experiment

Five students with ASD (5 males) were recruited from a primary school in Hong Kong (mean age 9.6 years old, standard deviation 0.55 years old). Their ages ranged from 8 to 10 years old. The selection criteria for the study of the students were (1) attending Grade 3 to Grade 5; (2) being able to speak Cantonese and reading and writing Traditional Chinese; and (3) having basic writing and reading abilities. Before the experiment was conducted, informed consent was obtained from the parents. Participation was completely voluntary and based on consent. The University's Institutional Review Board (IRB) approved the experimental protocol. We did not provide any remuneration to the participants.

Those five students learnt contextual training with Boon Boon with three contexts for three days. The students were required to complete Context 1 on Days 1 and 3 respectively, and Context 2 and Context 3 on Days 2 and 3, respectively. The three contexts were Supermarket, Restaurant and Transportation, which were designed to enhance students' word, sentence recognition, and sentence expansion skills through contextualized training. Each theme took approximately 10 minutes to complete. Each training time slot was set up with a maximum of three students. Participants in the same time slot were trained simultaneously in the same classroom, and the study was conducted while maintaining social distance from the premise.

4. Efficacy Verification

We used an Analysis of Variance (ANOVA) to investigate whether the humanoid robot had an impact on the learning engagement and motivation of students with ASD. The results showed that there was no significant difference in the motivation of students with ASD before and after learning with the robot.

Behavioural Engagement. Participants demonstrated a slight decrease in behavioural participation after training, $p = .39$ (pre-test: $M = 11.80$; post-test: $M = 10.40$).

Emotional Engagement. Participants' emotional performance remained unchanged after attending the training, $p = 1$ (pre-test: $M = 10.40$; post-test: $M = 10.40$).

Cognitive Engagement. Participants' post-training performance demonstrated a marginal decrease in performance compared to before training, but the difference was minimal, $p = .91$ (pre-test: $M = 12.00$; post-test: $M = 11.80$).

Intrinsic Motivation. Participants' post-training performance declined moderately compared to their performance before training, $p = .73$ (pre-test: $M = 10.40$; post-test: $M = 9.60$).

Table 1. Descriptive statistics of the questionnaire.

		Mean	Improvement (%)	SD	F-value	P-value
Behavioural Engagement	Pre	11.80	-11.86	2.59	0.82	0.39
	Post	10.40		2.30		
Emotional Engagement	Pre	10.40	0	1.67	0.00	1
	Post	10.40		1.95		
Cognitive Engagement	Pre	12.00	-1.67	2.24	0.02	0.91
	Post	11.80		2.95		
Intrinsic Motivation	Pre	10.40	-7.69	2.07	0.13	0.73
	Post	9.60		4.45		

Based on the results in Table 1, behavioural, cognitive engagement and intrinsic motivation of students with ASD slightly declined after three days of training. However, emotional engagement remained unchanged. The preliminary findings suggested that robot-assisted learning had no impact on the emotional engagement of students with ASD, but that there was less behavioural and cognitive desire to be actively engaged. As a result, the students with ASD prefer to learn with a human being rather than with a robot.

5. Conclusion and Perspective

We propose a new design for students with ASD to learn about social skills and social cognition, aiming to enhance the learning engagement of students with ASD by using a variety of contextual training. As a result of the experimental study, the behavioural engagement, cognitive engagement and intrinsic motivation of the students with ASD decreased slightly, which indicated that the children with ASD preferred to learn with a human rather than a machine. Therefore, we could simplify the interaction model in the future and improve the correct and timely feedback of the robot. Furthermore, adding some more game-based sessions, such as adding real-life animations or character dialogues, which make this training mode easier to follow and relaxed, in order to enhance the learning motivation and active participation of students with ASD.

References

- Autism spectrum disorder. (2023). *American Psychiatric Association (APA)* <https://www.apa.org.https://www.apa.org/topics/autism-spectrum-disorder>
- Bartl-Pokorny, K. D. and others. (2021). Robot-Based Intervention for Children With Autism Spectrum Disorder: A Systematic Literature Review. *IEEE Access*, 9, 165433–165450. <https://doi.org/10.1109/ACCESS.2021.3132785>
- Bellini, S. and others. (2007). A Meta-Analysis of School-Based Social Skills Interventions for Children With Autism Spectrum Disorders. *Remedial and Special Education*, 28(3), 153–162. <https://doi.org/10.1177/07419325070280030401>
- Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, 20(3), 775–803. <https://doi.org/10.1017/S0954579408000370>
- Filipek, P. and others. (2000). Practice parameter: Screening and diagnosis of autism: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Carolina Digital Repository (University of North Carolina at Chapel Hill)*. <https://doi.org/10.17615/0mve-3j41>
- Fung, K. Y. and others. (2024). Humanoid robot-empowered language learning based on self-determination theory.
- Fung, K. Y. and others. (2022). Designing a game for pre-screening students with specific learning disabilities in Chinese. In *Designing a game for pre-screening students with specific learning disabilities in Chinese*. Association for Computing Machinery.
- Fung, K. Y. and others. (2024). A Robot-Assisted Scenario Training for Students with ASD. *International Conference on Computers in Education*. <https://doi.org/10.58459/icce.2024.4891>
- Jones, C. R. G. and others. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, 11(1), 95–109. <https://doi.org/10.1002/aur.1873>
- Manning, K. D. and others. (2021). The micro revolution: effect of Bite-Sized Teaching (BST) on learner engagement and learning in postgraduate medical education. *BMC Medical Education*, 21(1), 69–69. <https://doi.org/10.1186/s12909-021-02496-z>
- Qidwai, U. and others. (2019). Humanoid Robot as a Teacher's Assistant: Helping Children with Autism to Learn Social and Academic Skills. *Journal of Intelligent & Robotic Systems*, 98(3–4), 759–770. <https://doi.org/10.1007/s10846-019-01075-1>
- Scassellati, B. and others. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, 3(21). <https://doi.org/10.1126/scirobotics.aat7544>
- Warren, Z. E. and others. (2015). Can Robotic Interaction Improve Joint Attention Skills? *Journal of Autism and Developmental Disorders*, 45(11), 3726–3734. <https://doi.org/10.1007/s10803-013-1918-4>

ProTeach: LLM Agent framework for Intelligent Tutoring System

Ivan LO¹, Victor CHENG², Chun Hung LI^{3*}

^{1,2,3}Autoba Limited, Hong Kong SAR, China

chli@autoba.net

Abstract: Preliminary use cases of large language models (LLMs) in educational settings have been widely focused on prompt-engineering to align the educational roles and expected effectiveness in teaching. To better tackle the challenges of an Intelligent Tutoring System (ITS) in helping students achieve educational goals, the emerging framework of LLM Agents is a promising direction. We developed an EFL (English as Foreign Language) teaching application according to the LLM Agentic framework that involves multiple intelligent modules reusable for similar use cases under the same teaching objective. Our system design provides a better solution to tackle multiple challenges in effective teaching style, students progress tracking, and increasing students' motivation for learning.

Keywords: Intelligent Tutoring System (ITS), Large Language Model Agent, AI enhanced education

1. Introduction

With the launch of ChatGPT in 2022, Generative AI and large language models (LLMs) have received huge interest in the education domain. Personalized learning and intelligent tutoring have become hot topics of research. Most of these systems are smart enough to differentiate between high or low performing students.

The students of low-performance groups often have less initiative to express their ideas and learning difficulties because they may lack self-confidence or are unable to articulate. On the other hand, high-performing students require less support or guidance. The use of LLM driven learning systems can evaluate and classify the strong and weak aspects of students and adapt the teaching materials and scaffolding strategies to match the student capabilities so that weak aspects can be precisely addressed while strong aspects can be further strengthened. For example, there are LLM learning systems that induce the scaffolding technique (Liu, Zhengyuan et al., 2024) in LLM responses through well-polished pedagogical instructions, as well as instruction-tuning (Park et al., 2024) with custom datasets to align LLM responses with the tutoring objective.

LLM Agents are newly emerging applications of LLM in solving tasks in an autonomous manner. The cores of the agents are LLMs that decide appropriate actions and tools to handle the user submitted tasks or questions. The external tools can be external memory, knowledge, function calling, or standardized procedures available to LLMs in order to complete tasks. They can greatly enhance the capabilities of LLMs which can only handle text tasks using LLMs' knowledge. For example, agents can obtain the latest world information through internet search tools or to obtain database records through database connection tools.

Past studies on Intelligent Tutoring Systems (ITS) have suggested modular approaches with core components are effective designs for education tasks. With this approach, module features can be standardized and hence be reused easily. This saves the costs and time in developing new systems or systems for new domains.

In this paper, we introduce a prototype called ProTeach for tutoring English as a Foreign Language (EFL). It is a complete LLM agentic framework with modular design for ITS. We propose a key feature of adjusting LLM tutoring mechanisms based on the state of student model, which allows differentiated teaching styles developed from user interactions.

2. Related Work

The Intelligent Tutoring System (ITS) (Nwana, 1990), also known as Computer Aided Instructions (CAI), was proposed in the 1980s. It refers to computer programs that are designed to provide tutoring to students and help them achieve their pre-defined learning objectives. ITS typically contains 3 functional modules: expert knowledge module, student model module, tutoring module, although in implementations the structure of the ITS could vary and may contain many more differentiated ones.

The advancement of LLMs in language understanding and knowledge retention has largely surpassed previous dialogue-based tutoring systems such as AutoTutor (Graesser, 1999) which requires human expertise to design questions and expected inputs. They conduct semantic analyses between the user inputs and the expected outputs, and produce responses according to their cosine similarity. It could be limited to handle only inputs close to pre-designed expectations.

There is some research that adapts LLMs for tutoring services, and in particular for tutoring EFL (Liu et al., 2024) (Park et al., 2024). Nevertheless, performance of current LLM-based tutoring systems highly depends on the input prompts which require personnel who are familiar with prompt engineering, teaching strategies, and course materials. Sometimes, it might take multiple trials to obtain a useful prompt. Through modularized design of the agentic framework, the development efforts can be minimized by module standardization and re-use.

Agentic framework can also extend ITS features by endowing LLMs capability of tool use and hence connecting with existing information systems or other agentic systems, such as analytic systems and administrative systems. This highly relieves the workload of tutors and educational stakeholders.

3. ProTeach: LLM Agent Framework for ITS

The novelty of ProTeach is a complete LLM agentic framework that specifies the modules required to function as competent tutors. The design of Proteach is illustrated in Figure 1 and 2 and explained below:

- (1) *LLM core module*
LLM decides the appropriate response and actions to be taken, including the use of tools. The system instructions contain information of available modules, tools and their appropriate usage.
- (2) *Student model representations*
Student models allow simulating students' responses, misconceptions, and knowledge boundaries. A minimal student model could be the student's learning journal, which is a text summarizing past learned concepts and struggles, and student's learning characteristics such as "weak in using English contractions, e.g. aren't, can't". This can be automatically generated and updated by the LLM when there is new conversation.
- (3) *External Knowledge Access*
External knowledge access in terms of retrievable books, dictionaries, articles, websites or other forms of knowledge would be important for the LLM to cross-check its own generated information.
- (4) *Custom tool sets retrieval*
A good illustration of ideas and concepts often requires graphical representations. Custom tools would include charts creation, graphing tools, note-taking tools.
- (5) *Memory and Progress Tracking*
Learning is defined as change in long-term memory. To ensure the learning objective is fulfilled, newly acquired concepts should be practiced to induce knowledge retrieval.
- (6) *Experience collection and gradients*
A dynamic LLM that is updatable from tutoring experience would increase its effectiveness through increase in interactions. Experience is collected from raw conversations, treating user input and student model representation as the state, the produced response as the action, and the next user message as the next state. It allows fine-tuning the LLM to optimize for the objective of student's learning.

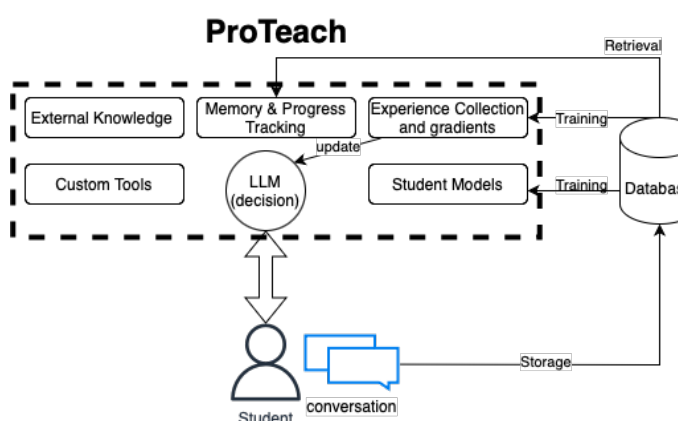


Figure 1. The ProTeach LLM Agent AI Tutor Framework and modules

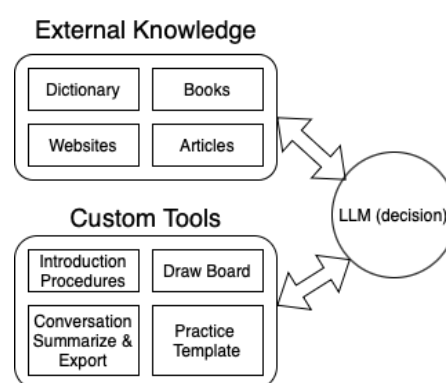


Figure 2. Sub-modules in External Knowledge & Custom Tools

In Figure 2, we illustrate the sub-modules in the External Knowledge module and the Custom Tools module. Sub-modules in External Knowledge provide retrieval of information in predefined formats to the LLM. They span different topics and are reusable.

In the Custom Tools module, it includes the introduction procedures for introducing a concept to learners without prior knowledge. The Draw Board is for creating diagrams and graphical representations. Conversation Summarize and Export sub-module allows LLM to consolidate a conversation and create customized notes in PDF format for the learner to review. Practice Template contains question formats (Fill in the Blanks, MCQ) that can be populated with a chosen topic and content.

4. ProTeach Application and Examples

We will illustrate the effectiveness of ProTeach through three illustrated examples. In the first example we illustrate how ProTeach adjusts its tutoring strategy based on predictions on the learned student model. In the second and third example, we illustrate the experience collection module that adjusts the tutoring strategy based on the student's capabilities.

Figure 3 shows an English exercise worksheet of an elementary school student. The red markings are the tutor's response. The student's response deviates from the tutor's response as this exercise is a drill for students to practise using contractions "can't, aren't, isn't" without explicitly introducing the term 'contraction' itself in the instruction.

✓ Can you swim? → No, I can't.
 否定式 'cannot' 的縮寫是 'can't'。'can't' 只有一個 'n'，不要拼錯。

A Match the questions with the answers. Write the correct letters in the boxes.

1. Is there any green tea?	☒ B	A No, they can't.
2. Can Sue climb the tree?	☒ D	B No, there isn't.
3. Are there any potatoes?	☒ C	C No, there aren't.
4. Can the children play tennis?	☒ A	D No, she can't.

B Give short answers to the questions. Use your own information.

- Are there any coloured pencils in your school bag?
 No, there aren't.
- Can you play table tennis?
 A little. Yes. No, I can't.

Figure 3. Exercise on the use of contractions

Example 1 - Use of contractions

LLM Analysis module inferred answer guidelines from image shown in Figure 3:

1. Use contractions (e.g. aren't, can't)

Memory and Progress Tracking module infer that 'contraction' in guideline 2 is not in the vocabulary of student learner and generate the update guideline 2:

2. Use short form to combine two words (e.g. aren't for are not, can't for cannot)

Student model module has recalled that the first language of the student is 'Traditional Chinese Hong Kong' and generates the Chinese guideline.

3. 使用縮寫形式來合併兩個單字 (例如: aren't 是 are not 的縮寫, can't 是 cannot 的縮寫)

The correct answer and the relevant guideline will be provided to the student and more importantly, ProTeach can learn from the student's response and proactively provide the important guideline to the student before they conduct the exercise in future. The model answer guideline in Figure 3 includes plural/singular agreement and will be implemented in a similar workflow.

Example 2 - Adjust strategy based on performance & student model

Student: I don't understand present perfect continuous tense.

ProTeach: The present perfect continuous tense is used to describe an action that started in the past and is still happening now or has recently stopped but has a connection to the present. It focuses on the duration of the action. (continued with Examples & Uses)

Student: So complicate I don't understand.

(This conversation is captured by the experience module : a weak student does not understand after a detailed explanation. This is a negative reinforcement to the adopted strategy which drives LLM to use a different tutoring mechanism for similar student models and inputs in the future.)

ProTeach: Let's try a simpler way. Imagine this: "I have been studying for two hours.". It means I started studying two hours ago and is still studying now (or just stopped). Try making a sentence like this: "I have been _____." Fill it with something you were doing todaying.

Student: I have been playing games.

ProTeach: Good job! That's the correct sentence.

5. Conclusions

In this paper, we proposed an LLM Agent framework as a modernized design framework for Intelligent Tutoring Systems. We believe that this framework is important to address the multiple challenges in adapting LLMs for language learning in lower grade students. Past research has placed their focus on optimizing generated instructions by prompt-engineering with pedagogical instructions. However, we believe a good system design with standardized, reusable modules weighs more than LLM reasoning capability in achieving education objectives. The limitations in our studies is we are illustrating the ideal functions and designs with the provided examples. LLM agents that are learning from conversation experience would be the next step to get around the limitation of human provided instructions.

References

- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4), 190--202.
- Chien, C. L., Huang, A. Y.Q., & Lu, O. H.T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments*, 10(11), 41.
- Graesser, A. C., Hastings, K. W., Hastings, P. W., Kreuz, R., & Tutoring Research Group, University of Memphis. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1), 35--51.
- Guo, Lu and Wang, Dong and Gu, Fei and Li, Yazheng and Wang, Yezhu and Zhou, Rongting. (2021). Evolution and trends in intelligent tutoring systems research: a multidisciplinary and scientometric view. *Asia Pacific Education Review*, 22(3), 441--461.
- Liu, Z and Yin, S and Lee, C and Chen, N (2024). Scaffolding language learning via multi-modal tutoring systems with pedagogical instructions. *2024 IEEE Conference on Artificial Intelligence (CAI)*, 1258--1265.
- Mustafa, M Y, et al. "A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda." *Smart Learning Environments* 11.1 (2024): 1-33.
- Nwana, H. S. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4), 251--277.
- Park, J Bae et al.. (2024). How to Align Large Language Models for Teaching English? Designing and Developing LLM based-Chatbot for Teaching English Conversation in EFL, Findings and Limitations.
- Stamper, J., Xiao, R., Hou, X. (2024). Enhancing llm-based feedback: Insights from intelligent tutoring systems and the learning sciences. *International Conference on Artificial Intelligence in Education*, 32--43.
- Venugopalan, Devika el al. (2025). Combining Large Language Models with Tutoring System Intelligence: A Case Study in Caregiver Homework Support. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 373--383.
- Wang, H el al. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. *Education and information technologies*, 28(7), 9113--9148.

The Opportunities and Challenges of Implementing Generative AI in Collaborative Learning: A Scoping Review

Xuechun SHI¹, Shen BA²

^{1,2}The Education University of Hong Kong, Hong Kong SAR, China
shixuechun2022@gmail.com, bas@eduhk.hk

Abstract: *The rise of Generative Artificial Intelligence (GenAI) has drawn increasing attention to its applications in education. The integration of GenAI in Collaborative Learning (CL) is a valuable area of research. This scoping review presents the application of GenAI in CL across various educational levels and learning activities. Furthermore, the research analyses that GenAI offers both opportunities and challenges in collaborative learning. The result shows that GenAI tools like conversational agents enhance student engagement and facilitate deeper cognitive processes by providing immediate feedback and scaffolding, fostering a richer collaborative environment. However, these advantages are tempered by concerns regarding the accuracy of AI-generated content, which can undermine trust and hinder participation. Ultimately, while GenAI presents exciting opportunities for enriching collaborative learning, its implementation must be approached with caution and needs continuous improvement to mitigate these challenges and enhance overall educational outcomes.*

Keywords: Generative Artificial Intelligence, Collaborative learning

1. Background

The rise of Generative Artificial Intelligence (GenAI) has sparked growing interest in its potential applications in education. GenAI, which includes technologies like ChatGPT and other AI-driven content generators, can assist students learning such as brainstorming ideas, generating text and images, and providing personalized learning support (Lo, 2023). One particularly promising area of exploration is Collaborative Learning (CL), where students work together to achieve common academic goals. The collaborative process is pivotal to fostering deep learning experiences through the negotiation of meanings and co-construction of knowledge among peers (Lar & Law, 2006). Furthermore, the integration of GenAI introduces new possibilities for supporting and enriching the learning process.

As the use of GenAI in CL expands, important questions arise regarding how these tools are being applied across different educational levels and learning activities. While researchers have begun investigating the role of GenAI in group-based learning, there is still a need for a more structured understanding of the specific types of GenAI being used, the contexts in which they are implemented, and how they support CL experiences. Additionally, while GenAI presents opportunities, such as enhancing creativity, streamlining group discussions, and providing instant feedback, it also brings challenges, including issues of accuracy, ethical concerns, and the potential for students to overly rely on AI-generated content (Lo, 2023). This study aims to explore these aspects by addressing two research questions: (1) How is GenAI being applied in CL across different educational levels and learning activities? (2) What are the key opportunities and limitations of using GenAI in CL?

2. Method

The methodology of this review follows the framework proposed by Arksey and O'Malley (2005) and further refined by Levac et al. (2010). This study selected five databases to screen the potential papers, which include Web of Science, ACM, EBSCOhost Research Databases, Scopus, and Wiley Online Library. When searching for articles from the databases, researchers considered two topics: CL and GenAI, and searched related terms, such as ("collaborative learning" OR "cooperative learning" OR "group learning" OR "peer-assisted learning") AND ("generative AI" OR "natural language interface*" OR "artificial intelligence agent*"). There are 1123 English papers from 2022 to the latest that match the search terms after removing duplicates.

The inclusion criteria are as follows: (1) Research must be an empirical study published as a journal article. (2) Studies should involve at least one generative AI tool to support learning. (3) Research should focus on CL, involving three or more subjects (For example, studies could include one person and two AI tools, or two people and one AI tool). Conversely, the exclusion criteria specify that any research that only reports the design of the GenAI without presenting empirical results related to learning outcomes will be excluded. Two researchers screened 10% of all papers respectively and checked the consistency. The inter-subjective

agreement showed good reliability (Cohen's $k = .60$). After discussing disagreements, researchers screened an additional 100 papers, improving reliability to Cohen's $k = .74$. Agreement increased from 87.5% to 98%, with discrepancies resolved by researchers. The remaining articles were then divided for screening. Finally, there are 13 studies included in the review.

3. Result

RQ1: How is GenAI being applied in CL across different educational levels and learning activities?

To conclude, the applications of GenAI in CL primarily involve embedding AI tools trained on large language models into learning platforms or chatbots, most of which are based on ChatGPT. Other content to answer the RQ1 can be found the Table 1 and Table 2.

Table 1. Demographic information of the selected articles

Publish year	N	Participant					Group size	N
		Number	N	Grade	N	Gender		
2025	4	<10	2	University	9	Equal ratio	7	Small (2-3)
2024	5	30<x<100	9	Primary	2	The ratio of females is	3	Medium (4-6)
2023	4	100<x<200	2	Secondary	1	much higher than males		Large (>6)
				Secondary school teachers	1	No mentioned	3	

Table 2. Summary of GenAI-related activities in CL

Type of Activities	Definition	Example	Article No.	N
Collaborative Problem-Solving Tasks	Teams work together to solve open-ended problems.	Database optimization; Designing advertising slogans.	1, 2, 3, 6, 7, 8, 10, 12	8
Role-Playing and Simulation	Learning through virtual scenarios or role-playing.	Tasks involved role-playing conversations with cartoon characters or famous people.	5	1
Knowledge Construction and Concept Mapping	Building knowledge networks or concept maps.	Defining and providing examples of semantic concepts; Comparing methods with historical mathematicians.	4, 8, 11, 13	4
Reflection and Metacognitive Tasks	Tasks involve self-reflection or strategic adjustments.	Discussed ChatGPT feedback, evaluated teaching practices; Reflection on design.	3, 5, 7	3
Language and Skill Training	Focused on language practice or skill enhancement.	Received worksheets with topics and prompts to guide their interactions with GenAI.	5	1

Note. The references of included articles with sequence numbers are in the appendix.

RQ2: What are the opportunities and limitations of using GenAI in CL?

The use of GenAI in CL presents several opportunities that enhance engagement, knowledge construction, and personalized learning experiences. First, GenAI-powered conversational agents and chatbots support collaboration by providing real-time feedback, scaffolding discussions, and improving cognitive engagement in group activities (Hu et al., 2025). These tools encourage students to articulate their thoughts more clearly and engage in deeper analysis, as seen in the studies on online collaborative writing and EFL speaking skills (Hu et al., 2025). Additionally, GenAI can facilitate structured reflective practices for educators, enabling them to refine their teaching strategies in collaborative settings (Arefian et al., 2024). In more immersive environments, GenAI-enhanced pedagogical agents (GPAs) integrated into augmented reality (AR) lessons significantly improve students' conceptual understanding and reduce cognitive load, demonstrating the potential of AI-driven interactive learning (Wei et al., 2025). Furthermore, GenAI can assist in synthesizing game levels tailored to different collaboration intensities, allowing students to experience varying degrees of teamwork and strategy development (Liu et al., 2023). Another advantage is the ability of GenAI to support culturally responsive learning, fostering cross-cultural collaboration in subjects like calculus, where students from different countries can engage in dialogues (Edwards et al., 2023).

However, the implementation of GenAI in CL also comes with notable limitations. One major concern is the accuracy and reliability of AI-generated content. Studies have shown that conversational agents may make errors in detecting student contributions, leading to trust issues and reduced participation in discussions (Do et al., 2023). Additionally, while GenAI can facilitate collaboration, students may become overly reliant on AI-generated responses rather than engaging in critical thinking or peer-to-peer interaction (Liu & Wang, 2024). Some studies also highlight the challenges of integrating AI in traditional classroom settings, where technological literacy varies among students and educators, potentially limiting AI's effectiveness (Canonigo, 2024). Finally, the effectiveness of GenAI in CL depends on the type of task and learning environment, with some studies showing no significant impact on behavioral engagement or writing performance despite AI

integration (Hu, et al., 2025). These findings suggest that while GenAI offers promising enhancements to CL, careful implementation and continuous refinement are necessary to maximize its benefits.

Appendix

For the references of included papers with sequence numbers see the link:
<https://drive.google.com/file/d/1rqCcJ6VfXJrR7yhiqfTJVkxakeUEwUlo/view?usp=sharing>

References

- Arefian, M. H., Çomoğlu, I., & Dikilitaş, K. (2024). Understanding EFL teachers' experiences of ChatGPT-driven collaborative reflective practice through a community of practice lens. *Innovation in Language Learning and Teaching*, 1–16. <https://doi.org/10.1080/17501229.2024.2412769>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Canonigo, A. M. (2024). Leveraging AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of Computer Assisted Learning*, 40(6), 3215–3229. <https://doi.org/10.1111/jcal.13065>
- Edwards, M. T., Yang, Z., & Lopez-Gonzalez, C. A. (2023). Fostering Culturally-Responsive Calculus Instruction: Enhancing Global Learning Experiences Through AI Integration. *Ohio Journal of School Mathematics*, 95, 39-47.
- Hu, W., Tian, J., & Li, Y. (2025). Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *The Internet and Higher Education*, 65, 100979-. <https://doi.org/10.1016/j.iheduc.2024.100979>
- Lai, M., & Law, N. (2006). Peer Scaffolding of Knowledge Building Through Collaborative Groups with Differential Learning Experiences. *Journal of Educational Computing Research*, 35(2), 123–144. <https://doi.org/10.2190/GW42-575W-Q301-1765>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the. Methodology. *Implementation Science: Iscus*, 5(1), 69–70. <https://doi.org/10.1186/1748-5908-5-69>
- Liu, H., Choi, M., Kao, D., & Mousas, C. (2023). Synthesizing Game Levels for Collaborative Gameplay in a Shared Virtual Environment. *ACM Transactions on Interactive Intelligent Systems*, 13(1), 1–36. <https://doi.org/10.1145/3558773>
- Liu, W., & Wang, Y. (2024). The Effects of Using AI Tools on Critical Thinking in English Literature Classes Among EFL Learners: An Intervention Study. *European Journal of Education*, 59(4). <https://doi.org/10.1111/ejed.12804>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education sciences*, 13(4), 410. Retrieved from <https://doi.org/10.3390/educsci13040410>
- Wei, X., Wang, L., Lee, L.-K., & Liu, R. (2025). Multiple Generative AI Pedagogical Agents in Augmented Reality Environments: A Study on Implementing the 5E Model in Science Education. *Journal of Educational Computing Research*, 63(2), 336–371. <https://doi.org/10.1177/07356331241305519>

The Role of Artificial Intelligence in Supporting Inclusive Education: Challenges and Opportunities for Teachers

Xiaoqing LIN^{1*}, Yi XIN^{2*}

^{1,2}Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong SAR, China
s1151953@s.eduhk.hk, s1151955@s.eduhk.hk

Abstract: *This paper explores the potential of artificial intelligence (AI) to enhance inclusive education for inclusive education practitioners. AI technologies offer the potential for personalized learning experiences through tailored lesson plans and adaptive learning strategies. AI also support teachers by providing personalized resources, assisting with lesson planning, and generating individualized education plans. However, research on effective AI integration in inclusive teacher education settings is limited, particularly regarding in pre-service and in-service training. Concerns about the quality and potential biases of AI-generated content require a future emphasis on AI literacy among inclusive practitioners. This paper suggests a collaborative approach between educators, AI developers, and researchers to ensure that AI tools effectively address the unique challenges faced by teachers, ultimately fostering the effectiveness of AI technology in inclusive education.*

Keywords: artificial intelligence, inclusive education, teacher education, professional development

1. Introduction

AI is an evolving field without a universal definition (Sheikh et al., 2023), but it generally refers to technologies that simulate human tasks (Baker & Smith, 2019). In education, AI has the potentials to transform by supporting teaching through personalized learning approaches and supporting inclusive education (IE) through tailored tools for students with special education needs (SEN). However, research on how IE teacher was trained and empowered through AI remains limited (Spulber, 2024). This paper discusses how AI supports IE teachers and identifies future directions for research.

2. AI as a Learning Platform

AI, as a learning platform, provides personalized learning for pre-service and in-service teachers, especially those with limited experience in teaching students with SEN (Yang & Bai, 2020). Generative AI platforms (e.g., MagicSchool and Curipod) help adapt lesson materials (Waterfield et al., 2024), and support instructional approaches through developing personalized contents (Marino et al., 2023). On the other hand, immersive technologies such as metaverse provides immersive learning environments, where teachers can practice IE pedagogies, and engage students with SEN (Khazanchi & Khazanchi, 2024) and supports students with SEN by providing individualized materials (Chițu et al., 2023). AI tools such as virtual whiteboards and 3D simulations can create inclusive and dynamic virtual classroom, allowing pre-service teachers to practice and adapt their methods, which increases lesson engagement and helps develop teaching strategies (Escotet, 2023; Chiu, 2023). Besides using AI platforms to enhance teachers' practical experiences, software applications such as AI algorithms and language analysis tools were found to enhance pre-service teachers' self-study through chatbots (Neumann et al., 2021) and assess teachers' professional growth (Yoo & Minjeong, 2020). Many institutions also offer flexible online IE related training, removing the constraints of real-time schedules and enhancing accessibility to training resources (McCoy & McNaughton, 2021).

3. AI as a Consultant

Beyond instruction, AI serves as a consultant offering real-time feedback, instructional resources, and personalized learning with aims to enhance personal growth and work efficacy (Khazanchi & Khazanchi, 2024). For example, pre-service IE teachers can use AI to practice writing individualized education plans (IEPs) (Thompson & Baumgartner, 2008) and design lesson plans for students with SEN (i.e., dyslexia) using feature of AI-generated feedback and customized contents generators (Johnston et al., 2024). Its adaptability allows for adjustments in learning experiences to suit the pacing of these teacher training programs (Shepherd et al., 2016).

At the same time, AI text generators (e.g., Bing Chat) have the potential to evaluate and filter high-quality students' present levels of academic achievement and functional performance and IEP components (Johnston et al., 2024), which assists pre-service teachers in effectively preparing for their major assignments by offering real-time feedback and support across various case sources (Escotet, 2023). For in-service teachers, AI helps improve their working efficiency and reduce paperwork burdens by streamlining the process of generating and storing IEPs electronically and confidentially using AI-powered tools (More & Hart Barnett, 2014). For teachers teaching students with learning disabilities (e.g., dyslexia), this function also helps refine and modify the original teaching materials without needing teachers to type every single word by themselves so that they will have more time to prepare the lessons.

4. AI as an Assistant

AI assists teachers in identifying, assessing, and supporting students with SEN. For the teachers who are unfamiliar with IE, AI tools can analyze student behaviors and identify learning difficulties through adaptive tutoring systems (Dutt et al., 2022; Toyokawa et al., 2023). AI also facilitates the efficacy of assessment, including grading and analysis (Marino et al., 2023). Instead of relying on paper-based exams, formative and summative assessments can be delivered in more diverse formats (i.e., gamified-based assessment); they help teachers measure knowledge, highlight the areas for improvement, and adjust the difficulty and focus of future questions to better match the teacher-learner's current abilities (Khazanchi & Khazanchi, 2024). The dynamic in assessing styles also aligns with the learning preference of many students with attention deficit hyperactivity disorder (ADHD) and learning difficulties (Iyer et al., 2023). Similarly, robotics is another representation of AI assistant that takes part in teachers' work by giving instructions more adaptively and rigorously without being influenced by states of mind (Hopcan et al., 2023). AI as an assistant helps teachers obtain more accurate results from AI-based assessment systems than traditional methods (Habib & Janae, 2024); thus, students' areas of strength and weakness could be analyzed more efficiently without requiring teachers to review each question by hands.

5. Future Needs

Despite growing interest in the use of AI in inclusive classrooms as a learning platform, consultant, and assistant, current research remains limited in both scope and depth. Much of the existing literature is either practical or review-based, with a significant lack of empirical, evidence-based studies focused on how AI support teachers specifically in IE settings. Where studies do exist, they tend to focus on students with SEN or teachers in general, overlooking the distinct challenges and experiences that IE teachers encounter, which limits the applicability and generalizability of research findings. These gaps may stem from the evolving and poorly defined nature of both AI and IE, as well as the complexity of conducting research in diverse educational settings. Practically speaking, concerns have also been raised regarding the quality and clarity of AI-generated content, as AI are often seen as helpful in reducing teacher workload, for example, drafting documents such as IEPs as mentioned before. However, questions remain about whether the content is appropriately tailored to individual students' needs and values.

Considering AI is still an evolving field with many unexplored areas, suggestions from both research and practical perspectives could be made to enlarge the potential of AI in promoting inclusive and equitable learning environments. First, teachers need a foundational understanding of how to use AI tools effectively and ethically. This includes skills in crafting effective prompts, recognizing technical limitations and ethical concerns of AI. AI literacy should be integrated into teacher education and professional development programs, supported by scientific frameworks to guide technology integration in teaching. Secondly, the development of effective AI tools for IE requires a collaborative ecosystem involving teachers, schools, and technical specialists to design AI tools tailored to IE challenges. Teachers should contribute insights from their classroom experience, helping to shape AI tools that respond to real IE classroom needs, and AI developers and researchers should ensure that AI-generated content meets high standards of quality and relevance. Therefore, more empirical research, such as case studies, qualitative research, and longitudinal studies, is needed, which has less reliance on control groups while still helping to explore the interplay between AI technology and teachers, ensuring the trend of promoting AI fits IE teachers' needs. More specifically, having pilot programs is a strategy to facilitate collaboration, which was implemented to test the effectiveness of AI tools and adjust based on teacher feedback.

References

Baker, T., & Smith, L. (2019). *Educ-AI-tion Rebooted ? Exploring the future of artificial intelligence in schools and colleges*. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

- Chițu, I. B., Tecău, A. S., Constantin, C. P., Tescașiu, B., Brătucu, T.-O., Brătucu, G., & Purcaru, I.-M. (2023). Exploring the opportunity to use virtual reality for the education of children with disabilities. *Children (Basel, Switzerland)*, 10(3), 436. <https://doi.org/10.3390/children10030436>
- Chiu, T. K. F. (2024). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*. <https://www.tandfonline.com/doi/abs/10.1080/10494820.2023.2253861>
- Dutt, S., J.Ahuja, N., & Kumar, M. (2022). An intelligent tutoring system architecture based on fuzzy neural network (FNN) for special education of learning disabled learners. *Education and Information Technologies*, 27, 1–21. <https://doi.org/10.1007/s10639-021-10713-x>
- Escotet, M. Á. (2023). The optimistic future of artificial intelligence in higher education. *PROSPECTS*, 1–10. <https://doi.org/10.1007/s11125-023-09642-z>
- Habib, H., & Janae, J. (2024). Breaking barriers: How AI is transforming special education classrooms. *Bulletin of Engineering Science and Technology*, 1(02), 86–108. <https://boengstech.com/index.php/bestec/article/view/21>
- Hopcan, S., Polat, E., Ozturk, M., & Ozturk, L. (2022). Artificial intelligence in special education: A systematic review. *Interactive Learning Environments*, 31, 1–19. <https://doi.org/10.1080/10494820.2022.2067186>
- Iyer, L. S., Chakraborty, T., Reddy, K. N., Jyothish, K., & Krishnaswami, M. (2023). AI-assisted models for dyslexia and dysgraphia: Revolutionizing language learning for children. In Kumar, A., Nayyar, A., Sachan, R. K., & Jain, R. (Eds.), *In AI-Assisted Special Education for Students With Exceptional Needs* (pp. 186–207). IGI Global. <https://doi.org/10.4018/979-8-3693-0378-8.ch008>
- Johnston, S., Jameson, J. M., O’Keeffe, B. V., & Raines, A. (2024). Teaching in the era of artificial intelligence: Reimagining activities and assignments in preservice special education teacher education programs. *Journal of Special Education Preparation*, 4(2), 38–50. <https://doi.org/10.33043/5z5b435y>
- Khazanchi, R., & Khazanchi, P. (2024). Generative AI to improve special education teacher preparation for inclusive classrooms. In Searson, M., Langran, E. & Trumble, J. (Eds.), *Exploring New Horizons: Generative Artificial Intelligence and Teacher Education* (pp.159-177). Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/primary/p/223928/>
- Marino, M. T., Vasquez, E., Dieker, L., Basham, J., & Blackorby, J. (2023). The future of artificial intelligence in special education technology. *Journal of Special Education Technology*, 38(3), 404–416. <https://doi.org/10.1177/01626434231165977>
- McCoy, A., & McNaughton, D. (2021). Effects of online training on educators’ knowledge and use of system of least prompts to support augmentative and alternative communication. *Journal of Behavioral Education*, 30(6). <https://doi.org/10.1007/s10864-020-09374-6>
- Mohammed, P. S., ‘Nell’ Watson, E. (2019). Towards inclusive education in the age of artificial intelligence: Perspectives, challenges, and opportunities. In Knox, J., Wang, Y., Gallagher, M. (Eds.) *Artificial Intelligence and Inclusive Education. Perspectives on Rethinking and Reforming Education*. (pp. 17-37). Springer, Singapore. https://doi.org/10.1007/978-981-13-8161-4_2
- More, C., & Hart Barnett, J. (2014). Developing individualized IEP goals in the age of technology: Quality challenges and solutions. *Preventing School Failure*, 58(2), 103–109. <https://doi.org/10.1080/1045988X.2013.782533>
- Neumann, A. T., Arndt, T., Köbis, L., Meissner, R., Martin, A., de Lange, P., Pengel, N., Klamma, R., & Wollersheim, H.-W. (2021). Chatbots as a tool to scale mentoring processes: Individually supporting self-study in higher education. *Frontiers in Artificial Intelligence*, 4, 668220–668220. <https://doi.org/10.3389/frai.2021.668220>
- Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial intelligence: Definition and background. In *Mission AI* (pp. 15–41). Springer, Cham. https://doi.org/10.1007/978-3-031-21448-6_2
- Shepherd, K. G., Fowler, S., McCormick, J., Wilson, C. L., & Morgan, D. (2016). The search for role clarity. *Teacher Education and Special Education*. 39(2), 83–97. <https://doi.org/10.1177/0888406416637904>
- Spulber, D. (2024). AI in inclusive education which differences in research trend. *Geopolitical, Social Security and Freedom Journal*, 7(1), 85–99. <https://doi.org/10.2478/gssfj-2024-0007>
- Thompson, S. A., & Baumgartner, L. (2008). Exploring portfolios in the elementary classroom with students with disabilities/exceptionalities: Timely or time-consuming? *Exceptionality Education International*, 18(3). <https://doi.org/10.5206/eei.v18i3.7631>
- Toyokawa, Y., Horikoshi, I., Majumdar, R., & Ogata, H. (2023). Challenges and opportunities of AI in inclusive education: A case study of data-enhanced active reading in Japan. *Smart Learning Environments*, 10(1), 67. <https://doi.org/10.1186/s40561-023-00286-2>
- Waterfield, D. A., Watson, L., & Day, J. (2024). Applying artificial intelligence in special education: Exploring availability and functionality of AI platforms for special educators. *Journal of Special Education Technology*. <https://doi.org/10.1177/01626434241257237>
- Yang, S., & Bai, H. (2020). The integration design of artificial intelligence and normal students’ Education. *Journal of Physics: Conference Series*, 1453(1), 25–27. <https://doi.org/10.1088/1742-6596/1453/1/012090>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Yoo, J. E., & Rho, M. (2020). Exploration of Predictors for Korean Teacher Job Satisfaction via a Machine Learning Technique, Group Mnet. *Frontiers in psychology*, 11, 441.
<https://doi.org/10.3389/fpsyg.2020.00441>

Translating Research on “The Animated Chinese History for the Curious Minds” for Primary School Children in Hong Kong

Wai-chun Cherry AU^{1*}, Ho-yin James PONG², Jong Hyuk David KANG³

^{1,2}Department of Literature and Cultural studies, The Education University of Hong Kong, Hong Kong SAR, China

³General Education Office, The Education University of Hong Kong, Hong Kong SAR, China
awaichun@eduhk.hk, s1147965@s.eduhk.hk, djhkang@eduhk.hk

Abstract: *The integration of animation in education, particularly in teaching Chinese history, has emerged as a significant area of exploration in Hong Kong. This paper examines the role of animation as a pedagogical tool within the Primary General Studies curriculum, highlighting its potential to enhance students' motivation and historical skills, mainly historical empathy. Utilizing a mixed-method sequential explanatory research design, data were collected from various stakeholders, including primary students, teachers, parents, and education officials, to synthesize findings regarding the effectiveness of animated resources in promoting historical literacy among schoolchildren. Survey results from four schools, involving 251 Primary Five students (Nboys = 127, Ngirls = 124), revealed that 85.3% of participants perceived animation as an effective learning tool, 84% were satisfied with the learning experience facilitated by animation, and 82% demonstrated historical empathy toward historical figures when animation was used in class. Additionally, all interviewees (N = 8) expressed enjoyment in learning Chinese history through animation, citing increased historical empathy, motivation, and interest. The findings align with the principles of Cognitive Load Theory and Multimedia Learning Theory, suggesting that animated videos, which integrate visual and auditory information, effectively enhance students' ability to manage cognitive resources and deepen their immersive engagement with historical figures. Further interviews with additional stakeholders will be incorporated into the primary research. This study contributes to the growing body of literature on the use of multimedia tools in education, offering insights into how animation can serve as a transformative medium for teaching history and fostering historical empathy among young learners.*

Keywords: Animation, Historical Empathy, Mixed-method Sequential Explanatory Research Design

1. Introduction

In the era of artificial intelligence (AI), technology has emerged as a robust device, significantly influencing various academic domains. Researchers are particularly interested in examining its influence and predicting its potential impact on human society. Extensive research has been conducted on the impact of AI across different educational fields. Nevertheless, there is a notable scarcity of literature addressing its influence on history education. The present study investigates the use of AI-generated videos and their impact on students' historical literacy, including attitudes, skills, and knowledge (ASK).

2. Literature Review

Cognitive Load Theory posits that learning is optimized when instructional materials are designed to reduce extraneous cognitive load (Sweller, 2012). Animated videos can present complex information visually engagingly, thus helping students manage their cognitive resources effectively. Mayer (2014) argues that people learn better through multimedia, as learning from words and pictures can be better than learning from mere words or text alone. This theoretical framework supports using animated videos in primary education, as combining visual and auditory information can enhance comprehension, retention, and attention spans among schoolers. Previous studies have shown that animated content captures students' attention more effectively than traditional instructional methods. Hsin, Li, and Tsai (2014) conducted a systematic literature review from 2003 to 2013 on how technologies influence young children's learning and found that technologies positively affect children's performance across developmental domains. They found that students reported higher levels of interest and engagement when learning using technology and animated videos compared to textbooks. Kosslyn et al. (2006) explain how mental imagery can enhance learning and memory retention, as evidenced by findings from cognitive neuroscience. Animations create more visual images for teaching materials, which could draw students' attention and raise their motivation to learn. Hence, students who learned through animation demonstrated greater understanding and recalled more learning content than those who received traditional instruction.

Historical empathy refers to understanding and appreciating individuals' perspectives, emotions, and contexts from the past. This skill is crucial for developing a deeper understanding of historical events and figures and fostering a meaningful connection to history. Davison (2017) proposed two types of empathy: emotional empathy, which involves the vicarious sharing of emotions, and cognitive empathy, which encompasses mental perspective-taking. When teaching historical empathy, key considerations may involve contextualization, fostering nuance, and creating opportunities for emotional engagement (Bartelds, et al., 2020). Karn (2024a; 2024b) demonstrated the potential for fostering two types of historical empathy through immersive experiences, such as via historical tours. By portraying characters' thoughts, feelings, and experiences, animation can help students to develop empathy for historical figures. This engagement allows students to "walk in the shoes" of historical figures, fostering a deeper understanding of their motivations and challenges.

Animated videos are a dynamic educational tool that can effectively present historical narratives, enriching students' immersive experiences (Zhuang, 2023). Furthermore, these videos simplify complex historical concepts, making them more accessible to primary students by integrating visual and auditory elements that enhance engagement and information retention. However, the issue of the relative benefits of using interactive pedagogy with animations compared to conventional methods remains an area requiring further research. The present study aims to examine the effectiveness of animated videos in fostering historical literacy, mainly focusing on historical empathy among primary schoolers.

3. Research Methodology

Research Question:

To what extent is the use of AI-generated animated videos effective in enhancing students' historical literacy, including attitudes, skills, and knowledge (ASK), in Chinese history lessons?

This study employed a mixed-method sequential explanatory research design to address the research question. The research was conducted in three stages as follows.

Stage 1: Collaborative lesson planning with teachers and the distribution of teaching resource packages to four target schools. (Resource link: <https://youtu.be/YebEmOm8Mtg?feature=shared>)

Stage 2: Administration of a pre-test before the trial lesson, followed by a post-test and survey after the lesson to students from the four target schools (N = 251 students).

Stage 3: Conducting focus group interviews with students from the four target schools (N = 8 students).

4. Results

4.1. Results from Quantitative Analysis

A total of 251 Primary Five students participated in this study, consisting of 127 males (50.6%) and 124 females (49.4%). The linguistic distribution of participants was predominantly Cantonese (71.9%), followed by Mandarin (14.6%), English (12.3%), and other languages (0.8%). The General Perception of the Animation Survey (GPAS) aims to assess students' learning, i.e., their satisfaction, motivation, engagement and knowledge. The reliability of the measurement scale used in this study was high, with a Cronbach's alpha coefficient of $\alpha = 0.903$ based on 10 items, indicating a strong internal consistency in the GPAS. Analysis revealed no significant gender differences in the ratings across all survey items. Seven items received favorable ratings of 80% or above. The highest-ranked item was G9 (I think that using animation in class is a very effective way of learning), with 85.3%. This was followed by G2 (I have learned scientific knowledge about Chinese history in class when using animation), with 84.4%, and G8 (I am satisfied with my learning experience using animation), with 83.8%. The survey results indicated that students were more engaged, actively participated in class activities and discussions, and had empathy toward the difficulties and situations faced by historical figures when animation was used in class. Overall, participants regarded the animated materials as effective in enhancing their learning and expressed satisfaction with the learning experience.

Table 1: The ranking, mean and SD, and % of agree and strongly agree preference of the GPAS (N = 251)

Ranking	Mean	S. D	Agree & Strongly agree (%)
G9: I think that using animation in class is a very effective way of learning.	3.32	0.91	85.3
G2: I have learned scientific knowledge about Chinese history in class when using animation.	3.22	0.94	84.4
G8: I am satisfied with my learning experience using animation.	3.31	0.92	83.8
G3: My interest in scientists in Chinese history has increased in class when using animation.	3.23	0.95	82.3
G7: I can have the empathy to experience the difficulties and situations faced by historical figures when using animation in class.	3.21	0.94	82.1
G5: I pay more attention to my studies when I use animation in class.	3.20	0.97	81.5
G10: I hope to learn more about the stories and deeds of Chinese scientists.	3.14	1.02	80.3
G4: I am more motivated to learn in class when using animation.	3.22	0.99	79.9
G6: I am more engaged and actively participate in class activities / answering questions / discussion when using animation in class	3.19	0.97	79.6
G1: I enjoy watching animated videos in class.	3.42	1.53	78.5

The results of the pre-test and post-test indicate that the trial lesson was effective in improving participants' historical knowledge, as demonstrated by an increase in the percentage of correct scores across seven items, as shown in Table 2. Given that the present study utilized nominal data (with options of "Yes" or "No"), the non-parametric related-sample McNemar Test was employed to assess significant differences between the pre-test and post-test results. Items 8 and 10 were excluded from further analysis, as their content was unrelated to the animated video used in the trial lesson. Most students answered correctly in Item 1 ("During the Ming Dynasty, missionaries were allowed to come to China to spread their faith") and Item 2 ("During the Ming Dynasty, Chinese people had the freedom to interact with foreigners"). The Chi-Square Test revealed no significant gender differences across all items except for Item 1. This analysis underscores the effectiveness of the trial lesson in enhancing students' historical knowledge. The findings also suggest that the animated video served as a valuable pedagogical tool, contributing to improved learning outcomes in Chinese history.

Table 2: Results of the pre-test and post-test in percentage of correct scores (n=251)

items	Correct % of pre-test	Correct % of post-test	Comparing post-test with pre-test (+ / -)	Significant difference ¹	Effect size ²
1. During the Ming Dynasty, missionaries were allowed to come to China to spread their faith.	76.5	88.5	+	.001***	0.275
2. During the Ming Dynasty, Chinese people had the freedom to interact with foreigners.	43.1	73.1	+	.032*	0.137
3. During the Ming Dynasty, there was no religious freedom for Chinese people.	50.0	52.7	+	.001***	0.239
4. During the Ming Dynasty, no Chinese people had foreign names.	57.3	45.5	-	NS	---
5. Chinese people created world maps during the Ming Dynasty.	46.9	52.7	+	NS	---
6. Xu Guangqi was a Chinese scientist.	74.6	72.7	-	NS	---
7. Xu Guangqi translated "Elements of Geometry" into Chinese.	70.4	79.6	+	NS	---
8. Cai Lun was the first in China to use a telescope to observe celestial bodies.	50.0	46.9	-	.017*	0.154
9. Western missionaries created mathematical terminology, such as the Chinese names for point, line, parallel line, triangle, and quadrilateral.	44.6	45.0	+	.022*	0.145
10. Sun Yat-sen collaborate with missionaries to improve the Chinese calendar	39.2	54.7	+	NS	---

Note 1: NS = Not Significant; *** Correlation is significant at the 0.01 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

Note 2: The effect size of Chi-Square (X^2) Test with a value of 0.1 is considered a small effect, 0.3 a medium effect, and 0.5 a large effect.

4.2. Results from Qualitative Analysis

Following the quantitative survey, the study proceeded to the qualitative phase, which involved conducting focus group interviews. A constant comparative analysis method was employed to analyze the interview data.

A total of eight Primary Five (Grade 5) students participated in the interviews, comprising six females and two males. The interview questions were structured around two themes:

I. Content of the Animated Video: This section focused on the students' perceptions and reflections on the content presented in the animated video used during the class.

II. Learning Experience with Interactive Pedagogy: The second part explored students' experiences regarding the interactive nature of the lesson and the role of animated video in enhancing their learning process.

All interviewees ($n = 8$) provided positive feedback on the use of animations in class, expressing enjoyment in watching the animated video. They described the video as interesting, lively, and visually appealing, with cute depictions of historical figures. Particularly, 38% of interviewees highlighted the video's effectiveness in helping them understand historical figures and acquire diverse knowledge. For instance, Student B from School 3 remarked, "When there's an animation, we feel happy because we can watch something funny and lively, and in the process (of the lesson), we can learn different knowledge."

When comparing the animated class with traditional teaching methods, all participants favored the former, citing the visuals as more engaging and easier to comprehend. They found the subject matter more interesting, which significantly improved their learning experience. All interviewees also agreed that watching the animation in the trial lesson heightened their interest in topics related to historical figures. For example, Student A from School 4 stated, "Watching animations could make it more engaging and spark our interest (in Chinese History)." Moreover, interviewees reported being deeply immersed in the animation, identifying with the historical figures' lives, experiences, growth, and achievements. They tended to consider issues from the perspective of the historical figure (Xu Guangqi) and how they navigated life challenges. This immersive experience facilitated the development of historical empathy among students, making it easier for them to connect with the historical context during the animated class. All participants would recommend the series "The Animated Chinese History for the Curious Minds" to other classes, particularly for junior grades, as they perceived the videos to be more suitable for younger students. These findings underscore the effectiveness of animated videos in fostering engagement, historical empathy, and a deeper understanding of Chinese history among primary school students.

5. Discussion

The present study highlights the significant potential of animated videos as an effective pedagogical tool in teaching Chinese history to primary schoolers in Hong Kong. The quantitative results indicate a strong connection between the use of animation and increased student engagement, motivation, and historical empathy. Notably, 85% of participants regarded animation as a very effective way of learning, nearly 80% reported active participation in animated class activities, and 82% expressed a sense of historical empathy toward historical figures. These results align with prior research (Bartelds, et al., 2020; Davison, 2017; Karn, 2024a; 2024b), which highlights the role of animation in fostering historical empathy among learners. Regarding the cognitive learning perspective, findings from the present study align with previous literature on learning theories, including the Cognitive Load Theory and Multimedia Learning Theory (Mayer, 2014; Sweller, 2012). By presenting complex historical narratives through engaging visual and auditory elements, animated videos help students manage their cognitive resources more efficiently, enhancing historical skills like critical thinking and problem-solving. The animated representations also allow students to immerse themselves in the perspectives of historical figures, fostering a more profound sense of historical empathy. Qualitative data further supports these findings, as students overwhelmingly preferred animated lessons over traditional methods, describing the former as more engaging and easier to understand. The incorporation of interactive elements, such as games and discussions, further enriches the learning experience and promotes higher-order thinking skills. Additionally, the results from the history tests demonstrated a significant improvement in students' historical knowledge following the animated lesson. This suggests that animated videos, when combined with interactive pedagogy, effectively promote students' affection, skills, and knowledge (ASK) in history education.

However, some considerations regarding implementing animated resources in the curriculum have been raised. One potential concern is the need to carefully select and produce animated content accurately representing authentic historical events and figures. Given the complexities of historical issues, it might be susceptible to producing biased or opinionated animations or narratives. This requires collaboration among historians, educators, and animators to create engaging and accurate content within a historical context. Moreover, ensuring equitable access to digital resources is essential for maximizing the benefits of animated learning tools across diverse educational settings.

6. Conclusion

In conclusion, this study lends support to evidence on the efficacy of animated videos as a pedagogical tool in teaching Chinese history to primary school students in Hong Kong. The integration of animation not only enhances student engagement and motivation but also fosters historical empathy, which is crucial for developing a nuanced understanding of history. The findings suggest that animated resources can serve as a valuable addition to the Primary General Studies curriculum, aligning with established learning theories. Future research could expand to include a larger and more diverse sample of students, as well as gather perspectives from teachers, parents, and education officials. Such a comprehensive approach would provide deeper insights into the impact of animation combined with interactive pedagogy on historical literacy. By addressing implementation challenges and ensuring the quality and accessibility of animated resources, educators can further enhance the effectiveness of animation in inspiring curiosity and fostering a deeper connection to Chinese history among students.

Acknowledgments

Dr. Kang, Jong Hyuk David, for his support and guidance while preparing the paper. He is currently the chief investigator of the project 'Title: Translating Research on Chinese History into Public Education in Hong Kong.'

References

- Bartelds, H., Savenije, G. M., & van Boxtel, C. (2020). Students' and teachers' beliefs about historical empathy in secondary history education. *Theory & Research in Social Education*, 48(4), 529–551. <https://doi.org/10.1080/00933104.2020.1808131>
- Davison, M. (2017). Teaching about the First World War today: Historical empathy and participatory citizenship. *Citizenship, Social and Economics Education*, 16(3), 148-156. <https://doi.org/10.1177/2047173417736906>
- Hsin, C. T., Li, M. C., & Tsai, C. C. (2014). The Influence of Young Children's Use of Technology on Their Learning: A Review. *Educational Technology & Society*. Vol.17 (4), 85-99. ISSN: EISSN-1436-4522.
- Karn, S. (2024a). Walking in their footsteps: Historical empathy and experiential learning on battlefield study tours. *Historical Encounters*, 11(1), 30-42. <https://doi.org/10.52289/hej11.103>.
- Karn, S. (2024b). Designing historical empathy learning experiences: a pedagogical tool for history teachers. *History Education Research Journal*, 21 (1), 6. DOI: <https://doi.org/10.14324/HERJ.21.1.06>.
- Kosslyn, S.M., Thompson, W.L., Ganis, G. (2006). *The Case for Mental Imagery*. Oxford Psychology Series. Oxford University Press: Oxford, UK; New York, NY, USA, 2006. <https://doi.org/10.1093/acprof:oso/9780195179088.001.0001>.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press <https://doi.org/10.1017/CBO9781139547369.005>.
- Sweller, J. (2012). Cognitive Load Theory: Recent Theoretical Advances. In J. L. Plass, R. Moreno, and R. Brünken (Ed.) *Psychology of Learning and Motivation* (pp. 29–47). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.004>.
- Zhuang, M. (2023). Animation of Experiment: The Science Education Film and Useful Animation in China. *Animation*, 18(2), 152-166. <https://doi.org/10.1177/17468477231182914>.

Strand

Learning Outcomes and AI-Driven Assessments

Fine-Grained Student Engagement Evaluation with Large Language

Model: A case in Collaborative learning

Sixu AN^{1*}, Haoli WANG², Qihao YANG³, Yu YANG⁴, Guandong XU⁵

^{1,2,3,4}Centre for Learning, Teaching and Technology, The Education University of Hong Kong, Hong Kong SAR, China

s1154340@s.edu.hk, whaoli@eduhk.hk, qyang@eduhk.hk, yanggy@eduhk.hk, gdxu@eduhk.hk

Abstract: *In recent years, the large language model (LLM) has performed well in education. With its outstanding language understanding ability, a large language model always works as a teaching assistant in educational applications. Collaborative learning (CL) is a useful learning form that can also be powered by LLM. One important problem in CL is engagement. Participants usually show different performances during CL. Engagement can be identified in three aspects. Traditionally, teachers use questionnaires or evaluation forms to measure students' engagement during the CL activity. Compared with the traditional engagement evaluation method, we propose an LLM-based method to evaluate student engagement from the three aspects based on the course dialogue content. We tested our method with two multi-agent construction class data. Our result shows that LLM can give evaluation results based on predefined definitions and evaluation standards. In future work, we will use more student behaviors to make the evaluation more accurate.*

Keywords: Large Language Model, Collaborative learning, Student Engagement

1. Introduction

Collaborative learning (CL) is a very important learning activity that improves student learning efficiency (Gokhale, 1995). In collaborative learning, students can work with each other to solve a complex problem and try to find the correct solution. However, one important problem in collaborative learning is engagement. Students in one group usually show different performances based on their cognitive level, personality, background, etc. Qureshi, M. A. et al. (2023) point out that students' engagement performance has an important impact on their learning performance. They try to investigate what factors are important for students' engagement performance in collaborative learning, including social interaction with peers, etc.

To explore student engagement performance more deeply, research starts to focus on how to measure student engagement performance in collaborative learning activities. There are many ways to explore the engagement performance. Fredricks et al. (2004) give definitions on the engagement performance. They think Engagement can be measured in three aspects:

1) Behavioral engagement (BE): students show performance that is conducted with positive learning outcomes, which can be identified as behavioral engagement. For example, students who pay attention to the tasks, follow the rules, and make an effort to finish the task can be regarded as engaged in the activity.

2) Cognitive engagement (CE): this level of engagement focuses more on the students' psychological effort to reflect on their thinking on the learning task. The students who show thinking on the problem can be classified as engaged cognitively in the learning process.

3) Emotional Engagement (EE): emotional engagement refers to students' learning experience during the process, including their academic feelings. For example, whether they feel enjoyable about the learning process.

In traditional collaborative learning activities, student engagement is usually classified as engaged or disengaged. This kind of measurement is coarse-grained, which cannot help the teacher identify what aspects the students should pay more attention to in order to improve their learning performance. With the advent of a large language model (LLM), artificial intelligence has performed well in question-answer tasks with humans. LLM can generate some human-like dialogue content based on what it has learned from its large amount of training data. A large number of tests show that it has great potential to understand human dialogue content and answer some relative questions based on its understanding. Based on its excellent performance in natural language understanding, it is possible to leverage the great language processing ability to analyze the student dialogue context and measure their engagement based on the dialogue content. This work tries to assess the students' engagement through three aspects by analyzing students' online course discussion dialogue with LLM.

2. Related Work

Engagement is important for teachers to measure students' learning performance. Much research has been carried out to assess the students' engagement. In the traditional assessment method, the teacher usually observes students' discussion behavior and judge students' engagement. Seidel, T et al. (2021) discuss what kind of diagnostic skills can be used when teachers observe students' engagement. In this work, they also want to infer students' characteristic features by observing engagement performance. The work done by Greene, B. A. et al (2004) uses path analysis to predict the students' cognitive engagement in their self-efficacy and their achievement goals. Li, T. et al. (2024) measure high school student engagement during the science learning process. They proposed the initial instrument of student engagement in science learning (SESL), which included 20 items in four dimensions (cognitive, emotional, behavioral, and agentic engagement). Xu, B. et al. (2024) used a mixed methods approach in the Chinese collaborative learning engagement scale. They carried out a qualitative study to assess Chinese student engagement, which proves consistent with student engagement conception in Western countries. However, existing methods are mostly based on handcrafted scales or questionnaires for engagement measurement. To evaluate students' performance, teachers usually observe students' behaviors and what they say in the course. Then, they will give an evaluation based on what they have observed. In this work, we propose an automatic engagement evaluation method based on the LLM, which can work as a teaching aid tool.

3. Method

You are a teacher who analyzes student discussion dialogue in terms of the three dimensions of engagement: Behavioral Engagement, Cognitive Engagement, and Emotional Engagement the definitions of the three aspects are defined as follows:

Behavioral Engagement: Identify instances where students demonstrate positive learning behaviors, such as paying attention to the tasks, following rules, and making efforts to complete the task. Note specific examples where students actively participate and contribute to the activity.

Cognitive Engagement: Look for evidence of students reflecting on their thinking related to the learning task. Highlight moments where students articulate their thought processes, analyze problems, or demonstrate critical thinking about the subject matter. Provide examples that show how students engage cognitively in the learning process.

Emotional Engagement: Assess the emotional aspects of the student's learning experiences during the discussion. Identify any expressions of enjoyment, motivation, or enthusiasm about the learning process. Consider how students respond to challenges and their overall feelings toward the task. Provide specific examples that illustrate their emotional engagement.

Figure 1. Engagement Evaluation Prompt.

Currently, the language model focuses on the language data. In this work, we only use discussion content provided to LLM for engagement evaluation. A prompt is designed to clarify the evaluation process so that LLM can carry out the evaluation process. Typical student engagement assessment is usually based on some specific standards. We organized the standards into LLM prompts. First, the definition details of cognitive engagement, behavioral engagement, and emotional engagement are well described in the prompt, as shown in Figure 1. In our research, we expect that the LLM can evaluate student engagement performance in a quantitative format. Thus, we use a scale table to indicate the LLM and how to mark the student engagement level based on their behavior during the discussion. The details are summarized in Table 1. The standards in Table 1 are transferred into word descriptions and prompts. Thus, our prompt tells the LLM how to evaluate the students in the three aspects and what performance can be taken at different levels.

Table 1. Engagement Evaluation Standard

Engagement Score	Description
0	No engagement
1-3	Very Limited engagement with only little attention or effort in the discussion.
4-6	Moderate engagement: the student makes some effort in the discussion process with some reflection.
7-9	High engagement: students actively participate in the discussion with strong understanding and positive feelings for the whole learning process.
10	Exceptional engagement: students are fully immersed in the discussion with careful thinking and enthusiastic expression.

4. Experiment

To evaluate the effectiveness of the designed prompt for engagement measurement, we choose two typical online discussion dialogues for engagement performance evaluation. For the two discussions, one focuses on the social problem, and the other one is about the math problem. For these two different problems, the students always show different levels of behavioral tendency. Specifically, students who have social problems always use critical thinking and analysis. They always show their diverse thinking pattern based on their personal view and social background. Compared with social problems, in the math problem, students usually show logical thinking and analysis. They usually use formulas and deductions for problem-solving. One problem is that it is very hard for us to collect real discussions from true class with several students working together for problem solutions. Thus, we leverage the ability of LLM to create a multi-agent simulation for the class scene with question discussion. The result is taken as the first discussion conversation, which is a discussion between two students and one teacher. The discussion content is about the 2009 US economic stimulus, and students discuss its impact and effectiveness. The dialogue content is shown in Table 2.

To evaluate the general ability of the LLM for engagement evaluation, the second dialogue content is selected from the dataset from the work of Liu, J (2024). Their dataset is about the Socratic study process between teachers and students with different cognitive states. The dataset is also generated by the LLM model based on a multi-agent discussion mechanism. During the study process, the student learned math problems with the teacher, showing students different levels of cognitive thinking. Finally, by the Socratic teaching process, the students understand how to solve math problems. The details of the second conversation used for model evaluation are summarized in Table 3.

The LLM is asked to score engagement for each student in the conversation from three dimensions based on the scale standard in Table 1.

Table 2. Conversation 1 Dialogue Content

Speaker	Utterance
Mary	I think the 2009 US economic stimulus package is a good plan to combat the recession. We should talk about its effectiveness in stimulating economic growth and creating jobs.
Mary	I think the 2009 US economic stimulus package is an interesting topic. how did it affect the unemployment rate, though?
Jennifer	I don't really understand how the 2009 US economic stimulus package works. Did it really create jobs? @Mary, can you explain its impact on unemployment?
Mary	Honestly, I am not entirely sure how the 2009 US economic stimulus package affected unemployment. However, I think we should research its impact on specific industries and how it helped stimulate economic growth in certain areas. @Jennifer, I'm glad you brought that up; it's definitely something we should explore further.
Mary	I think we should also consider the potential drawbacks of the stimulus package, such as increasing national debt and potential misuse of funds. @Jennifer, I don't think we have enough information to fully understand its impact on unemployment yet. We should research more and discuss further.
Jennifer	I think we need to research the 2009 US economic stimulus package's impact on specific industries and how it helped stimulate economic growth in those areas. @Mary, I agree with you that considering the potential drawbacks is also important. But I'm still a bit lost. Can we summarize what we have so far and what we need to find out?

Table 3. Conversation 2 Dialogue Content

Speaker	Utterance
Teacher	How many clips did Natalia sell in May?
Student A	I'm confused. Does \"half as many\" mean the same number as April?
Teacher	Not exactly; \"half as many\" means half the amount. If Natalia sold 48 clips in April, what would be half of 48? That would tell us how many she sold in May. Can you solve that?
Student B	I'm still not sure. Is it like 48 divided by 2?
Teacher	Exactly, that's right! If we divide 48 by 2, we get the number of clips Natalia sold in May. What number do you get when you do that?
Student C	Oh, I get it now. So 48 divided by 2 is 28.
Teacher	Very close! Though there seems to be a little mistake. When you divide 48 by 2, do you get 28 or another number? Try again and see if you get a different result.
Student C	Oh, I'm sorry. I must have made a calculation error. Let me try again. Yeah, 48 divided by 2 is 24, not 28.
Teacher	That's correct! Natalia sold 24 clips in May. Now that we know how many clips she sold in both April and May, how do we find the total number of clips she sold in those two months?
Student C	We add the two numbers together right? So, 48 plus 24 equals 72. So, Natalia sold 72 clips in April and May.
Teacher	That's right! Good job! So, if Natalia sold 48 clips in April and 24 clips in May, she did indeed sell 72 clips in total. Well done!

5. Results

Table 4 and Table 5 summarize the evaluation results from LLM. The base model used for engagement measurement is Llama-3 -70b (Dubey, A. et al 2024).

Table 4. Conversation 1 Evaluation Result

	BE	CE	EE
Participant 1 (Mary)	7	6	6
Participant 2(Jennifer)	7	7	6

Table 5. Conversation 2 Evaluation Result

	BE	CE	EE
Student A	8	8	7
Student B	7	7	6
Student C	9	9	8

We also make the Llama-3 -70b explain the reason why it gives this rating for each student. From this reason, it can be identified that LLM can understand the multiparty dialogue with a clear summary. In conversation 1, Jennifer shows study progress from a low understanding of the topic trying to propose some questions, which shows that she has good performance on behavioral Engagement. For cognitive engagement, Jennifer can identify the potential direction, which shows that she has taken thinking activity about the study topic. For emotional engagement, Jennifer shows some extent of interest in her attitude toward studying. Thus, Jennifer is marked for 7,7,6 for her engagement performance. Compared with conversation 1, the content of conversation 2 is shorter. During the conversation, student C showed three instances of interaction with the teacher. Some important hints can be found in the conversation to mark the students. First, student C can find his own mistakes and proactively correct errors, which shows high behavioral engagement. Under the teacher's guidance, students recalculate and get the correct answer, indicating high cognitive engagement. For emotional engagement, Student C expressed his apologies after finding his mistakes and was willing to recalculate, showing an emotional attitude to face challenges and self-correction. Thus, student C has been marked for a high score of 9,9,8.

6. Conclusion

This work proves that LLM possesses the ability to evaluate student engagement from three aspects. Compared with binary engagement classification, which only classified the student as either engaged or disengaged, this fine-grained engagement evaluation can offer teachers more comprehensive information to enhance student engagement in the course.

The LLM can analyze the students' performance based on their discussion histories. However, it also identifies some potential problems. First, the rating score for all students ranges from 6-9, which is a moderate to high engagement level. No students have been marked as low engagement. One possible reason is that some important behavior indicators for low engagement, for example, distractions or off-topic comments, rarely appear in a recorded discussion text. In this case, the LLM may focus on visible participation, including

speaking or having some responding. The model will ignore some other forms of engagement, such as quiet reflection or disengagement, which is very important in evaluating students' engagement performance. Thus, for more accurate evaluation in future work, it is necessary to consider more aspects of student behaviors for engagement evaluation.

Secondly, another problem is that the evaluation effectiveness can be influenced by the context of the discussion. Specifically, if discussion content is relatively simple or lacks depth, student' performance can not reflect their real engagement. For example, for some basic knowledge discussions, the student may show high engagement even if their understanding and thinking are not deep. Therefore, in order to improve the evaluation accuracy, future research should consider how to better capture and assess the actual participation of students in discussions at different difficulty levels.

Acknowledgements

We thank all the people who support our research.

References

- Gokhale, A. A. (1995). Collaborative Learning Enhances Critical Thinking. *Journal of Technology Education*, 7, 22-30. <http://scholar.lib.vt.edu/ejournals/JTE/v7n1/gokhale.jte-v7n1.html>
- Qureshi, M. A., Khaskheli, A., Qureshi, J. A., Raza, S. A., & Yousufi, S. Q. (2023). Factors affecting students' learning performance through collaborative learning and engagement. *Interactive Learning Environments*, 31(4), 2371-2391.
- Halpin, P. F., von Davier, A. A., Hao, J., & Liu, L. (2017). Measuring student engagement during collaboration. *Journal of Educational Measurement*, 54(1), 70-84.
- Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia-social and behavioral sciences*, 31, 486-490.
- Qiu, R. G. (2019). A systemic approach to leveraging student engagement in collaborative learning to improve online engineering education. *International Journal of Technology Enhanced Learning*, 11(1), 1-19.
- Seidel, T., Schnitzler, K., Kosel, C., Stürmer, K., & Holzberger, D. (2021). Student characteristics in the eyes of teachers: Differences between novice and expert teachers in judgment accuracy, observed behavioral cues, and gaze. *Educational Psychology Review*, 33, 69-89.
- Xu, B., Stephens, J. M., & Lee, K. (2024). Assessing student engagement in collaborative learning: Development and validation of new measure in China. *The Asia-Pacific Education Researcher*, 33(2), 395-405.
- Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary educational psychology*, 29(4), 462-482.
- Li, T., He, P., & Peng, L. (2024). Measuring high school student engagement in science learning: an adaptation and validation study. *International Journal of Science Education*, 46(6), 524-547.
- Liu, J., Huang, Z., Xiao, T., Sha, J., Wu, J., Liu, Q., ... & Chen, E. SocraticLM: Exploring Socratic Personalized Teaching with Large Language Models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- Darr, C. W. (2012). Measuring student engagement: The development of a scale for formative use. In *Handbook of research on student engagement* (pp. 707-723). Boston, MA: Springer US.
- Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., ... & Ganapathy, R. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Developing a VR-based Immersive Learning Platform with a Multimodal AI Agents System for Improving University Students' Academic Presentation and Self-Regulated Learning

Yanjie SONG^{1*}, Jiachen FU², Kaiyi WU³, Chiu Tsun TSANG⁴
^{1,2,3,4}The Education University of Hong Kong, Hong Kong SAR, China
ysong@eduhk.hk, fjiachen@eduhk.hk, kwu@eduhk.hk, stevetsang@eduhk.hk

Abstract: Recent technological advancements have paved the way for Artificial Intelligence (AI) and Virtual Reality (VR) applications across various fields, including education. However, leveraging AI and VR to enhance students' academic presentation proficiency and self-regulated learning remains relatively underexamined. This article introduces an immersive learning platform built upon "LearningverseVR" that integrates multiple intelligent agents and VR technologies to facilitate self-regulated learning. The platform features two core components: (1) a multimodal AI agents' framework with generative AI, providing students with guided presentation preparation and real-time feedback; and (2) a VR-based immersive environment that simulates realistic presentation scenarios. By employing VR headsets, the platform captures various behavioral cues (facial expressions, gaze, gestures, and body language), enabling comprehensive analysis of presentation performance. A pilot study involving seven participants was conducted: three participants used the platform with multiple AI agents without VR (Group A) and four used the platform with VR and multimodal AI agents (Group B). Data collection includes questionnaires about student's perceptions and self-regulated learning, and semi-structured interviews. Results show that this platform offers novel perspectives on integrating AI and VR to support self-regulated learning in academic presentation contexts, paving the way for more extensive research in higher education settings.

Keywords: AI, VR-based immersive learning platform, Self-regulated learning, Academic presentation

1. Introduction

Academic presentations are fundamental to success in higher education and professional domains, yet many students struggle with linguistic hurdles (e.g., grammar, vocabulary) and performance anxieties (Mak, 2021; Roslim et al., 2023). Additionally, they may experience anxiety when making presentations. Self-regulated learning is increasingly recognized as a crucial skill for the 21st century, which promotes the transfer of knowledge across different contexts, which is vital for effective presentations (Zyl & Mentz, 2022). Despite widespread use of AI in education, there is a lack of comprehensive and complex exploration of its potential in developing students' academic presentation competency and self-regulated learning in immersive learning environments.

Although AI-enhanced tools have been explored to address certain facets of presentation skills, most tools focus on isolated features and do not account for the broader behavioral cues essential to effective presentations (Chen et al., 2022).

Recent advances in AI and VR present an opportunity to bridge these gaps. Generative AI, for instance, can provide targeted feedback on presentation content, but there is a dearth of platforms that integrate multimodal AI agents within VR to capture and analyze a wide range of learner behaviors in real time. In response, this paper introduces a VR-based immersive learning platform that combines multiple AI agents for multimodal feedback with SRL strategies to enhance university students' academic presentation competency. A pilot study was conducted to investigate learners' perceptions of the platform and its impact on self-regulated learning, offering preliminary insights into its potential effectiveness and directions for further refinement.

2. Design a multimodal AI agents' system

A multimodal AI agent is an advanced AI technology capable of processing and integrating multiple types of data, such as text, images, and audio, to perform complex tasks and interactions (Chen & Li, 2024). Multimodal AI agents can process and learn from diverse data types, including natural language, visual, and audio inputs, enabling them to perform complex tasks and make informed decisions (Chen & Li, 2024). Multimodal AI agents in education are transforming the learning landscape by integrating various data modalities to enhance educational experiences. For example, MusicARLtrans Net is a multimodal agent that

uses reinforcement learning to optimize music teaching strategies (Chang et al., 2024). It integrates auditory, visual, and textual data to provide personalized feedback, significantly improving learning outcomes and user satisfaction in music education. The Agent Collaboration Network (ACN) framework exemplifies a multimodal AI agent system, where multiple specialized agents collaborate to enhance response quality, personalization, and interactivity in AI search engines (Shi et al., 2024). In this project, we have designed and developed a multimodal AI agent framework to integrate multiple intelligent agents (see Figure 1). With the multimodal AI agent framework, the platform can collect and analyze diverse types of data generated by learners during their study process.

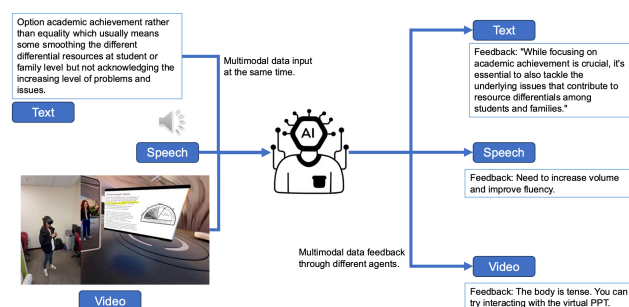


Figure 1. A sample of the multimodal AI agents system

Rather than relying solely on a collection of multiple intelligent agents, this framework leverages multimodal AI agents collecting and analysing data pertinent to all six dimensions of the presentation. By synthesizing these diverse data streams, the system can generate richer, context-aware feedback, thereby offering more targeted guidance for presenters.

3. Development of a VR-based immersive learning platform

To develop an immersive learning environment for academic presentations, we designed a 3D speech space based on LearningverseVR (Song et al., 2024) using Unity. Within this virtual reality (VR) environment, learners can monitor their facial expressions, gestures, and body language through a virtual full-length mirror, and utilize a virtual presentation board for displaying their slides (see Figure 2). By using a VR headset, the platform can collect a comprehensive range of multimodal data during the presentation process, such as voice, facial expressions, gaze, gestures, and body language.



Figure 2. The avatar's gaze, facial expressions, gestures, and body language accurately mirrored their own.

These multimodal elements are vital for evaluating presentation performance. Research has shown that analyzing diverse data streams like voice, posture, gestures, gaze, expressions) and auditory (vocal tone, volume, pacing), provides more accurate and detailed feedback than relying on a single modality alone (Chen et al., 2015). For example, a visual analytics system can parse rhetorical modes, body postures, and gestures in TED talks, demonstrating how integrated multimodal data facilitates in-depth understanding of a presenter's strengths and weaknesses (Wu & Qu, 2018). By integrating these features within an immersive VR environment, learners receive realistic, immediate feedback, enabling them to refine their delivery and gradually overcome common public speaking challenges.

In summary, integrating VR not only enables the capture of multimodal data in real-time but also heightens learners' sense of presence, thereby potentially improving their engagement with and reflection on public speaking tasks.

4. A pilot study

A pilot study was conducted to understand students' perceptions of the VR-based immersive learning platform with the multimodal AI agents' system and its impact on their self-regulated learning. The research questions are:

1. What are the students' perceptions of the VR-based immersive learning platform with the multimodal AI agents' system?
2. What are students perceived self-regulated learning after using the VR-based immersive learning platform with the multimodal AI agents' system.

This study adopts a design-based research (DBR) approach (Barab & Squire, 2016), characterised by iterative cycles of design, implementation, evaluation, and redesign within educational contexts. Design-based research is particularly appropriate for this scenario as it allows for the systematic refinement of technological interventions while simultaneously developing context-specific learning theories (Wang & Hannafin, 2005). Our pilot study represents the first iteration in this design-based research process, where we implement the initial design of our VR-based immersive learning platform with multimodal AI agents and collect data regarding its effectiveness. Findings from this iteration will inform subsequent redesigns, creating a cyclical improvement process that aligns with the iterative nature of both design-based research and self-regulated learning.

4.1. Research procedure design

We invited seven undergraduate participants. Three participants used the platform with multiple AI agents but without VR and a multimodal AI agents' system (Group A). Four participants used the platform with VR and multimodal AI agents (Group B). The duration of this pilot was 150 minutes and Figure 3 shows the research procedure.

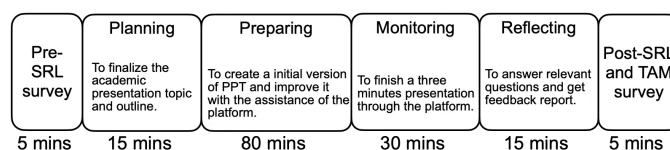


Figure 3. Research procedure

In the Planning stage, a planning agent assists learners in developing initial outlines, defining objectives, and organizing presentation topics. The planning agent establishes learning objectives through guided dialogue, encoding parameters into a shared semantic graph that synchronizes with downstream agents. In the early stages, cognitive strategy consistency is prioritized over performative enhancement. The feedback agent conducts concurrent visual-linguistic analysis using computer vision (CV) slide segmentation and NLP script evaluation. This feedback report might highlight weaknesses in slide organization, suggest improved visuals, or recommend clearer phrasing. Otherwise, a preparing assistant agent assists learners in optimizing PPT slides and presentation scripts. Building on the feedback report provided, this agent engages in iterative, multi-round dialogues with learners. It offers interactive suggestions and personalized tips for refining presentation content. The assistant agent then adopts a phased dialogue strategy, first resolving critical issues through directive questioning before collaboratively refining structural components, thereby preventing cognitive overload while supporting self-regulated learning. For instance, if a script is too lengthy or lacks a compelling narrative flow, the preparing assistant agent prompts the learner with targeted questions and sample revisions, thereby promoting the SRL processes of self-reflection and strategy use. In the Monitoring stage, all participants will give a 3-minute academic presentation (see Figure 5).

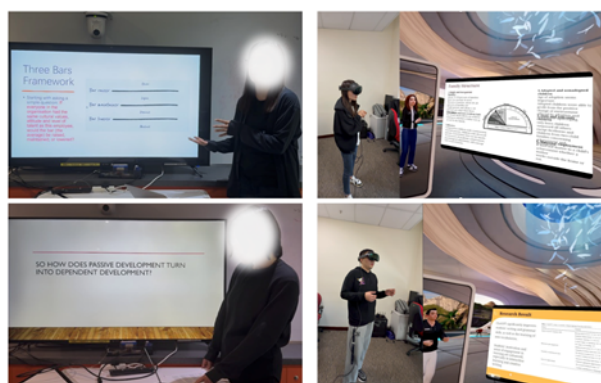


Figure 5. Different forms of presentation delivery between Group A (left) and Group B (right)

Group A will conduct an academic presentation using a computer in a real meeting room. Group B makes an academic presentation in a VR environment and can monitor its expression, gaze, gesture, and body language in real time through the multimodal agent's system. Group B's VR presentation environment implements real-time multimodal fusion through a weighted decision layer. When detecting conflicting signals (e.g., appropriate gestures but poor slide-visual synchronization), the platform will give recommendations. In the Reflecting stage, participants are required to answer questions generated by the platform and receive a feedback report based on specified presentation criteria. This helps participants better reflect on their learning processes and performance. And this enables learners to pinpoint specific self-regulated learning gaps, such as maintaining overconfidence in slide clarity despite persistent system alerts about text overcrowding or misplaced diagram annotations.

4.2. Data Collection and Data analysis

In terms of data collection, this pilot study was conducted with questionnaires including pre- and post SRL survey (Barnard et al., 2009), and post survey of students' perceptions of the VR-based platform adapted from TAM (Al-Adwan et al., 2023). These surveys used five-point Likert scale (from 1-strongly disagree to 5-strongly agree). Descriptive statistics were employed to summaries the SRL performance of participants on both the pre-survey and post-survey between two groups. And the post survey of TAM for perception between the two groups was compared.

A Self-Regulated Learning (SRL) survey was administered to both Group A (3 participants) and Group B (4 participants) before and after the intervention. In addition to SRL, a Technology Acceptance Model (TAM) questionnaire was used to collect participants' perceptions across different dimensions.

5. Preliminary findings and discussion

Regarding students perceived SLR, Group A's mean SRL score increased from 2.92 (pre-test) to 3.17 (post-test); individual scores ranged from 2.16 to 3.33. Group B mean SRL score rose from 3.18 (pre-test) to 3.60 (post-test), with individual scores ranging from 2.70 to 4.05. Given the small sample sizes, only descriptive methods were employed to observe trends. The findings suggest that the VR-based immersive learning platform augmented by multimodal AI agents holds promise for enhancing both academic presentation skills and self-regulated learning. Participants in both Group A ($n = 3$) and Group B ($n = 4$) demonstrated an increasing trend in mean scores from pre- to post-intervention, indicating increased confidence and autonomy in managing their learning processes.

The results of the students' perceptions of the VR-based immersive environment show that Group B's mean scores are higher than Group A's mean scores, with Behavioral Intention increasing from 2.90 to 3.88, Perceived Ease of Use from 3.00 to 3.81, Self-Efficacy from 3.10 to 3.75, Perceived Enjoyment from 2.60 to 3.17, Perceived Cyber Risks from 3.00 to 3.56, Personal Innovativeness from 3.20 to 3.58, and Perceived Usefulness from 3.60 to 3.75. The results imply that learners became more inclined to use the platform and felt more capable of navigating it effectively. Notably, Perceived Usefulness increased only modestly, a finding echoed in semi-structured interviews where some students perceived the multimodal analysis and feedback reports to be still under refinement.

Participants reported that VR-based presentation scenarios and instant AI-driven feedback enhanced their preparedness and motivation. Several emphasized that immersive rehearsal reduced their anxiety by approximating the dynamics of real presentation settings, allowing them to practice essential behaviors such as maintaining eye contact, managing gestures, and using appropriate vocal projection.

Overall, these results suggest that the project is viewed more positively in terms of ease of use, enjoyment, security, and user intention, aligning with participants' qualitative feedback regarding an improved user experience in the VR-based academic presentation environment.

6. Conclusion

These preliminary results align with the project's aim of integrating AI and VR technologies to foster students' academic presentation competency and self-regulated learning. While the limited sample size constrains the generalizability of the findings, the observed increases in SRL and positive TAM responses underscore the platform's feasibility and potential value. Further research with larger cohorts is planned to confirm these initial outcomes and refine the platform's design for broader educational contexts.

Acknowledgements

The study was funded by the Fund for Innovative Technology-in-Education (FITE) (Ref. 05404 and 05408) (2023-26), and Joint Centre for Artificial Intelligence (JC_AI) (Ref. 02190) (2024-26), University Grant Committee, Hong Kong.

References

- Al-Adwan, A. S., Li, N., Al-Adwan, A., Abbasi, G. A., Albelbisi, N. A., & Habibi, A. (2023). Extending the technology acceptance model (TAM) to Predict University Students' intentions to use metaverse-based learning platforms. *Education and Information Technologies*, 28(11), 15381-15413.
- Barab, S., & Squire, K. (2016). Design-based research: Putting a stake in the ground. In *Design-based research* (pp. 1-14). Psychology Press.
- Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *The internet and higher education*, 12(1), 1-6.
- Chang, J., Wang, Z., & Yan, C. (2024). MusicARLtrans Net: a multimodal agent interactive music education system driven via reinforcement learning. *Frontiers in Neurorobotics*, 18, 1479694.
- Chen, J., Lai, P., Chan, A., Man, V., & Chan, C. H. (2022). AI-assisted enhancement of student presentation skills: Challenges and opportunities. *Sustainability*, 15(1), 196.
- Chen, L., Leong, C. W., Feng, G., Lee, C. M., & Somasundaran, S. (2015, September). Utilizing multimodal cues to automatically evaluate public speaking performance. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 394-400). IEEE.
- Chen, W., & Li, Z. (2024). Octopus v3: Technical report for on-device sub-billion multimodal ai agent. *arXiv preprint arXiv:2404.11459*.
- Mak, S. H. Y. (2021). Analysing the needs of EFL/ESL learners in developing academic presentation competence. *RELC Journal*, 52(3), 379-396.
- Roslim, N., Nimehchisalem, V., Abdullah, M. H. T., & Razali, N. M. (2023). Students' perceptions of personality traits, presentation skills, and audience factors in their online presentations. *Studies in English Language and Education*, 10(2), 777-788.
- Shi, Y., Xu, M., Zhang, H., Zi, X., & Wu, Q. (2024, October). A learnable agent collaboration network framework for personalized multimodal ai search engine. In *Proceedings of the 2nd International Workshop on Deep Multimodal Generation and Retrieval* (pp. 12-20).
- Song, Y., Wu, K., & Ding, J. (2024). Developing an immersive game-based learning platform with generative artificial intelligence and virtual reality technologies—"LearningverseVR". *Computers & Education: X Reality*, 4, 100069.
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4), 5-23.
- Wu, A., & Qu, H. (2018). Multimodal analysis of video collections: Visual exploration of presentation techniques in TED Talks. *IEEE Transactions on Visualization and Computer Graphics*, 26(7), 2429-2442.
- Zyl, S., & Mentz, E. (2022). Deeper Self-Directed Learning for the 21st Century and Beyond (pp. 50-77).

A meta-review on the role of AI in English writing: A recent trend

Hetian YU¹, Lan YANG^{2*}

The Education University of Hong Kong, Hong Kong SAR, China
s1146961@s.edu.hk, yanglan@eduhk.hk

Abstract: *This meta-review evaluated the expanding use of generative artificial intelligence in English writing education, using Pekrun's Control-Value Theory (CVT) to assess its educational impact. This analysis includes 19 review articles published between 2020 and 2025, highlighting a growing focus on automated writing evaluation tools, including ChatGPT and other AI applications in English writing education. Collectively, these review articles examined a total of 719 studies. Our findings show a concentrated effort on quantifying the effects of AI tools on students' writing performance. However, studies documented in these review papers lack a deep exploration of the psychological impacts on students, particularly in terms of their perceived control and value. This review not only identifies these gaps but also offers practical suggestions and directions for future research for more effectively incorporating these technologies into English writing instruction, aiming to enhance both educational outcomes and student engagement.*

Keywords: AI, English writing, control-value theory, meta-review

1. Introduction

The advent of generative AI has stormed the education community worldwide. Over the past five years, there has been surge in research on the integration of artificial intelligence (AI) in English writing (Lu et al., 2024; Shi et al., 2025; Yan, 2023). The benefits of AI on learners have been widely researched, with the potential to improve writing performance, enhance motivation, and promote self-regulated writing (Su et al., 2023). Given the rapid accumulation of research on reviewing the impacts of AI in English writing, it is essential to conduct a meta-review on previous review papers to synthesize existing research on this topic, contributing to provide clear directions for future research and English writing instruction.

2. Methodology

This review was conducted to summarize the research objectives and the major findings of previous review studies and to provide directions for future research (Pham et al., 2014). Five essential steps are included: Forming research questions, searching studies, selecting relevant studies, coding the data, and summarizing the results.

2.1. Step 1: Forming research questions

In this review, we searched two widely used databases (Scopus and Web of Science) to identify existing review studies with an explicit focus on AI in English writing. This study was guided by the two research questions as follows:

RQ1: What are the objectives of the review studies on AI in English writing?

RQ2: What are the major findings of the review studies on AI in English writing?

2.2. Step 2: Literature search

This review aimed to include all peer-reviewed review articles published between 2020 and 2025. Specifically, we searched two databases using the following keywords: (AI* OR Artificial Intelligence*) AND (L2 writing* OR English writing*) AND (Systematic review* OR Meta-analysis*). After several rounds of screening titles, keywords, and abstracts, we identified 22 articles. Figure 1 presents the procedure of literature selection.

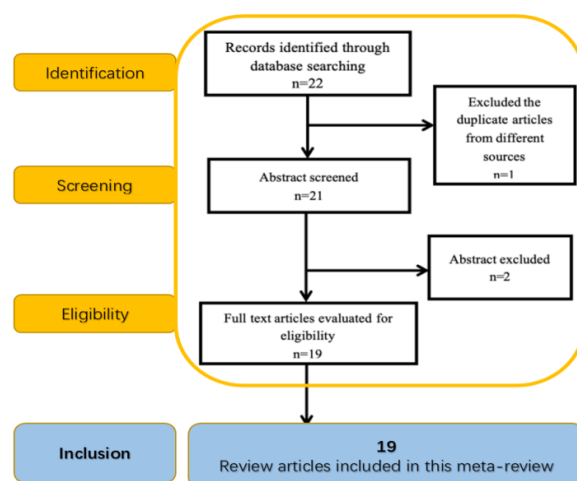


Figure1. Flowchart of Literature Selection Following the PRISMA Approach (Moher et al., 2009)

2.3. Step 3: Inclusion and exclusion criteria

To ensure the articles retrieved were appropriate to the stated objectives, the inclusion and exclusion criteria were adopted. First, the articles had to be published in peer-reviewed journals. Second, they had to be published between 2020 and 2025. Third, the articles needed to be review studies on AI in English writing. Fourth, the articles had to be published in English. Lastly, abstracts of the articles needed to be available online. Accordingly, we excluded book chapters, conference papers, and dissertations. We also excluded journal articles published before 2020, articles published in languages other than English. After applying these inclusion and exclusion criteria, a total of 19 articles remained for our further analysis. These references have also been listed in the references with asterisk (*). Collectively, these review articles examined a total of 719 studies; however, some of these studies may have been included in multiple reviews.

2.4. Step 4. Coding the selected articles

A thematic analysis was then conducted to code the research objectives and the major findings of the review studies based on the Control-Value Theory (Pekrun, 2006). The first author and the corresponding author collaboratively worked a coding scheme in light of the CVT to ensure a consistent framework for analyzing the 19 review papers using the CVT lens. They then discussed the results to resolve discrepancies and achieve consistency, ensuring the reliability and validity of the analysis. According to the theoretical model, we categories research findings into four themes. The first category, environment, refers to the AI tools used in English writing (e.g., ChatGPT). Control-value appraisals pertain to individuals' perceptions of their competence in using AI and the perceived value of AI tools (e.g., flexibility and convenience provided by ChatGPT). Emotions refer to the feelings of individuals when using AI tools (e.g., enjoyment). Learning outcome focuses on the impacts of AI on learners (e.g., improvements in English writing performance).

2.5. Step 5: Summarizing the results

We visualized the results in Figure 2.

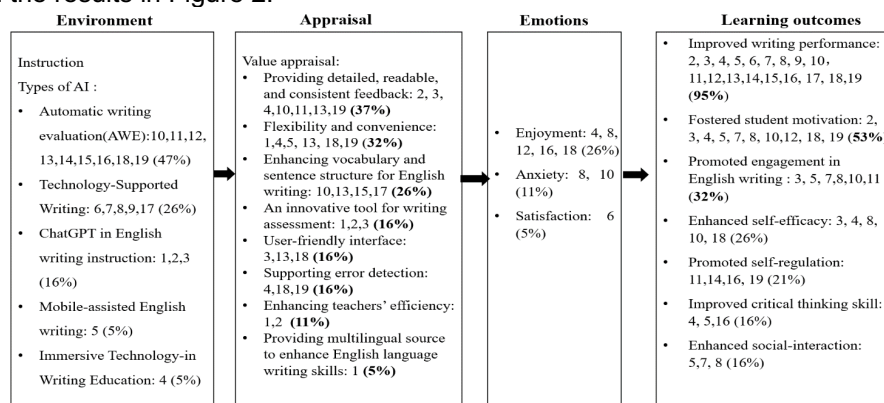


Figure 2. Mapping Key Findings from 19 Review Papers onto the Control-Value Theory (CVT) Model

3. Results

Using the CVT model to frame the key findings from 19 review papers on AI in English writing, we identified meaningful patterns that would inform educational practices in English writing education. The CVT model emphasizes the influence of perceived control, values, emotions, and their cumulative effect on learning outcomes, providing a comprehensive theoretical lens through which to analyze and interpret the integration of AI tools in writing instruction.

The results showed these studies documented in the reviewed review papers explored various AI-assisted tools, including Automatic Writing Evaluation (AWE) (9 studies), ChatGPT (3 studies), computer-assisted technology (5 studies), mobile-assisted technology (1 study), and immersive technology (1 study). The increasing emergence of innovative AI tools, such as the metaverse, suggests a need for future research to expand beyond AWE and explore other advanced AI applications in English writing.

3.1. Research Focus and Gaps

A significant majority (18 studies) concentrated on the impact of AI on English writing outcomes, while limited attention was given to learners' perceptions (2 studies, 11%) and emotional experiences (1 study, 5%) when using AI. This highlights a gap in understanding the psychological aspects of AI-assisted writing, warranting further research on how affective factors influence learning outcomes.

3.2. Impact on Learning Outcomes and Emotional Experience

Most review studies emphasized learning outcomes, with over 95% reporting improvements in writing performance through AI integration. Additionally, AI was found to enhance learners' writing motivation, self-efficacy, and engagement. However, only six studies examined learners' emotional experiences, indicating that AI in English writing often evokes enjoyment and reduces anxiety. Future research should extend beyond performance-based outcomes to explore learners' affective and cognitive engagement with AI tools, which could inform the design of more effective AI-assisted writing activities.

3.3. Learner Competence and AI-Assisted Writing

These reviews identified key benefits of AI, such as providing immediate feedback and aiding in language error correction. However, no studies in these reviews examined learners' perceived competence in using AI for writing. Given AI's positive impact on self-efficacy, future research should investigate how learners' perceived competence shapes their engagement and effectiveness in AI-assisted writing.

4. Limitations Identified in the Meta-Review and Future Directions for Research

The meta-review of AI in English writing observed at least four key limitations: First, there is a narrow focus on AI's influence on self-efficacy, with minimal attention to broader control aspects such as perceived competence and autonomy in writing tasks. Second, studies predominantly assess immediate performance improvements, overlooking deeper value appraisals like relevance and long-term benefits of AI tools in writing. Third, emotional considerations are limited to basic responses like anxiety and enjoyment, ignoring a broader emotional spectrum (Pekrun, 2006, 2024) crucial for understanding engagement in writing. Lastly, there is a lack of integration between control beliefs, value appraisals, and actual writing outcomes, which is essential for a comprehensive understanding of AI's educational impact on English writing.

Future research should address these limitations by integrating the focus on expanding AI tools and exploring their developmental impacts of AI in English writing. This should include investigating how AI affects learners at different educational levels, conducting intervention studies to enhance AI literacy, and improve writing performance. The emergence of AI tools like ChatGPT, VR, and others within the metaverse offers additional resources to extend research beyond traditional platforms to include newer, innovative technologies. This involves examining learners' perceptions and comparing the impacts of different types of AI tools on learning outcomes, which can provide deeper insights into the varied effectiveness of these technologies in English writing. In addition to expanding the types of AI tools studied, future research should address the broader spectrum of emotional, cognitive, and motivational factors influencing learning with AI. The control-value theoretical lens will be a useful guidance. This includes investigating how perceived competence and autonomy affect students' engagement and proficiency with AI-enhanced writing tasks, understanding the control aspects more deeply, and examining students' valuation of AI writing tools in terms of their relevance and long-term utility to enhance motivational impacts. Moreover, it is crucial to broaden the study of emotional responses beyond typical emotions like anxiety and enjoyment to include emotions such as frustration,

boredom, or pride, which may influence writing engagement directly (Li et al., 2023; Pekrun, 2024). Developing conceptual models that integrate control beliefs, value appraisals, and emotional responses based on previous research could better predict learning outcomes and refine AI educational tools. These models should aim not only to enhance writing performance as a product but also to foster students' positive control and value appraisals, engaging them in the English writing process and forming a positive learning experience.

5. Theoretical and Practical Implication

Theoretically, integrating AI tools in English writing education, as analyzed through Control-Value Theory (CVT), offers significant theoretical implications. CVT posits that students' motivation and achievement are influenced by their perceived control over and the value they attribute to learning tasks. AI tools like ChatGPT and VR enhance perceived control by providing immediate, personalized feedback, allowing students to adjust their learning strategies effectively. This boosts their engagement and perceived autonomy in the learning process. Additionally, by making writing tasks more interactive and relevant, these technologies increase the intrinsic value of learning activities, which CVT suggests is crucial for deepening student engagement. Also, the continuous feedback loop facilitated by AI can positively influence students' emotional responses, fostering feelings of accomplishment and improving overall learning outcomes. Practically, based on insights from this meta-review, English educators are encouraged to strategically integrate AI tools in writing instruction through three targeted practices: Firstly, by incorporating AI technologies like ChatGPT and VR into the curriculum to enhance personalized and adaptive learning experiences, thereby boosting student engagement and autonomy. Secondly, educators can harness the power of AI to provide immediate, formative feedback, while also supervising and enriching this feedback to address complex writing skills and ensure accuracy. Lastly, a *collaborative learning ecosystem* (see Tan et al., 2022, for a systematic review) should be established between educators and students, where students engage with AI tools for initial feedback, perform critical analysis, and undertake revisions, with educators guiding the iterative use of AI to foster critical and creative thinking skills. This comprehensive approach has its potential to not only streamline feedback processes but also significantly enhance students' writing proficiency and self-regulatory abilities.

6. Conclusion

This meta-review underscores the benefits and limitations of current AI applications in English writing education, guided by the CVT framework. Future research should explore expanding AI tool capabilities to support not just grammatical accuracy but also creative and critical writing skills. Additionally, understanding the full spectrum of emotional impacts from AI use could inform the development of more supportive and empathetic AI educational tools. Finally, investigating a wider range of AI technologies and teaching approaches could help achieve broader educational objectives, enhancing both immediate writing skills and long-term learner development. This balanced approach is crucial as the scope of this meta-review is inherently limited to existing studies and should be interpreted with caution regarding AI's potential effects.

References

- Asad, M. M., Shahzad, S., Shah, S. H. A., Sherwani, F., & Almusharraf, N. M. (2024). ChatGPT as artificial intelligence-based generative multimedia for English writing pedagogy: challenges and opportunities from an educator's perspective. *The International Journal of Information and Learning Technology*, 41(5), 490-506. <https://doi.org/10.1108/IJILT-02-2024-0021>
- Biju, L. G., & Vijayakumar, S. (2023). Systematic review on the impact of technology-supported writing environment for developing writing skills. *Computer-Assisted Language Learning Electronic Journal*, 24(3), 290-313.
- Chen, Y., Li, M., Huang, C., Cukurova, M., & Ma, Q. (2024). A systematic review of research on immersive technology-enhanced writing education: The current state and a research agenda. *IEEE Transactions on Learning Technologies*, 17, 919-938. <https://doi.org/10.1109/TLT.2023.3341420>
- Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, 29(11), 1451-14203. <https://doi.org/10.1007/s10639-023-12402-3>
- Dizon, G., & Gayed, J. M. (2024). A systematic review of Grammarly in L2 English writing contexts. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2397882>
- Fan, N., & Ma, Y. (2022). The effects of Automated Writing Evaluation (AWE) feedback on students' English writing quality: A Systematic Literature Review, 28, 53-73. *Language Teaching Research Quarterly*. <https://doi.org/10.32038/ltrq.2022.28.03>

- Xu, G., Looi, C. K., McMinn, S., Cheng, G., (Eds.). (2025). *Proceedings of International Conference on AI and Education 2025 (AIEdu2025)*. Hong Kong SAR: The Education University of Hong Kong.
- Fu, Q.-K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning*, 37(1-2), 179-221. <https://doi.org/10.1080/09588221.2022.2033787>
- Huang, X., Xu, W., Li, F., & Yu, Z. (2024). A meta-analysis of effects of automated writing evaluation on anxiety, motivation, and second language writing Skills. *The Asia-Pacific Education Researcher*, 33(4), 957-976. <https://doi.org/10.1007/s40299-024-00865-y>
- Ibrahim, K., & Kirkpatrick, R. (2024). Potentials and implications of ChatGPT for ESL writing instruction. *International Review of Research in Open and Distance Learning*, 25(3), 394-409. <https://doi.org/10.19173/irrodl.v25i3.7820>
- Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 123, Article 103332. <https://doi.org/10.1016/j.system.2024.103332>
- Khamraeva Shavkatovna, D., Hashim, H., & Jamaludin, K. A. (2024). Mobile learning for transforming English writing skills: A systematic review (2013–2023). *International Journal of Interactive Mobile Technologies*, 18(19), 4-17. <https://doi.org/10.3991/ijim.v18i19.47729>
- Mohsen, M. A. (2022). Computer-Mediated Corrective Feedback to Improve L2 Writing Skills: A Meta-Analysis. *Journal of Educational Computing Research*, 60(5), 1253-1276. <https://doi.org/10.1177/07356331211064066>
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: a three-level meta-analysis. *Interactive Learning Environments*, 32(2), 727-744. <https://doi.org/10.1080/10494820.2022.2096642>
- Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, 38(2), 599-620. <https://doi.org/10.1111/jcal.12635>
- Seyyedrezaei, M. S., Amiryousefi, M., Gimeno-Sanz, A., & Tavakoli, M. (2024). A meta-analysis of the relative effectiveness of technology-enhanced language learning on ESL/EFL writing performance: retrospect and prospect. *Computer Assisted Language Learning*, 37(7), 1771-1805. <https://doi.org/10.1080/09588221.2022.2118782>
- Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: a review study. *Interactive Learning Environment*, 32(6), 2538-2566. <https://doi.org/10.1080/10494820.2022.2153145>
- Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36-57. <https://doi.org/10.58304/ijts.20240304>
- Xu, Z., Banerjee, M., Ramirez, G., Zhu, G., & Wijekumar, K. (2019). The effectiveness of educational technology applications on adult English language learners' writing quality: a meta-analysis. *Computer Assisted Language Learning*, 32(1-2), 132-162. <https://doi.org/10.1080/09588221.2018.1501069>
- Zhang, R., & Zou, D. (2023). A review of research on technology-enhanced peer feedback for second language writing based on the activity theory framework. *Education and Information Technologies*, 28(6), 6727-6753. <https://doi.org/10.1007/s10639-022-11469-8>

The EdUHK x HKUST Joint International Conference on
AI and EDUCATION



<https://aiedu2025.eduhk.hk/>

20 - 21 August 2025