Luo *et al.*
International Journal of Educational Technology in Higher Education (2025) 22:42
https://doi.org/10.1186/s41239-025-00540-2

International Journal of Educational Technology in Higher Education

REVIEW ARTICLE

Open Access

Design and assessment of AI-based learning tools in higher education: a systematic review

Jihao Luo[1], Chenxu Zheng[2], Jiamin Yin[3*] and Hock Hai Teo[1]

*Correspondence:
Jiamin Yin
yinjiamin@rmbs.ruc.edu.cn
[1]Department of Information Systems and Analytics, School of Computing, National University of Singapore, Singapore, Singapore
[2]Discipline of Business Information Systems, The University of Sydney Business School, The University of Sydney, Sydney, Australia
[3]Department of Management Science and Engineering, Renmin Business School, Renmin University of China, Beijing, P.R. China

Abstract
Artificial intelligence (AI)-based learning tools are increasingly integrated in higher education, offering benefits such as personalized learning experiences, real-time feedback, and increased flexibility. However, effective design and implementation strategies for these tools are not well established. This study addresses this gap through a systematic literature review with two main objectives: (1) to summarize the design features of AI-based learning tools currently employed in higher education, focusing on aspects such as algorithm types, training datasets, modes of information presentation, and their roles in the learning process; and (2) to assess their impacts on college students' cognitive, skill-based, and affective learning outcomes. Our review encompasses 63 peer-reviewed articles published between January 2014 and April 2024. Notably, approximately half of the reviewed studies employ publicly available AI systems for instructional purposes ($n = 32$), while the other half develop proprietary AI-based learning tools ($n = 31$). 26 studies use AI techniques to generate and deliver multimodal learning materials. Moreover, we identify three primary roles of AI in higher education: assessment and evaluation ($n = 45$), personalized feedback and recommendations ($n = 46$), and intelligent tutoring ($n = 26$). In terms of learning outcomes, although AI tools generally improve cognitive knowledge acquisition and affective outcomes, their effectiveness in developing cognitive process and skills varies significantly. Lastly, we provide recommendations for optimizing the design and implementation of AI-based learning tools in higher education and outline promising directions for future research in this field.

Keywords Artificial intelligence, System implementation, Higher education, Systematic literature review, PRISMA

Introduction

Recent advancements in artificial intelligence (AI) have profoundly reshaped higher education, enhancing and sometimes surpassing human capabilities in complex instructional tasks (Bates et al., 2020; Labadze et al., 2023). Innovations in natural language processing (NLP), machine learning, and generative pre-trained transformers (GPT) have facilitated the development of sophisticated AI-based learning tools that provide personalized instructional content and feedback, effectively addressing the diverse needs of college students (Chen et al., 2023). The AI market in higher education generated approximately USD 2 billion in revenue in 2022, with projections indicating a robust compound annual growth rate of 36.0% from 2022 to 2030 (Grand View Research, 2022).

Despite the rapid evolution of AI-based learning tools, challenges remain in their design and assessment. In particular, these tools often present information in ways that learners struggle to interpret properly. Unlike traditional computer-assisted instructional tools that follow predefined decision rules, many AI tools operate as "black boxes," obscuring their decision-making processes (Bates et al., 2020; Glikson & Woolley, 2020; Prinsloo, 2020). This opacity can impede understanding of essential knowledge concepts, particularly in disciplines that require complex reasoning, such as mathematics, physics, and medicine (Almarode & Vandas, 2019). Consequently, it is vital to understand how these tools generate, process, and present learning materials. However, existing studies in this field have largely overlooked these critical design aspects of AI-based learning tools (Bond et al., 2024; Crompton & Burke, 2023; Ouyang et al., 2022; Zawacki-Richter et al., 2019).

Moreover, the use of AI-based learning tools may lead to unintended consequences for college students, who face the dual challenges of mastering disciplinary knowledge while preparing for professional practice (Dall'Alba, 1994). This requires active engagement in their studies, the development of cognitive processes and skills, and the cultivation of positive attitudes and motivation toward learning. Yet, excessive dependence on AI tools can result in procrastination, memory erosion, and declined academic performance, ultimately stunting personal growth (Abbas et al., 2024). Current reviews of AI in higher education often provide a superficial overview of the affected learning outcomes (Bond et al., 2024; Chu et al., 2022; Ouyang et al., 2022), leaving a significant gap in our understanding of AI's broader impact across various disciplines. Therefore, a comprehensive assessment of AI-based learning tools is imperative for developing strategies that promote their responsible and effective use in higher education.

To address these gaps, this study aims to systematically review empirical research on the design and assessment of AI-based learning tools in higher education. To our knowledge, this is the first large-scale synthesis addressing the following two research questions:

RQ 1 How are AI-based learning tools designed to support college students' learning?

RQ 2 How do AI-based learning tools impact college students' learning outcomes?

To answer these questions, we first provide an overview of state-of-the-art AI-based learning tools, examining their algorithms, training datasets, information presentation modes, and roles in learning. We then apply Kraiger et al.'s framework (Kraiger et al., 1993; Wan et al., 2012) to assess how these tools influence three key dimensions of learning outcomes: cognitive, skill-based, and affective. Cognitive outcomes involve the acquisition of disciplinary knowledge and the development of cognitive processes. Skill-based outcomes focus on the progression from skill acquisition to proficiency, while affective outcomes reflect learners' attitudes, values, and motivation.

The remainder of the paper is organized as follows. We begin with the theoretical background, followed by the methodology for our systematic literature review. Next, we present the results from our analysis of 63 peer-reviewed journal articles, synthesize these findings, and propose a framework for future research. Finally, we conclude the paper.

Theoretical background

The definition of artificial intelligence (AI) has evolved alongside technological advancements. Initially coined by McCarthy in 1956 (McCarthy et al., 1956), AI was described as “the science and engineering of making intelligent machines.” Minsky (1968) later expanded this definition by emphasizing AI’s capacity to solve problems typically associated with human intelligence. Today, AI is characterized as a field focused on developing “rational agents” (Berente et al., 2021) that act adaptively to their environments (Newell & Simon, 1976; Russell & Norvig, 2010) and perform cognitive tasks such as problem-solving, decision-making, and learning (Baker & Smith, 2019). It encompasses various algorithms, including machine learning, deep learning, neural networks, and NLP. Recent breakthroughs in large language models (LLMs) have further expanded AI’s ability in comprehending human inquiries and processing text, images, and audio (Zhao et al., 2024).

The application of AI in higher education is promising due to the knowledge-intensive and cognitively demanding nature of this sector (Dall’Alba, 1994). AI has the vast potential to help college students develop critical competencies essential for professional success, attracting considerable scholarly interest. Several reviews have examined AI’s role in higher education. Zawacki-Richter et al. (2019) systematically reviewed 146 studies from 2007 to 2018 and delineated four core functionalities of AI: profiling and prediction, assessment and evaluation, adaptive systems and personalization, and intelligent tutoring systems. Ouyang et al. (2022) synthesized findings from 32 studies on AI in online higher education between 2011 and 2020, emphasizing AI’s capabilities in recommending educational resources, predicting learning outcomes, automating assessments, and enhancing learning experiences. In addition, Crompton and Burke (2023) identified five key AI applications from 138 studies conducted between 2016 and 2022: assessment, prediction, AI assistants, intelligent tutoring, and learning management. Bond et al. (2024) conducted a meta-analysis of 66 reviews published between 2018 and 2023 and found that AI is predominantly utilized for adaptive systems and personalization in higher education.

However, existing reviews lack detailed descriptions of the design features of advanced AI-based learning tools, limiting our understanding of how to effectively develop these tools for college students. Furthermore, these reviews offer a narrow analysis of AI’s impact on learning outcomes. For example, Ouyang et al. (2022) mainly focused on the effects of AI on academic performance and engagement, while Chu et al. (2022) merely cataloged the frequency of learning outcomes affected by AI. These limitations stress the need for a more comprehensive understanding of the effects of AI on learning outcomes in higher education. To address these gaps, we conducted an extensive review of the literature on AI-based learning tools in higher education, focusing on their design features and potential impacts on cognitive, skill-based, and affective learning outcomes.

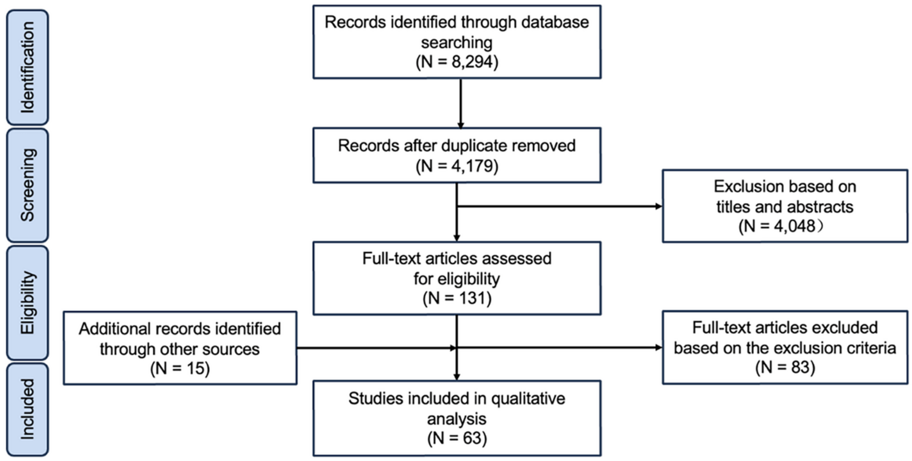

Fig. 1 Flow diagram of the systematic literature search based on the PRISMA statement

Table 1 Search strategy

Category	Included Key Terms
AI	("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning")
Higher education	("undergraduate*" OR "graduate*" OR "university student*" OR "college student*")
Learning outcomes	("learning outcome*" OR "learning performance" OR "achievement*" OR "assessment*" OR "evaluation" OR "gain*" OR "improv*" OR "enhanc*")

Methods

This study employs a systematic literature review, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Keele, 2007; Moher et al., 2009; Page et al., 2021). The established framework provides a structured, step-by-step approach for conducting systematic reviews, encompassing the development of search strategies, the application of inclusion and exclusion criteria, and the documentation of the screening process through a flow diagram (see Fig. 1). Adhering to these guidelines ensures transparency, minimizes selection bias, and promotes comprehensive coverage of the relevant literature.

Literature search

The literature search was conducted in April 2024 across ten electronic databases to capture a broad interdisciplinary scope, including general multidisciplinary databases (i.e., Web of Science, Scopus), alongside discipline-specific databases in education and psychology (i.e., ERIC, PsycInfo), business and management (i.e., Business Source Ultimate, ABI/INFORM), science and engineering (i.e., Engineering Village), and healthcare (i.e., PubMed, CINAHL, MedLine). By integrating both multidisciplinary and domain-specific databases, we ensured a comprehensive representation of diverse scholarly perspectives across various fields.

We restricted our search to English-language, peer-reviewed journal articles published between January 2014 and April 2024. We identified the relevant publications using three groups of keywords related to AI (Díaz & Nussbaum, 2024; Yin et al., 2021), higher education (Baig & Yadegaridehkordi, 2023; Crompton & Burke, 2023), and learning outcomes (Wei et al., 2021), respectively (see Table 1). Boolean operators were employed to combine these keywords, and the search was conducted by title, abstract, and keywords. The initial search yielded 8,294 journal articles.

Inclusion and exclusion criteria

Studies were included in the review if they: (1) involved a specific AI-based learning tool used as the primary instructional tool; (2) empirically evaluated the impact of the tool on learning outcomes; (3) included college students (undergraduates or postgraduates) as participants; and (4) were published in English language, peer-reviewed journals between January 2014 and April 2024.

The exclusion criteria were as follows: (1) studies utilizing non-AI technologies for instruction, such as virtual reality or robotics; (2) studies that focused solely on the technical development of AI tools without actual implementation in higher education; (3) studies examining only user perceptions of AI tools; (4) studies not involving college students as participants; (5) studies published in languages other than English; and (6) studies published in non-peer-reviewed sources, including books, book chapters, dissertations, commentaries, keynote presentations, panel discussions, work-in-progress articles, and perspective papers.

Selection of articles

Figure 1 illustrates the article selection process based on PRISMA guidelines. The initial search yielded 8,294 journal articles, from which we removed 4,115 duplicates. Following this, two student coders (ZY and YT) were recruited to independently screen the titles and abstracts of the remaining 4,179 studies for eligibility. Discrepancies were included in the full-text review for further assessment (Noetel et al., 2021). Subsequently, two additional student coders (CM and XQ) independently evaluated the full-text articles against the inclusion and exclusion criteria, documenting reasons for exclusions. The inter-rater reliability between the two coders was assessed using Cohen's kappa (κ) (Cohen, 1960), with a score of 0.81 indicating almost perfect agreement (Landis & Koch, 1977). At this stage, discrepancies were resolved through discussion or consultation with a third coder (JH). This process resulted in 131 articles for full-text review, among which 48 met our inclusion criteria. Additionally, one coder (JH) conducted forward and backward citation searches to identify fifteen additional relevant studies (Müller & Wulf, 2020; Pigott & Polanin, 2020). Ultimately, a total of 63 articles were included in the systematic review.

Data extraction and analysis

Following article selection, we developed a predefined data extraction form to systematically gather information from the included articles (see Fig. 2). This form captured three key areas: (1) study characteristics, including identifiers, methodology, and participant characteristics; (2) design elements of AI-based learning tools, encompassing AI features, modes of AI-enabled information presentation, and the roles of AI in the learning process; and (3) assessment of these tools, with a focus on learning outcomes and key findings.

Study characteristics

- Study identifiers (i.e., authors, year of publication, journal title, article title, and country of study)
- Methodology (i.e., study design, comparison groups, instructional context, study discipline, and study duration)
- Participant characteristics (i.e., gender, stage of education, and participant discipline)

Design elements of AI-based learning tools

- AI features (i.e., name of the AI systems/software, source of training data, and AI algorithms)
- Modes of AI-enabled information presentation
- Roles of AI in the learning process (i.e., assessment and evaluation, personalized feedback and recommendation, and intelligent tutoring)

Assessment of AI-based learning outcomes

- Learning outcomes (i.e., cognitive, skill-based, and affective)
- Key findings

Fig. 2 Components of the data-charting form

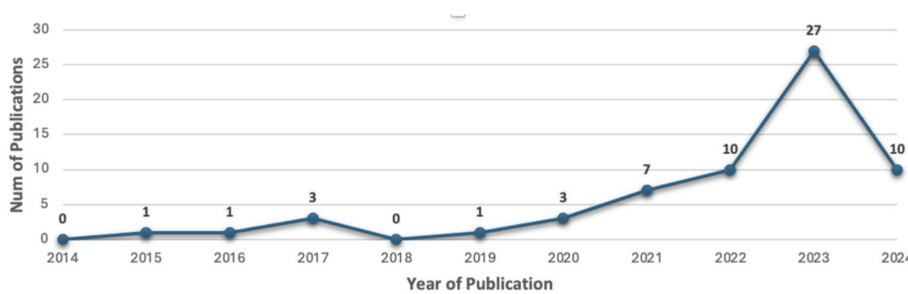

Fig. 3 Distribution of the included articles from January 2014 and April 2024

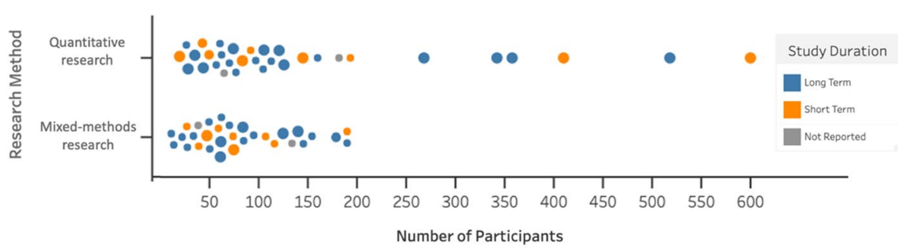

Fig. 4 Overview of methodology and participant characteristics

Results

In this section, we first present an overview of the characteristics of the included studies, followed by an analysis addressing our two research questions.

Overview of study characteristics

Table A.1 summarizes the study characteristics, including identifiers, methodologies, and participant demographics. Figure 3 displays the distribution of publication years for the included articles, revealing a notable increase in published studies since 2014, particularly after 2021. This upward trend reflects a growing interest in AI-based learning tools in higher education.

We further analyzed the methodologies employed in the reviewed studies, as detailed in Fig. 4. The dot size represents the strength of evidence (U.S. Department of Education, 2023), with randomized experimental designs categorized as strong and quasi-experimental designs as moderate. Studies are color-coded by duration: long-term (blue), short-term (orange), and not reported (grey). Of the 63 reviewed studies, 32 (50.8%) utilized quantitative research methods, while 31 (49.2%) employed mixed-methods. Among these, 22 studies (34.9%) implemented fully randomized designs, including fourteen incorporating a pretest-posttest design. One study adopted a randomized crossover trial with a pretest-posttest design. The remaining 40 studies (63.5%) were quasi-experimental, with 36 incorporating a pretest-posttest design. Moreover, most studies (59/63, 93.7%) reported their research duration: eighteen studies (30.5%) were short-term (typically lasting no more than five hours), while 41 (69.5%) were long-term, spanning from two weeks (Saravia-Rojas et al., 2024) to four years (An & Wang, 2023).

Regarding participant characteristics, 26 studies (41.2%) involved more than 100 participants, while seventeen studies (27.0%) had fewer than 50 participants. A significant majority (54/63, 85.7%) reported the participants' academic programs: 46 studies (85.2%) focused exclusively on undergraduates, four (6.3%) targeted graduates, and four (6.3%) included both groups.

We also categorized the instructional contexts of the reviewed studies. Language learning was the most prominent field of study ($n = 16$), followed closely by medical education ($n = 15$) and general education topics such as team collaboration and research ethics ($n = 15$). Other fields included computer science ($n = 9$), arts ($n = 2$), psychology ($n = 1$), business ($n = 1$), biology ($n = 1$), chemistry ($n = 1$), mathematics ($n = 1$), and technology education ($n = 1$). Over two-thirds ($n = 46$) of the studies were conducted within specific higher education courses: twenty studies (43.5%) covered entire courses, eleven (23.9%) focused on one or a few specific classes, and fifteen (32.6%) were structured as training sessions. The remaining seventeen studies (27.0%) were standalone experiments conducted outside of higher education courses.

RQ 1: how are AI-based learning tools designed to support college students' learning?

This section addresses our first research question regarding the design elements of AI-based learning tools in higher education. We focus on three key design dimensions: (1) Technological dimension, examining the development of these tools, including the algorithms and training datasets used; (2) Content presentation dimension, which involves how AI enables information presentation within the tools; and (3) Pedagogical dimension, exploring the roles these tools play in facilitating learning (see Table A.2 for a detailed overview).

AI features

Our analysis of AI features revealed that approximately half of the reviewed studies (32/63, 50.8%) employed publicly available AI systems for instructional purposes (see Table 2). Within this group, seventeen studies (53.1%) leveraged general-purpose AI systems such as ChatGPT (e.g., Chang et al., 2024; Essel et al., 2024; Urban et al., 2024), iFlyRec (Jiang et al., 2022, 2023), and Replika (Lin & Mubarok, 2021). The remaining fifteen studies focused on AI tools tailored for specific instructional activities, including nine for medical training (e.g., An & Wang, 2023; Liu et al., 2016; Simsek-Cetinkaya & Cakir, 2023), five for language learning (e.g., Dizon & Gayed, 2021; El Shazly, 2021; Rokhayani et al., 2022), and one for music learning (Lv, 2023).

Table 2 Overview of AI features

Category	Example Articles
Publicly available AI-based learning tools	
General-purpose AI tools (n = 17)	e.g., Escalante et al. (2023); Hakiki et al. (2023); Vicente-Yagüe-Jara et al. (2023)
Tailored AI systems for specific instructional activities (n = 15)	**Medical training (n = 9)**: e.g., Chang et al. (2022); Lee et al. (2022); Liu et al. (2016); Simsek-Cetinkaya and Cakir (2023)
	Language learning (n = 5): e.g., Dizon and Gayed (2021); El Shazly (2021); Rokhayani et al. (2022)
	Music learning (n = 1): Lv (2023)
Self-developed AI-based learning tools	
Developed from scratch (n = 23)	e.g., Howard et al. (2017); Sun et al. (2023); Tegos and Demetriadis (2017); Zheng et al. (2023e, 2023f)
Developed using low-code and no-code platforms (n = 8)	e.g., Kaiss et al. (2023); Liaw et al. (2023); Wolfe et al. (2015)

In contrast, 31 studies developed proprietary AI-based learning tools for targeted instructional objectives (e.g., Conati et al., 2021; Howard et al., 2017; Koć-Januchta et al., 2020). Among these, 23 studies (36.5%) built their tools from scratch, with twenty providing detailed descriptions of the AI techniques used. This statistic is unsurprising, as educators are often overwhelmed by heavy teaching workloads and lack the time and resources needed to develop AI tools. The most commonly used algorithm was NLP ($n = 15$), including techniques such as Bidirectional encoder representations from transformers (BERT), the Natural Language Toolkit, and latent semantic analysis. Other frequently used machine learning models comprised neural networks ($n = 8$), random forests ($n = 2$), Bayesian methods ($n = 2$) and K-nearest neighbors ($n = 2$). Additionally, eight studies developed conversational agents using low-code and no-code platforms, such as Google Dialogflow ($n = 4$), ManyChat ($n = 1$), FlowXO ($n = 1$), Landbot ($n = 1$), and AutoTutor ($n = 1$), enabling instructors with minimal coding expertise to develop their own chatbots.

Lastly, we examined the training datasets used in the self-developed AI tools. Eighteen studies collected historical student learning data from the same or related courses (e.g., Afzaal et al., 2024; Conati et al., 2021; Zheng et al., 2021), while eleven sourced online learning materials (e.g., Chiu et al., 2022; Essel et al., 2022; Howard et al., 2017). Notably, two studies did not disclose information regarding their training data (Koć-Januchta et al., 2020; Sun et al., 2023).

Modes of AI-enabled information presentation

A total of 37 studies utilized AI techniques to generate and deliver learning materials in a single modality. Of these, 36 focused on verbal information (e.g., Chang et al., 2022, 2024; Darban, 2023; Dizon & Gayed, 2021), while one study presented learning content through graphs and charts (Sun et al., 2023). The remaining 26 studies employed AI tools to provide multimodal learning materials (e.g., Afzaal et al., 2024; Liaw et al., 2023; Wolfe et al., 2015). For example, Lin and Mubarok (2021) leveraged the AI-powered chatbot Replika to enhance college students' English speaking skills through text and voice interactions. Similarly, An and Wang (2023) implemented an AI-based diagnosis and recognition system in gastroscopy training, which offered real-time verbal cues for missed detections, visualized high-risk lesions on gastroscopy images, and highlighted unexamined areas using a stomach anatomical diagram.

Roles of AI in higher education

Our review identified three primary roles of AI in higher education: assessment and evaluation, personalized feedback and recommendations, and intelligent tutoring, with some AI tools fulfilling multiple roles simultaneously.

A core function of AI-based learning tools is to facilitate assessment and evaluation by gauging student understanding and engagement, automating grading, and delivering feedback. Among the studies reviewed, 45 (71.4%) utilized AI tools for these purposes (e.g., Cheng, 2017; Dai & Wu, 2023; Mahapatra, 2024). For example, Sun et al. (2023) employed an AI-enabled radar chart visualization tool to display performance scores for graduate students in an online research ethics course. Furthermore, Zheng et al. (2022a) used BERT to analyze discussion scripts from online collaborative learning sessions, evaluate learner engagement, and provide targeted feedback to improve collaborative knowledge development.

We also identified 46 studies that examined AI-based learning tools capable of providing personalized feedback and recommendations (e.g., Darban, 2023; Kavadella et al., 2024; Rokhayani et al., 2022). These tools track learning progress and offer real-time, tailored guidance. For example, Koć-Januchta et al. (2020) augmented biology education with an AI-enriched textbook, allowing undergraduates to ask questions through a dialog box and retrieve related questions and answers. This tool also recommended exploration questions to deepen students' understanding of essential biology concepts. In another study, Afzaal et al. (2024) introduced a dashboard that integrated adaptive feedback and counterfactual explanations in an undergraduate programming course, enabling students to track their learning progress toward personalized learning objectives.

Finally, 26 studies leveraged advanced AI tools, such as chatbots, to curate and deliver learning content (e.g., Chang et al., 2022, 2024; Urban et al., 2024). These chatbots enabled learners to assess their understanding while delivering personalized learning materials tailored to their specific needs. For example, Han et al. (2022) utilized Landbot to develop a chatbot that organized learning materials based on learners' mastery of prerequisite knowledge. Likewise, Zhang et al. (2023) introduced a chatbot designed to improve argumentative writing skills, which engaged learners with multimedia content, posed questions about logical fallacies, and provided tailored responses and suggestions based on their answers.

RQ 2: how do AI-based learning tools impact college students' learning outcomes?

Building on Kraiger et al.'s (1993) framework, our second research objective is to evaluate how AI-based learning tools influence cognitive, skill-based, and affective learning outcomes (see Table A.3 for details). Figure 5 summarizes the learning outcomes assessed, with dot size indicating assessment frequency for each outcome. Figure 6 illustrates the relationships among AI tool development, their roles in the learning process, and the associated learning outcomes.

Cognitive learning outcomes

Cognitive learning outcomes encompass not only the acquisition of disciplinary knowledge but also the development of cognitive processes essential for knowledge acquisition

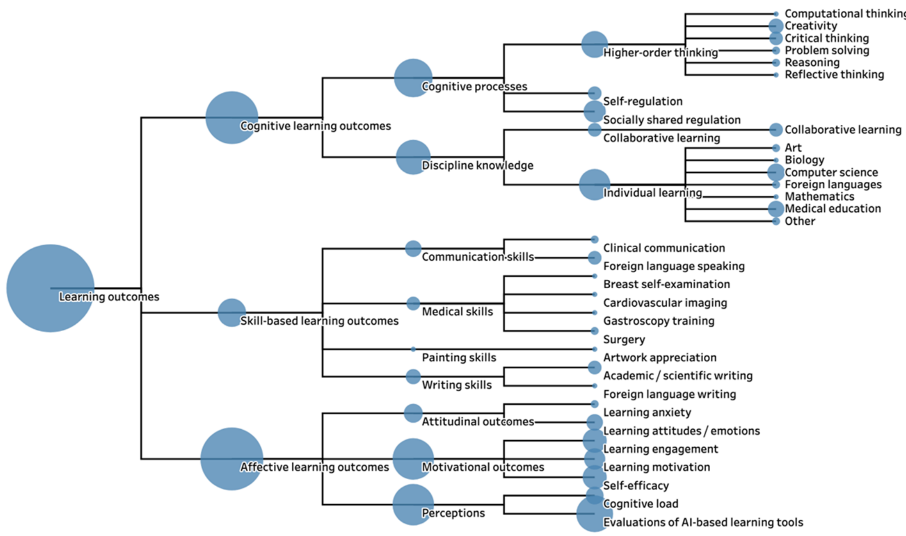

Fig. 5 Overview of learning outcomes measured in the reviewed studies

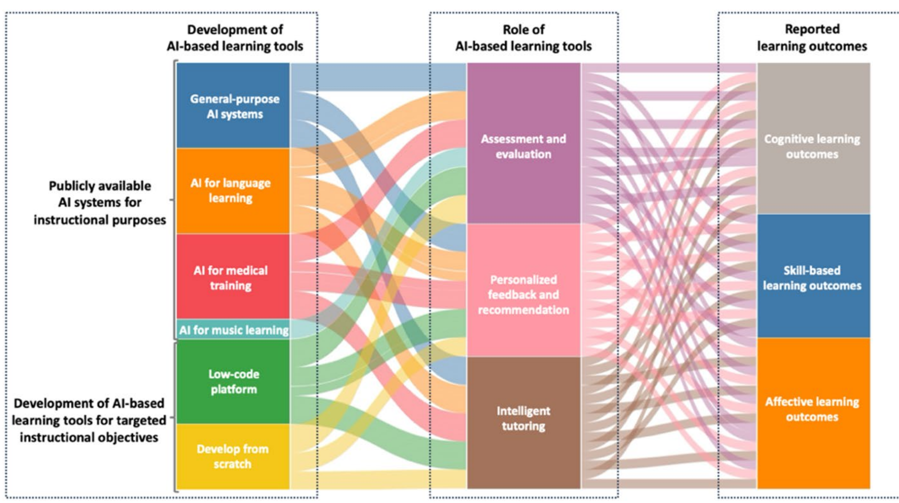

Fig. 6 Mapping the development, roles, and learning outcomes of AI-based learning tools

(Kraiger et al., 1993; Wan et al., 2012). A summary of these outcomes is presented in Table 3.

Disciplinary knowledge

Among the 28 studies examining disciplinary knowledge acquisition, 23 focused on individual learning across various fields, including computer science (Kaiss et al., 2023), medical education (Kavadella et al., 2024), foreign languages (Rokhayani et al., 2022), art (Chiu et al., 2022), mathematics (Hasan & Khan, 2023), and biology (Koć-Januchta et al., 2020). Of these, 21 studies found that AI-based learning tools produced knowledge outcomes comparable to or exceeding those of conventional tutoring methods (e.g., Chiu et al., 2022; Darban, 2023; Wolfe et al., 2015). For example, Hasan and Khan (2023) showed that AI-driven tailored messaging significantly improved students' pass rates in a 12-week discrete structures course. Conversely, two studies reported limited benefits from AI tools (Han et al., 2022; Huang et al., 2023). To illustrate, Huang et al. (2023)

Table 3 Cognitive learning outcomes

Category	Theme		Example Articles
Discipline knowledge	Individual learning	Computer science ($n=8$)	Afzaal et al. (2024); Conati et al. (2021)
		Medical education ($n=7$)	Chang et al. (2022); Han et al. (2022)
		Foreign languages ($n=2$)	Rokhayani et al. (2022); Wei (2023)
		Art ($n=2$)	Chiu et al. (2022); Lv (2023)
		Mathematics ($n=1$)	Hasan and Khan (2023)
		Biology ($n=1$)	Koć-Januchta et al. (2020)
	Collaborative learning ($n=5$)		Zheng et al., (2021, 2023d)
Cognitive processes	Higher-order thinking	Creativity ($n=6$)	Niloy et al. (2024); Vicente-Yagüe-Jara et al. (2023)
		Critical thinking ($n=5$)	Chang et al. (2024); Essel et al. (2024)
		Problem solving ($n=2$)	Chang et al. (2024); Yilmaz and Karaoglan Yilmaz (2023)
		Reasoning ($n=2$)	Han et al. (2022); Tegos and Demetriadis (2017)
		Computational thinking ($n=1$)	Yilmaz and Karaoglan Yilmaz (2023)
		Reflective thinking ($n=1$)	Essel et al. (2024)
	Self-regulation ($n=5$)		Rokhayani et al. (2022); Sun et al. (2023)
	Socially shared regulation ($n=12$)		Tegos and Demetriadis (2017); Zheng et al. (2021)

found that AI-enabled personalized video recommendations only benefited moderately motivated learners, as highly motivated learners had already mastered the content while less motivated ones were reluctant to engage with the system.

On the other hand, five studies demonstrated that AI-based learning tools effectively facilitated knowledge acquisition in collaborative learning environments (e.g., Tegos & Demetriadis, 2017; Zheng et al., 2024a). For example, Zheng et al. (2024a) revealed that an AI-based learning system integrating adaptive feedback and personalized recommendations consistently outperformed the system relying solely on adaptive feedback.

Cognitive processes

28 studies investigated how AI-based learning tools shape cognitive processes, specifically higher-order thinking ($n=13$), self-regulation ($n=5$), and socially shared regulation ($n=12$).

Higher-order thinking refers to cognitive abilities related to information analysis, evaluation, and synthesis (Hopson et al., 2001; Schoenfeld, 1999), encompassing creativity (Niloy et al., 2024), critical thinking (Essel et al., 2024), reasoning (Han et al., 2022), and computational thinking (Yilmaz & Karaoglan Yilmaz, 2023). Findings on the development of higher-order thinking were mixed. Specifically, Urban et al. (2024) observed that college students using ChatGPT produced more creative solutions to ill-defined problems than their peers working independently. However, another study reported that students assisted by chatbots generated stronger arguments but exhibited no improvement in the complexity of their writing structures compared to those aided by human tutors (Guo et al., 2023). Additional research has raised concerns about declined deep thinking (Essel et al., 2022), diminished critical thinking (Guo & Lee, 2023), and the imposition of rigid writing patterns (Mahapatra, 2024).

Self-regulation pertains to the conscious management of thoughts, behaviors, and feelings to achieve personal goals (Jin et al., 2023; Weiner et al., 2012). Five studies consistently reported positive impacts of AI tools on self-regulation (Afzaal et al., 2024; Han et al., 2022; Rokhayani et al., 2022; Sun et al., 2023; Wei, 2023). For example, Afzaal et al. (2024) demonstrated that an AI-powered dashboard integrating feedback and counterfactual explanations boosted goal-setting, planning, progress monitoring, and self-evaluation abilities among programming undergraduates.

Socially shared regulation involves peers collectively managing learning activities (Hadwin et al., 2011). Twelve studies in our review documented AI's potential to cultivate socially shared regulation (e.g., Tegos & Demetriadis, 2017; Zheng et al., 2022a, 2022b, 2023b, 2023c). Tegos and Demetriadis (2017), for example, developed MentorChat, an AI-driven dialogue system that provided real-time guidance and visualized students' reasoning processes, significantly enhancing socially shared regulation among undergraduates engaged in collaborative web interface evaluations.

Skill-based learning outcomes

Skill-based learning outcomes include the progression from initial skill acquisition to effortless automaticity (Kraiger et al., 1993). As shown in Table 4, seven reviewed studies examined the promise and challenges of AI in enhancing communication skills, yielding mixed evidence. For example, Liu et al. (2016) demonstrated the effectiveness of EQClinic in enhancing clinical communication skills, whereas Shorey et al. (2023) found that an AI-based virtual counseling application negatively affected nursing students' communication performance compared to a previous cohort that had not used the application.

Six studies explored the role of AI in developing writing skills, with five reporting positive outcomes (Cheng, 2017; Dizon & Gayed, 2021; Escalante et al., 2023; Mahapatra, 2024; Zhang et al., 2023). Escalante et al. (2023) found that students who received feedback from ChatGPT improved their academic writing quality and efficiency comparably to those guided by human tutors. However, not all findings were favorable; for example, Saravia-Rojas et al. (2024) noted that ChatGPT hindered dental students' academic writing performance compared to traditional bibliographic search methods.

Another six studies addressed the development of medical skills, again with mixed results (e.g., An & Wang, 2023; Yang & Shulruf, 2019; Zhao et al., 2020). Fazlollahi et al. (2022), for example, found that a virtual operative assistant (VOA) improved surgical performance and skill transfer compared to remote human guidance. However, a subsequent study by Fazlollahi et al. (2023) showed that while VOA-based training improved procedural safety, it also decreased movement and efficiency metrics.

Lastly, one study investigated AI's potential in artistic skill development (Chiu et al., 2022). In this study, students equipped with a deep learning-based art learning system

Table 4 Skill-based learning outcomes

Category	Theme	Example Articles
Communication skills	Foreign language speaking ($n=5$)	El Shazly (2021); Jiang et al. (2022); Lin and Mubarok (2021)
	Clinical communication ($n=2$)	Liu et al. (2016); Shorey et al. (2023)
Writing skills	Academic writing ($n=5$)	Cheng (2017); Nazari et al. (2021); Zhang et al. (2023)
	Foreign language writing ($n=1$)	Dizon and Gayed (2021)
Medical skills	Surgery ($n=3$)	Fazlollahi et al. (2022), (2023); Yang and Shulruf (2019)
	Breast self-examination ($n=1$)	Simsek-Cetinkaya and Cakir (2023)
	Cardiovascular imaging ($n=1$)	Zhao et al. (2020)
	Gastroscopy training ($n=1$)	An and Wang (2023)
Painting skills	Artwork appreciation ($n=1$)	Chiu et al. (2022)

outperformed their traditionally instructed peers in all painting dimensions, including pattern, space, color, and theme.

Affective learning outcomes

50 studies assessed affective learning outcomes, which can be classified into three categories: attitudinal ($n=9$), motivational ($n=29$), and user perceptions of AI-based learning tools ($n=37$) (see Table 5 for details).

Attitudinal outcomes pertain to emotional learning, reflecting learners' feelings, interpersonal relationships, and situational management (Kraiger et al., 1993). Among the nine studies in this category, seven reported favorable findings (e.g., Chiu et al., 2022; Nazari et al., 2021; Wolfe et al., 2015). For example, Chen and Yu (2020) developed a deep neural network model that paired students with suitable entrepreneurial projects to enhance their entrepreneurial attitudes. On the flip side, two studies warned that AI tools could exacerbate learning anxiety (El Shazly, 2021; Simsek-Cetinkaya & Cakir, 2023).

Motivational outcomes are closely tied to learners' drive to achieve learning goals (McCombs, 1984; Ross et al., 2018), a critical intrinsic factor driving effective learning (Lei et al., 2024). Our review encompassed 29 studies examining AI's impact on two core motivational outcomes: learning motivation and engagement ($n=22$), and self-efficacy ($n=14$). Eleven of these studies documented enhanced learning motivation through AI-based learning tools (e.g., Darban, 2023; Huang et al., 2023; Yilmaz & Karaoglan Yilmaz, 2023), and fourteen studies noted improved engagement (e.g., Hasan & Khan, 2023; Lin & Mubarok, 2021; Lv, 2023).

Self-efficacy, defined as students' beliefs in their ability to attain learning goals (Kolil et al., 2020; Pajares, 1996), was assessed in fourteen studies, with thirteen yielding positive outcomes (e.g., Essel et al., 2022, 2024; Liaw et al., 2023). Notably, Essel et al. (2024) observed that undergraduates using ChatGPT exhibited increased self-efficacy than those using traditional databases and search engines. Yet, one study expressed concern that 60% of students feared chatbot usage might undermine their writing self-efficacy (Zhang et al., 2023).

Finally, 31 studies investigated user perceptions of AI-based learning tools. Eight studies specifically addressed cognitive load, with six indicating that comparable cognitive demands between AI-supported and traditional instruction (e.g., Fazlollahi et al., 2022; Koć-Januchta et al., 2020) and two reporting reduced cognitive load with AI interventions (Urban et al., 2024; Zheng et al., 2024b). Moreover, 23 studies highlighted student satisfaction with AI, praising features such as instant feedback and flexible learning

Table 5 Affective learning outcomes

Category	Theme	Example Articles
Attitudinal outcomes	Learning attitudes and emotions ($n=7$)	Chiu et al. (2022); Nazari et al. (2021)
	Learning anxiety ($n=2$)	El Shazly (2021); Simsek-Cetinkaya and Cakir (2023)
Motivational outcomes	Learning motivation ($n=11$)	Wolfe et al. (2015); Yilmaz and Karaoglan Yilmaz (2023)
	Learning engagement ($n=14$)	Lv (2023); Nazari et al. (2021)
	Self-efficacy ($n=14$)	Essel et al. (2022); Han et al. (2022); Liaw et al. (2023)
Perceptions of AI-based learning tools	Cognitive load ($n=8$)	Koć-Januchta et al. (2020); Zheng et al. (2021)
	User evaluations ($n=31$)	Mahapatra (2024); Rokhayani et al. (2022)

options. Nevertheless, concerns surfaced regarding the reliability of AI tools, particularly when they delivered outdated, incorrect, or irrelevant information (Essel et al., 2022, 2024). Moreover, Mahapatra (2024) further cautioned against growing dependency on AI among college students, which coincided with a decline in motivation for independent thinking.

Discussion

AI-based learning tools hold immense promise for transforming higher education. However, their misuse can adversely affect learning outcomes (Niloy et al., 2024; Shorey et al., 2023). To fully harness the benefits of AI, it is imperative for stakeholders to cultivate a thorough understanding of these tools prior to their integration into educational frameworks. Our review seeks to address this imperative through two primary research objectives.

First, we aim to identify the key design elements of AI-based learning tools as described in the literature. In this pursuit, we observed that the advent of low-code and no-code platforms has significantly diminished the technical barriers for instructors, thus facilitating broader adoption of AI within higher education. Our findings further reveal that these tools primarily serve three core functions: assessing and evaluating students' learning outcomes, delivering personalized feedback and recommendations, and acting as intelligent tutors by tailoring instructional content to the unique progress and needs of individual learners.

Our second objective is to assess the impact of AI-based learning tools on college students' learning outcomes. A considerable portion of studies have addressed cognitive ($n = 44$) and affective learning outcomes ($n = 50$), with skill-based learning outcomes ($n = 20$) have received relatively scant attention. While many of the reviewed AI tools demonstrated their efficacy in enhancing cognitive knowledge and affective learning, their impact on cognitive process development and skill acquisition were more mixed and indefinite. These mixed outcomes underscore the necessity of aligning AI designs with specific learning objectives and highlight importance of continuous evaluation to ensure these tools effectively support diverse learning needs.

Design-to-evaluation framework

Building on our findings, we propose an iterative "design-to-evaluation" framework for the effective integration of AI in higher education (see Fig. 7). This framework emphasizes the reciprocal relationship between design and evaluation: effective designs must undergo holistic evaluations, and insights gained from these evaluations should inform ongoing design improvements. Specifically, we advocate for five core design principles: (1) human-centered designs that prioritize the learning needs of college students, (2) multimodal designs that enrich the learning experience, (3) transparent designs that mitigate the risks of misinformation, (4) inclusive designs that accommodate the needs of marginalized and disadvantaged students, and (5) ethical designs that address algorithmic biases and safeguard data privacy (see Table 6). Lastly, we provide actionable recommendations for the design and evaluation of AI-based learning tools and outline promising avenues for future research.

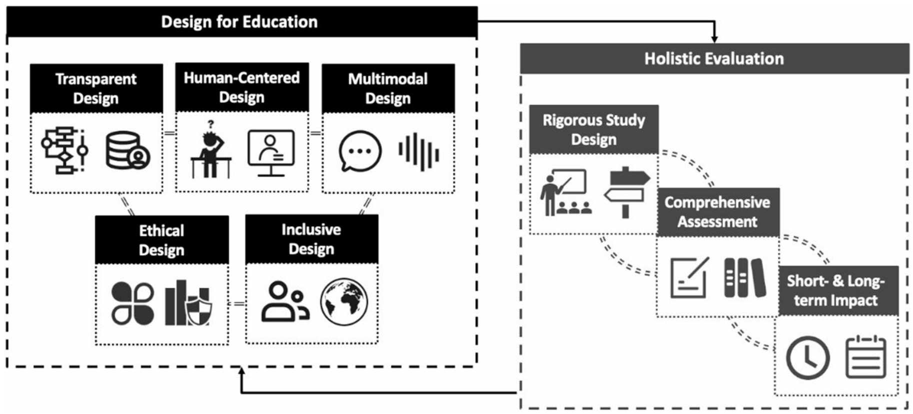

Fig. 7 Overview of the design-to-evaluation loop framework

Table 6 Future research directions for designing AI for higher education

Principle	Research Direction	Sample Future Research Questions
Human-centered design	Development of AI-based learning tools tailored to diverse student learning traits	1. What learning characteristics should be prioritized in the design of AI-based learning tools to effectively address the unique preferences and needs of college students? 2. How does tailoring AI-generated instructional materials to align with college students' learning traits influence their learning outcomes?
	Development of AI-based learning tools that incorporate human tutor insights	1. How can AI-based learning tools effectively integrate the knowledge and feedback provided by human tutors (e.g., personalized feedback, guidance, and instruction)? 2. How do college students perceive and respond to AI tools that simulate the teaching strategies of human tutors?
Multimodal design	Appropriate integration of multimodal content in AI-based learning tools	1. What modalities (e.g., text, images, audio, video) should be integrated into AI-based learning tools to effectively support the acquisition of complex knowledge and skills? 2. How do different modalities facilitate knowledge and skill acquisition within AI-based learning environments?
Transparent design	Implementation of transparent AI-based learning tools to facilitate learning	1. How do college students' interactions with transparent versus non-transparent AI tools affect their comprehension of learning materials? 2. What effect do transparent AI tools have on college students' critical thinking and reliance on AI-generated instructional content?
Inclusive design	Support for equity and inclusion in AI-based learning tools	1. How can AI-based learning tools be designed to provide meaningful and equitable educational support for marginalized and disadvantaged students? 2. How can AI-based learning tools be tailored to support students with disabilities, ensuring an equitable learning environment for all learners?
Ethical design	Mitigation of algorithmic biases in the development of AI-based learning tools	1. How can AI-based learning tools be designed to incorporate college students' backgrounds and experiences to reduce algorithmic bias and foster learning among underrepresented students? 2. What strategies can be employed to identify and mitigate algorithmic biases related to race, gender, and ethnicity in learning materials and assessment processes?
	Protection of data security, privacy, and fundamental human rights	1. How can data governance frameworks be structured to balance adaptive personalization with the protection of student privacy, autonomy, and informed consent in AI-based learning environments? 2. How can AI-based learning tools ensure transparent and ethical data practices that uphold students' human rights?

Designing AI for education

In this subsection, we present five core principles to guide the design of AI-based learning tools: human-centeredness, multimodality, transparency, inclusivity, and ethical responsibility (see Table 6).

Human-centered design

A human-centered approach is critical to ensuring AI-based learning tools align with the diverse learning needs and characteristics of college students (Boy, 2017; El-Sabagh, 2021). Currently, many existing AI tools personalize learning content primarily based on student performance metrics (e.g., Kavadella et al., 2024; Koć-Januchta et al., 2020; Vicente-Yagüe-Jara et al., 2023). However, they often overlook intrinsic learner traits, such as conscientiousness or learning styles, despite evidence that incorporating these factors can significantly enhance engagement and outcomes (Conati et al., 2021; Kaiss et al., 2023). In addition, we highlight the limitations of one-size-fits-all AI tools; for instance, AI-driven personalized video recommendations tend to disproportionately benefit moderately motivated learners, leaving highly motivated and less motivated learners underserved (Huang et al., 2023). Future AI designs must therefore integrate learner characteristics to ensure meaningful and equitable learning experiences for all students (Karimi, 2016; Tlili et al., 2016).

Furthermore, we underscore the importance of embedding human pedagogical expertise into AI tools. Expert human tutors excel at delivering learning materials strategically based on comprehensive assessments of student knowledge, cognition, and motivation (Burroughs et al., 2019). Unfortunately, our review reveals that current general-purpose AI-based learning tools primarily offer real-time, contextually isolated answers to student inquiries (e.g., Chang et al., 2024; Essel et al., 2024; Urban et al., 2024), which may undermine deeper cognitive engagement, stifle critical thinking, and foster an overdependence on AI (Essel et al., 2022; Niloy et al., 2024). To address these challenges, future research should prioritize integrating human instructional expertise into AI tools to enhance their pedagogical depth and strategic instructional capabilities.

Multimodal design

We recommend integrating multimodal content (e.g., text, images, audio, and interactive elements) into AI-based learning tools (Moreno & Mayer, 2007). Engaging multiple sensory channels enriches learning experiences, deepens knowledge comprehension and retention, and strengthens real-world problem-solving skills (Issa et al., 1999; Mayer, 1997, 2024). For example, textual content supports self-paced learning, audio materials benefit auditory learners, and visual elements such as images and animations help illustrate complex or esoteric knowledge concepts (Jo, 2024; Lange & Costley, 2020). Nonetheless, simply combining multiple modalities is insufficient; each modality must be carefully selected, synchronized, and tailored to specific learning tasks and objectives (Sarter, 2006; Xu et al., 2023). Our review indicates that current multimodal AI tools often inadequately support the development of complex skills (e.g., Fazlollahi et al., 2023; Shorey et al., 2023; Simsek-Cetinkaya & Cakir, 2023). Therefore, future efforts should strategically integrate and optimize multimodal features to support complex knowledge and skill acquisition more effectively.

Transparent design

Transparency is essential to mitigate misinformation risks and promote critical thinking. Research has shown that explaining the motivations ("why") and processes ("how") behind adaptive hints can foster learning for less conscientious learners (Conati et al., 2021). However, many of the AI tools we reviewed rely on opaque, black-box algorithms, which encourage passive consumption of AI-generated content rather than promoting critical reflection (e.g., Fazlollahi et al., 2023; Shorey et al., 2023; Simsek-Cetinkaya & Cakir, 2023). This opacity is compounded by large language models (LLMs) that produce seemingly plausible yet unverified content (Yadkori et al., 2024). As a result, college students may accept AI-generated content uncritically (Guo & Lee, 2023). To counteract this, we advocate for the design of transparent AI tools that elucidate their reasoning processes, which can empower students to critically assess learning materials, sharpen their understanding of core concepts, and strengthen their analytical reasoning (Conati et al., 2021; Ma et al., 2023).

Inclusive design

To promote educational equity, AI tools must be designed to accommodate diverse and marginalized student populations (Dumont & Ready, 2023). Currently, the advent of AI-based learning tools has the potential to exacerbate educational inequities by widening access gaps (Hardman, 2023). To elaborate, our review found that only four studies were conducted in countries with low or medium Human Development Index (HDI) rankings (Essel et al., 2022, 2024; Mahapatra, 2024; Niloy et al., 2024), and none of the reviewed studies addressed the needs of students with disabilities. Thus, we recommend that AI tools should be tailored to meet the learning needs of students in underdeveloped regions. Additionally, these tools should incorporate accessibility features (e.g., text-to-speech, adaptive interfaces, etc.) to foster an inclusive learning environment for all students.

Ethical design

Ethical considerations must underpin the design of AI-based learning tools, addressing critical issues such as algorithmic bias, privacy, and human rights (Hardman, 2023; Williamson et al., 2020; Williamson & Eynon, 2020). AI tools trained on large-scale datasets often inadvertently encode and propagate harmful stereotypes and biases related to gender, race, and ethnicity (Bender et al., 2021; Dixon-Román et al., 2020; Perrotta & Selwyn, 2020). For example, female students from historically marginalized groups, such as African American and Latina communities, frequently experience reduced cognitive engagement in computing fields, partly due to implicit biases embedded in instructional materials and the absence of culturally responsive pedagogies (Hoffman et al., 2019; Jiang et al., 2024; Noble, 2018). To combat these biases, it is essential to establish rigorous data auditing processes and advance algorithm designs that incorporate students' demographic, cultural, and experiential perspectives during the development of AI-based learning tools.

Moreover, personalized AI-based learning tools routinely collect sensitive student data, including demographic information, behavioral patterns, and academic records, and real-time monitoring, often without consent or adequate privacy safeguards (Afzaal et al., 2024; Conati et al., 2021). Commercial AI platforms, in particular, pose heightened risks of data misuse and exploitation (Pangrazio et al., 2023; Williamson et al., 2023), threatening to erode students' autonomy and diminish their agency over their personal learning trajectories. To safeguard student privacy and uphold their rights to informed choice and autonomy, future AI tools must carefully balance personalization with stringent data protection measures. Implementing end-to-end encryption, anonymization

Table 7 Future research directions for holistic evaluations of AI-based learning tools

Principle	Research Direction	Sample Future Research Questions
Rigorous study design	Improvement of methodological rigor	What guidelines should be established to improve study rigor in research examining the efficacy of AI-based learning tools?
Comprehensive assessment	Assessment of AI-based learning tools across diverse disciplines and learning outcomes	1. How do AI-based learning tools impact learning outcomes across various disciplines? 2. How can we tailor large language models (LLMs) to produce high-quality domain-specific content, particularly in STEM fields? 3.What strategies can be implemented to optimize AI-based learning tools for enhancing cognitive processes such as reasoning, critical thinking, and creativity among college students? 4. How can AI-based learning tools be utilized to support the development of complex skills in fields such as design, engineering, and negotiation? 5. How can AI-based learning tools be deployed to ensure equity, data security and privacy, and environmental sustainability?
Short- & Long-term impacts	Evaluation of short-term and long-term impacts of AI-based learning tools	1. What are the effects of AI-based learning tools on short-term versus long-term knowledge retention? 2. How can AI-based learning tools be optimized to support the transfer of knowledge and skills to new problem-solving scenarios?

techniques, and transparent consent mechanisms will be essential to prevent unauthorized data access and misuse, thereby ensuring ethical integrity in AI-augmented education (Selwyn et al., 2023; Williamson & Eynon, 2020).

Toward a holistic evaluation of AI-based learning tools

Beyond thoughtful design, we advocate for rigorous and holistic evaluation frameworks to systematically assess the efficacy, impact, and sustainability of these tools (see Fig. 7; Table 7).

Establishing evidence through rigorous study designs

Rigorous study designs are essential to establish the efficacy of AI-based learning tools in higher education. However, our review found that over half of the studies ($n = 40$) employed quasi-experimental designs without randomization (e.g., Dizon & Gayed, 2021; Mahapatra, 2024; Rokhayani et al., 2022; Zhang et al., 2023). Furthermore, many of these studies were conducted in offline learning settings, where participants could discuss learning materials with each other, raising concerns about potential contamination (e.g., El Shazly, 2021; Hakiki et al., 2023). While such design flaws may be justifiable given real-world constraints, they can undermine the reliability and validity of the findings (Duflo & Kremer, 2003; Ranganathan & Aggarwal, 2018). We hence call for the implementation of carefully designed randomized experiments in future research to rigorously evaluate the effectiveness of AI-based learning tools.

Comprehensive assessments of AI-based learning tools

Our review reveals a significant disciplinary imbalance in the scholarly attention paid to AI-based learning tools, with language instruction receiving considerable focus (e.g., Cheng, 2017; Dai & Wu, 2023; Escalante et al., 2023; Mahapatra, 2024) compared to STEM fields (e.g., Afzaal et al., 2024; Koć-Januchta et al., 2020). This disparity arises, in part, from the natural alignment of large language models (LLMs) with language education. Conversely, the effective application of these models in STEM disciplines is hampered by deficiencies in logical reasoning (Wu et al., 2023) and concerns regarding the quality of domain-specific learning materials (Pal et al., 2024). To fully unlock the potential of AI in higher education, more research is needed to investigate AI's potential across a broader spectrum of academic fields beyond language instruction.

Additionally, the studies we reviewed predominantly focused on cognitive ($n = 44$) and affective learning outcomes ($n = 50$), leaving skill-based outcomes ($n = 20$) largely overlooked. This notable gap highlights the pressing need to explore how AI might nurture complex, real-world competencies such as design thinking, engineering expertise, and negotiation prowess. Moreover, mixed findings regarding the efficacy of AI in enhancing cognitive processes, such as reasoning (Han et al., 2022), critical thinking (Essel et al., 2022), and creativity (Niloy et al., 2024), suggest that existing AI tools remain insufficiently attuned to the nuanced demands of these higher-order cognitive competencies. Accordingly, designing AI tools that better support cognitive development and skill acquisition represents a fruitful avenue for future research.

Ethical vigilance must permeate every stage of AI tool development and deployment, ensuring rigorous monitoring and evaluation to mitigate biases, safeguard data privacy, and promote environmental sustainability. Feedback from both students and instructors, especially from marginalized communities, should be systematically gathered and analyzed to identify and rectify hidden algorithmic biases (Hardman, 2023; Williamson & Eynon, 2020). Also, rather than relying solely on static security protocols, stakeholders should embrace adaptive security measures and conduct regular audits to proactively detect and resolve emerging threats (Janssen et al., 2020). Ethical evaluations must also encompass environmental and sustainable considerations, acknowledging the substantial computational resources required to train and deploy large-scale AI models (Bender et al., 2021). Balancing these environmental costs against the learning benefits will encourage universities and colleges to implement energy-efficient AI tools. Overall, embedding these multidimensional ethical evaluations is essential for fostering trust, equity, and environmental responsibility within AI-augmented learning environments.

Examining both short-term and long-term effects

We call for a careful evaluation of both the immediate and enduring impacts of AI-based learning tools in higher education. Existing literature underscores the critical importance of long-term knowledge retention for meaningful learning and the successful transfer of acquired competencies to new problem-solving contexts (Lindsey et al., 2014). Yet, the majority of studies we reviewed prioritized short-term learning outcomes (e.g., Howard et al., 2017; Kaiss et al., 2023; Rokhayani et al., 2022), with only four exceptions incorporating delayed assessments to evaluate knowledge (Zheng et al., 2023a, 2023b) and skill retention (Dai & Wu, 2023; Mahapatra, 2024). Given the pivotal role of knowledge retention in human learning (Custers, 2010), future research should integrate immediate and delayed assessments to quantify the efficacy of AI tools in promoting long-term knowledge and skill acquisition.

Moreover, inconsistencies between short-term and long-term outcomes emerged in our review (e.g., Mahapatra, 2024; Urban et al., 2024). For instance, Urban et al. (2024) documented enhanced creativity following a brief, 30-minute interaction with ChatGPT, while Mahapatra (2024) cautioned against rigid writing patterns and increased dependency on ChatGPT after a month-long intervention. Such discrepancies highlight the need for further inquiry into optimizing AI-based learning tools to facilitate knowledge retention and transfer.

Recommendations for the responsible integration of AI in higher education

Integrating AI-based learning tools into higher education is a multifaceted endeavor, which demands thoughtful deliberation throughout the entire design and evaluation lifecycle. Here, we propose several recommendations aimed at ensuring the effective, ethical, and responsible deployment of AI in educational contexts.

First and foremost, we urge adherence to the iterative "design-to-evaluation" framework. At the design stage, AI tools must embody: (a) human-centered principles that are attentive to students' diverse preferences, needs, and learning styles, (b) multimodal learning content that enrich educational experiences and heighten student engagement, (c) transparent reasoning processes that mitigate misinformation risks and cultivate user trust, (d) inclusive approaches that thoughtfully accommodate the diverse backgrounds and lived experiences of all students, and (e) ethical considerations that proactively mitigate algorithmic biases and safeguard student privacy, security and fundamental human rights. By embracing these design tenets, AI-based learning tools can empower college students, granting them greater autonomy and agency over their learning journeys.

Second, rigorous and holistic evaluation frameworks must accompany AI integration to assess their efficacy, impact, and sustainability. Robust, randomized experimental designs set within authentic educational settings will enhance the validity, reliability, and generalizability of findings. Moreover, evaluations should span diverse academic disciplines and encompass multiple dimensions of learning outcomes across both short and long terms to provide a holistic picture of the pedagogical effectiveness of these tools.

Meanwhile, ethical considerations must permeate every stage of AI design, deployment, and governance. Stakeholders must proactively identify and mitigate biases embedded within algorithms and training datasets to prevent reinforcement of existing social inequalities or unintended harm to marginalized groups (Bender et al., 2021; Williamson & Eynon, 2020). Concurrently, governments and regulatory bodies should establish comprehensive legal frameworks to safeguard student privacy and fundamental human rights, particularly as commercial entities increasingly shape AI-augmented education (Williamson et al., 2023). Such frameworks should emphasize explicability, combining intelligibility (clarity on how AI systems operate) and accountability (clarity on who is responsible for their outcomes), to foster transparency, public trust, and ethical governance in educational AI (Floridi & Cowls, 2022). Furthermore, attention must be paid to the environmental implications of AI's energy consumption, ensuring its sustainable development and responsible deployment (Bender et al., 2021).

Lastly, we contend that AI-based learning tools should serve as supportive allies rather than dominant arbiters of educational practice (Williamson et al., 2023). To maximize AI's potential, fostering AI literacy among both students and instructors is imperative (Allen & Kendeou, 2024; Ng et al., 2021). For students, enhanced AI literacy encourages critical reflection and active engagement, counteracting passive consumption of AI-generated content (Allen & Kendeou, 2024; Long & Magerko, 2020; Williamson et al., 2020). For instructors, cultivating AI literacy helps resist overreliance on AI tools that often prioritize narrow metrics, such as standardized test scores and narrowly defined skillsets, over learners' unique strengths, needs, and developmental trajectories (Williamson et al., 2020). Therefore, instructors must retain pedagogical agency and embrace pedagogical practices that nurture holistic learner growth (Williamson et al., 2023). In sum, integrating established AI literacy frameworks in educational practice will empower both learners and educators to critically evaluate and meaningfully utilize AI tools in alignment with broader educational aims and values (Allen & Kendeou, 2024; Ng et al., 2021).

Limitations and directions for future research

This review acknowledges several limitations and presents several avenues for future research. First, we focused exclusively on AI-based learning tools in higher education, excluding articles in other educational settings such as K-12 education, lifelong learning, and informal education. Moreover, we restricted our review to peer-reviewed journal articles published in English, potentially overlooking relevant studies published in conference proceedings or non-English sources. Future research should expand our scope to include these additional perspectives. Second, our evaluation primarily emphasized the impact of AI tools from the perspective of college students. A fruitful direction for future research lies in exploring instructors' lived experiences with AI and examining how these technologies shape their pedagogical practices and identities. Finally, although we have discussed ethical considerations broadly, we did not explicitly and systematically review studies addressing the ethical implications of AI, which is essential for ensuring responsible and equitable AI deployment.

Conclusion

AI-based learning tools bear tremendous potential for higher education. Our systematic review examines their design and effectiveness, demonstrating significant improvements in cognitive knowledge and affective learning outcomes for college students. Yet, their impact on higher-order cognitive processes and skill development presents a more complex, nuanced picture. Drawing upon these insights, we propose a clear research roadmap to guide future design and evaluation efforts, with the ultimate goal of better equipping students for their future careers.

Moving forward, research should closely examine how college students and instructors interact with AI-based learning tools across diverse disciplinary contexts. Such investigations promise richer insights into the adoption and implementation of these emerging technologies in higher education. Also, interdisciplinary collaboration is vital for developing ethical AI literacy frameworks that prioritize transparency, accountability, and equity. Finally, concerted policy initiatives are critical to ensuring that AI tools are not only pedagogically effective but also equitable, trustworthy, and responsive to the diverse needs and aspirations of both students and instructors.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41239-025-00540-2.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
JL: Conceptualization; Formal analysis; Methodology; Writing - Original Draft; Writing - Review & Editing; CZ: Project administration; Formal analysis; Funding; Methodology; Writing - Review & Editing; JY: Conceptualization; Data curation; Funding; Project administration; Methodology; Writing - Review & Editing; TH: Conceptualization; Funding; Project administration; Supervision; Writing - Review & Editing.

Funding
This research is supported by the National University of Singapore, Singapore under its Humanities and Social Sciences Seed Fund, the National Natural Science Foundation of China (Grant Number: 72301279; 72442016; 72472151), and The University of Sydney, Australia under its Internal Fund.

Data availability
The datasets used and/or analyzed during the current study (the bibliography of included studies) are available from the corresponding author upon request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 8 February 2025 / Accepted: 20 May 2025

Published online: 11 July 2025

References

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, *21*(1), 10. https://doi.org/10.1186/s41239-024-00444-7

Afzaal, M., Zia, A., Nouri, J., & Fors, U. (2024). Informative feedback and explainable AI-Based recommendations to support students' Self-regulation. *Technology Knowledge and Learning*, *29*(1), 331–354. https://doi.org/10.1007/s10758-023-09650-0

Allen, L., & Kendeou, K. (2024). ED-AI lit: An interdisciplinary framework for AI literacy in education. *Policy Insights from the Behavioral and Brain Sciences*, *11*(1), 3–10. https://doi.org/10.1177/23727322231220339

Almarode, J., & Vandas, K. (2019). *Clarity for learning: Five essential practices that empower students and teachers*. Corwin.

An, P., & Wang, Z. (2023). Application value of an artificial intelligence-based diagnosis and recognition system in gastroscopy training for graduate students in gastroenterology: A preliminary study. *Wiener Medizinische Wochenschrift*. https://doi.org/10.1007/s10354-023-01020-w

Baig, M. I., & Yadegaridehkordi, E. (2023). Flipped classroom in higher education: A systematic literature review and research challenges. *International Journal of Educational Technology in Higher Education*, *20*(1), 61. https://doi.org/10.1186/s41239-023-00430-5

Baker, T., & Smith (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. Accessed September 23, 2024.

Bates, T., Cobo, C., Mariño, O., & Wheeler, S. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education*, *17*(1), 42. https://doi.org/10.1186/s41239-020-00218-x

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Special issue editor's comments: Managing artificial intelligence. *Management Information Systems Quarterly*, *45*(3), 1433–1450.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-023-00436-z

Boy, G. A. (2017). *The handbook of Human-Machine interaction: A Human-Centered design approach*. CRC.

Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). A review of the literature on teacher effectiveness and student outcomes. In N. Burroughs, J. Gardner, Y. Lee, S. Guo, I. Touitou, K. Jansen, & W. Schmidt (Eds.), *Teaching for excellence and equity: Analyzing teacher characteristics, behaviors and student outcomes with TIMSS* (pp. 7–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-16151-4_2

Chang, C. Y., Hwang, G., & Gau, M. (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology*, *53*(1), 171–188. https://doi.org/10.1111/bjet.13158

Chang, C. Y., Yang, C. L., Jen, H. J., Ogata, H., & Hwang, G. H. (2024). Facilitating nursing and health education by incorporating ChatGPT into learning designs. *Educational Technology & Society*, *27*(1), 215–230. https://www.jstor.org/stable/48754852

Chen, Z., & Yu, X. (2020). Adoption of human personality development theory combined with deep neural network in entrepreneurship education of college students. *Frontiers in Psychology*, *11*, 1346. https://doi.org/10.3389/fpsyg.2020.01346

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, *25*(1), 161–182. https://doi.org/10.1007/s10796-022-10291-4

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, *34*, 18–27. https://doi.org/10.1016/j.iheduc.2017.04.002

Chiu, M. C., Hwang, G. J., Hsia, L. H., & Shyu, F. M. (2022). Artificial intelligence-supported Art education: A deep learning-based system for promoting university students' Artwork appreciation and painting outcomes. *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2022.2100426

Chu, H. C., Hwang, G. H., Tu, Y. F., & Yang, K. H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.7526
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104
Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, *298*, 103503. https://doi.org/10.1016/j.artint.2021.103503
Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), 22. https://doi.org/10.1186/s41239-023-00392-8
Custers, E. J. F. M. (2010). Long-term retention of basic science knowledge: A review study. *Advances in Health Sciences Education*, *15*(1), 109–128. https://doi.org/10.1007/s10459-008-9101-y
Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, *36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272
Dall'Alba, G. (1994). The role of teaching in higher education: Enabling students to enter a field of study and practice. *Learning and Instruction*, *3*(4), 299–313. https://doi.org/10.1016/0959-4752(93)90021-Q
Darban, M. (2023). The future of virtual team learning: Navigating the intersection of AI and education. *Journal of Research on Technology in Education*, 1–17. https://doi.org/10.1080/15391523.2023.2288912
Díaz, B., & Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education*, *217*, 105071. https://doi.org/10.1016/j.compedu.2024.105071
Dixon-Román, E., Nichols, T., Philip,and, & Nyame-Mensah, A. (2020). The racializing forces of/in AI educational technologies. *Learning Media and Technology*, *45*(3), 236–250. https://doi.org/10.1080/17439884.2020.1667825
Dizon, G., & Gayed, J. (2021). Examining the impact of grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, *17*(2), 74–92. https://doi.org/10.29140/jaltcall.v17n2.336
Duflo, E., & Kremer, M. (2003). Use of randomization in the evaluation of development effectiveness. *Evaluating development effectiveness* (pp. 205). Transaction.
Dumont, H., & Ready, D. D. (2023). On the promise of personalized learning for educational equity. *Npj Science of Learning*, *8*(1), 1–6. https://doi.org/10.1038/s41539-023-00174-x
El Shazly, R. (2021). Effects of artificial intelligence on english speaking anxiety and speaking performance: A case study. *Expert Systems*, *38*(3), e12667. https://doi.org/10.1111/exsy.12667
El-Sabagh, H. A. (2021). Adaptive e-learning environment based on learning styles and its impact on development students' engagement. *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-021-00289-4
Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2
Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, *19*(1), 57. https://doi.org/10.1186/s41239-022-00362-6
Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large Language models (LLMs). *Computers and Education: Artificial Intelligence*, *6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198
Fazlollahi, A. M., Bakhaidar, M., Alsayegh, A., Yilmaz, R., Winkler-Schwartz, A., Mirchi, N., Langleben, I., Ledwos, N., Sabbagh, A. J., Bajunaid, K., Harley, J. M., & Del Maestro, R. F. (2022). Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: A randomized clinical trial. *JAMA Network Open*, *5*(2), e2149008. https://doi.org/10.1001/jamanetworkopen.2021.49008
Fazlollahi, A. M., Yilmaz, R., Winkler-Schwartz, A., Mirchi, N., Ledwos, N., Bakhaidar, M., Alsayegh, A., & Del Maestro, R. F. (2023). AI in surgical curriculum design and unintended outcomes for technical competencies in simulation training. *JAMA Network Open*, *6*(9), e2334658. https://doi.org/10.1001/jamanetworkopen.2023.34658
Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the City* (pp. 535–545). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119815075.ch45
Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057
Grand View Research (2022). *AI In Education Market Size & Share Report, 2022–2030*. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-education-market-report. Accessed September 23, 2024.
Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, *100*(12), 4876–4883. https://doi.org/10.1021/acs.jchemed.3c00505
Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, *203*, 104862. https://doi.org/10.1016/j.compedu.2023.104862
Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of self-regulation of learning and performance* (pp. 65–84). Routledge/Taylor & Francis Group.
Hakiki, M., Fadli, R., Samala, A. D., Fricticarani, A., Dayurni, P., Rahmadani, K., & Astiti, A. D. (2023). Exploring the impact of using Chat-GPT on student learning outcomes in technology learning: The comprehensive experiment. *Advances in Mobile Learning Educational Research*, *3*(2), 859–872. https://doi.org/10.25082/AMLER.2023.02.013
Han, J. W., Park, J., & Lee, H. (2022). Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: A quasi-experimental study. *BMC Medical Education*, *22*(1), 830. https://doi.org/10.1186/s12909-022-03898-3
Hardman, D. P. (2023, June 29). The AI-Education Divide [Substack newsletter]. *Dr Phil's Newsletter, Powered by DOMS™AI*. https://drphilippahardman.substack.com/p/the-ai-education-divide. Accessed September 16, 2024.
Hasan, M. R., & Khan, B. (2023). An AI-based intervention for improving undergraduate STEM learning. *PLOS ONE*, *18*(7), e0288844. https://doi.org/10.1371/journal.pone.0288844
Hoffman, B., Morelli, R., & Rosato, J. (2019). Student engagement is key to broadening participation in CS. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, *1123*, 1129. https://doi.org/10.1145/3287324.3287438

Hopson, M. H., Simms, R. L., & Knezek, G. A. (2001). Using a Technology-Enriched environment to improve Higher-Order thinking skills. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2001.10782338

Howard, C., Jordan, P., Di Eugenio, B., & Katz, S. (2017). Shifting the load: A peer dialogue agent that encourages its human collaborator to contribute more to problem solving. *International Journal of Artificial Intelligence in Education*, *27*(1), 101–129. https://doi.org/10.1007/s40593-015-0071-y

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, *194*, 104684. https://doi.org/10.1016/j.compedu.2022.104684

Issa, R. R. A., Cox, R. F., & Killingsworth, C. F. (1999). Impact of Multimedia-Based instruction on learning and retention. *Journal of Computing in Civil Engineering*, *13*(4), 281–290. https://doi.org/10.1061/(ASCE)0887-3801(1999)13:4(281)

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Jiang, S., McClure, J., Tatar, C., Bickel, F., Rosé, C. P., & Chao, J. (2024). Towards inclusivity in AI: A comparative study of cognitive engagement between marginalized female students and peers. *British Journal of Educational Technology*, *55*(6), 2557–2573. https://doi.org/10.1111/bjet.13467

Jiang, Jong, M. S. Y., Wu, N., Shen, B., Chai, C. S., Lau, W. W. F., & Huang, B. (2022). Integrating automatic speech recognition technology into vocabulary learning in a flipped english class for Chinese college students. *Frontiers in Psychology*, *13*, 902429. https://doi.org/10.3389/fpsyg.2022.902429

Jiang, Jong, M. S., Lau, W. W., Chai, C., & Wu, N. (2023). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computer Assisted Learning*, *39*(1), 125–140. https://doi.org/10.1111/jcal.12732

Jin, S. H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00406-5

Jo, H. (2024). From concerns to benefits: A comprehensive study of ChatGPT usage in education. *International Journal of Educational Technology in Higher Education*, *21*(1), 35. https://doi.org/10.1186/s41239-024-00471-4

Kaiss, W., Mansouri, K., & Poirier, F. (2023). Effectiveness of an adaptive learning chatbot on students' learning outcomes based on learning styles. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(13), 250–261. https://doi.org/10.3991/ijet.v18i13.39329

Karimi, S. (2016). Do learners' characteristics matter? An exploration of mobile-learning adoption in self-directed learning. *Computers in Human Behavior*, *63*, 769–776. https://doi.org/10.1016/j.chb.2016.06.014

Kavadella, A., Dias Da Silva, M. A., Kaklamanos, E. G., Stamatopoulos, V., & Giannakopoulos, K. (2024). Evaluation of ChatGPT's Real-Life implementation in undergraduate dental education: Mixed methods study. *JMIR Medical Education*, *10*, e51344. https://doi.org/10.2196/51344

Keele (2007). *Guidelines for performing systematic literature reviews in software engineering* Vol. 5. Technical report, ver. 2.3 ebse technical report. ebse.

Koć-Januchta, M. M., Schönborn, K. J., Tibell, L. A. E., Chaudhri, V. K., & Heller, H. C. (2020). Engaging with biology by asking questions: Investigating students' interaction and learning with an artificial Intelligence-Enriched textbook. *Journal of Educational Computing Research*, *58*(6), 1190–1224. https://doi.org/10.1177/0735633120921581

Kolil, V. K., Muthupalani, S., & Achuthan, K. (2020). Virtual experimental platforms in chemistry laboratory education and its impact on experimental self-efficacy. *International Journal of Educational Technology in Higher Education*, *17*(1), 30. https://doi.org/10.1186/s41239-020-00204-3

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*(2), 311–328. https://doi.org/10.1037/0021-9010.78.2.311

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lange, C., & Costley, J. (2020). Improving online video lectures: Learning challenges created by media. *International Journal of Educational Technology in Higher Education*, *17*(1), 16. https://doi.org/10.1186/s41239-020-00190-6

Lee, Y. F., Hwang, G. J., & Chen, P. Y. (2022). Impacts of an AI-based Chabot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational Technology Research and Development*, *70*(5), 1843–1865. https://doi.org/10.1007/s11423-022-10142-8

Lei, H., Chen, C., & Luo, L. (2024). The examination of the relationship between learning motivation and learning effectiveness: A mediation model of learning engagement. *Humanities and Social Sciences Communications*, *11*(1), 1–11. https://doi.org/10.1057/s41599-024-02666-6

Liaw, S. Y., Tan, J. Z., Lim, S., Zhou, W., Yap, J., Ratan, R., Ooi, S. L., Wong, S. J., Seah, B., & Chua, W. L. (2023). Artificial intelligence in virtual reality simulation for interprofessional communication training: Mixed method study. *Nurse Education Today*, *122*, 105718. https://doi.org/10.1016/j.nedt.2023.105718

Lin, C. J., & Mubarok, H. (2021). Learning analytics for investigating the Mind Map-Guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, *24*(4), 16–35. https://www.jstor.org/stable/48629242

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' Long-Term knowledge retention through personalized review. *Psychological Science*, *25*(3), 639–647. https://doi.org/10.1177/0956797613504302

Liu, C., Lim, R. L., McCabe, K. L., Taylor, S., & Calvo, R. A. (2016). A Web-Based telehealth training platform incorporating automated nonverbal behavior feedback for teaching communication skills to medical students: A randomized crossover study. *Journal of Medical Internet Research*, *18*(9), e246. https://doi.org/10.2196/jmir.6299

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376727

Lv, H. Z. (2023). Innovative music education: Using an AI-based flipped classroom. *Education and Information Technologies*, *28*(11), 15301–15316. https://doi.org/10.1007/s10639-023-11835-0

Ma, S., Lei, Y., Wang, X., Zheng, C., Shi, C., Yin, M., & Ma, X. (2023). Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3544548.3581058

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, *11*(1), 9. https://doi.org/10.1186/s40561-024-00295-9

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*(1), 8. https://doi.org/10.1007/s10648-023-09842-1

McCarthy, J., Minsky, M. L., Rochester, N., Corporation, I. B. M., & Shannon, C. E. (1956). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–12.

McCombs, B. L. (1984). Processes and skills underlying continuing intrinsic motivation to learn: Toward a definition of motivational skills training interventions. *Educational Psychologist*, *19*(4), 199–218. https://doi.org/10.1080/00461528409529297

Minsky, M. (1968). *Semantic information processing*. The MIT Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and Meta-Analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Müller, F. A., & Wulf, T. (2020). Technology-supported management education: A systematic review of antecedents of learning effectiveness. *International Journal of Educational Technology in Higher Education*, *17*(1), 47. https://doi.org/10.1186/s41239-020-00226-x

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, *7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Acm*, *19*(3), 113–126. https://doi.org/10.1145/360018.360022

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is Chatgpt a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, *40*(2), 919–930. https://doi.org/10.1111/jcal.12929

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU. https://doi.org/10.2307/j.ctt1pwt9w5

Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video improves learning in higher education: A systematic review. *Review of Educational Research*, *91*(2), 204–236. https://doi.org/10.3102/0034654321990713

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology*, *134*, 103–112. https://doi.org/10.1016/j.jclinepi.2021.02.003

Pajares, F. (1996). Self-Efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578. https://doi.org/10.2307/1170653

Pal, S., Bhattacharya, M., Lee, S. S., & Chakraborty, C. (2024). A Domain-Specific Next-Generation large Language model (LLM) or ChatGPT is required for biomedical engineering and research. *Annals of Biomedical Engineering*, *52*(3), 451–454. https://doi.org/10.1007/s10439-023-03306-x

Pangrazio, L., Selwyn, Neil, & Cumbo, B. (2023). A patchwork of platforms: Mapping data infrastructures in schools. *Learning Media and Technology*, *48*(1), 65–80. https://doi.org/10.1080/17439884.2022.2035395

Perrotta, C.,and, & Selwyn, N. (2020). Deep learning goes to school: Toward a relational Understanding of AI in education. *Learning Media and Technology*, *45*(3), 251–269. https://doi.org/10.1080/17439884.2020.1686017

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-Quality Meta-Analysis in a systematic review. *Review of Educational Research*, *90*(1), 24–46. https://doi.org/10.3102/0034654319877153

Prinsloo, P. (2020). Of 'black boxes' and algorithmic decision-making in (higher) education– A commentary. *Big Data & Society*, *7*(1), 2053951720933994. https://doi.org/10.1177/2053951720933994

Ranganathan, P., & Aggarwal, R. (2018). Study designs: Part 1– An overview and classification. *Perspectives in Clinical Research*, *9*(4), 184–186. https://doi.org/10.4103/picr.PICR_124_18

Rokhayani, Rukmini, Hartono, & Mujiyanto (2022). Integrating technology in online learning based on Computer-Mediated communication artificial intelligence to improve students' achievement. *Journal of Higher Education Theory and Practice*, *22*(15). https://doi.org/10.33423/jhetp.v22i15.5575

Ross, B., Chase, A. M., Robbie, D., Oates, G., & Absalom, Y. (2018). Adaptive quizzes to increase motivation, engagement and learning outcomes in a first year accounting unit. *International Journal of Educational Technology in Higher Education*, *15*(1), 30. https://doi.org/10.1186/s41239-018-0113-2

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Pearson.

Saravia-Rojas, M. Á., Camarena-Fonseca, A. R., León-Manco, R., & Geng-Vivanco, R. (2024). Artificial intelligence: ChatGPT as a disruptive didactic strategy in dental education. *Journal of Dental Education*, *jdd.13485*. https://doi.org/10.1002/jdd.13485

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, *36*(5), 439–445. https://doi.org/10.1016/j.ergon.2006.01.007

Schoenfeld, A. H. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher*, *28*(7), 4–14. https://doi.org/10.3102/0013189X028007004

Selwyn, N., Campbell, Liz, & Andrejevic, M. (2023). Autoroll: Scripting the emergence of classroom facial recognition technology. *Learning Media and Technology*, *48*(1), 166–179. https://doi.org/10.1080/17439884.2022.2039938

Shorey, S., Ang, E. N. K., Ng, E. D., Yap, J., Lau, L. S. T., Chui, C. K., & Chan, Y. H. (2023). Evaluation of a Theory-Based virtual counseling application in nursing education. *CIN: Computers Informatics Nursing*, *41*(6), 385–393. https://doi.org/10.1097/CIN.0000000000000999

Simsek-Cetinkaya, S., & Cakir, S. K. (2023). Evaluation of the effectiveness of artificial intelligence assisted interactive screen-based simulation in breast self-examination: An innovative approach in nursing students. *Nurse Education Today*, *127*, 105857. https://doi.org/10.1016/j.nedt.2023.105857
Sun, J. C. Y., Tsai, H. E., & Cheng, W. K. R. (2023). Effects of integrating an open learner model with AI-enabled visualization on students' self-regulation strategies usage and behavioral patterns in an online research ethics course. *Computers and Education: Artificial Intelligence*, *4*, 100120. https://doi.org/10.1016/j.caeai.2022.100120
Tegos, & Demetriadis (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Educational Technology & Society*. http://www.jstor.org/stable/jeductechsoci.20.1.99
Tlili, A., Essalmi, F., Jemni, M., Kinshuk, & Chen, N. S. (2016). Role of personality in computer based learning. *Computers in Human Behavior*, *64*, 805–813. https://doi.org/10.1016/j.chb.2016.07.043
U.S. Department of Education. (2023). *Artificial intelligence and the future of teaching and learning*. U.S. Department of Education, Office of Educational Technology.
Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031
Vicente-Yagüe-Jara, M. I., López-Martínez, O., Navarro-Navarro, V., & Cuéllar-Santiago, F. (2023). Writing, creativity, and artificial intelligence. ChatGPT in the university context. *Comunicar*, *31*(77). https://doi.org/10.3916/C77-2023-04
Wan, Z., Compeau, D., & Haggerty, N. (2012). The effects of Self-Regulated learning processes on E-Learning outcomes in organizational settings. *Journal of Management Information Systems*, *29*(1), 307–340. https://doi.org/10.2753/MIS0742-1222290109
Wei, L. (2023). Artificial intelligence in Language instruction: Impact on english learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955
Wei, X., Saab, N., & Admiraal, W. (2021). Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review. *Computers & Education*, *163*, 104097. https://doi.org/10.1016/j.compedu.2020.104097
Weiner, I. B., Reynolds, W. M., & Miller, G. E. (2012). *Handbook of psychology, educational psychology*. Wiley.
Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning Media and Technology*, *45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995
Williamson, B., Macgilchrist, Felicitas, & Potter, J. (2023). Re-examining AI, automation and datafication in education. *Learning, Media and Technology*, *48*(1), 1–5. https://doi.org/10.1080/17439884.2023.2167830
Williamson, B., Bayne, Sian, & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. *Teaching in Higher Education*, *25*(4), 351–365. https://doi.org/10.1080/13562517.2020.1748811
Wolfe, C. R., Reyna, V. F., Widmer, C. L., Cedillos, E. M., Fisher, C. R., Brust-Renck, P. G., & Weil, A. M. (2015). Efficacy of a Web-Based intelligent tutoring system for communicating genetic risk of breast cancer: A Fuzzy-Trace theory approach. *Medical Decision Making*, *35*(1), 46–59. https://doi.org/10.1177/0272989X14535983
Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122–1136. https://doi.org/10.1109/JAS.2023.123618. IEEE/CAA Journal of Automatica Sinica.
Xu, W., Wu, Y., & Ouyang, F. (2023). Multimodal learning analytics of collaborative patterns during pair programming in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 8. https://doi.org/10.1186/s41239-022-00377-z
Yadkori, Y. A., Kuzborskij, I., György, A., & Szepesvári, C. (2024). *To Believe or Not to Believe Your LLM* (arXiv:2406.02543). arXiv. http://arxiv.org/abs/2406.02543
Yang, Y. Y., & Shulruf, B. (2019). An expert-led and artificial intelligence system-assisted tutoring course to improve the confidence of Chinese medical interns in suturing and ligature skills: A prospective pilot study. *Journal of Educational Evaluation for Health Professions*, *16*, 7. https://doi.org/10.3352/jeehp.2019.16.7
Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*, 100147. https://doi.org/10.1016/j.caeai.2023.100147
Yin, J., Ngiam, K. Y., & Teo, H. H. (2021). Role of artificial intelligence applications in Real-Life clinical practice: Systematic review. *Journal of Medical Internet Research*, *23*(4), e25759. https://doi.org/10.2196/25759
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education– where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0
Zhang, R., Zou, D., & Cheng, G. (2023). Chatbot-based training on logical fallacy in EFL argumentative writing. *Innovation in Language Learning and Teaching*, *17*(5), 932–945. https://doi.org/10.1080/17501229.2023.2197417
Zhao, X. X., Wu, S. P., Wang, J. Y., Gong, X. Y., He, X. R., Xi, M. J., & Yuan, W. F. (2020). Comparison of multiple quantitative evaluation indices of theoretical knowledge and clinical practice skills and training of medical interns in cardiovascular imaging using blended teaching and the case resource network platform (CRNP). *Medical Science Monitor*, *26*. https://doi.org/10.12659/MSM.923836
Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., & Wen, J. R. (2024). *A Survey of Large Language Models* (arXiv:2303.18223). arXiv. https://doi.org/10.48550/arXiv.2303.18223
Zheng, L., Zhong, L., Niu, J., Long, M., & Zhao, J. (2021). Effects of personalized intervention on collaborative knowledge building, group performance, socially shared metacognitive regulation, and cognitive load in Computer-Supported collaborative learning. *Educational Technology & Society*, *24*(3), 174–193. https://www.jstor.org/stable/27032864
Zheng, L., Niu, J., & Zhong, L. (2022a). Effects of a learning analytics-based real-time feedback approach on knowledge elaboration, knowledge convergence, interactive relationships and group performance in CSCL. *British Journal of Educational Technology*, *53*(1), 130–149. https://doi.org/10.1111/bjet.13156
Zheng, L., Zhong, L., & Niu, J. (2022b). Effects of personalised feedback approach on knowledge building, emotions, co-regulated behavioural patterns and cognitive load in online collaborative learning. *Assessment & Evaluation in Higher Education*, *47*(1), 109–125. https://doi.org/10.1080/02602938.2021.1883549

Zheng, L., Zhong, L., & Fan, Y. (2023a). An automated analysis of topic distributions and features approach to promoting group performance, collaborative knowledge Building and socially shared regulation in online collaborative learning. *Australasian Journal of Educational Technology*, 114–132. https://doi.org/10.14742/ajet.7995

Zheng, L., Zhong, L., & Fan, Y. (2023b). An immediate analysis of the interaction topic approach to promoting group performance, knowledge convergence, cognitive engagement, and coregulation in online collaborative learning. *Education and Information Technologies*, *28*(8), 9913–9934. https://doi.org/10.1007/s10639-023-11588-w

Zheng, L., Fan, Y., Chen, B., Huang, Z., LeiGao, & Long, M. (2023c). An AI-enabled feedback-feedforward approach to promoting online collaborative learning. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12292-5

Zheng, L., Long, M., Chen, B., & Fan, Y. (2023d). Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: An automated assessment and feedback approach based on knowledge graphs. *International Journal of Educational Technology in Higher Education*, *20*(1), 46. https://doi.org/10.1186/s41239-023-00415-4

Zheng, L., Long, M., Niu, J., & Zhong, L. (2023e). An automated group learning engagement analysis and feedback approach to promoting collaborative knowledge building, group performance, and socially shared regulation in CSCL. *International Journal of Computer-Supported Collaborative Learning*, *18*(1), 101–133. https://doi.org/10.1007/s11412-023-09386-0

Zheng, L., Niu, J., Long, M., & Fan, Y. (2023f). An automatic knowledge graph construction approach to promoting collaborative knowledge building, group performance, social interaction and socially shared regulation in CSCL. *British Journal of Educational Technology*, *54*(3), 686–711. https://doi.org/10.1111/bjet.13283

Zheng, L., Fan, Y., Gao, L., Huang, Z., Chen, B., & Long, M. (2024a). Using AI-empowered assessments and personalized recommendations to promote online collaborative learning performance. *Journal of Research on Technology in Education*, 1–27. https://doi.org/10.1080/15391523.2024.2304066

Zheng, L., Fan, Y., Huang, Z., & Gao, L. (2024b). Impacts of three approaches on collaborative knowledge building, group performance, behavioural engagement, and socially shared regulation in online collaborative learning. *Journal of Computer Assisted Learning*, *40*(1), 21–36. https://doi.org/10.1111/jcal.12860

Publisher's note