A Meta Systematic Review of Artificial Intelligence in Higher Education: A call for increased ethics, collaboration, and rigour

9 authors, including:

Melissa Bond
University College London
128 PUBLICATIONS **4,578** CITATIONS

SEE PROFILE

Nina Bergdahl
Halmstad University
55 PUBLICATIONS **608** CITATIONS

SEE PROFILE

Emily Oxley
University of Glasgow
28 PUBLICATIONS **51** CITATIONS

SEE PROFILE

Sin Wang Chong
University of St Andrews
116 PUBLICATIONS **620** CITATIONS

SEE PROFILE

A Meta Systematic Review of Artificial Intelligence in Higher Education: A call for increased ethics, collaboration, and rigour

Melissa Bond – corresponding author
melissa.bond@ucl.ac.uk
EPPI Centre, University College London
Knowledge Center for Education, University of Stavanager
National Institute of Teaching

Hassan Khosravi
h.khosravi@uq.edu.au
Institute for Teaching and Learning Innovation, The University of Queensland

Maarten De Laat
Maarten.DeLaat@unisa.edu.au
Centre for Change and Complexity in Learning, Education Futures, University of South Australia

Nina Bergdahl
nina.bergdahl@hh.se
Halmstad University
Stockholm University

Violeta Negrea
v.negrea@niot.org.uk
National Institute of Teaching

Emily Oxley
e.oxley@niot.org.uk
National Institute of Teaching

Phuong Pham
phuong.pham@mymail.unisa.edu.au
Centre for Change and Complexity in Learning, Education Futures, University of South Australia

Sin Wang Chong
swc5@st-andrews.ac.uk
International Education Institute, University of St Andrew's
National Institute of Teaching

George Siemens
George.Siemens@unisa.edu.au
Centre for Change and Complexity in Learning, Education Futures, University of South Australia

A Meta Systematic Review of Artificial Intelligence in Higher Education: A call for increased ethics, collaboration, and rigour

Abstract

Although the field of Artificial Intelligence in Education (AIEd) has a substantial history as a research domain, never before has the rapid evolution of AI applications in education sparked such prominent public discourse. Given the growing AIEd literature base in higher education, it is important to ensure that the field has a solid research and conceptual grounding as AI adoption increases. This review of reviews is the first comprehensive meta review to explore the scope and nature of AIEd in higher education (AIHEd) research, by synthesising secondary research (e.g., systematic reviews), indexed in the Web of Science, Scopus, ERIC, EBSCOHost, IEEE Xplore, ScienceDirect and ACM Digital Library, or captured through snowballing in OpenAlex, ResearchGate and Google Scholar. Reviews were included if they synthesised applications of AI solely in formal higher or continuing education, were published in English between 2018 and July 2023, were journal articles or full conference papers, and if they had a method section. 66 publications were included for data extraction and synthesis in EPPI Reviewer, which were predominantly systematic reviews (66.7%), published by authors from North America (27.3%), conducted in teams (89.4%) in mostly domestic-only collaborations (71.2%). Findings show that these reviews mostly focused on AIHEd generally (47.0%) or Profiling and Prediction (28.8%) as thematic foci, however key findings indicated a predominance of the use of Adaptive Systems and Personalisation in higher education. Research gaps identified suggest a need for greater ethical, methodological, and contextual considerations within future research, alongside interdisciplinary approaches to AIHEd application. Suggestions are provided to guide future primary and secondary research.

Keywords

Artificial Intelligence, AIEd, AI, evidence synthesis, tertiary review, research methods, quality assessment, intelligent tutoring systems, adaptive systems, prediction, personalisation, automatic assessment

Introduction

Artificial Intelligence (AI) has existed as a distinct discipline since the 1956 Dartmouth Workshop and its adoption in education, particularly with the early introduction of intelligent tutoring systems, has become a substantive research domain (AIEd). Despite the growing realisation of the potential for AI within education, influenced by educational evidence-based policy, including education departments and international organisations (e.g., OECD, 2021), it has arguably only now transitioned from work in labs to active practice in classrooms, and broken through the veil of public discourse. The introduction of OpenAI's ChatGPT[1] and DALL-E[2], for example, has both captured the public's imagination and anxiety in equal measure (Bozkurt et al., 2023), requiring schools, universities, and organisations to respond to generative AI's growing capabilities, with increasing numbers of publicly available AI chatbots on the horizon (e.g., Google's Bard[3] and Meta's LLaMA[4]). The uptake of these tools has given rise to a debate in education about readiness, ethics, trust, impact, and value add of AI, as well as the need for governance, regulation,

---

[1] https://chat.openai.com/
[2] https://openai.com/dall-e-2
[3] https://blog.google/technology/ai/bard-google-ai-search-updates/
[4] https://ai.meta.com/blog/large-language-model-llama-meta-ai/

research and training to cope with the speed and scale at which AI is transforming teaching and learning. Globally, governments are putting measures in place to respond to this unfolding phenomenon. For example, in Europe they introduced the EU AI Act, which they claim is the world's first comprehensive AI law[5] with a focus on regulating AI in general not just in educational settings. Australia established a taskforce to outline a framework for generative artificial intelligence in schools[6] and in the United States, the Department of Education calls for an AI bill of rights to develop a comprehensive approach towards the adoption of AI in education[7]. Needless to say, these actions must be based on a solid foundation of research and conceptual grounding. Even though there is a vibrant AIEd research community, much of this foundational work is still in development. This tertiary review[8], which is the first of its kind in AIEd, provides the foundation for future conceptualisation and utilisation of AI in higher education.

Contribution of this review

Whilst evidence synthesis is a welcome approach to gaining insight into effective applications of AI in education, there is a risk of 'research waste' in every field of research due to a duplication of efforts, by conducting reviews on the same or similar topics (Grainger et al., 2020; Siontis & Ionnadis, 2018). This can occur when researchers do not give enough consideration to work that has already been published, costing valuable time, effort, and money (Robinson et al., 2021). To help avoid research waste, and to map the state of the AIEd field in higher education (AIHEd), this review is the first to undertake a tertiary review approach (Kitchenham et al., 2009). A tertiary review is a type of research that synthesises evidence from secondary studies, such as systematic reviews, and is sometimes known as a review of reviews or as an overview (Sutton et al., 2019). This method allows researchers to gain an overarching meta view of a field through a systematic process, identifying and analysing types of evidence and key characteristics, exploring how research has been conducted, and identifying gaps in the literature to better guide future field development (Polanin et al., 2017). Given the current interest around the uptake of generative AI, taking stock of existing research, and providing suggestions for where the field might go in the future, is critical at this time.

Research questions

Against this background, the following research question and sub questions guide this review:

1. What is the nature and scope of AIEd evidence synthesis in AIHEd?
    a. What kinds of evidence syntheses are being conducted?
    b. In which conference proceedings and academic journals are AIHEd evidence syntheses published?
    c. What is the geographical distribution of authorship and authors' affiliations?
    d. How collaborative is AIHEd evidence synthesis?
    e. What technology is being used to conduct AIHEd evidence synthesis?
    f. What is the quality of evidence synthesis exploring AIHEd?
    g. What main applications are explored in AIHEd secondary research?
    h. What are the key findings of AIHEd research?
    i. What are the benefits and challenges reported within AIHEd reviews?
    j. What research gaps have been identified in AIHEd secondary research?

---

[5] https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
[6] https://education.nsw.gov.au/about-us/strategies-and-reports/draft-national-ai-in-schools-framework
[7] https://www.ed.gov/news/press-releases/us-department-education-shares-insights-and-recommendations-artificial-intelligence
[8] Otherwise known as a review of reviews (see Kitchenham et al., 2009; Sutton et al., 2019)

Literature Review

Artificial Intelligence in Education (AIEd)

The evolution of AIEd can be traced back several decades, exhibiting a rich history of intertwining educational theory and emergent technology (Doroudi, 2022). As the field matured through the 1990s and into the 2000s, research began to diversify and deepen, exploring varied facets of AIEd such as intelligent tutoring systems (Woolf, 2010), adaptive learning environments (Desmarais & Baker, 2012) as well as supporting collaborative learning environments (Dillenbourg & Jermann, 2007). In the last decade, the synergies between AI technologies and educational practices have further intensified, propelled by advancements in machine learning, natural language processing, and cognitive computing. This era explored innovative applications, including chatbots for student engagement, automated grading and feedback, predictive analytics for student success, and various adaptive platforms for personalised learning. Yet, amid the technological strides, researchers also continued to grapple with persisting challenges and new dilemmas such as ensuring ethical use (Holmes et al., 2021), enhancing system transparency and explainability (Khosravi et al., 2022), and navigating the pedagogical implications of increasingly autonomous AI systems in educational settings (Han et al., 2023).

To gain further understanding of the applications of AI in higher education, and to provide guidance to the field, Zawacki-Richter et al. (2019) developed a typology (see. Fig 1), classifying research into four broad areas; Profiling and prediction, intelligent tutoring systems, assessment and evaluation and adaptive systems and personalisation.

Profiling and Prediction: This domain focuses on employing data-driven approaches to make informed decisions and forecasts regarding students' academic journeys. It includes using AI to optimise admissions decisions and course scheduling, predict and improve dropout and retention rates, and develop comprehensive student models to evaluate and enhance academic achievement by scrutinising patterns and tendencies in student data.

Intelligent Tutoring Systems (ITS): This domain leverages AI to enrich teaching and learning experiences by providing bespoke instructional interventions. The systems work by teaching course content, diagnosing students' strengths and weaknesses and offering automated, personalised feedback, curating appropriate learning materials, facilitating meaningful collaboration among learners, and providing insights from the teacher's perspective to improve pedagogical strategies.

Assessment and Evaluation: This domain focuses on the potential of AI to automate and enhance the evaluative aspects of the educational process. It includes leveraging algorithms for automated grading, providing immediate and tailored feedback to students, meticulously evaluating student understanding and engagement, ensuring academic integrity, and implementing robust mechanisms for the evaluation of teaching methodologies and effectiveness.

Adaptive Systems and Personalisation: This domain explores the use of AI to mould educational experiences that are tailored to individual learners. This involves tailoring course content delivery, recommending personalised content and learning pathways, supporting teachers in enhancing learning design and implementation, utilising academic data to monitor, guide, and support students effectively, and representing knowledge in intuitive and insightful concept maps to facilitate deeper understanding.
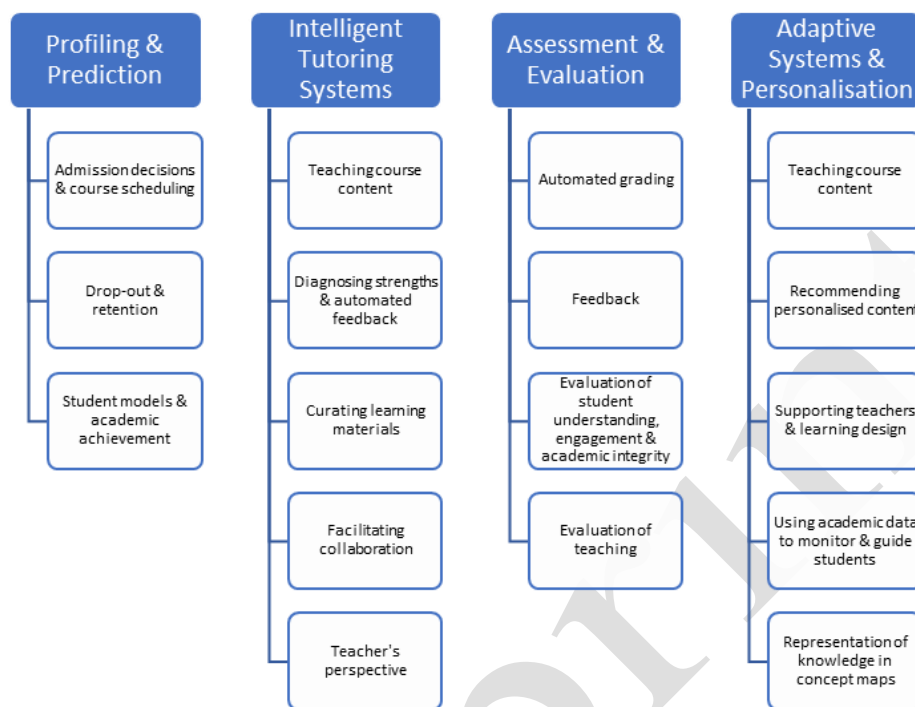
Figure 1. Zawacki-Richter et al.'s (2019) original AIEd typology

Prior AIEd syntheses in higher education

There has been a proliferation of evidence synthesis conducted in the field of EdTech, particularly within the past five years (Zawacki-Richter, 2023), with the rising number of secondary research resulting in the need for tertiary reviews (e.g., Lai & Bower, 2020; Tamim et al., 2011). The interest in AIEd has also been increasing (e.g., Chen et al., 2022), for example the first phase of a systematic review of pedagogical agents by Sikström et al. (2022), included an umbrella review of six reviews and meta-analyses, and Daoudi's (2022) review of learning analytics and serious games included at least four literature reviews. Furthermore, according to Google Scholar[9], the AIHEd review by Zawacki-Richter et al. (2019) has been cited 1100 times since it was published, with the article accessed over 185,000 times and appearing six times in written news stories[10], indicating a wide-ranging public interest in AIHEd.

Prior AIHEd review of reviews have so far also taken place within secondary research (e.g., systematic reviews), rather than as standalone reviews of reviews such as this one. Saghiri et al. (2022), for example, included an analysis of four systematic reviews in their scoping review of AI applications in dental education, de Oliveira et al. (2021) included eight reviews in their systematic review of educational data mining for recommender systems, and Sapci & Sapci (2020) included five reviews in their systematic review of medical education. However, by synthesising both primary and secondary studies within the one review, there is a risk of study duplication, and authors need to be particularly careful to ensure that a primary study identified for inclusion is not also included in one of the secondary studies, to ensure that the results presented are accurate, and the review conducted to a high quality.

---

[9] As of 1st October 2023, https://scholar.google.com/scholar?oi=bibs&hl=en&cites=6006744895709946427
[10] According to the journal website on Springer Open (see Zawacki-Richter et al., 2019)

Evidence synthesis methods

Literature reviews (or narrative reviews) are the most commonly known form of secondary research; however, a range of evidence synthesis methods have increasingly emerged, particularly from the field of health care. In fact, Sutton et al. (2019) identified 48 different review types, which they classified into seven review families (see Table 1). Although part of the traditional review family, literature reviews have increasingly been influenced by the move to more systematic approaches, with many now including method sections, whilst still using the 'literature review' moniker (e.g., Alyahyan & Düştegör, 2020). Bibliometric analyses have also emerged as a popular form of evidence synthesis (e.g., Linnenluecke et al., 2020; Zheng et al., 2022), which analyse bibliographic data to explore research trends and impact. Whilst not included in the Sutton et al. (2019) framework, their ability to provide insight into a field arguably necessitates their inclusion as a valuable form of evidence synthesis.

Table 1. Review families

Traditional review family	Critical review, integrative review, narrative review, narrative summary, state of the art review
Rapid review family	Rapid reviews, rapid evidence assessment, rapid realist synthesis
Purpose specific review family	Content analysis, scoping review, mapping review
Systematic review family	Meta-analysis, systematic review
Qualitative review family	Qualitative evidence synthesis, qualitative meta-synthesis, meta-ethnography
Mixed methods review family	Mixed methods synthesis, narrative synthesis
Review of review family	Review of review, umbrella review

Evidence synthesis quality

It is crucial that any type of evidence synthesis reports the methods used in complete detail (aside from those categorised in the 'traditional review family'), to enable trustworthiness and replicability (Chalmers et al., 2023; Gough et al., 2012). Guidance for synthesis methods have been available for more than a decade (e.g., Moher et al., 2009; Rader et al., 2014) and are constantly being updated as the methodology advances (e.g., Rethlefsen et al., 2021; Tricco et al., 2018). However, issues of quality when undertaking evidence synthesis persist. Chalmers et al. (2023), for example, analysed the quality of 307 reviews in the field of Applied Linguistics against the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidelines (Shamseer et al., 2015), and did find that most of the information expected in any research report were present; background, rationale, objectives, and a conclusion. However, only 43% included the search terms used to find studies, 78% included the inclusion/exclusion criteria, 53% explained how studies were selected, and 51% outlined the data collection process.

Another popular quality assessment tool is the Database of Abstracts and Reviews of Effects (DARE) tool (Centre for Reviews and Dissemination, 1995), which was used by Kitchenham et al. (2009) in a computer science tertiary review; a methodology that has since been heavily adopted by researchers across a range of disciplines, including computer science, social sciences, and education[11]. The authors used the DARE tool to assess the quality of 20 computer science systematic reviews based on four criteria:

1. Are the review's inclusion and exclusion criteria described and appropriate?

---

[11] As of 1st October 2023, it has been cited 2,341 times according to Scopus and 4,508 times according to Google Scholar.

2. Is the literature search likely to have covered all relevant studies?
3. Did the reviewers assess the quality/validity of the included studies?
4. Were the basic data/studies adequately described?

Kitchenham et al. (2009) found that, although only 35% of studies scored 2 out of 4 or lower, few assessed the quality of the primary studies that had been included in the review. The average score overall was 2.6 out of 4, increasing in quality across 2004-2007, with a Spearman correlation of 0.51 ($p < 0.023$).

In the field of EdTech, Lai & Bower (2020) conducted a tertiary review by also adopting Kitchenham et al.'s (2009) quality assessment method, critically analysing 73 reviews to uncover the technologies, themes, general findings, and quality of secondary research that has been conducted. They found that there was very little consistency in how articles were organised, with only six papers (8.2%) explicitly defining quality assessment criteria. The average total quality score was 2.7 out of 4 (SD = 0.59), with only four reviews receiving full marks. There was, however, a slight increase in review quality over time, rising from 2.5 in 2010 to 2.9 in 2018. Likewise, in a tertiary mapping review of 446 EdTech evidence syntheses (Buntins et al., In peer review), 44% ($n = 192$) provided the full search string, 62% ($n = 275$) included the inclusion/exclusion criteria, 37% ($n = 163$) provided the data extraction coding scheme, and only 26% of systematic reviews conducted a quality assessment. Similar findings were reported in an umbrella review of 576 EdTech reviews (Zawacki-Richter, 2023), where 73.4% did not conduct a quality appraisal, and only 8.1% achieved a quality score above 90 (out of 100).

Method

Therefore, in order to map the state of the AIEd field, explore the quality of evidence synthesis conducted, and with a view to suggest future primary and secondary research (Sutton et al., 2019), a tertiary review was conducted (Kitchenham et al., 2009; Lai & Bower, 2020), with the reporting here guided by the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA, Page et al., 2021; see OSF[12]) for increased transparency. As with other rigorous forms of evidence synthesis such as systematic reviews (Sutton et al., 2019), this tertiary review was conducted using explicit, pre-defined criteria and transparent methods of searching, analysis, and reporting (Gough et al., 2012; Zawacki-Richter et al., 2020). All search information can be found on the OSF[13].

Search strategy and study selection
The review was conducted using an iterative search strategy and was developed based on a previous review of research on AIHEd (Zawacki-Richter et al., 2019) and a tertiary mapping review of methodological approaches to conducting secondary research in the field of EdTech (Buntins et al., In peer review). The initial search was conducted on 13 October 2022, with subsequent searches conducted until 18 July 2023 to ensure the inclusion of extant literature (see OSF for search details[14]). The platforms and databases searched were the Web of Science, Scopus, ERIC, EBSCOHost (all databases), IEEE Xplore, Science Direct and ACM Digital Library, as these have been found particularly useful for evidence synthesis (e.g., Gusenbauer & Haddaway, 2019). The OpenAlex platform (Priem et al., 2022) was also searched, which indexes approximately 209 million publications, and was accessed through evidence synthesis software EPPI Reviewer version 6 (Thomas et al., 2023). This included conducting a citation search, bibliography search and bidirectional checking of bi-citations and recommendations on identified included items. Items were also added manually (see Fig. 3) by finding them through ResearchGate or social media throughout

---

[12] https://osf.io/y2afk/?view_only=5dc18e0e4b3846bb9a4d9697c892485e
[13] https://osf.io/y2afk/?view_only=5dc18e0e4b3846bb9a4d9697c892485e
[14] https://osf.io/y2afk/?view_only=5dc18e0e4b3846bb9a4d9697c892485e

the reviewing process until July 2023. Additional searches were conducted in Google Scholar for the terms "artificial intelligence" AND "systematic review" AND "education", with the first 50 returned result pages (500 items) searched for pertinent literature.

Search string

A search string was developed (see Fig. 2) based on the search strings from the two previous reviews (Buntins et al., In peer review; Zawacki-Richter et al., 2019), focusing on forms of AI, formal teaching and learning settings, and variations of evidence synthesis. Whilst some tertiary reviews focus on one form of secondary research (e.g., meta-analyses; Higgins et al., 2012), it was decided to include any form of evidence synthesis as the goal of this review was to map the field, irrespective of the secondary research approach used.

AI	"artificial intelligence" OR "machine intelligence" OR "intelligent support" OR "intelligent virtual reality" OR "chat bot*" OR "machine learning" OR "automated tutor" OR "personal tutor*" OR "intelligent agent*" OR "expert system" OR "neural network" OR "natural language processing" OR "intelligent tutor*" OR "adaptive learning system*" OR "adaptive educational system*" OR "adaptive testing" OR "decision trees" OR "clustering" OR "logistic regression" OR "adaptive system*"
AND	
Education sector	"higher education" OR college* OR undergrad* OR graduate OR postgrad* OR "K-12" OR kindergarten* OR "corporate training*" OR "professional training*" OR "primary school*" OR "middle school*" OR "high school*" OR "elementary school*" OR "vocational education" OR "adult education" OR "workplace learning" OR "corporate academy"
AND	
evidence synthesis	"systematic review" OR "scoping review" OR "narrative review" OR "meta-analysis" OR "evidence synthesis" OR "meta-review" OR "evidence map" OR "rapid review" OR "umbrella review" OR "qualitative synthesis" OR "configurative review" OR "aggregative review" OR "thematic synthesis" OR "framework synthesis" OR "mapping review" OR "meta-synthesis" OR "qualitative evidence synthesis" OR "critical review" OR "integrative review" OR "integrative synthesis" OR "narrative summary" OR "state of the art review" OR "rapid evidence assessment" OR "qualitative research synthesis" OR "qualitative meta-summary" OR "meta-ethnography" OR "meta-narrative review" OR "mixed methods synthesis" OR "scoping study" OR "systematic map"

Figure 2. Tertiary review search string

Inclusion/exclusion criteria and screening

The search strategy yielded 5,609 items (see Fig. 3), which were exported as .ris or .txt files and imported into the evidence synthesis software EPPI Reviewer (Thomas et al., 2023). Following the automatic removal of 449 duplicates within the software, 5,160 items remained to be screened on title and abstract, applying the inclusion and exclusion criteria (see Table 1). Studies were included if they were a form of secondary research on artificial intelligence applications within formal education settings, with an explicit method section and had been published after January 2018. Owing to time and the project scope, studies were only included if they had been published in the English language, and were either a peer-reviewed journal article, book chapter or conference paper. Although reviews have already started being published on the topic of generative AI, and ChatGPT in particular (e.g., İpek et al., 2023; Lo, 2023), the decision was made to exclude these from this sample, as these AI developments arguably represent the next stage of AI evolution in teaching and learning (Bozkurt & Sharma, 2023; Wu et al., 2023).
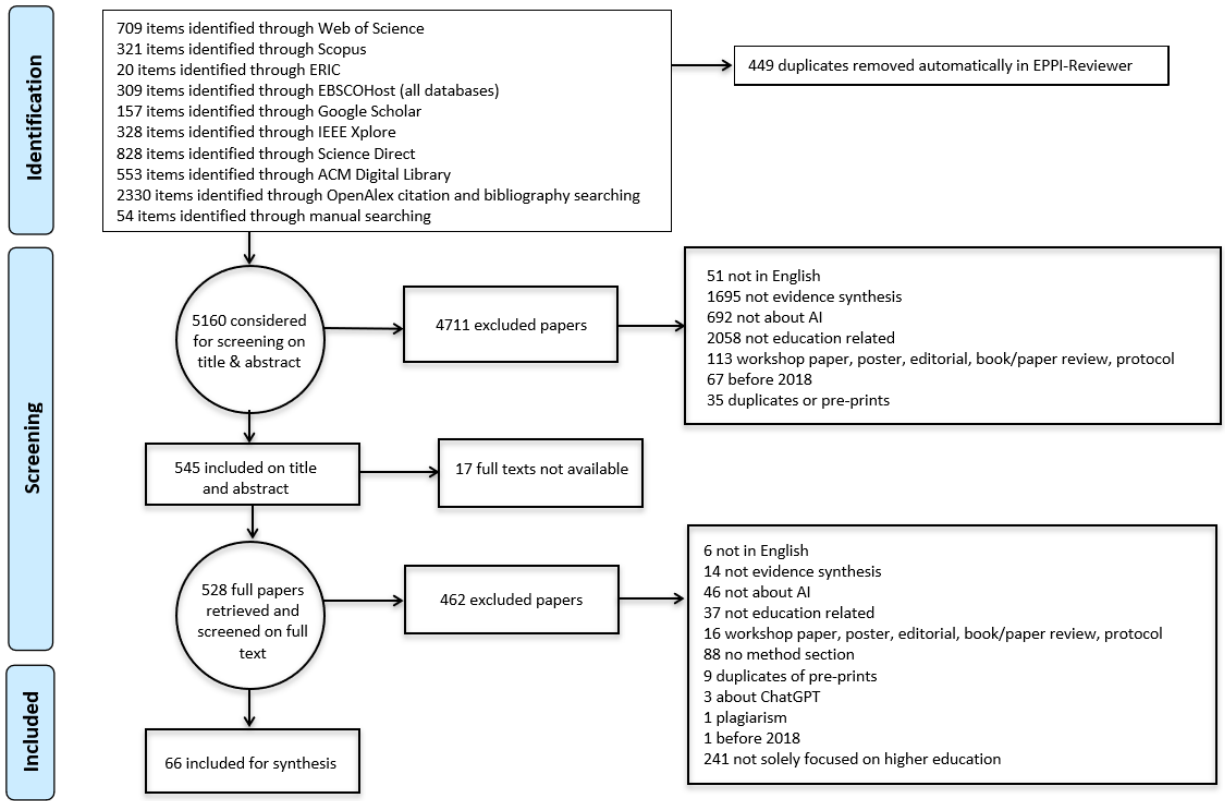
Identification

709 items identified through Web of Science
321 items identified through Scopus
20 items identified through ERIC
309 items identified through EBSCOHost (all databases)
157 items identified through Google Scholar
328 items identified through IEEE Xplore
828 items identified through Science Direct
553 items identified through ACM Digital Library
2330 items identified through OpenAlex citation and bibliography searching
54 items identified through manual searching

449 duplicates removed automatically in EPPI-Reviewer

Screening

5160 considered for screening on title & abstract

4711 excluded papers

51 not in English
1695 not evidence synthesis
692 not about AI
2058 not education related
113 workshop paper, poster, editorial, book/paper review, protocol
67 before 2018
35 duplicates or pre-prints

545 included on title and abstract

17 full texts not available

528 full papers retrieved and screened on full text

462 excluded papers

6 not in English
14 not evidence synthesis
46 not about AI
37 not education related
16 workshop paper, poster, editorial, book/paper review, protocol
88 no method section
9 duplicates of pre-prints
3 about ChatGPT
1 plagiarism
1 before 2018
241 not solely focused on higher education

Included

66 included for synthesis

Figure 3. Meta review PRISMA diagram

Table 2. Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
Published Jan 2018 to 18 July 2023	Published before Jan 2018
Applications of AI in education	Not about artificial intelligence
Formal teaching and learning setting	Informal learning/not formally recognised
Journal articles or conference papers	Editorials, book chapters, meeting abstracts, workshop papers, posters, book reviews, dissertations
Secondary research with a method section	Primary research or literature review with no formal method section
English language	Not in English

To ensure inter-rater reliability between members of the research team, following lengthy discussion and agreement on the inclusion and exclusion criteria by all authors, two members of the team (MB and PP) double screened the first 100 items, resulting in almost perfect agreement (Cohen's k=0.89) (McHugh, 2012). After the two disagreements were reconciled, the remaining 5,060 items were screened on title and abstract by the same authors, resulting in 4,711 items excluded. To continue ensuring inter-rater reliability at the screening on full text stage of 545 studies, three rounds of comparison coding were conducted (50, 30 and 30 items). The same two members of the team (MB and PP) responsible for screening the remaining items, again achieved almost perfect agreement (Cohen's k=0.85) (McHugh, 2012), with 307 evidence syntheses identified across all education levels for data extraction and synthesis. The reviews that only focus on higher education (or also continuing education) were then identified (*n* = 66) and will be the sole

focus of the synthesis in this article. It should be noted that a further 32 reviews were identified that include a focus on higher education in some way (see OSF[15]), i.e., the results are combined with other study levels such as K-12, but it was decided not to include them in this article, to ensure that all results pertain to higher education.

Data extraction

The data extracted for this tertiary review were slightly modified from those used by Buntins et al. (In peer review) and Zawacki-Richter et al. (2019), and included publication and authorship information (e.g. publication type and name, number of authors, author affiliation), review type (as self-declared by the authors and informed by the typology by Sutton et al., 2019), review focus (e.g. AIEd in general or specific type of AI as per Zawacki-Richter et al.'s, 2019 typology), specific educational and participant context (e.g. undergraduates, Health & Welfare), methodological characteristics (e.g. databases used and number of included studies), key findings and research gaps identified (see OSF[16] for the full coding scheme). All data were extracted manually and input into EPPI Reviewer (Thomas et al., 2023), including author affiliations and countries, owing to issues identified in EdTech research with missing metadata in the Web of Science (Bond, 2018). Where the author information was not directly provided on either the PDF or the journal website, the code 'Not mentioned' was assigned. An initial five studies were coded by all authors, to ensure agreement on the coding scheme, although the key findings and research gaps were coded inductively.

To answer sub-question 1f about the quality of AIHEd secondary research, the decision was made to use the Database of Abstracts and Reviews of Effects (DARE) tool (Centre for Reviews and Dissemination, 1995), which has been used in previous tertiary reviews (e.g., Kitchenham et al., 2009; Tran et al., 2021). Although the authors acknowledge the AMSTAR 2 tool as an effective quality assessment tool for systematic reviews (Shea et al., 2017), the present review includes any kind of evidence synthesis, as long as it has a method section. Therefore, the decision was made to use a combination of four DARE criteria (indicated by D; as used by Lai & Bower, 2020), alongside items from the AMSTAR 2 tool, and further bespoke criteria, as developed by Buntins et al. (In peer review):

1. Are there any research questions, aims or objectives? (AMSTAR 2)
2. Were inclusion/exclusion criteria reported in the review and are they appropriate? (D)
3. Are the publication years included defined?
4. Was the search adequately conducted and likely to have covered all relevant studies? (D)
5. Was the search string provided in full? (AMSTAR 2)
6. Do they report inter-rater reliability? (AMSTAR 2)
7. Was the data extraction coding scheme provided?
8. Was a quality assessment undertaken? (D)
9. Are sufficient details provided about the individual included studies? (D)
10. Is there a reflection on review limitations?

The questions were scored as per the adapted method used by Kitchenham et al. (2009, p. 9) and Tran et al. (2021, Figure S1). The scoring procedure was Yes = 1, Partly = 0.5 and No = 0 (see Fig. 3). However, it should be noted that certain types of evidence synthesis do not always need to include a quality assessment (e.g., scoping, traditional literature, and mapping reviews, see Sutton et al., 2019) and so these were coded as 'not applicable' (N/A) in the coding scheme and scored 1. It should also be noted that the quality appraisal was also not used to eliminate studies from the corpus in this case, but rather to answer one of the sub research questions. Due to this, a quality indicator was used in the inclusion/exclusion criteria instead, namely if a review did not have an identifiable method section it would be excluded, as it was reasoned that

---

these were not attempting to be systematic at all. An overall score was determined out of 10 and items determined as critically low (0-2.5), low (3-4.5), medium (5-7), high (7.5-8.5) or excellent (9 – 10) quality; a similar approach used by other reviews (e.g., Urdaneta-Ponte et al., 2021).

Criterion	Score	Interpretation
RQs	Yes	RQs, aims or objectives are explicitly and clearly defined.
	Partly	Some mention of objectives or aims are alluded to.
	No	No RQs, aims or objectives are identifiable.
Inclusion/exclusion	Yes	The criteria used are explicitly defined in the paper.
	Partly	The criteria are implicit.
	No	The criteria are not defined and cannot be readily inferred.
Publication years	Yes	The publication years are clearly stated, e.g. 2010-2020.
	Partly	The publication years state from a year OR until a year.
	No	The publication years are not defined at all.
Search coverage	Yes	4 or more digital libraries searched and included additional search strategies (e.g. snowballing) OR identified and referenced all pertinent journals.
	Partly	3 or 4 digital libraries searched with no extra search strategies OR searched a defined but restricted set of journals and conference proceedings.
	No	2 digital libraries searched or an extremely restricted set of journals.
Search string	Yes	The search string was reported in full and is replicable.
	Partly	Examples of keywords only were given.
	No	The search terms were not provided in any form.
Inter-rater reliability	Yes	An inter-rater reliability value is reported (e.g. Cohen's kappa).
	Partly	Mention is made of how disagreements were reconciled.
	No	No inter-rater reliability is mentioned.
Data extraction	Yes	The full coding scheme was provided.
	Partly	Examples are provided, but not the full list.
	No	The coding scheme was not provided at all.
Quality assessment	Yes	Explicitly defined quality criteria extracted for each study.
	Partly	The research question involved quality issues that are addressed by the study.
	No	No explicit quality assessment of individual papers has been attempted.
Study description	Yes	Information is presented about each paper.
	Partly	Only summary information is presented about individual papers.
	No	The results for individual studies are not specified.
Review limitations	Yes	Yes, there is a specific identifiable limitations section.
	Partly	There is some mention of limitations.
	No	There is no limitations section or reflection on limitations

Figure 3. Quality assessment criteria

To answer sub-questions 1g and 1h, the evidence syntheses in the corpus were coded using Zawacki-Richter et al.'s (2019) typology of Profiling and Prediction, Assessment and Evaluation, Adaptive Systems and Personalisation, and Intelligent Tutoring Systems as a starting point. Studies were coded as 'General AIEd' if they claimed to be searching for any applications of AI in education (e.g., Chu et al., 2022). It should also be noted that, whilst reviews might have said they were focused on 'General AIEd' and were therefore coded as such under 'Focus of AI review', their findings might have focused specifically on 'Assessment

and Evaluation' and 'Intelligent Tutoring Systems', which were then coded as such under 'AI Topics and Key Findings'. For example, Alkhalil et al.'s (2021) mapping review of big data analytics in higher education was coded as 'Profiling and Prediction' and 'Adaptive Systems and Personalisation' under 'Focus of AI review', but they also discussed the use of big data in evaluating teachers and learning material to aid quality assurance processes, which meant that their results were also coded under 'Assessment and Evaluation' in the 'AI Topics and Key Findings' section of the data extraction coding tool.

Data synthesis and interactive evidence & gap map development

A narrative synthesis of the data was undertaken (Petticrew & Roberts, 2006), including a tabulation of the included studies (see Appendix A), to provide an overview of the AIHEd field. Further tables are provided throughout the text, or included as appendices, accompanied by narrative descriptions. To provide further visual overviews, and to provide publicly accessible resources to the field beyond that which this article can provide, interactive evidence and gap maps were produced for each research question, using the EPPI Mapper application (Digital Solution Foundry & EPPI Centre, 2023). To do this, a JSON report of all included studies and associated coding were exported from EPPI Reviewer (Thomas et al., 2023) and imported into the EPPI Mapper application, where display options were chosen. The HTML files were then uploaded to the project page and are available to access and download[17]. An openly accessible web database of the included studies is also available[18], which allows users to view the data in an interactive way through crosstabulation and frequency charts, with direct links to included studies, as well as to save and export the data. This was created using the EPPI Visualiser app, which is located within EPPI Reviewer[19].

Limitations

Whilst every attempt was made to conduct this meta review as rigorously and transparently as possible, there are some limitations that should be acknowledged. Firstly, the protocol was not pre-registered within an official systematic review repository, such as Prospero[20], as this is not a medical study and is a review of reviews. However, all search information is openly accessible on the OSF[21] and in the future, the authors will make use of an organisation such as the International Database of Education Systematic Reviews[22], which is now accepting protocols from any education discipline. Only the first 500 records in Google Scholar were considered, as opposed to the 1000 records recommended by Haddaway et al. (2015), although OpenAlex was also used to supplement this. Further individual academic journals could also have been manually searched, such as *Computers & Education: Artificial Intelligence*, as well as literature published in languages other than English, to reduce language bias (Stern & Kleijnen, 2020). Furthermore, the quality assessment tool that was developed is not perfect, and it could be argued that the distance between yes, no, and partly cannot be assumed to be equal. However, the two tools are widely used, and this approach has been used in the field previously (e.g., Kitchenham et al., 2009; Tran et al., 2021). Choosing not to include reviews synthesising research on generative AI is certainly a limitation, however given that primary and secondary research on the topic are currently only at a nascent stage, it would be preferable to see the field develop somewhat first, so that effective insights can be drawn.

---

[17] https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3917
[18] https://eppi.ioe.ac.uk/eppi-vis/login/open?webdbid=322
[19] For more information about EPPI Mapper and creating interactive evidence gap maps, as well as using EPPI Visualiser, see https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3790
[20] https://www.crd.york.ac.uk/PROSPERO/
[21] https://osf.io/y2afk/?view_only=5dc18e0e4b3846bb9a4d9697c892485e
[22] https://idesr.org/

Findings

General publication characteristics

Of the 66 evidence syntheses solely focused on AIEd in higher education (AIHEd) identified, the majority were published as journal articles (81.8%, $n = 54$), as opposed to conference papers ($n = 12$), but only 57.6% are available open access[23]. Although there has been an exponential growth in the interest in AIEd (Chen et al., 2022; OECD, 2023), there was a slight reduction in the number published in 2020 before rising again (see Fig. 5)[24]. This is likely due to the impact of the COVID-19 pandemic, and it is interesting to note that 12 had already been published in 2023 up to mid-July.
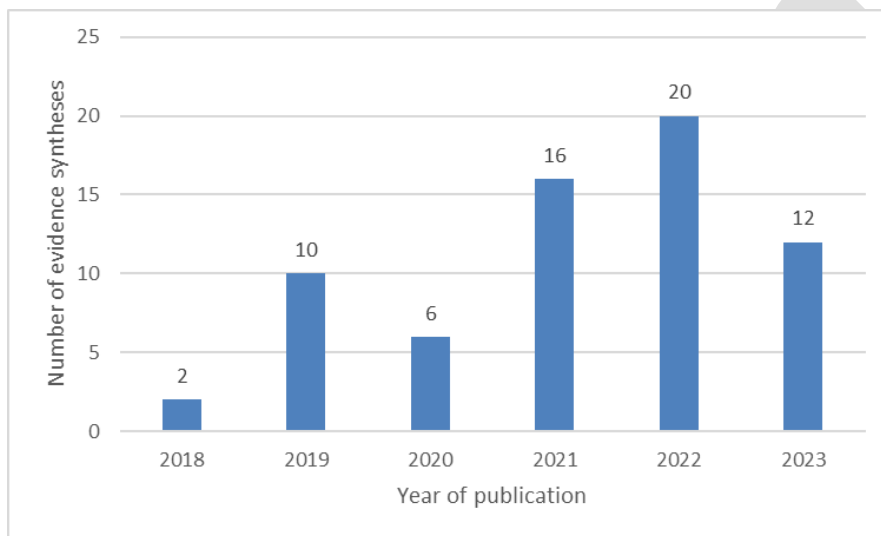

Figure 5. Number of higher education evidence syntheses published by year

Although many reviews synthesised research across multiple settings, there were a small number that focused on AIHEd in specific disciplines or with particular groups of participants, for example Health & Welfare ($n = 14$), STEM ($n = 4$), online or blended learning ($n = 5$), foreign language learning ($n = 2$), pre-service teachers (Salas-Pilco et al., 2022), students with disabilities (Fichten et al., 2021), and undergraduate students (Lee et al., 2021). Six evidence syntheses had a specific geographical focus, with three centred on research conducted within specific countries: India (Algabri et al., 2021; Bhattacharjee, 2019) and Saudi Arabia (Alotaibi & Alshehri, 2023). The other three focused on research from within the regions of Africa (Gudyanga, 2023; Maphosa & Maphosa, 2020) and Latin America (Salas-Pilco & Yang, 2022).

What kinds of evidence syntheses are being conducted in AIHEd?

There were eight different types of evidence syntheses conducted in AIHEd (see Appendix B), as identified by their authors. Systematic literature reviews were by far the most popular type, accounting for two thirds of the corpus (66.7%, $n = 44$), followed by scoping reviews (12.1%, $n = 8$). There were two reviews where authors conducted both a systematic review and a meta-analysis (Fahd et al., 2022; Fontaine et al., 2019),

---

[23] See Appendix A for a tabulated list of included study characteristics and https://eppi.ioe.ac.uk/eppi-vis/login/open?webdbid=322 for the interactive web database.
[24] It should be noted that Cardona et al. (2023) was originally published in 2020 but has since been indexed in a 2023 journal issue. Their review has been kept as 2020.

and two reviews where authors identified their work as a mapping review and a systematic review (del Gobbo et al., 2023; Zhong, 2022).

In which conferences and academic journals are AIHEd evidence syntheses published?
AIHEd evidence syntheses were published in 42 unique academic journals and 11 different conference proceedings (see Appendix C). The top conference was the International Conference on Human-Computer Interaction ($n = 2$), with all other conferences publishing one paper each. The top seven journals were *Education and Information Technologies* ($n = 4$), *International Journal of Educational Technology in Higher Education* ($n = 4$), *Education Sciences* ($n = 3$), *Interactive Learning Environments* ($n = 2$), *Technology, Knowledge and Learning* ($n = 2$), *Sustainability* ($n = 2$), and *JMIR Medical Education* ($n = 2$). All of these journals have published systematic reviews (see Appendix D), although other types have been published as well, with the exception of *Technology, Knowledge and Learning* and *Sustainability*.

What are AIHEd evidence synthesis authors' institutional and disciplinary affiliations?
The AIHEd evidence syntheses in this corpus were written by authors from 110 unique institutions, with the top seven most productive institutions located in five different continents (see Appendix E). The most productive institution in each continent were the University of Toronto (North America, n = 5), The Independent Institute of Education (Africa, n = 3), Central China Normal University and Fu Jen Catholic University (Asia, n = 2 each), Sultan Qaboos University (Middle East, n = 2), and the University of Newcastle (Oceania, n = 2). The European and the South and Central American institutions all had one publication each.

Although Crompton & Burke (2023) have reported a rise in the number of Education affiliated authors in AIEd primary research, more than half of evidence synthesis in this corpus have been published by first authors from STEM affiliated backgrounds (56.1%), with Computer Science & IT (30.3%, n = 20) authors the most prolific (see Appendix F). Education affiliated authors do still represent 25.8%, which is encouraging, and six publications did not mention the disciplinary affiliation of their authors. Researchers from Education and Computer Science & IT have published more of a range of evidence synthesis than the other disciplines, although still with a heavy skew towards systematic reviews (71% and 75% respectively). Another interesting finding is that Health, Medical & Physical Education researchers have published twice as many scoping reviews (n = 7) as they have systematic reviews (n = 3) in this corpus, which may perhaps be due to the longer history of evidence synthesis in that discipline (Sutton et al., 2019).

What is the geographical distribution of AIHEd evidence synthesis authorship?
The authorship of AIHEd secondary research has been quite evenly spread between authors from North America (27.3%), Europe (24.2%) and Asia (22.7%), followed by the Middle East (13.6%; see Appendix G). In line with previous EdTech research (e.g., Bond et al., 2019), there was far less representation from South and Central America (4.5%). Authorship was spread across 32 different countries (see Appendix I), with arguably less dominance by the United States than two other recent EdTech tertiary reviews (Buntins et al., In peer review; Zawacki-Richter, 2023) have found. Whilst it was the most productive country (see Table 3), the United States was closely followed by Canada and Australia. Furthermore, all continents aside from South and Central America are represented in the top nine most productive countries.

Table 3. Top nine most productive countries

Rank	Country	Count	Percentage
1	United States	11	16.7%
2	Canada	9	13.6%
3	Australia	7	10.6%
4	South Africa	6	9.1%
5	China	5	7.6%
6	Saudi Arabia	4	6.1%
=	Spain	4	6.1%
7	Germany	3	4.5%
=	India	3	4.5%

When the geographical distribution is viewed by evidence synthesis type (see Appendix H), researchers in Africa, North America, Oceania, the Middle East, and Europe have used a wider range of secondary research approaches, although European and Oceanian authors have heavily favoured systematic reviews (75%).

How collaborative is AIHEd evidence synthesis?

AIHEd evidence synthesis is almost always published collaboratively (89.4%, $n = 59$), particularly in teams of two, three or four researchers (see Appendix I), with 21 authors of a scoping review the largest number in one publication (Charow et al., 2021). African and Middle Eastern researchers have published more as single authors (29% and 22% of publications from those regions). Co-authorship, however, tends to occur in domestic collaborations (71.2%), with only 18.2% of publications internationally co-authored. Rates of domestic co-authorship are particularly high in Oceania (75%) and Europe (69%). The highest rate of international research collaboration is found in South & Central America and the Middle East (33% of cases respectively). Bibliometric reviews (50%), integrative reviews (50%) and meta-analyses (33%) have the highest rates of international co-authorship, although these are also some of the lowest numbers of evidence synthesis produced. Interestingly, systematic reviews are almost exclusively undertaken by researchers located within the same country (70.5%), with all eight scoping reviews published by domestic research collaborations.

What technology is being used to conduct AIHEd evidence synthesis?

51.5% of reviews ($n = 34$) did not report using any kind of digital evidence synthesis tool in their article to conduct their review (see Appendix J) and of those that did, only 12.1% ($n = 8$) reported using some kind of evidence synthesis software, which have integrated machine learning functionality (e.g., deduplication, priority screening, snowball searching) to assist in making the review process more transparent and efficient. The most popular of these were EPPI Reviewer ($n = 3$)[25] and Covidence ($n = 3$)[26]. AIHEd secondary researchers have mostly used spreadsheets (16.7%) and reference management software (16.7%) to manage their reviews, with authors of critical reviews, literature reviews and systematic reviews the least likely to report whether a tool was used at all.

---

[25] https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4&
[26] https://www.covidence.org/

AIHEd evidence synthesis quality

The AIHEd reviews in the corpus were assessed against 10 quality assessment criteria (see Table 4), based on the DARE (Centre for Reviews and Dissemination, 1995; Kitchenham et al., 2009) and AMSTAR 2 (Shea et al., 2017) tools, as well as the method by Buntins et al. (In peer review). Almost all studies provided explicit information about their research questions, aims or objectives (92.4%), the inclusion/exclusion criteria (77.3%) and the publication years of literature included in the review (87.9%). Whilst 68.2% of reviews provided the exact search string used, there were still 25.8% ($n = 17$) that only provided some of the words used to find the included studies. The most concerning findings were that 31.8% of studies only searched in one or two databases, 51.5% did not report anything about inter-rater reliability or how screening and coding decisions were decided between review teams, only 24.2% provided their exact data extraction coding scheme, 45.5% did not undertake any form of quality assessment, and 34.8% did not reflect at all upon the limitations of their review.

Table 4. Quality assessment for AIHEd corpus (n = 66)

Criteria	Yes	Partly	No	N/A
Are there any research questions, aims or objectives?	92.4%	6.1%	1.5%	
Were inclusion/exclusion criteria provided in the method section?	77.3%	19.7%	3.0%	
Are the publication years included defined in the title, abstract or method section?	87.9%	7.6%	4.5%	
Was the search adequately conducted and likely to have covered all relevant studies?	51.5%	16.7%	31.8%	
Was the search string reported in full?	68.2%	25.8%	6.1%	
Do they report inter-rater reliability?	22.7%	21.2%	51.5%	4.5%[27]
Is the data extraction coding schema provided?	24.2%	45.5%	30.3%	
Is some form of quality assessment applied?	15.2%	13.6%	45.5%	25.8%[28]
Are sufficient details provided about the individual included studies?	28.8%	42.4%	21.2%	7.6%[29]
Is there a reflection on review limitations?	40.9%	24.2%	34.8%	

The reviews were given an overall quality assessment score out of 10 (see Fig. 6), averaging 6.57 across the corpus. Looking at the quality over time (see Appendix K), it is encouraging to see that the percentage of 'critically low' and 'low quality' studies being conducted appears to be reducing. Meta-analyses and scoping reviews were predominantly coded as 'high quality' or 'excellent quality', with far more variability in the quality of systematic reviews. Conference papers were lower quality than journal articles, with only 8% of conference papers receiving a 'high quality' rating and none receiving 'excellent quality'. This may, however, be partially owing to the limitations on word count that conference proceedings impose. For example, the most prolific conference in this corpus, the Human Computer Interaction Conference[30], accepts paper submissions of up to 20 pages including references. Given the often-lengthy reference list required by an evidence synthesis paper, this restricts the depth of information that can be provided.

To gain greater insight into methodological quality, each review was coded on whether a specific method or approach was followed (see Appendix K). Although 18.2% ($n = 12$) of publications did not provide a reference to a specific approach followed, including some that said they followed the PRISMA guidelines (e.g., Page et al., 2021) but did not cite them, 29 different publications were referenced. Of these, the

---

[27] Bibliometric analyses
[28] Bibliometric analyses, literature reviews, integrative reviews, scoping reviews, mapping reviews
[29] Bibliometric analyses
[30] https://2024.hci.international/papers

original (Liberati et al., 2009; Moher et al., 2009) and the updated PRISMA guidelines (Moher et al., 2015; Page et al., 2021) were referenced as a primary approach by 33.3% (*n* = 22), not including the scoping review PRISMA-S guidelines (Tricco et al., 2018) in a further four. However, authors from an Education disciplinary background were slightly more likely to use PRISMA than those from Computer Science, who preferred to follow the guidance of Kitchenham and colleagues (Kitchenham, 2004; Kitchenham & Charters, 2007; Kitchenham et al., 2009; Kitchenham et al., 2010).
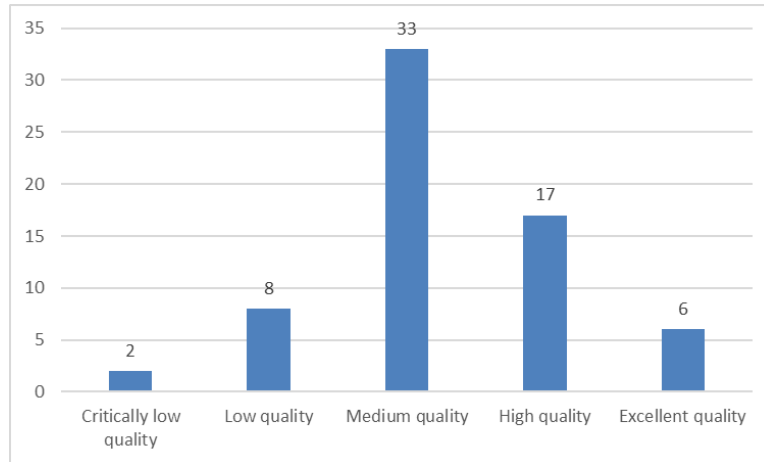

Figure 6. Overall quality assessment

AIEd applications in higher education

The reviews were categorised using Zawacki-Richter et al.'s (2019) classification (profiling and prediction; intelligent tutoring systems; adaptive systems and personalisation; assessment and evaluation; see Fig. 1), depending upon their purported focus within the title, abstract, keywords or search terms, with any reviews not specifying a particular focus categorised as 'General AIEd' (see Table 5). Most of the reviews (47%, *n* = 31) fell under the latter category and explored a range of AI applications. This was followed by reviews focusing on profiling and prediction (e.g., Abu Saa et al., 2019) and adaptive systems and personalisation (e.g., Fontaine et al., 2019). Reviews focused specifically on assessment and evaluation (e.g., Banihashem et al., 2022) and intelligent tutoring systems (e.g., Crow et al., 2018) were rare.

Table 5. AI applications as the primary focus across reviews (n = 66)

Review focus	*n*	%
General AIEd	31	47.0%
Profiling and Prediction	19	28.8%
Adaptive Systems and Personalisation	18	27.3%
Assessment and Evaluation	3	4.5%
Intelligent Tutoring Systems	1	1.5%

Key findings in AIEd higher education evidence synthesis

The student life-cycle (Reid, 1995) was used as a framework to identify AI applications at the micro level of teaching and learning (e.g., tutoring, assessment), as well as at the institutional and administrative level (e.g., admissions, graduation). Almost all the reviews included research focused on academic support services at the teaching and learning level (n = 64, 97.0%)[31], with only 39.3% (n = 26) specifically locating research relating to institutional and administrative services, all of which also included research at the micro level. A lower level of focus on administration was also found by Crompton & Burke (2023), where only 11% of higher education research focused on managers, despite AI being useful for personalising the university experience for students in regards to admissions, examinations and library services (Algabri et al., 2021; Zawacki-Richter et al., 2019), exploring trends across large datasets (Zhang et al., 2023), and for quality assurance (Kirubarajan et al., 2022; Manhiça et al., 2022; Rabelo et al., 2023).

The key findings of the reviews were also classified into the four main thematic AI application areas (see Table 6). More than half of the reviews (54.5%) discussed applications related to adaptive systems and personalisation, closely followed by profiling and prediction (48.5%). The key findings will now be synthesised, according to the four application topics (see Fig. 1).

Table 6. Number of AI applications in key findings across studies (n = 66), multiple mentions possible

AI applications	n	%
Adaptive Systems and Personalisation	36	54.5%
Profiling and Prediction	32	48.5%
Assessment and Evaluation	26	39.4%
Intelligent Tutoring Systems	14	21.2%

Adaptive Systems and Personalisation

All the reviews on adaptive systems (n = 36) are situated at the teaching and learning level, with only 12 reviews (33.3%) reporting findings for the administrative and institutional level. The 36 reviews were then further classified into six subcategories: chatbots/virtual assistants (n = 20), providing personalised content (n = 14), facial recognition/mood detection (n = 9), recommender systems/course scheduling (n = 5), and robots (n = 3). Li et al.'s (2021) review also focused on the challenges currently faced by adaptive learning research, which were classed into technical and privacy issues. They found that research is still at a nascent stage, with a gap between theory and practice, and that further interdisciplinary approaches are needed, alongside the collection and sharing of massive data that adheres to privacy considerations. Andersen et al.'s (2022) scoping review of adaptive learning in nursing education suggests that further attention also needs to be paid to learning design, alongside further qualitative research.

Chatbots/virtual assistants

Chatbots were referred to in various ways throughout the literature, including as virtual assistants, virtual agents, intelligent helpers, voice assistants, conversational agents, and intelligent helpers (Chaka, 2023;

---

[31] Two bibliometric studies (Gudyanga, 2023; Hinojo-Lucena et al., 2019) focused on trends in AI research (countries, journals etc) and did not specify particular applications.

Crompton & Burke, 2023). Virtual patient apps have become increasingly used within nursing, dental and medical contexts (Buchanan et al., 2021; Frangoudes et al., 2021; Harmon et al., 2021; Saghiri et al., 2022; Zhang et al., 2023), with Hwang et al.'s (2022) review of 112 AI-supported nursing education articles finding that intelligent agents were the most used AI system (53% of studies). Research measured the effectiveness of chatbots on student learning outcomes, critical thinking, empathy, communication skills and satisfaction (Chaka, 2023; Frangoudes et al., 2021), with a review of English as a foreign language literature (Klímová & Ibna Seraj, 2023) finding chatbots having a particularly positive influence on developing speaking skills (intonation, stress, and fluency), possibly in part due to feelings of reduced anxiety (Zhai & Wibowo, 2023). Virtual assistants can be particularly useful to enhance accessibility for visually and hearing-impaired students, through automatic speech recognition, text to speech and sign language interpretation (Fichten et al., 2021), as well as to help detect anxiety and depressive symptoms in students (Salas-Pilco & Yang, 2022). There is potential to use chatbots in a more institution-wide role, for example to collate opinions about teaching and the institution (Sourani, 2019) or to scale mentoring of students who are on field placements (Salas-Pilco et al., 2022), and one review found that students prefer chatbots to other communication methods (Hamam, 2021). Further development is suggested on the evaluation of chatbots, such as their effectiveness on affective and social aspects of learning (Algabri et al., 2021; Frangoudes et al., 2021).

Providing personalised content
The use of personalised learning through adaptive systems was identified in 14 reviews, which particularly highlighted the benefits of customising learning to support students (e.g., Algabri, 2021), although Fontaine et al.'s (2019) meta-analysis of 21 Health & Welfare studies found that adaptive learning only had a statistically significant effect on learning skills, rather than on building factual knowledge. Fariani et al.'s (2022) review of 39 personalised learning studies found that personalised teaching materials were the most widely used (49%), followed by learning paths (29%), learning strategies (17%) and learning environments (5%), with 49% using machine learning algorithms and 51% measuring the impact of implementing personalisation on learning. Zhong's (2022) review of 41 studies found that 54% used learning traits to structure learning content, with macro the most popular sequencing approach (24%). Further studies are needed to explore how personalisation impacts affective aspects such as motivation, engagement, and interest (Alamri, 2021; Fariani et al., 2022), with primary research needing to provide more explicit information about the algorithms and architecture used (Fontaine et al., 2019).

Facial recognition/mood detection/neurophysiological
Five studies (10%) in Kirubarajan et al.'s (2022) scoping review used motion tracking systems to assess kinematics while students operated. Face tracker software has been used to manage student attendance (Salas-Pilco & Yang, 2022), determine whether students are accurately interpreting ECGs (Zhang et al., 2023), and to analyse students' emotions during clinical simulations, which can enable educators to tailor simulations to student needs more effectively (Buchanan et al., 2021). Li et al. (2021) explored a large amount of adaptive learning literature and suggest that research providing real insight into understanding students' psychological emotions and cognition is currently at a nascent stage. However, Darvishi et al. (2022) suggest that neuro measurements can help fill this gap by providing further insight into learner mental states and found that facial measurements had a higher adoption rate than EEGs, although cognitive constructs were measured in more EEG studies. 66% ($n = 6$) of the reviews reporting the use of neurophysiological AI, stressed the need for further ethical considerations when undertaking such research

in the future, including obtaining participant consent (Salas-Pilco & Yang, 2022), more transparent development of AI and clearer reporting of study design (Kirubarajan et al., 2022). Darvishi et al. (2022) suggested that propensity-score matching could be used in order to conduct quasi-experimental studies more ethically.

Recommender systems/course scheduling

Five reviews located studies on the use of recommender systems (RSs) in higher education, including Rabelo et al. (2023), who argue that administrators could make more use of RSs to help retention, including recommending subjects and courses. Banihashem et al. (2022)'s systematic review on the role of learning analytics to enhance feedback reported a few studies where systems had guided students and recommended course material, Zawacki-Richter et al. (2019) found three studies, including one suggesting pedagogical strategies for educators (see Cobos et al., 2013), and Manhiça et al. (2022) found two studies referring to RSs in their systematic review, although unfortunately further details were not provided, possibly owing to the conference paper size restraints. Providing much needed insight, Urdaneta-Ponte et al.'s (2021) systematic review focused solely on recommender systems in HE and included 98 studies. The most commonly used development techniques were collaborative filtering, followed by RSs that combine different techniques. Most RSs suggested learning resources (37.76%) and courses (33.67%), with very few studies looking at RSs for other purposes. 78% of studies focused on students, and therefore future research could explore the perceptions of educators and other stakeholders. Urdaneta-Ponte et al. (2021) suggest that further investigation is needed of algorithms that are based on a semantic approach, as well as further development of hybrid systems. They also suggest that user information could be explored along with information from different sources, such as social media, to build more complete profiles, however this raises questions about the ethics of obtaining this kind of data.

Robots

Only three reviews mentioned the use of robots within higher education. In Chaka's (2023) literature review, 38% ($n = 10$) of studies focused on the use of robotics, the majority of which explored how robots could be used to enhance the teaching and learning of undergraduate students, with one study exploring the use of a robot-assisted instructional package to help teach students with intellectual disabilities how to write messages (see Pennington et al., 2014). Five studies (18.5%) in Buchanan et al.'s (2021) nursing scoping review pertained to robots, with one study suggesting that there would be an increased presence of humanoid robots and cyborgs in the future to complement high-fidelity simulators. Maphosa & Maphosa (2021) called for further primary research on the development and application of intelligent robots, although Chaka (2023) pointed out that barriers to further HE implementation will need to be overcome, including challenges with infrastructure and technology, educator acceptance, and curricula being "robotics-compliant" (p. 34).

Profiling and Prediction

All of the reviews pertaining to profiling and prediction included a focus on teaching and learning ($n = 32$), with just over half ($n = 17$, 53.1%) detailing examples of AI support at the administrative level. The 32 reviews were further classified into six subcategories: dropout/retention ($n = 25$), academic achievement/learning outcomes ($n = 24$), admissions/timetabling ($n = 6$), career paths/placement ($n = 4$), student satisfaction ($n = 3$), and diagnostic prediction ($n = 3$).

Dropout/retention

The use of AI to predict dropout and/or assist with student retention was mentioned as a focus in 25 reviews (37.9%). While implementing a system to detect students at risk of failing is seen as a popular goal (Liz-Domínguez et al., 2019), and that AI's recommendation systems can predict student outcomes with high accuracy (Maphosa & Maphosa, 2021), there is limited evidence that dropout prediction models are being effectively used within institutions (McConvey et al., 2023). From a data utilisation perspective, Li et al. (2022) indicate that factors such as individual characteristics, family background, gender, age, and race/ethnicity play a significant role in influencing student motivation and persistence. Additionally, they report that a student's prior knowledge holds notable significance in determining dropout rates and performance (Cardona et al., 2023). McConvey et al. (2023) suggest that several research endeavours have successfully incorporated students' social media activity, financial data, interactions with learning management systems, and student survey responses in their predictive models. They further emphasise that demographic data and LMS activity are commonly used as predictors in this realm. In terms of algorithms, several reviews (e.g., Fahd et al., 2022; Hellas et al., 2018) report that classifiers are preferred over regression algorithms, especially for dropout and failure risks, as the outputs are categorical variables. In terms of broader implications and considerations, there is an emphasis on combining human assessment with machine learning for more nuanced decision-making (Manhiça et al., 2022) and that care should be taken when using protected attributes (like minority status or family income) in predictions, as they might amplify existing inequities in higher education (McConvey et al., 2023).

Academic achievement/learning outcomes

24 reviews reported findings associated with predicting academic performance, course selection, course completion, engagement, and academic success, including many reviews looking generally at AIHEd finding high numbers of studies focused on performance prediction to support learning outcomes (e.g., Ouyang et al., 2022; Sghir et al., 2022). Seven reviews purely focused on the use of AI to predict academic performance in HE (Abu Saa et al., 2019; Fahd et al., 2022; Ifenthaler & Yau, 2020; Zulkifli et al., 2019), with some reviews specialising in specific disciplines (STEM; Hellas et al., 2018; Moonsamy et al., 2021) and study levels (undergraduates; Alyahyan & Düştegör, 2020). The features commonly used for prediction can be categorised into demographic (age, gender, etc.), personality (self-efficacy, self-regulation, etc.), academic (previous performance, high school performance, etc.), behavioural (log data, engagement), and institutional (teaching approach, high school quality) (Abu Saa et al., 2019). Alyahyan & Düştegör (2020) report that prior-academic achievement, student demographics, e- learning activity and psychological attributes are the most common factors reported and that the top two factors (prior academic achievement and student demographics) were present in 69% of included literature. Hellas et al. (2018) discovered that the techniques used to predict academic outcomes can be categorised into several approaches: Classification (utilising supervised learning like Naive Bayes and Decision Trees), Clustering (which involves unsupervised learning and data partitioning), Statistical methods (such as correlation and regression), Data mining (for pinpointing features and trends), and other assorted techniques. The review highlighted that linear regression models are commonly used for predictive purposes. In contrast, classification methods often involve comparing different algorithms, resulting in a range of predictive outcomes. Future research should ensure that a detailed description is provided on what is being predicted, how and why (Hellas et al., 2018), could be deepened by more diverse study design, such as longitudinal and large-scale studies (Ifenthaler & Yau, 2020) with multiple data collection techniques (Abu Saa et al., 2019), in a more diverse

array of contexts (e.g., Fahd et al., 2022; Sghir et al., 2022), especially developing countries (e.g., Pinto et al., 2023).

Admissions/timetabling

The use of AI to assist with admissions, course booking behaviour, timetabling, and thesis allocation have seen significant advances in higher education, which was reported in six reviews (9.1%), although they only reported on a considerably small number of studies; for example, Zawacki-Richter et al. (2019) found seven studies (4.8%), Sghir et al. (2022) found three studies (4.1%), and Otoo-Arthur & van Zyl (2020) two studies (3.6%). Alam & Mohanty (2022) suggest that applications can be sorted with a 95% accuracy rate when using the support vector machine method. While the use of AI can potentially liberate administrative staff from routine tasks to handle more intricate cases (Zawacki-Richter et al., 2019), it also introduces bias, as the approaches have been shown to give prospective students from certain geographic locations an advantage in the college admissions process (Alam & Mohanty, 2022). The surge in data from learning management systems (LMS) and self-serve course registration has boosted research in these sectors, and algorithms targeting course selection, program admission, and pathway advising can have significant and sometimes restrictive effects on students (McConvey et al., 2023). It might restrict or overly influence student choices and inadvertently narrow down diverse learning paths and experiences.

Career paths/placement

Four reviews reported findings pertaining to the use of AI to assist with career paths and placements. Although McConvey et al. (2023) reported that 18% ($n = 7$) of the papers in their review were related to pathway advising, the number of studies researching this remains quite low, with Alkhalil et al. (2021) finding that managing large volumes of data was the main challenge when using AI to support student career pathways. Pinto et al. (2023) reported that some researchers have employed ML based approaches to predict the employability of college graduates to develop study plans that match the demands of the labour market. Salas-Pilco & Yang (2022) highlight that upon graduation, while students anticipate employability, many face challenges securing jobs. AI's role in predicting employability outcomes emphasises the necessity of offering guidance to graduates, ensuring quality in higher education, and understanding graduates' behavioural patterns to better support their career trajectories.

Student satisfaction

A small number of studies have explored using AI to predict student satisfaction, which was only mentioned in three reviews. Ouyang et al. (2022) highlighted a paper in their review (Hew et al., 2020), which analysed the course features of 249 randomly sampled MOOCs, and 6,393 students' perceptions were examined to understand what factors predicted student satisfaction. They found that the course instructor, content, assessment, and time schedule played significant roles in explaining student satisfaction levels. Pinto et al. (2023) highlighted findings from two studies; the first (Abdelkader et al., 2022) posited that feature selection (FS) increased the predictive accuracy of their ML model, allowing them to predict student satisfaction with online education with nearly perfect accuracy, and the second (Ho et al., 2021) investigated the most important predictors in determining the satisfaction of undergraduate students during the COVID-19 pandemic using data from Moodle and Microsoft Teams, which was also included in Rangal-de Lázaro & Duart (2023)'s review. The results showed that random forest recursive feature elimination improved the predictive accuracy of all the ML models.

Diagnostic prediction

Three reviews on AI applications in nursing and medical education (Buchanan et al., 2021; Hwang et al., 2022; Lee et al., 2021) discussed the prevalence of research on AI for diagnosis/prognosis prediction. Whilst all three reviews reported increasing use, they particularly highlighted the implications that this has for higher education curricula, which was also echoed by other medical reviews in the corpus (e.g., Burney & Ahmad, 2022). Lee at al. (2021) stressed the need for an evidence-informed AI curriculum, with an emphasis on ethical and legal implications, biomedical knowledge, critical appraisal of AI systems, and working with electronic health records. They called for an evaluation of current AI curricula, including changes in student attitudes, AI knowledge and skills. Buchanan et al. (2021) suggest that ethical implications, digital literacy, predictive modelling, and machine learning should now be part of any nursing curriculum, which Charow et al. (2021), Grunhut et al. (2021), Harmon et al. (2021) and Sapci & Sapci (2020) argue should be designed and taught by multidisciplinary teams. Further collaboration between educators and AI developers would also be a way forward (Zhang et al., 2023).

Assessment and Evaluation

Three reviews focused specifically on assessment and evaluation, including plagiarism (Albluwi, 2019), online learning (Del Gobbo et al., 2023), and the role of learning analytics with feedback (Banihashem et al., 2022). The systematic review by Crompton & Burke (2023) found that assessment and evaluation was the most common use of AIHEd, and the algorithm most frequently applied in nursing education for assessment and evaluation in Hwang et al.'s (2022) systematic review was natural language parsing (18.75%). All of the reviews containing findings about assessment and evaluation ($n = 26$) pertain to teaching and learning research, with 10 (38.5%) reporting on the use of AI to assist evaluation at the administrative level. Here, AI has been used to evaluate student outcomes to determine admission decisions (Alam & Mohanty, 2022), to inform faculty and institutional quality assurance measures (Alkhalil et al., 2021; Bearman et al., 2022; Crompton & Burke, 2023; Kirubarajan et al., 2022; Manhiça et al. 2022; Rabelo et al., 2023; Sghir et al., 2022; Zhang et al., 2023), and to analyse the impact of university accreditation on student test performance, as well as academic research performance and scientific productivity (Salas-Pilco & Yang, 2022). However, there remain many concerns about how institutions are storing and using teaching and learning data (see section below, *Research Gaps*), and therefore further data regulations and a greater emphasis on ethical considerations are needed (Bearman et al., 2022; Ullrich et al., 2022).

The 26 Assessment and Evaluation reviews were further classified into six subcategories: the evaluation of student understanding, engagement, and academic integrity ($n = 17$), automated grading and online exams ($n = 14$), automated feedback ($n = 10$), evaluation of teaching ($n = 5$), evaluation of learning material ($n = 5$), and the evaluation of universities ($n = 2$).

Evaluation of student understanding, engagement, and academic integrity

17 reviews (25.8%) included primary studies that evaluated AI's impact on learning effectiveness and behaviour (Chu et al., 2022), engagement (Rabelo et al., 2023; Sghir et al., 2022), plagiarism (Albluwi, 2019), reflections and higher order thinking (Crompton & Burke, 2023), often through LMS data (Manhiça et al. 2022), with a view to identifying students at risk and to enable earlier interventions (Banihashem et al., 2022). However, studies that provided explicit details about the actual impact of AI on student learning were rather rare in many of the reviews (e.g., two studies in Rangel-de Lázaro & Duart, 2023; three studies in Zawacki-Richter et al., 2019), and Hwang et al. (2022) found very few studies that explored AI's effect

on cognition and affect in nursing education, with further research suggested to explore the acquisition of nursing knowledge and skills, such as the use of AI to evaluate handwashing techniques and to evaluate nursing student emotions during patient interaction, as reported by Buchanan et al. (2021). This area seems to be slightly more advanced in medical education research, as Kirubarajan et al. (2022) found 31 studies that used AI to evaluate the surgical performance of trainees, including suturing, knot tying and catheter insertion (see also Burney & Ahmad, 2022; Sapci & Sapci, 2020). Zhang et al. (2023) point out, however, that machine learning can only classify surgical trainees into novices and experts through operations on virtual surgical platforms, and therefore some students might be able to deceive the algorithms. Here Albluwi (2019) stresses the need for more emphasis on integrating academic integrity and AI ethics into the curriculum.

Automated grading and online exams

Automatic assessment was found to be the most common use of AIHEd in Crompton and Burke's (2023) systematic review (18.8%, *n* = 26), which contrasts with small numbers found in other reviews, exploring the use of automated essay evaluation systems (AES; Ouyang et al., 2022) and remotely proctored exams (Pinto et al., 2023; Rangel-de Lázaro & Duart, 2023). AES use in the studies found by Zawacki-Richter et al. (2019) were mostly focused on undergraduate students and were used within a range of disciplines, as opposed to the heavy STEM focus reported by del Gobbo et al. (2023), who found the two most used approaches to be term frequency-inverse document frequency (TF-IDF) and Word Embeddings. Although automatic grading has been found to lessen teacher workload (e.g., Salas-Pilco et al., 2022), Alam and Mohanty (2022) suggest that using AES in small institutions would be challenging, owing to the large number of pre-scored exams required for calibration, and although automatic grading has been used for a wide range of tasks, from short answer tests to essays (Burney & Ahmad, 2022), they found that AES might not be appropriate for all forms of writing.

Automated feedback

Most of the 10 reviews (15.2%) identified only a small number of studies that evaluated the impact of automated feedback on students, including on academic writing achievement (Rangel-de Lázaro & Duart, 2023; Zawacki-Richter et al., 2019), on reflection (Salas-Pilco et al., 2022), and on self-awareness (Ouyang et al., 2022). Two studies in the scoping review by Kirubarajan et al. (2022) reported two studies who used real-time feedback using AI for modelling during surgery. Manhiça et al. (2022) also found two studies exploring automated feedback, but unfortunately did not provide any further information about them, which gives further weight to the potential of more research need in this area.

Evaluation of teaching

Five reviews (7.6%) found a small number of studies where AI had been used to evaluate teaching effectiveness. This was done by using data mining algorithms to analyse student comments, course evaluations and syllabi (Kirubarajan et al., 2022; Salas-Pilco & Yang, 2022; Zawacki-Richter et al., 2019), with institutions now being able to identify low-quality feedback given by educators and to flag repeat offenders (Zhang et al., 2023). Rabelo et al. (2023) argue, however, that management should make more use of this ability to evaluate teaching quality.

Evaluation of learning material

Five reviews (7.6%) mentioned the use of AI to evaluate learning materials, such as textbooks (Crompton & Burke, 2023), particularly done by measuring the amount of time students spend accessing and using them in the LMS (Alkhalil et al., 2021; Rabelo et al., 2023; Salas-Pilco et al., 2022). In Kirubarajan et al.'s (2022) scoping review on surgical education, nine studies used AI to improve surgical training materials by, for example, categorising surgical procedures.

Intelligent Tutoring Systems (ITS)

As with the adaptive systems and personalisation reviews, all of the ITS reviews included research at the teaching and learning milieu ($n = 14$), with only two reviews (14.3%) reporting a specific use of ITS at the administrative level. Alotaibi & Alshehri (2023) reported the use of intelligent academic advising, where students are provided with individualised guidance and educational planning, and Zawacki-Richter et al. (2019) reported examples of AI to support university career services, including an interactive intelligent tutor to assist new students (see Lodhi et al., 2018). Although there were more reviews classified under *adaptive systems and personalisation* that focused on the administrative level, previous reviews have commented on the lack of reporting of ITS use in higher education (e.g., Crompton & Burke, 2023), and therefore this perhaps represents an area for future exploration. One review (Crow et al., 2018) focused solely on the role of ITS in programming education, and found that no standard combination of features had been used, and suggested that future research could evaluate individual features or compare the implementation of different systems.

The 14 ITS reviews were further classified into six subcategories; diagnosing strengths/providing automated feedback ($n = 8$), teaching course content ($n = 8$), student ITS acceptance ($n = 4$), curating learning materials ($n = 3$), facilitating collaboration between learners ($n = 2$), and academic advising ($n = 2$; mentioned above).

Diagnosing strengths/providing automated feedback

Eight reviews (12.1%) reported on findings of ITS diagnosing strengths and gaps, suggesting learning paths and providing automated feedback (Salas-Pilco & Yang, 2022), which can help reduce educator workload (Alam & Mohanty, 2022) and ensure that students receive timely information about their learning (Crompton & Burke, 2023). ITS were the second most researched AI application (20%, $n = 10$) in Chu et al.'s (2022) systematic review of the top 50 most cited AIHEd articles in the Web of Science, with the greatest focus being on students' learning behaviour and affect. Rangal-de Lázaro and Duart (2023) reported that this was also the focus in three studies in the fields of Business and Medicine.

Teaching course content

Eight reviews (12.1%) also mentioned the role of ITS in teaching course content. Most prevalent was the use of ITS in the medical and scientific fields, for example, as virtual patient simulators or case studies to nursing, medical or dental students and staff (Buchanan et al., 2021; Hwang et al., 2021; Saghiri et al., 2022). In scientific settings, students performed experiments using lab equipment, with support tailored to their needs (Crompton & Burke, 2023). Personalised tutoring was also frequently mentioned in addition to teaching content. Rangal-de Lázaro and Duart (2023) discussed the use of an interactive tutoring component for a Java programming course throughout the Covid-19 pandemic. Intelligent feedback and hints can be

embedded into programming tasks, helping with specific semantic or syntactic issues (Crow et al., 2018), and specifically tailored hints and feedback were also provided on tasks to solve problems (Zawacki-Richter et al., 2019).

Student ITS acceptance

Student acceptance of ITS was addressed in four reviews (6.1%), including Rangal-de Lázaro and Duart (2023) who found five papers focused on Engineering Education (4.7% of studies). Chu et al. (2022) found that the most frequently discussed ITS issues were related to affect ($n = 17$, 41.5%) with the most common topics being student attitudes ($n = 6$, 33.33%) and opinions of learners or learning perceptions ($n = 6$, 33.33%), followed by emotion ($n = 3$, 18.75%). Technology acceptance model or intention of use, self-efficacy or confidence, and satisfaction or interest were less discussed. Harmon et al. (2021) found a limited amount of evidence of positive effects of AI on learning outcomes in their review on pain care in nursing education. The reactions of participants varied and were affected by many factors, including technical aspects (e.g., accessibility or internet speed), a lack of realism, poor visual quality of nonverbal cues, and the ability to ask avatars a question. Saghiri et al. (2022) examined artificial intelligence (AI) and virtual teaching models within the context of dental education and evaluated students' attitudes towards VR in implant surgery training, where they also found current ITS capacity to impact on student acceptance, suggesting that future tools need to account for differentiation of oral anatomy.

Curating learning materials

Three reviews (4.5%) addressed the use of material curation when using ITS. Zawacki-Richter et al. (2019) found three studies (2.1%) that discussed this function, which relate to the presentation of personalised learning materials to students, and only one study was identified by Zhang et al. (2023). Crow et al. (2018) concluded that when designing systems to intelligently tutor programming, it would be valuable to consider linking supplementary resources to the intelligent and adaptive component of the system and have suggested this for future ITS development.

Facilitating collaboration between learners

Two reviews (3.0%) discussed findings related to ITS facilitating collaboration, which can help by, for example, generating questions and providing feedback on the writing process (Alam & Mohanty, 2022). Zawacki-Richter et al. (2019) only found two primary studies that explored collaborative facilitation and called for further research to be undertaken with this affordance of ITS functionality.

Benefits and challenges within AIHEd

The evidence syntheses that focused on a range of AI applications or on AI more generally ($n = 31$; see Appendix E) were also coded inductively for any benefits and challenges that had been identified in the primary studies they reviewed. It should be noted that two reviews only considered the affordances of AIHEd (Crompton & Burke, 2023; Rangal-de Lázaro & Duart, 2023), four reviews did not mention any benefits, and six reviews did not mention any challenges, which for four reviews was due to their bibliometric nature (Gudyanga, 2023; Hinojo-Lucena et al.; Maphosa & Maphosa, 2021; Ullrich et al., 2022).

Benefits of using AI in higher education

Twelve different benefits were identified across the 31 reviews (see Appendix L), the most prominent of which was *personalised learning* reported in 12 studies (see Table 7), followed closely by *greater insight into student understanding, positive influence on learning outcomes*, and *reduced planning and administration time for teachers* identified in 32.3% of reviews. The top six benefits will be discussed below.

Table 7. Top six benefits across studies (n = 31), multiple mentions possible

Benefits	*n*	%
Personalised learning	12	38.7%
Greater insight into student understanding	10	32.3%
Positive influence on learning outcomes	10	32.3%
Reduced planning and administration time for teachers	10	32.3%
Greater equity in education	7	22.6%
Precise assessment & feedback	7	22.6%

In the context of *personalised learning*, Zawacki-Richter et al. (2019) and Sourani (2019) noted the adaptability of AI to create personalised learning environments, enabling the customisation of educational materials to fit individual learning needs (Algabri et al., 2021; Buchanan et al., 2021), and thereby support student autonomy by allowing learning at an individual pace (Alotaibi, 2023; Bearman et al., 2022). Diagnostic and remedial support is another focus, particularly in tailoring learning paths based on knowledge structures, which can facilitate early interventions for potentially disengaged students (Alam & Mohanty, 2022; Chu et al., 2022). Interestingly, ten reviews found or mentioned the ability of AI to *positively influence learning outcomes* (e.g., Alotaibi & Alshehri, 2023; Fichten et al., 2021), yet few reviews in this corpus provided real evidence of impact (as mentioned above in *Assessment and Evaluation*). AI was identified, however, as enhancing learning capabilities and facilitating smoother transitions into professional roles, especially in nursing and medicine (Buchanan et al., 2021; Hwang et al., 2022; Sapci & Sapci, 2020), alongside stimulating student engagement (Chaka, 2023) and honing specific skills such as writing performance through immediate feedback systems (Ouyang et al., 2022).

Several reviews highlighted that AI could automate routine tasks and thereby *reduce planning and administrative tasks* (e.g., Alam & Mohanty, 2022). For instance, AI-powered chatbots and intelligent systems facilitate lesson planning and handle student inquiries, which streamlines the administrative workflow (Algabri et al., 2021), and automated grading systems can alleviate workload by assessing student performance (e.g., Crompton & Burke, 2023). Several reviews highlighted the role of machine learning and analytics in *enhancing our understanding of student behaviours to support learning* (e.g., Alotaibi & Alshehri, 2023) and, complementing this, Ouyang et al. (2022), Rangal-de Lázaro & Duart (2023), and Salas-Pilco & Yang (2022) found primary research that focused on the utility of predictive systems. These systems are designed for the early identification of learning issues among students and offer guidance for their academic success. Finally, and relating to gaining greater insight into student understanding, reviews

identified studies that focus on analysing student interactions and providing adaptive feedback (e.g., Manhiça et al., 2022). This is complemented by Alam & Mohanty (2022), who highlighted the role of machine learning in classifying patterns and modelling student profiles. The significance of predictive analytics is further supported by reviews such as Salas-Pilco et al. (2022) and Ouyang et al. (2022), which discuss their utility in enabling timely interventions and guiding academic success.

Seven reviews mentioned the ability of AI to enable *greater equity in education*, with the evolving role of universities in community development a potential contributing factor (Alotaibi & Alshehri, 2023). Authors argue that AI has the potential to offer cheaper, more engaging, and more accessible learning options (Alam & Mohanty, 2022; Algabri et al., 2021), for example the use of expert systems to offer guidance to students who lack access to human advisors (Bearman et al., 2022), which might also help alleviate social isolation in distance education (Chaka, 2023). AI as an accessibility tool was also discussed in regard to innovations such as the 'Smart Cane' in India (Bhattacharjee, 2019), as well as the potential to help alleviate academic stress for students with disabilities (Fichten et al., 2021). AI also provides the possibility of forming diverse study groups, thereby enriching, and diversifying the educational experience (Manhiça et al., 2022).

Regarding *precise assessment and feedback*, Algabri et al. (2021) describe how AI can not only speed up grading but also make it more objective and error-free, with analytics tools that help educators monitor student progress. Similarly, Ouyang et al. (2022) note the effectiveness of automated essay evaluation systems in providing immediate, impactful feedback to improve student writing. This aligns with Kirubarajan et al. (2022), who reported high accuracy levels in AI-based assessments distinguishing between novice and expert surgeons, and Zhang et al. (2023) add that machine learning can extract objective indicators for automated skills assessment. On the other hand, Charow et al. (2021) and Saghiri et al. (2022) discuss the relevance of AI in healthcare, providing tools for data-driven decision-making and individualised feedback, respectively. These studies collectively indicate that AI holds promise for making educational assessments more precise, timely, and tailored to individual needs.

Challenges of using AI in higher education
17 different challenges were identified across the 31 reviews (see Appendix L), which despite being more in number than the benefits, were mentioned in fewer studies. The most frequently mentioned challenge was a *lack of ethical consideration*, reported in nine studies (see Table 8), followed closely by *curriculum development, infrastructure, lack of teacher technical knowledge* and *shifting authority*, identified in 22.6% of reviews. These top five challenges will be discussed below.

The *ethical implications* in the application of AI in education range from the need for professional preparedness to systemic lapses in research rigour, including a lack of adherence to ethical procedures when collecting data for primary research (e.g., Salas-Pilco & Yang, 2022). Grunhut et al. (2021) and Lee et al. (2021) highlighted the ethical challenges that medical professionals face, especially when interpreting AI predictions, with Zhang et al. (2023) suggesting that the premature use of AI tools in medical education could lead to ethical and legal issues. At the instructional level, Chu et al. (2022) explored the ethical challenges in balancing human and machine-assisted learning, suggesting that educators need to consciously reflect on these issues when incorporating AI into their teaching methods.

Table 8. Top five challenges across studies (n = 31), multiple mentions possible

Challenges	*n*	%
Lack of ethical consideration	9	29.0%
Curriculum development	7	22.6%
Infrastructure	7	22.6%
Lack of teacher technical knowledge	7	22.6%
Shifting Authority	7	22.6%

In relation to the challenges of integrating AI into education, *curriculum development issues* and *infrastructural problems* span from broad systemic concerns to specific educational contexts. According to Ouyang et al. (2022), there is a disconnect between AI technology and existing educational systems, and suggest the need for more unified, standardised frameworks that incorporate ethical principles and advocate for the development of multidisciplinary teams (Charow et al., Lee et al., 2021), with a stronger focus on more robust and ethically aware AI curricula (e.g., Grunhut et al., 2021). Furthermore, despite its potential, a country may lag behind in both AI research and infrastructure in the field of education (Bhattacharjee, 2019). When it comes to infrastructural challenges, the reviews predominantly mention technical, literacy, and financial barriers (Alotaibi & Alshehri, 2023; Charow et al., 2021), such as the high costs associated with developing virtual programming and the necessity for high-speed internet (Harmon et al., 2021).

With the potential to slow AI curriculum development and application efforts, several reviews mentioned a *lack of teacher technical knowledge*, reporting that many educators were either unaware of AI's potential or would need new technological skills and knowledge in order to effectively use it (Alotaibi & Alshehri, 2023; Bhattacharjee, 2019; Chu et al., 2022), which was recognised as a significant hindrance to effective operations (Grunhut et al., 2021; Lee et al., 2021). However, also pertinent here, is the recognition that faculty generally have limited time to integrate AI effectively into the curriculum (Charow et al., 2021), which is not helped by the fear of losing their jobs or being replaced by AI (Alotaibi & Alshehri, 2023; Bearman et al., 2022). Charow et al. (2021) emphasise the need for educators and medical professionals to see AI as augmenting rather than replacing their roles, and there is therefore a need for the development of AI understanding, otherwise Lee et al. (2021) point out that misunderstanding could lead to a *shift in authority* from clinicians to AI systems. Overcoming resistance to change and solving various challenges, including those of an ethical and administrative nature, were thereby identified as pivotal for successful AIHEd integration (Sourani, 2019).

What research gaps have been identified?

Each review in this corpus (*n* = 66) was searched for any research gaps that had been identified within the primary studies, which were then coded inductively (see Appendix A). More than 30 different categories of research suggestions emerged (see Appendix M), with the top ten research gap categories found in more than 10% of the corpus (see Table 9). The most prominent research issue (in 40.9% of studies) relates to the need for further ethical consideration and attention within AIHEd research as both a topic of research and as an issue in the conduct of empirical research, followed closely by the need for a range of further empirical research with a greater emphasis on methodological rigour, including research design and

reporting (36.4%). AIHEd reviews also identified the need for future primary research with a wider range of stakeholders (21.2%), within a more diverse array of countries (15.2%) and disciplines (16.7%).

Table 9. Top ten research gaps across studies (n = 66), multiple mentions possible

Research gaps	*n*	**%**
Ethical implications	27	40.9%
More methodological approaches needed	24	36.4%
More research in Education needed	22	33.3%
More research with a wider range of stakeholders	14	21.2%
Interdisciplinary approaches required	11	16.7%
Research limited to specific discipline areas	11	16.7%
More research in a wider range of countries, esp. developing	10	15.2%
Greater emphasis on theoretical foundations needed	9	13.6%
Longitudinal studies recommended	8	12.1%
Research limited to a few topics	8	12.1%

Ethical implications

Eight reviews found that primary research rarely addressed privacy problems, such as participant data protection during educational data collection (Alam & Mohanty, 2022; Fichten et al., 2021; Li et al., 2021; Manhiça et al., 2022; Otoo-Arthur & van Zyl, 2020; Salas-Pilco et al., 2022; Salas-Pilco & Yang, 2022; Zawacki-Richter et al., 2019), and that this necessitates the need for the creation or improvement of ethical frameworks (Zhai & Wibowo, 2023), alongside a deeper understanding of the social implications of AI more broadly (Bearman et al., 2022). Educating students about their own ethical behaviour and the ethical use of AI also emerged as an important topic (Albluwi, 2019; Buchanan et al., 2021; Charow et al., 2021; Lee et al., 2021; Salas-Pilco & Yang, 2022), with the need for more evaluation and reporting of current curriculum impact, especially in the fields of Nursing and Medicine (e.g., Grunhut et al., 2021). Potential topics of future research include:

- Student perceptions of the use of AI in assessment (del Gobbo et al., 2023);
- How to make data more secure (Ullrich et al., 2022);
- How to correct sample bias and balance issues of privacy with the affordances of AI (Saghiri et al., 2022; Zhang et al., 2023); and
- How institutions are storing and using teaching and learning data (Ifenthaler & Yau, 2020; Maphosa & Maphosa, 2021; McConvey et al., 2023; Rangal-de Lázaro & Duart, 2023; Sghir et al., 2023; Ullrich et al., 2022).

Methodological approaches

Aside from recognising that further empirical research is needed (e.g., Alkhalil et al., 2021; Buchanan et al., 2021), more rigorous reporting of study design in primary research was called for, including ensuring

that the number of participants and study level is reported (Fichten et al., 2021; Harmon et al., 2021). Although there is still a recognised need for AIHEd quasi-experiments (Darvishi et al., 2022) and experiments, particularly those that allow multiple educational design variations (Fontaine et al., 2019; Hwang et al., 2022; Zhang et al., 2023; Zhong, 2022), a strong suggestion has been made for more qualitative, mixed methods and design-based approaches (e.g., Abu Saa et al., 2019), alongside longitudinal studies (e.g., Zawacki-Richter et al., 2019) and larger sample sizes (e.g., Zhang et al., 2023). Further potential approaches and topics include:

- The use of surveys, course evaluation surveys, network access logs, physiological data, observations, interviews (Abu Saa et al., 2019; Alam & Mohanty, 2022; Andersen et al., 2022; Chu et al., 2022; Hwang et al., 2022; Zawacki-Richter et al., 2019);
- More evaluation of the effectiveness of tools on learning, cognition, affect, skills etc rather than focusing on technical aspects like accuracy (Albluwi, 2019; Chaka, 2023; Crow et al., 2018; Frangoudes et al., 2021; Zhong, 2022);
- Multiple case study design (Bearman et al., 2022; Ullrich et al., 2022);
- Cross referencing data with external platforms such as social media data (Rangal-de Lázaro & Duart, 2023; Urdaneta-Ponte et al., 2021); and
- A focus on age and gender as demographic variables (Zhai & Wibowo, 2023).

Study contexts

In regard to stakeholders who should be included in future AIHEd research, reviews identified the need for more diverse populations when training data (e.g., Sghir et al., 2022), such as underrepresented groups (Pinto et al., 2023) and students with disabilities (Fichten et al., 2021), to help ensure that their needs are reflected in AI development. Further primary research with postgraduate students (Crompton & Burke, 2023), educators (Alyahyan & Düştegör, 2020; del Gobbo et al., 2023; Hamam, 2021; Sourani, 2019), and managers/administrators (e.g., Ullrich et al., 2022) has also been called for.

More research is needed within a wider range of contexts, especially developing countries (e.g., Pinto et al., 2023), such as India (Bhattacharjee, 2019) and African nations (Gudyanga, 2023; Maphosa & Maphosa, 2020), to better understand how AI can be used to enhance learning in under-resourced communities (Crompton & Burke, 2023). Multiple reviews also stressed the need for further research in disciplines other than STEM (e.g., Chaka, 2023), including Social Sciences (e.g., Alyahyan & Düştegör, 2020), Visual Arts (Chu et al., 2022) and hands-on subjects such as VET education (Fariani et al., 2022), although there were still specific areas of need identified in nursing (Hwang et al., 2022) and dentistry (Saghiri et al., 2022) for example. The state of AIHEd research within Education itself is an issue (Alam & Mohanty, 2022; Zawacki-Richter et al., 2019), and suggestions for more interdisciplinary approaches have been made, in order to improve pedagogical applications and outcomes (e.g., Kirubarajan et al., 2022). Potential further research approaches include:

- Student perceptions of effectiveness and AI fairness (del Gobbo et al., 2023; Hamam, 2021; Otoo-Arthur & van Zyl, 2020);
- Combining student and educator perspectives (Rabelo et al., 2023);
- Low level foreign language learners and chatbots (Klímová & Seraj, 2023);
- Non formal education (Urdaneta-Ponte et al., 2021); and
- Investigating a similar dataset with data retrieved from different educational contexts (Fahd et al., 2022)

Discussion

By using the framework of Zawacki-Richter et al. (2019), this tertiary review of 66 AIHEd evidence syntheses found that most reviews report findings on the use of adaptive systems and personalisation tools, followed by profiling and prediction tools. However, owing to the heavy predominance of primary AIHEd research in STEM and Health & Welfare courses, as in other EdTech research (e.g., Lai & Bower, 2019), AI applications and presence within the curriculum appear to be at a more mature stage in those rather than in other disciplines. Furthermore, insights into how AI is being used at the postgraduate level, as well as at the institutional and administrative level, remain limited.

This review of reviews confirms that the benefits of AI in higher education are multifold. Most notably, AI facilitates personalised learning, which constitutes approximately 38.7% of the identified advantages in the reviewed studies. AI systems are adaptable and allow learning materials to be tailored to individual needs, thereby enhancing student autonomy, and enabling early interventions for disengaged students (Algabri et al., 2021; Alotaibi & Alshehri, 2023; Bearman et al., 2022). Other significant benefits include the positive influence on learning outcomes, reduced administrative time for educators, and greater insight into student understanding. AI not only enhances traditional academic outcomes but also aids in professional training and specific skill development (Buchanan et al., 2021; Hwang et al., 2022; Sapci & Sapci, 2020). However, the adoption of AI in higher education is not without challenges. The most frequently cited concern is the lack of ethical consideration in AI applications, followed by issues related to curriculum development and infrastructure. Studies indicate the need for substantial financial investment and technical literacy to fully integrate AI into existing educational systems (Charow et al., 2021; Alotaibi & Alshehri, 2023). Moreover, there is a noted lack of educator technical knowledge and fears regarding job displacement due to AI, which require attention (Alotaibi, 2023; Bearman et al., 2022).

In contrast to previous reviews in the field of EdTech (e.g., Bodily et al., 2019; Zawacki-Richter & Latchem, 2018), and previous EdTech tertiary reviews (Buntins et al., In peer review; Zawacki-Richter, 2023), authors conducting AIHEd evidence synthesis represent a wide range of countries, with the top six most productive countries from six different continents. Despite this, there is still less research emerging from Oceania, Africa and, in particular, from South and Central America, although in the case of the latter, it is possible that this is due to authors publishing in their own native language rather than in English (Marin et al., 2023). Related to the issue of global reach, only 57.6% of evidence synthesis in this sample were published open access, as opposed to 88.3% of higher education EdTech research published during the pandemic (Bond et al., 2021). This limits not only the ability of educators and researchers from lower resourced institutions to read these reviews, but it decreases its visibility generally, thereby increasing the likelihood that other researchers will duplicate effort and conduct similar or the same research, leading to 'research waste' (Grainger et al., 2020; Siontis & Ioannidis, 2018). Therefore, to move the AIHEd field forward, we are calling for a focus on three particular areas, namely ethics, collaboration, and rigour.

A call for increased ethics
There is a loud and resounding call for an enhanced focus on ethics in future AIHEd research, with 40.9% of reviews in this corpus indicating that some form of ethical considerations are needed. Whilst this realisation is not lost on the AIEd field, with at least four evidence syntheses published specifically on the topic in the last two years (Guan et al., 2023; Mahmood et al., 2022; Rios-Campos et al., 2023; Yu & Yu,

2023)[32], this meta review indicates that the issue remains pressing. Future primary research must ensure that lengthy consideration is given to participant consent, data collection procedures, and data storage (Otoo-Arthur & van Zyl, 2020). Further consideration must also be given to the biases that can be perpetuated through data (Zhang et al., 2023), as well as embedding ethical AI as a topic throughout the HE curriculum (Grunhut et al., 2021).

There is also a need for more ethical consideration when conducting evidence synthesis. This review uncovered examples of evidence synthesis that stated the 'use' of the PRISMA guidelines (Page et al., 2021), for example, but that did not cite it in the reference list or cited it incorrectly, as well as secondary research that used the exact methodology and typology of Zawacki-Richter et al. (2019), ending up with very similar findings, but that did not cite the original article at all. Further to this, one review was excluded from the corpus, as it plagiarised the entire Zawacki-Richter et al. (2019) article. Whilst concerns are growing over the use and publication of generative AI produced summaries that plagiarise whole sections of text (see Kalz, 2023), ensuring that we conduct primary and secondary research as rigorously and transparently as possible is our purview as researchers, and is vitally needed if we are to expand and enhance the field.

A call for increased collaboration

The findings of this review highlighted the need for collaboration in four key areas: the development of AI applications, designing and teaching AI curriculum, researching AIHEd, and conducting evidence syntheses. To translate future AI tools into practice and meet community expectations, there is a need to include intended users in their development (Harmon et al., 2021; McConvey et al., 2021), which Li et al. (2021) also suggest could include the collection and sharing of massive data across disciplines and contexts, whilst adhering to considerations of privacy. Multidisciplinary teams should then be brought together, including data scientists, educators, and students, to ensure that AI curricula are robust, ethical and fit for purpose (Charow et al., 2021; Sapci & Sapci, 2020). In the case of medical education, health professionals and leaders, as well as patients, should also be involved (Grunhut et al., 2021; Zhang et al., 2023).

To evaluate the efficacy of AI applications in higher education, interdisciplinary research teams should include a range of stakeholders from diverse communities (Chu et al., 2022; Crompton & Burke, 2023; Hwang et al., 2021), for example linking computer scientists with researchers in the humanities and social sciences (Ullrich et al., 2022). Finally, in terms of evidence synthesis authorship, the large amount of domestic research collaborations indicates that the field could benefit from further international research collaborations, especially for authors in Oceania and Europe, as this might provide more contextual knowledge, as well as help eliminate language bias when it comes to searching for literature (Rangal-de Lázaro & Duart, 2023). A large proportion of authors from Africa and the Middle East also published as single authors (29% and 22% respectively). By conducting evidence synthesis in teams, greater rigour can be achieved through shared understanding, discussion, and inter-rater reliability measures (Booth et al., 2013). It should be noted here, however, that less than half of the reviews in this corpus (43.9%, $n = 29$) did not report any inter-rater agreement processes, which, although this is better than what was found in previous meta reviews of EdTech research (Buntins, et al., In peer review; Zawacki-Richter, 2023), represents the beginning of a much-needed discussion on research rigour.

---

[32] These are not included in this corpus, as they include results from other educational levels.

A call for increased rigour

The prevailing landscape of AIHEd research evidences a compelling call for enhanced rigour and methodological robustness. A noticeable 65% of reviews are critically low to medium quality, signalling an imperative to recalibrate acceptance criteria to strengthen reliability and quality. The most concerning findings were that 31.8% of studies only searched in one or two databases, only 24.2% provided their exact data extraction coding scheme (compared to 51% in Chalmers et al., 2023 and 37% in Buntins et al. (In peer review), 45.5% did not undertake any form of quality assessment, and 34.8% did not reflect at all upon the limitations of their review. Furthermore, over half of the reviews (51.5%) did not report whether some form of digital evidence synthesis tool was used to conduct the review. Given the affordances in efficiency that machine learning can bring to evidence synthesis (e.g., Stansfield et al., 2022; Tsou et al., 2020), as well as the enhanced transparency through visualisation tools such as EPPI Visualiser, it is surprising that the AIHEd community has not made more use of them (see Zhang & Neitzel, 2023). These inconsistencies and the lack of using any methodological guidance, or the frequent recourse to somewhat dated (yet arguably seminal) approaches by Kitchenham and colleagues (2004, 2007, 2009) – prior to the first and subsequently updated PRISMA guidelines (Moher et al., 2009; Page et al., 2021) – underscore an urgent necessity for contemporary, stringent, and universally adopted review guidelines within AIEd, but also within the wider field of EdTech (e.g., Jing et al., 2023) and educational research at large (e.g., Chong et al., 2023; Ellaway, 2014).

Conclusion

This tertiary review synthesised the findings of 66 AIHEd evidence syntheses, with a view to map the field and gain an understanding of authorship patterns, research quality, key topics, common findings, and potential research gaps in the literature. Future research will explore the full corpus of 307 AIEd evidence syntheses located across various educational levels, providing further insight into applications and future directions, alongside further guidance for the conduct of evidence synthesis. While AI offers promising avenues for enhancing educational experiences and outcomes, there are significant ethical, methodological, and pedagogical challenges that need to be addressed to harness its full potential effectively.

Declarations

---

[33] https://eppi.ioe.ac.uk/eppi-vis/login/open?webdbid=322

[34] https://osf.io/y2afk/?view_only=5dc18e0e4b3846bb9a4d9697c892485e

References

* Indicates that the article is featured in the corpus of the review

Abdelkader, H. E., Gad, A. G., Abohany, A. A., & Sorour, S. E. (2022). An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students During the COVID-19 pandemic. *IEEE Access, 10*, 6286–6303. https://doi.org/10.1109/ACCESS.2022.3143035

* Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques. *Technology, Knowledge and Learning*, *24*(4), 567–598. https://doi.org/10.1007/s10758-019-09408-7

* Alam, A., & Mohanty, A. (2022). Foundation for the Future of Higher Education or 'Misplaced Optimism'? Being Human in the Age of Artificial Intelligence. In M. Panda, S. Dehuri, M. R. Patra, P. K. Behera, G. A. Tsihrintzis, S.-B. Cho, & C. A. Coello Coello (Eds.), Innovations in Intelligent Computing and Communication (pp. 17–29). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-031-23233-6_2

* Algabri, H. K., Kharade, K. G., & Kamat, R. K. (2021). Promise, Threats, And Personalization In Higher Education With Artificial Intelligence. *Webology, 18*(6), 2129–2139.

* Alkhalil, A., Abdallah, M. A., Alogali, A., & Aljaloud, A. (2021). Applying Big Data Analytics in Higher Education: A Systematic Mapping Study. *International Journal of Information and Communication Technology Education, 17*(3), 29–51. https://doi.org/10.4018/IJICTE.20210701.oa3

Allman, B., Kimmons, R., Rosenberg, J., & Dash, M. (2023). Trends and Topics in Educational Technology, 2023 Edition. *TechTrends: Linking Research & Practice to Improve Learning, 67*(3), 583–591. https://doi.org/10.1007/s11528-023-00840-2

* Alotaibi, N. S., & Alshehri, A. H. (2023). Prospers and Obstacles in Using Artificial Intelligence in Saudi Arabia Higher Education Institutions—The Potential of AI-Based Learning Outcomes. *Sustainability, 15*(13), 10723. https://doi.org/10.3390/su151310723

* Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education, 17*(1), 1–21. https://doi.org/10.1186/s41239-020-0177-7

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

* Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, *37*, 100489. https://doi.org/10.1016/j.edurev.2022.100489

* Bhattacharjee, K. K. (2019). Research Output on the Usage of Artificial Intelligence in Indian Higher Education - A Scientometric Study. In 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 916–919). IEEE. https://doi.org/10.1109/ieem44572.2019.8978798

Bodily, R., Leary, H., & West, R. E. (2019). Research trends in instructional design and technology journals. *British Journal of Educational Technology, 50*(1), 64–79. https://doi.org/10.1111/bjet.12712

Bond, M. (2018). Helping doctoral students crack the publication code: An evaluation and content analysis of the Australasian Journal of Educational Technology. *Australasian Journal of Educational Technology*, *34*(5), 168–183. https://doi.org/10.14742/ajet.4363

Bond, M., Bedenlier, S., Marín, V. I., & Händel, M. (2021). Emergency remote teaching in higher education: mapping the first global online semester. *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-021-00282-x

Bond, M., Zawacki-Richter, O., & Nichols, M. (2019). Revisiting five decades of educational technology research: A content and authorship analysis of the British Journal of Educational Technology. *British Journal of Educational Technology*, *50*(1), 12–63. https://doi.org/10.1111/bjet.12730

Booth, A., Carroll, C., Ilott, I., Low, L. L., & Cooper, K. (2013). Desperately seeking dissonance: Identifying the disconfirming case in qualitative evidence synthesis. *Qualitative Health Research*, *23*(1), 126–141. https://doi.org/10.1177/1049732312466295

Bozkurt, A., & Sharma, R. C. (2023). Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning. *Asian Journal of Distance Education, 18*(1). http://dx.doi.org/10.5281/zenodo.7755273

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., Koutropoulos, A., Toquero, C. M., Singh, L., Tlili, A., Lee, K., Nichols, M., Ossiannilsson, E., Brown, M., Irvine, V., Raffaghelli, J. E., Santos-Hermosa, G., Farrell, O., Adam, T., Thong, Y. L., Sani-Bozkurt, S., Sharma, R. C., Hrastinski, S., & Jandrić, P. (2023). Speculative futures on ChatGPT and generative Artificial Intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, 18(1), 1-78. http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/709/394

* Buchanan, C., Howitt, M. L., Wilson, R., Booth, R. G., Risling, T., & Bamford, M. (2021). Predicted Influences of Artificial Intelligence on Nursing Education: Scoping Review. *JMIR Nursing, 4*(1), e23933. https://doi.org/10.2196/23933

Buntins, K., Bedenlier, S., Marín, V. I., Händel, M., & Bond, M. (In peer review). Methodological approaches to evidence synthesis in educational technology: A tertiary systematic mapping review.

* Burney, I. A., & Ahmad, N. (2022). Artificial Intelligence in Medical Education: A citation-based systematic literature review. *Journal of Shifa Tameer-E-Millat University, 5*(1), 43–53. https://doi.org/10.32593/jstmu/Vol5.Iss1.183

* Cardona, T., Cudney, E. A., Hoerl, R., & Snyder, J. (2023). Data Mining and Machine Learning Retention Models in Higher Education. *Journal of College Student Retention: Research, Theory and Practice*, *25*(1), 51-75. https://doi.org/10.1177/1521025120964920

Centre for Reviews and Dissemination (UK). (1995). *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews.* https://www.ncbi.nlm.nih.gov/books/NBK285222/. Accessed 4 January 2023.

* Chaka, C. (2023). Fourth industrial revolution—a review of applications, prospects, and challenges for artificial intelligence, robotics and blockchain in higher education. *Research and Practice in Technology Enhanced Learning, 18*, 2, 1-39. https://doi.org/10.58459/rptel.2023.18002

Chalmers, H., Brown, J., & Koryakina, A. (2023). Topics, publication patterns, and reporting quality in systematic reviews in language education. Lessons from the international database of education systematic reviews (IDESR). *Applied Linguistics Review*, *0*(0). https://doi.org/10.1515/applirev-2022-0190

* Charow, R., Jeyakumar, T., Younus, S., Dolatabadi, E., Salhia, M., Al-Mouaswas, D., Anderson, M., Balakumar, S., Clare, M., Dhalla, A., Gillan, C., Haghzare, S., Jackson, E., Lalani, N., Mattson, J., Peteanu, W., Tripp, T., Waldorf, J., Williams, S., . . . Wiljer, D. (2021). Artificial Intelligence Education Programs for Health Care Professionals: Scoping Review. *JMIR Medical Education*, *7*(4), e31043. https://doi.org/10.2196/31043

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology and Society*, *25*(1), 28–47. https://www.jstor.org/stable/10.2307/48647028

Chong, S. W., Bond, M., & Chalmers, H. (2023). Opening the methodological black box of research synthesis in language education: where are we now and where are we heading? *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2022-0193

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, 38(3), 22–42. https://doi.org/10.14742/ajet.7526

Cobos, C., Rodriguez, O., Rivera, J., Betancourt, J., Mendoza, M., León, E., & Herrera-Viedma, E. (2013). A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes. *Information Processing and Management, 49*(3), 607–625. https://doi.org/10.1016/j.ipm.2012.12.002

* 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). https://doi.org/10.1186/s41239-023-00392-8

* Crow, T., Luxton-Reilly, A., & Wuensche, B. (2018). Intelligent tutoring systems for programming education. In R. Mason & Simon (Eds.), *Proceedings of the 20th Australasian Computing Education Conference* (pp. 53–62). ACM. https://doi.org/10.1145/3160489.3160492

Daoudi, I. (2022). Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda. *Education and Information Technologies, 27*(8), 11237–11266. https://doi.org/10.1007/s10639-022-11087-4

* de Oliveira, T. N., Bernardini, F., & Viterbo, J. (2021). An Overview on the Use of Educational Data Mining for Constructing Recommendation Systems to Mitigate Retention in Higher Education. In 2021 IEEE Frontiers in Education Conference (FIE) (pp. 1–7). IEEE. https://doi.org/10.1109/FIE49875.2021.9637207

* Del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning. A systematic mapping. *Studies in Educational Evaluation, 77*, 101258. https://doi.org/10.1016/j.stueduc.2023.101258

Desmarais, M. C., & Baker, R. S. D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. User Modeling and User-Adapted Interaction, 22, 9-38.

Digital Solution Foundry, & EPPI-Centre. (2023). *EPPI-Mapper* (Version 2.2.3) [Computer software]. UCL Social Research Institute, University College London. http://eppimapper.digitalsolutionfoundry.co.za/#/

Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives (pp. 275-301). Boston, MA: Springer US.

Doroudi, S. (2022). The intertwined histories of artificial intelligence and education. *International Journal of Artificial Intelligence in Education,* 1-44.

* Fahd, K., Venkatraman, S., Miah, S. J., & Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies, 27*(3), 3743–3775. https://doi.org/10.1007/s10639-021-10741-7

* Fichten, C., Pickup, D., Asunsion, J., Jorgensen, M., Vo, C., Legault, A., & Libman, E. (2021). State of the Research on Artificial Intelligence Based Apps for Post-Secondary Students with Disabilities. *Exceptionality Education International, 31*(1), 62–76. https://doi.org/10.5206/EEI.V31I1.14089

* Fontaine, G., Cossette, S., Maheu-Cadotte, M.-A., Mailhot, T., Deschênes, M.-F., Mathieu-Dupuis, G., Côté, J., Gagnon, M.-P., & Dubé, V. (2019). Efficacy of adaptive e-learning for health professionals and students: A systematic review and meta-analysis. *BMJ Open*, *9*(8), e025252. https://doi.org/10.1136/bmjopen-2018-025252

* Frangoudes, F., Hadjiaros, M., Schiza, E. C., Matsangidou, M., Tsivitanidou, O., & Neokleous, K. (2021). An Overview of the Use of Chatbots in Medical and Healthcare Education. In P. Zaphiris & A. Ioannou (Eds.), Lecture Notes in Computer Science. Learning and Collaboration Technologies: Games and Virtual Environments for Learning (Vol. 12785, pp. 170–184). Springer International Publishing. https://doi.org/10.1007/978-3-030-77943-6_11

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews*. SAGE.

Grainger, M. J., Bolam, F. C., Stewart, G. B., & Nilsen, E. B. (2020). Evidence synthesis for tackling research waste. *Nature Ecology & Evolution*, *4*(4), 495–497. https://doi.org/10.1038/s41559-020-1141-6

* Grunhut, J., Wyatt, A. T., & Marques, O. (2021). Educating Future Physicians in Artificial Intelligence (AI): An Integrative Review and Proposed Changes. Journal of Medical Education and Curricular Development, 8, 23821205211036836. https://doi.org/10.1177/23821205211036836

* Gudyanga, R. (2023). Mapping education 4.0 research trends. *International Journal of Research in Business and Social Science, 12*(4), 434–445. https://doi.org/10.20525/ijrbs.v12i4.2585

Gusenbauer, M., & Haddaway, N. R. (2020). Which Academic Search Systems are Suitable for Systematic Reviews or Meta-Analyses? Evaluating Retrieval Qualities of Google Scholar, PubMed and 26 other Resources. *Research Synthesis Methods*, *11*(2), 181–217. https://doi.org/10.1002/jrsm.1378

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS ONE, 10*(9), e0138237. https://doi.org/10.1371/journal.pone.0138237

* Hamam, D. (2021). The New Teacher Assistant: A Review of Chatbots' Use in Higher Education. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), Communications in Computer and Information Science. HCI International 2021 - Posters (Vol. 1421, pp. 59–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-78645-8_8

Han, B., Nawaz, S., Buchanan, G., & McKay, D. (2023, June). Ethical and Pedagogical Impacts of AI in Education. In International Conference on Artificial Intelligence in Education (pp. 667-673). Cham: Springer Nature Switzerland.

* Harmon, J., Pitt, V., Summons, P., & Inder, K. J. (2021). Use of artificial intelligence and virtual reality within clinical simulation for nursing pain education: A scoping review. *Nurse Education Today, 97,*104700. https://doi.org/10.1016/j.nedt.2020.104700

* Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., & Liao, S. N. (2018). Predicting Academic Performance: A Systematic Literature Review. In ITiCSE 2018 Companion, Proceedings Companion of the 23rd Annual ACM

Conference on Innovation and Technology in Computer Science Education (pp. 175–199). Association for Computing Machinery. https://doi.org/10.1145/3293881.3295783

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education, 145*, 103724. https://doi.org/10.1016/j.compedu.2019.103724

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The impact of digital technology on learning: A summary for the Education Endowment Foundation.* Education Endowment Foundation. https://eric.ed.gov/?id=ED612174

* Hinojo-Lucena, F.-J., Aznar-Diaz, I., Romero-Rodríguez, J.-M., & Cáceres-Reche, M.-P. (2019). Artificial Intelligence in Higher Education: A Bibliometric Study on its Impact in the Scientific Literature. *Education Sciences, 9*. https://doi.org/10.3390/educsci9010051

Ho, I. M., Cheong, K. Y., & Weldon, A. (2021). Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. *PLoS ONE, 16*(4). https://doi.org/10.1371/journal.pone.0249423

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 1-23.

* Hwang, G.-J., Tang, K.-Y., & Tu, Y.-F. (2022). How artificial intelligence (AI) supports nursing education: profiling the roles, applications, and trends of AI in nursing education research (1993–2020). *Interactive Learning Environments*, 1–20. https://doi.org/10.1080/10494820.2022.2086579

* Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research & Development, 68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

İpek, Z. H., Gözüm, A. İ. C., Papadakis, S., & Kallogiannakis, M. (2023). Educational Applications of the ChatGPT AI System: A Systematic Review Research. Educational Process International Journal, 12(3). https://doi.org/10.22521/edupij.2023.123.2

Jing, Y., Wang, C., Chen, Y., Wang, H., Yu, T., & Shadiev, R. (2023). Bibliometric mapping techniques in educational technology research: A systematic literature review. *Education and Information Technologies.* Advance online publication. https://doi.org/10.1007/s10639-023-12178-6

Kalz, M. (2023). *AI destroys principles of authorship. A scary case from educational technology publishing.* https://kalz.cc/2023/09/15/ai-destroys-principles-of-authorship.-a-scary-case-from-educational-technology-publishing

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence, 3, 100074.

* Kirubarajan, A., Young, D., Khan, S., Crasto, N., Sobel, M., & Sussman, D. (2022). Artificial Intelligence and Surgical Education: A Systematic Scoping Review of Interventions. *Journal of Surgical Education, 79*(2), 500–515. https://doi.org/10.1016/j.jsurg.2021.09.012

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*. Keele. Software Engineering Group, Keele University. https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Kitchenham, B., & Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering: Technical Report EBSE 2007-001.* Keele University and Durham University.

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, *51*(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering – A tertiary study. *Information and Software Technology*, *52*(8), 792–805. https://doi.org/10.1016/j.infsof.2010.03.006

* Klímová, B., & Ibna Seraj, P. M. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. *Frontiers in Psychology, 14*, 1131506. https://doi.org/10.3389/fpsyg.2023.1131506

Lai, J. W., & Bower, M. (2019). How is the use of technology in education evaluated? A systematic review. *Computers & Education*, *133*, 27–42. https://doi.org/10.1016/j.compedu.2019.01.010

Lai, J. W., & Bower, M. (2020). Evaluation of technology use in education: Findings from a critical analysis of systematic literature reviews. *Journal of Computer Assisted Learning*, *36*(3), 241–259. https://doi.org/10.1111/jcal.12412

* Lee, J., Wu, A. S., Li, D., & Kulasegaram, K. M. (2021). Artificial Intelligence in Undergraduate Medical Education: A Scoping Review. *Academic Medicine, 96*(11S), S62-S70. https://doi.org/10.1097/ACM.0000000000004291

* Li, F., He, Y., & Xue, Q. (2021). Progress, Challenges and Countermeasures of Adaptive Learning: A Systematic Review. *Educational Technology and Society, 24*(3), 238–255. https://eric.ed.gov/?id=EJ1305781

* Li, C., Herbert, N., Yeom, S., & Montgomery, J. (2022). Retention Factors in STEM Education Identified Using Learning Analytics: A Systematic Review. *Education Sciences, 12*(11), 781. https://doi.org/10.3390/educsci12110781

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ (Clinical Research Ed.)*, *339*, b2700. https://doi.org/10.1136/bmj.b2700

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, *45*(2), 175–194. https://doi.org/10.1177/0312896219877678

* Liz-Domínguez, M., Caeiro-Rodríguez, M., Llamas-Nistal, M., & Mikic-Fonte, F. A. (2019). Systematic Literature Review of Predictive Analysis Tools in Higher Education. *Applied Sciences, 9*(24), 5569. https://doi.org/10.3390/app9245569

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Lodhi, P., Mishra, O., Jain, S., & Bajaj, V. (2018). StuA: An intelligent student assistant. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*(2), 17–25. https://doi.org/10.9781/ijimai.2018.02.008.

* Manhica, R., Santos, A., & Cravino, J. (2022). The use of artificial intelligence in learning management systems in the context of higher education : Systematic literature review. In 2022 17th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1–6). IEEE. https://doi.org/10.23919/CISTI54924.2022.9820205

* Maphosa, M., & Maphosa, V. (2020). Educational data mining in higher education in sub-saharan africa. In K. M. Sunjiv Soyjaudah, P. Sameerchand, & U. Singh (Eds.), Proceedings of the 2nd International Conference on Intelligent and Innovative Computing Applications (pp. 1–7). ACM. https://doi.org/10.1145/3415088.3415096

Marin, V.I., Buntins, K., Bedenlier, S., & Bond, M. (2023). Invisible borders in educational technology research? A comparative analysis. *Education Technology Research & Development, 71,* 1349-1370. https://doi.org/10.1007/s11423-023-10195-3

* McConvey, K., Guha, S., & Kuzminykh, A. (2023). A Human-Centered Review of Algorithms in Decision-Making in Higher Education. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1–15). ACM. https://doi.org/10.1145/3544548.3580658

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ (Clinical Research Ed.)*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews, 4(1), 1. https://doi.org/10.1186/2046-4053-4-1

* Moonsamy, D., Naicker, N., Adeliyi, T. T., & Ogunsakin, R. E. (2021). A Meta-analysis of Educational Data Mining for Predicting Students Performance in Programming. *International Journal of Advanced Computer Science and Applications, 12*(2), 97–104. https://doi.org/10.14569/IJACSA.2021.0120213

OECD. (2021). *AI and the Future of Skills, Volume 1: Capabilities and Assessments*. OECD Publishing, Paris. https://doi.org/10.1787/5ee71f34-en

OECD.AI. (2023). AI publications by country. Visualisations powered by JSI using data from OpenAlex., accessed on 27/9/2023, www.oecd.ai .

* Otoo-Arthur, D., & van Zyl, T. (2020). A Systematic Review on Big Data Analytics Frameworks for Higher Education - Tools and Algorithms. In EBIMCS '19, Proceedings of the 2019 2nd International Conference on E-Business, Information Management and Computer Science. Association for Computing Machinery. https://doi.org/10.1145/3377817.3377836

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pennington, R., Saadatzi, M. N., Welch, K. C., & Scott, R. (2014). Using Robot-Assisted Instruction to Teach Students with Intellectual Disabilities to Use Personal Narrative in Text Messages. *Journal of Special Education Technology, 29*(4), 49–58. https://doi.org/10.1177/016264341402900404

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, *18*(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Peters, M. D. J., Marnie, C., Colquhoun, H., Garritty, C. M., Hempel, S., Horsley, T., Langlois, E. V., Lillie, E., O'Brien, K. K., Tunçalp, Ö., Wilson, M. G., Zarin, W., & Tricco, A. C. (2021). Scoping reviews: Reinforcing and advancing the methodology and application. *Systematic Reviews*, *10*(1), 263. https://doi.org/10.1186/s13643-021-01821-3

Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences*. Blackwell Publishing.

* Pinto, A. S., Abreu, A., Costa, E., & Paiva, J. (2023). How Machine Learning (ML) is Transforming Higher Education: A Systematic Literature Review. *Journal of Information Systems Engineering and Management, 8*(2), 21168. https://doi.org/10.55267/iadt.07.13227

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2017). Overviews in Education Research. *Review of Educational Research*, *87*(1), 172–203. https://doi.org/10.3102/0034654316631117

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. ArXiv. https://arxiv.org/abs/2205.01833

* Rabelo, A., Rodrigues, M. W., Nobre, C., Isotani, S., & Zárate, L. (2023). Educational data mining and learning analytics: a review of educational management in e-learning. *Information Discovery and Delivery*. Advance online publication. https://doi.org/10.1108/idd-10-2022-0099

Rader, T., Mann, M., Stansfield, C., Cooper, C., & Sampson, M. (2014). Methods for documenting systematic review searches: A discussion of common issues. *Research Synthesis Methods*, *5*(2), 98–115. https://doi.org/10.1002/jrsm.1097

* Rangel-de Lázaro, G., & Duart, J. M. (2023). You Can Handle, You Can Teach It: Systematic Review on the Use of Extended Reality and Artificial Intelligence Technologies for Online Higher Education. *Sustainability, 15*(4), 3507. https://doi.org/10.3390/su15043507

Reid, J. (1995). Managing learner support. In F. Lockwood (Ed.), Open and distance learning today, (pp. 265–275). London: Routledge

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). Prisma-S: An extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, *10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Robinson, K. A., Brunnhuber, K., Ciliska, D., Juhl, C. B., Christensen, R., & Lund, H. (2021). Evidence-Based Research Series-Paper 1: What Evidence-Based Research is and why is it important? *Journal of Clinical Epidemiology*, *129*, 151–157. https://doi.org/10.1016/j.jclinepi.2020.07.020

* Saghiri, M. A., Vakhnovetsky, J., & Nadershahi, N. (2022). Scoping review of artificial intelligence and immersive digital tools in dental education. *Journal of Dental Education, 86*(6), 736–750. https://doi.org/10.1002/jdd.12856

* Salas-Pilco, S., Xiao, K., & Hu, X. (2022). Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review. *Education Sciences, 12*(8), 569. https://doi.org/10.3390/educsci12080569

* Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: a systematic review. *International Journal of Educational Technology in Higher Education, 19*(1). https://doi.org/10.1186/s41239-022-00326-w

* Sapci, A. H., & Sapci, H. A. (2020). Artificial Intelligence Education and Tools for Medical and Health Informatics Students: Systematic Review. *JMIR Medical Education, 6*(1), e19285. https://doi.org/10.2196/19285

* Sghir, N., Adadi, A., & Lahmer, M. (2022). Recent advances in Predictive Learning Analytics: A decade systematic review (2012-2022). *Education and Information Technologies, 28*, 8299–8333. https://doi.org/10.1007/s10639-022-11536-0

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ (Clinical Research Ed.)*, *350*, g7647. https://doi.org/10.1136/bmj.g7647

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). Amstar 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare

interventions, or both. *BMJ (Clinical Research Ed.)*, *358*, j4008.
https://doi.org/10.1136/bmj.j4008

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education, 188*, 104564. https://doi.org/10.1016/j.compedu.2022.104564

Siontis, K. C., & Ioannidis, J. P. A. (2018). Replication, Duplication, and Waste in a Quarter Million Systematic Reviews and Meta-Analyses. *Circulation. Cardiovascular Quality and Outcomes*, *11*(12), e005212. https://doi.org/10.1161/CIRCOUTCOMES.118.005212

* Sourani, M. (2019). Artificial Intelligence: A Prospective or Real Option for Education? *Al Jinan* الجنان, 11(1), Article 23. https://digitalcommons.aaru.edu.jo/aljinan/vol11/iss1/23

Stansfield, C., Stokes, G., & Thomas, J. (2022). Applying machine classifiers to update searches: Analysis from two case studies. *Research Synthesis Methods, 13*(1), 121-133. https://doi.org/10.1002/jrsm.1537

Stern, C., & Kleijnen, J. (2020). Language bias in systematic reviews: You only get out what you put in. *JBI Evidence Synthesis, 18*(9), 1818–1819. https://doi.org/10.11124/JBIES-20-00361

Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. *Health Information and Libraries Journal*, *36*(3), 202–222. https://doi.org/10.1111/hir.12276

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning. *Review of Educational Research, 81*(1), 4–28. https://doi.org/10.3102/0034654310393361

Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., Bond, M., & Koryakina, A. (2023). *EPPI Reviewer: Advanced software for systematic reviews, maps and evidence synthesis* [Computer software]. EPPI Centre Software. UCL Social Research Institute. London. https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4

Tran, L., Tam, D. N. H., Elshafay, A., Dang, T., Hirayama, K., & Huy, N. T. (2021). Quality assessment tools used in systematic reviews of in vitro studies: A systematic review. *BMC Medical Research Methodology*, *21*(1), 101. https://doi.org/10.1186/s12874-021-01295-w

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). Prisma Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Tsou, A.Y., Treadwell, J.R., Erinoff, E. *et al.* (2020). Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. *Systematic Reviews, 9*, 73. https://doi.org/10.1186/s13643-020-01324-7

* Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation Systems for Education: Systematic Review. *Electronics*, *10*(14), 1611. https://doi.org/10.3390/electronics10141611

* Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media & Technology*, *45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Woolf, B. P. (2010). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann.

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.

Zawacki-Richter, O. (2023). Umbrella Review in ODDE. Herbsttagung der Sektion Medienpädagogik (DGfE), 22 September 2023.

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (Eds.). (2020). *Systematic Reviews in Educational Research*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27602-7

* 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). https://doi.org/10.1186/s41239-019-0171-0

* Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence, 4*, 100134. https://doi.org/10.1016/j.caeai.2023.100134

Zhang, Q., & Neitzel, A. (2023). Choosing the Right Tool for the Job: Screening Tools for Systematic Reviews in Education. *Journal of Research on Educational Effectiveness*, 1–27. https://doi.org/10.1080/19345747.2023.2209079

* Zhang, W., Cai, M., Lee, H. J., Evans, R., Zhu, C., & Ming, C. (2023). AI in Medical Education: Global situation, effects and challenges. *Education and Information Technologies*. Advance online publication. https://doi.org/10.1007/s10639-023-12009-8

Zheng, Q., Xu, J., Gao, Y., Liu, M., Cheng, L., Xiong, L., Cheng, J., Yuan, M., OuYang, G., Huang, H., Wu, J., Zhang, J., & Tian, J. (2022). Past, present and future of living systematic review: A bibliometrics analysis. *BMJ Global Health*, *7*(10). https://doi.org/10.1136/bmjgh-2022-009378

* Zhong, L. (2022). A systematic review of personalized learning in higher education: learning content structure, learning materials sequence, and learning readiness support. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2022.2061006

* Zulkifli, F., Mohamed, Z., & Azmee, N. A. (2019). Systematic research on predictive models on students' academic performance in higher education. *International Journal of Recent Technology and Engineering, 8*(2 SI 3), Article B10610782S319, 357–363. https://doi.org/10.35940/ijrte.B1061.0782S319

Please cite as:

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S.W., & Siemens, G. (2023). A Meta Systematic Review of Artificial Intelligence in Higher Education: A call for increased ethics, collaboration, and rigour. *Pre-print*.