International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

Machine-Assisted Thematic Mapping of AI in Education Discourse: A Systematic Review from 2005-2024

INPLASY202560042
doi: 10.37766/inplasy2025.6.0042
Received: 9 June 2025
Published: 10 June 2025

Kaur, H; Pavlova, M.

Corresponding author:
Harmandeep Kaur

harman.deep8991@gmail.com

Author Affiliation:
The Hong Kong Polytechnic University.

ADMINISTRATIVE INFORMATION

Support - There are no sources of financial support.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202560042

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 10 June 2025 and was last updated on 10 June 2025.

INTRODUCTION

Review question / Objective This article is looking at the ways AI has been discussed and defined in education over the period of 2005-2024, namely through the following qualitative research questions:
1. In what ways has AIEd been perceived/ discussed till now?
2. What are the future directions and emerging trends in AI research within the educational domain?

Rationale This article aims to explore the discussions of Artificial Intelligence in Education (AIEd) by analysing the scope of debate and its applications in different educational areas.

Condition being studied The incorporation of Artificial Intelligence (AI) in education is being studied. The study explores in what ways AI is being discussed, implemented, and envisioned in the education sector.

METHODS

Search strategy
Search Procedure
The search terms were constrained to "AI in higher education", "AI in education", "Artificial intelligence in higher education", and "Artificial intelligence in education" on JSTOR, ERIC, ScienceDirect, IEEE, Taylor & Francis, and Google Scholar. The search process across the databases was conducted by only one review, the primary author, with the “peer-reviewed” filter activated. The compilation looked holistically at different aspects of AIEd, studying its applications for both teaching and learning, and beyond as in educational administration. Both within and outside class engagement in knowledge dissemination and acquisition were accounted for, including AI-facilitated teacher administrative duties such as paper marking, and student self-revisions. A limitation to acknowledge is our restriction to only the literature in English. It is ideal for future reviews to include articles from languages other than English.

Participant or population The search based on the above-mentioned criteria identified 70 peer-reviewed journal articles. All journal articles discussed how schools, educators and students adopted AI in teaching, learning, and educational administration.

Intervention This is not applicable to our study.

Comparator This is not applicable to our study.

Study designs to be included The review included a range of studies including observational studies, experimental studies, descriptive studies, and even other reviews; any academic articles studying, suggesting, or ideating the adoption of AIEd.

Eligibility criteria Selection Criteria
The selection criteria were for the journal articles to be peer-reviewed and to include the words "Artificial Intelligence/AI", and/ or "Education", "Teaching", "Learning", "Teacher", "Student", or "Pedagogy" in the title. If only either one was included, the article was scanned through to see traces of the other within the larger body of the text, so as to decide whether to include it. For instance, the word "Artificial Intelligence" may sometimes be absent from the title, but present within the article body. This was accepted. Names of such articles that were approved included "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence" and "Using Intelligent Tutor Technology to Implement Adaptive Support for Student Collaboration". This was essential because terms such as learning analytics and data mining that are often associated with AI, and singularly present in article headings, are components of itself that but in isolation do not suffice to represent AI's integration into educational processes. True AI requires the recruitment of these elements into larger automated decision-making systems, thus the assumed unsuitability of academic studies mentioning firstly only the components, and secondly without presenting their incorporation in the production of machine judgements.
The last date of search was 1 October 2024. The search based on the above-mentioned criteria identified 70 peer-reviewed journal articles.

Information sources Information sources were peer-reviewed academic articles from the following databases: JSTOR, ERIC, ScienceDirect, IEEE, Taylor & Francis, and Google Scholar.

Main outcome(s) Findings reveal a trajectory in AIEd discourse that transitions from initial technological optimism to more nuanced implementation considerations. In addition to teacher preparedness and professional development remaining an expanding area of research, other key themes developing across periods include the expanding scope of AI from a pedagogical tool to curricular content, infrastructure and environmental requirements for AI adoption, and ethical considerations for AI adoption. The findings indicate that future AIEd discourse will likely centre on institutes discussing and developing frameworks to evaluate specialized and potentially transformative general-purpose technology in the field of education, establishing collaborative decision-making processes to enhance the acceptance of AIEd amongst key stakeholders, and creating sustainable models for the continuous professional development of teachers.

Additional outcome(s) NA.

Data management Data and records were maintained by the primary author. Due to no sensitive information being stored, the method of storage was not encrypted. The author maintained a Microsoft Excel log indicating the authors, date of publication, database, title, and doi of each academic article. A separate log was maintained for academic articles from 2005-2011, 2012-2018, and 2019-2024. The first record was made on 8 January 2024, and the last one on 1 October 2024.

Quality assessment / Risk of bias analysis To assess the certainty and reliability of the machine-generated dendrograms and clusters, the entire data analysis process was repeated thrice in its entirety. This involved verifying the chronological arrangement of abstracts within their respective five-year categories, reapplying lemmatization and stop-word removal, reasserting the cosine similarity metric, and retransforming the pre-processed text via the BoW model. This was always followed by the hierarchical clustering being re-run with cluster similarity threshold set between 1.1 and 1.2. All three instances yielded identical dendrograms with the same thematic clusters, thereby affirming the robustness and reproducibility of the outcomes.

Strategy of data synthesis Data Analysis
The first stage of data analysis involved chronologically categorizing the studies into three five-year phases, namely 2005-2011 (n=23), 2012-2018 (n=19), and 2019-2024 (n=28). Although AIEd research predates 2005, this temporal

boundary was selected for its representation of a significant inflection point in the field's development; Williamson and Eynon (2020), for instance, regard the period as one where "new research fields of educational data mining and learning analytics began to emerge". Additionally, the period of early 2000s marks a quantitative expansion in AIEd research transitioning from relatively sparse contributions to a more robust body of work in recent years (Guan, 2020; Kaya, 2024), making it an ideal starting point for this study.

Orange Data Mining

In the second stage, the abstracts for each of the defined time periods were separately subjected to hierarchical clustering using the cosine similarity metric, with Orange Data Mining 3.37 (Demšar et al., 2013). Since all studies were incorporated into the machine synthesis process, no additional eligibility assessment was required at this stage. The text preprocessing involved lemmatization, a Natural Language Processing (NLP) technique that reduced words into their base forms, such as converting “runs”, “running”, and “ran” into the root form “run”, and removal of stop words such as “the”, “is”, and “and”. This standardization process aimed to improve the accuracy of the semantic analysis of the texts/ the clustering process by eliminating both the morphological variations and non-informative terms.

The pre-processed text was transformed using the Bag of Words (BoW) widget which converted textual content into numerical vectors, facilitating the analysis of abstracts based on their lexical composition. The implementation of cosine distance in this stage enabled the identification of thematic similarities by focusing on the relative distribution of terms across texts instead of their absolute word frequencies. This ensured that texts with similar conceptual structures would be grouped together even if they differed in overall length or word count. The output of this process was a series of dendrograms, each identifying clusters of abstracts with thematic similarities. To ensure consistency and interpretability across the analysis periods, the cluster similarity threshold was set between 1.1 and 1.2. These values, closer to 1, indicated a high degree of thematic similarity among the texts and consistently facilitated the formation of three distinct clusters per period (2005-2011, 2012-2018, and 2019-2024).

Subgroup analysis This was not applicable as the subject of most academic articles were Artificial Intelligence (AI) systems. The human subjects were diverse, ranging from teachers, students, curriculum developers, and even institutions exploring possibilities with AI in education (AIEd) for enhanced teaching and learning experiences. The focus of this study was not on the human subjects but the perceived benefits and potential of, and concerns regarding AIEd.

Sensitivity analysis The cluster similarity threshold for the hierarchical cluster dendrograms (the outputs of this study) was set between 1.1 and 1.2. The value, closer to 1, indicated a high degree of thematic similarity between texts grouped together. The implementation of cosine similarity metric also enabled the identification of thematic similarities by focusing on the relative distribution of terms across texts instead of their absolute word frequencies. This ensured that texts with similar conceptual structures would be grouped together even if they differed in overall length or word count.

Language restriction A limitation to acknowledge is our restriction to only the literature in English.

Country(ies) involved Hong Kong.

Other relevant information NA.

Keywords Artificial intelligence; education; educational technology; systematic literature review; ChatGPT; DeepSeek.

Dissemination plans The review is expected to be published in a Humanities and Social Sciences academic journal, focusing on contemporary education discourse. The audience are expected to be educators who reflect on such studies in discussing and deciding how to shape pedagogy and institutions to not only identify and meet contemporary teacher and learner needs, but also leverage on technology for effective teaching and learning.

Contributions of each author

Author 1 - Harmandeep Kaur - Author 1 drafted the original manuscript, conducted the methodology, edited the manuscript, and produced data visualizations.
Email: harman.deep8991@gmail.com
Author 2 - Margarita Pavlova - Author 2 edited the manuscript, and supervised the design and implementation of the study.