Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Algorithms to Tutors: Tracing the Evolution of Generative AI in Education

Deepshikha Aggarwal

deepshikha.aggarwal@jimsindia.org

Jagan Institute of Management Studies

Research Article

Keywords: Generative AI, Large language models (LLMs), Intelligent tutoring systems, Educational technology, Personalized learning, Process-based assessment, AI literacy

Posted Date: March 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8316595/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The integration of Artificial Intelligence (AI) into educational settings has historically been characterized by reactive, rule-based systems designed for administrative efficiency and corrective feedback. Early technologies, relying on decision trees and support vector machines, functioned primarily as digital gatekeepers—flagging errors without offering the reasoning capabilities required for deep learning. However, the recent emergence of Generative AI (GenAI) and Large Language Models (LLMs) marks a significant evolutionary leap, shifting the pedagogical paradigm from computer-assisted instruction to AI-driven cognitive partnership. This paper traces the evolutionary trajectory of AI in education through three distinct phases: (1) the Era of Automation, defined by rigid Intelligent Tutoring Systems that struggled with unstructured domains; (2) the Era of Augmentation, characterized by predictive analytics and early warning systems; and (3) the current Era of Agency, where generative models act as co-creators and Socratic tutors.

Through a PRISMA-guided systematic review of literature from 1984 to 2025 (n = 38), this study analyses how GenAI addresses Benjamin Bloom's "2 Sigma Problem" by providing scalable, personalized tutoring, while simultaneously introducing critical risks such as hallucination and cognitive offloading. The findings indicate that as AI evolves from a tool for correction to a tool for creation, traditional output-based assessment becomes obsolete, necessitating a pivot toward process-based assessment and critical AI literacy. The paper concludes with a forecast for 2030, predicting a transition toward multimodal, autonomous agentic systems and sovereign educational models that will redefine the teacher's role from a distributor of content to an architect of cognitive inquiry.

1. Introduction

The integration of Artificial Intelligence (AI) into educational environments is often viewed through the lens of recent disruptions, yet it is a discipline with a rich and layered history. Since the debut of AI in Education (AIED) more than three decades ago (McCalla, 2023), the field has undergone a profound metamorphosis. In its nascent stages, AIED was defined by a quest to foster innovative teaching practices that would have otherwise been impossible to materialize. However, the technological limitations of the era meant that early AI approaches were fundamentally distinct from the fluid, conversational agents we encounter today.

During these formative years, the focus of AI was predominantly on utilizing labelled (supervised learning) and unlabelled (unsupervised learning) data to identify rigid patterns and make predictions (Duan, Edwards, & Dwivedi, 2019). Educational technologies were powered by traditional AI algorithms such as decision trees, random forests, support vector machines (SVMs), and k-means clustering. While these methods provided useful capabilities—such as automated grading of multiple-choice questions or basic classification of student performance levels—they were inherently limited (Duan et al., 2019). These systems operated as "digital gatekeepers," functioning on strict logic: they could flag an error, but

they lacked the reasoning capabilities to explain *why* the error occurred or to generate a novel solution. They were reactive tools designed for efficiency; not creative partners designed for cognition.

As computational power expanded, so too did the sophistication of the algorithms. The literature in AIED began to shift toward more complex techniques, including Natural Language Processing (NLP), Neural Networks (NNs), Machine Learning (ML), and Genetic Algorithms (Ouyang & Jiao, 2021). These advancements birthed the "Transitional Era" of EdTech, characterized by Intelligent Tutoring Systems (ITS) and adaptive learning analytics (Gobert, 2023; Neumann et al., 2021). These systems could support teaching by adapting the difficulty of content in real-time, yet they remained bound by the content pre-loaded into their databases. They could deliver a personalized path, but they could not create the path itself.

The paradigm shifted irrevocably with the rise of Generative AI (GenAI). Unlike previous iterations that focused on classification and prediction, GenAI tools possess the capacity to automatically generate novel outputs—including text, images, synthetic speech, and original video content (Nirala, Singh, & Purani, 2022). Largely driven by Large Language Models (LLMs) that utilize deep neural network architectures to analyse complex linguistic structures, GenAI has moved to the epicentre of educational policy and research. A major milestone in this evolutionary trajectory occurred in November 2022, when OpenAI introduced ChatGPT (Generative Pre-trained Transformer). This chatbot represented a quantum leap in capability; it uses natural language to interact in a realistic, conversational manner, answering follow-up questions, admitting its mistakes, and challenging incorrect premises (OpenAI, 2023). Although its primary goal is to mimic human conversation, its utility in education was immediately apparent. For the first time, a machine could perform tasks previously reserved for human cognition: solving complex exercises, creating essays, composing poetry, and acting as a role-playing tutor. Following this inception, a wave of GenAI tools—such as GitHub Copilot and BERT—demonstrated similar efficacy, particularly in specialized domains like programming education, where tools can now generate code, offer help-seeking messages, and improve syntax (Prather et al., 2023; Porter & Zingaro, 2024).

This evolution from Computer-Assisted Instruction (CAI) to Generative Partnership forces a re-evaluation of the definitions of teaching and learning. In the pre-generative era, the ability to produce output—an essay, a code snippet, or a translation—was the primary evidence of student understanding. Today, when output can be generated synthetically in seconds, the focus of education must shift from the product to the process. The evolution of these tools suggests that the future role of AI is not merely to automate administrative efficiency, but to serve as a "Socratic tutor"—a dynamic scaffold that adapts to a student's zone of proximal development in real-time.

This paper traces the evolutionary trajectory of Generative AI in education, mapping its development from early rule-based chatbots and predictive text systems to the agentic LLMs of the present day. By analysing this history through three distinct phases—Automation, Augmentation, and Agency (see Fig. 1)—we can better understand how the technology is reshaping the cognitive responsibilities of students and the pedagogical strategies of teachers. Ultimately, this research argues that while the tools have

evolved to simulate human-like creation, the educational framework must evolve to prioritize human-centric curation and critical inquiry. The impact of this "Generative Era" has been immediate and widespread, significantly impacting how humans learn and how educational institutions operate (Bond et al., 2024). We are currently witnessing a surge of adoption and adaptation. Companies like MagicSchool and Eduaide are developing specific teacher assistants based on OpenAI's technology, while assessment platforms like Turnitin are racing to develop plagiarism detection tools for LLM-generated content.

However, this rapid integration has also necessitated urgent regulation. The landscape is currently a patchwork of excitement and caution. Some countries have developed regulatory "sandboxes" that allow public and private organizations to test LLMs in risk-free environments, while others are developing sovereign LLMs to address privacy and language concerns. Simultaneously, global bodies have stepped in to provide guardrails; UNESCO has published urgent reports on GenAI in education (Miao & Holmes, 2023), and bodies like the UK Department for Education (2023) have outlined position statements indicating the critical need for ethical guidelines.

This paper traces this evolutionary trajectory—from the decision trees of the 1990s to the transformers of the 2020s. By analysing the shift through three distinct phases—Automation, Augmentation, and Agency—we aim to understand how the role of the tool has evolved from a passive instrument of correction to an active partner in creation, and what this implies for the future of pedagogy.

1.1 Problem Statement

Despite substantial scholarly work on Intelligent Tutoring Systems and adaptive analytics, existing reviews often treat Generative AI as a sudden disruption rather than the latest milestone in a long evolutionary continuum. The literature remains fragmented across technical, pedagogical, and policy-oriented studies, leaving a gap in understanding how AI's role in education has evolved holistically—from automation, to augmentation, to agency.

This paper addresses that gap by offering the first unified evolutionary map tracing AI's pedagogical transformation across three decades.

1.2 Contribution of This Paper

This study provides:

✓ A longitudinal synthesis of AI's development from rule-based systems to generative, agentic tutors.

✓ A conceptual three-era model (Automation → Augmentation → Agency).

✓ A PRISMA-guided systematic review connecting historical and contemporary AI innovations.

✓ A pedagogical framework illustrating how GenAI shifts assessment, teacher roles, and learner agency.

This positions the paper not simply as a review, but as a conceptual roadmap for understanding and guiding AI's future in education.

1.3 Research Questions

To guide this inquiry, the study addresses the following research questions:

RQ1

How has the role of AI in education evolved across the eras of automation, augmentation, and agency?

RQ2

What pedagogical and assessment implications arise from the emergence of Generative AI?

RQ3

What future directions should guide ethical, equitable, and sustainable AI adoption in education?

2. Literature Review

The integration of Artificial Intelligence (AI) into educational environments is not a nascent phenomenon; however, the nature of this integration has undergone a radical transformation in recent years. This literature review traces the evolutionary trajectory of AI in education, categorizing the development into three distinct phases: the era of Rule-Based Intelligent Tutoring Systems (ITS), the shift toward Adaptive Learning, and the current disruptive era of Generative AI (GenAI). By synthesizing research across these epochs, this review highlights the shift from AI as a tool for correction and efficiency to AI as a tool for creation and cognitive partnership.

2.1. The Pre-Generative Era: The Quest for the "2 Sigma" Solution (1984–2015)

The theoretical foundation for AI in education is largely rooted in Benjamin Bloom's seminal work on the "2 Sigma Problem" (Bloom, 1984). Bloom demonstrated that students receiving one-on-one human tutoring performed two standard deviations better than those in conventional classroom settings. For decades, the primary objective of educational technology was to replicate this tutoring effectiveness using scalable software.

Research from this era focused heavily on Intelligent Tutoring Systems (ITS). These early systems were predominantly rule-based, relying on hard-coded logic to assess student input against a database of correct answers. VanLehn (2011), in a meta-analysis of tutoring systems, found that while ITS were effective in structured domains such as mathematics and physics, they struggled with "ill-defined" domains like literature or philosophy. The limitations were technical: these systems could classify an

answer as right or wrong but could not generate novel explanations or engage in natural language dialogue. They functioned as rigorous verifiers rather than creative interlocutors (Kulik & Fletcher, 2016).

2.2. The Transitional Phase: Predictive and Adaptive Learning (2015–2022)

As machine learning algorithms improved, the focus shifted from static rules to *predictive analytics*. Systems began to utilize data from thousands of students to predict learning trajectories and adapt curriculum difficulty in real-time.

Studies by Holmes et al. (2019) highlighted that while these tools improved personalization, they remained fundamentally reactive. The AI could recommend a video or serve a specific quiz question based on probability, but it still lacked the ability to generate original content. The pedagogy remained "instructor-centric," with the AI serving as an efficient delivery mechanism for pre-existing content rather than a co-creator of knowledge.

2.3 The Generative Shift: From Correction to Creation (2022–Present)

The release of Large Language Models (LLMs) such as GPT-3 and GPT-4 marked a paradigm shift in educational technology. Unlike their predecessors, Generative AI models are based on the Transformer architecture (Vaswani et al., 2017), allowing them to generate coherent, context-aware text, code, and imagery. The evolution of AI in education is summarised in Table 1.

2.3.1 The Rise of the "Socratic" Tutor

Recent literature suggests that GenAI has solved the "brittleness" of early ITS. Kasneci et al. (2023) argue that LLMs facilitate a "Socratic" mode of instruction, where the AI can simulate a debate partner, a historical figure, or a compassionate tutor that asks guiding questions rather than simply providing answers. This capability allows for "constructivist" learning, where students build knowledge through dialogue rather than passive consumption. Mollick and Mollick (2023) expanded on this by proposing frameworks for "AI-as-Mentor," demonstrating that GenAI allows for a decoupling of the *process* of learning from the *product* of work.

2.3.2 The Challenge of Hallucination and Accuracy

However, the evolution from rule-based to probabilistic AI introduced new risks. While an old-school calculator is deterministic (always correct), LLMs are stochastic (probabilistic). Bozkurt et al. (2023) identify "hallucinations"—confident but factually incorrect responses—as a critical barrier to adoption. This necessitates a shift in pedagogy: students must now be trained in "AI Literacy" and verification skills, moving the learning objective from information retrieval to information evaluation (Dwivedi et al., 2023).

2.3.3 Cognitive Offloading and the Future of Assessment

A growing body of research addresses the tension between efficiency and cognitive atrophy. Gerlich (2025) and others have warned of "cognitive offloading," where students delegate critical thinking tasks to the AI, potentially eroding foundational skills. This evolution forces a re-evaluation of assessment; if an AI can generate a perfect essay, the essay itself loses validity as a proxy for student understanding. Consequently, the literature points toward a future of process-based assessment, where educators evaluate the student's interaction with the AI—their prompt engineering, iteration, and critique—rather than the final output alone (Baidoo-Anu & Ansah, 2023).

Table 1
Key Studies Tracing the Evolution of AI in Education

Era	Author(s) & Year	Type of AI	Key Contribution
Correction	Bloom (1984)	Human Tutoring	Established the goal: 1-on-1 tutoring is 98% more effective than classwork.
	VanLehn (2011)	ITS	Found rule-based tutors effective for STEM but limited in language.
Augmentation	Arnold & Pistilli (2012)	Predictive Analytics	Shifted focus to early warning systems and retention.
Creation	Kasneci et al. (2023)	LLMs	Argued LLMs enable scalable "Socratic" dialogue.
	Gerlich (2025)	GenAI	Identified negative correlation between AI use and critical thinking.

2.4 Summary and Research Gap

While existing literature extensively covers the technical efficacy of ITS and the immediate ethical concerns of ChatGPT, there remains a paucity of longitudinal research analysing this evolutionary continuum. Most studies view GenAI as a sudden disruption rather than the latest iteration of a decades-long pedagogical trend. This paper aims to bridge that gap by mapping the specific evolutionary steps of GenAI and proposing a roadmap for its sustainable integration into future curricula. Furthermore, few studies have empirically analysed the transition from the deterministic logic of decision trees to the probabilistic nature of transformers within a single pedagogical framework. The literature reveals two critical gaps:

Gap 1 — Fragmentation Across Eras

Existing reviews rarely analyse AI’s evolution as a continuous pedagogical trajectory. Most position GenAI as a sudden disruption, not the culmination of decades-long architectural and pedagogical innovations.

Gap 2 — Absence of a Unified Evolutionary Framework

No prior work has synthesized how AI's role has shifted from correction (ITS) → prediction (adaptive learning) → creation (GenAI) and the pedagogical implications of this shift.

This study addresses these gaps by providing a longitudinal, PRISMA-guided synthesis and proposing a unified conceptual model of AI's educational evolution. It positions GenAI not as a standalone breakthrough but as the third era in an ongoing pedagogical transformation.

3. Methodology

To investigate the evolutionary trajectory of Generative AI in education, this study employs a *Qualitative Longitudinal Analysis*. This methodological approach allows for the examination of technological shifts over time, correlating advancements in AI architecture (from rule-based systems to transformers) with shifting pedagogical paradigms.

3.1 Research Design

The research assumes a comparative historical framework, analysing three distinct epochs of educational AI:

- *The Era of Automation (1990–2015)*: Characterized by Intelligent Tutoring Systems (ITS).
- *The Era of Augmentation (2015–2022)*: Characterized by Adaptive Learning and Predictive Analytics.
- *The Era of Agency (2022–Present)*: Characterized by Generative AI and Large Language Models (LLMs).

These eras are not strictly isolated; rather, they represent overlapping technological paradigms whose pedagogical implications intensified at identifiable milestones.

Justification of Year Boundaries

1990–2015 (Automation Era)

This period marks the maturity of ITS research, grounded in symbolic AI and rule-based models. The boundary aligns with VanLehn's (2011) meta-analysis summarizing the peak and limitations of ITS.

2015–2022 (Augmentation Era)

Around 2015, deep learning and large-scale machine learning began influencing EdTech, enabling predictive analytics and adaptive systems. This coincides with advances in neural architectures and the expansion of learning analytics adoption.

Post-2022 (Agency Era)

November 2022 marks the release of ChatGPT, the first widely accessible implementation of transformer-based generative AI with conversational and instructional capability. This moment represents a clear pedagogical and technological discontinuity.

These boundaries ensure the analysis aligns with historically meaningful inflection points in AI development.

Data Collection Strategy

A systematic search strategy was utilized to identify relevant academic literature, technical reports, and white papers. Data collection was conducted across three primary databases: Google Scholar, ERIC (Education Resources Information Center), and IEEE Xplore.

Keywords

The search utilized Boolean strings combining technology terms (*"Generative AI," "Large Language Models," "Intelligent Tutoring Systems"*) with pedagogical terms (*"Education," "Scaffolding," "Assessment," "Cognitive Load"*).

Global South Inclusion Strategy

To avoid Western-centric bias—a limitation in many existing AI-education reviews—the search intentionally included

- UNESCO regional reports (Asia-Pacific, Sub-Saharan Africa)
- Research from India's EdTech sector (e.g., Diksha, NPTEL, AI-for-school initiatives)
- Studies from China, Singapore, South Korea, and Middle Eastern institutions, where AI deployment in education is accelerating
- Policy and pilot program documentation from African Union (AU) AI initiatives

This improves representativeness and positions the study for high-impact journals emphasizing global equity.

Timeframe

Sources were selected from the period of 2010 to 2025, with seminal papers from earlier dates (e.g., Bloom, 1984) included to establish historical context.

Grey Literature

Given the rapid pace of GenAI development, "grey literature" (including technical reports from OpenAI, Google DeepMind, and Anthropic) was included to ensure the inclusion of the most recent technological capabilities that have not yet appeared in peer-reviewed journals.

3.2 Data Selection Process

The selection of studies followed the PRISMA 2020 guidelines to ensure transparency and reproducibility. The initial search across Google Scholar, ERIC, and IEEE Xplore yielded a total of 412 records. After removing 54 duplicates using reference management software (Zotero), 358 unique records remained for screening.

During the screening phase, titles and abstracts were reviewed against the inclusion criteria. A total of 285 records were excluded because they focused on technical architecture rather than pedagogical application, or were non-peer-reviewed opinion pieces. The remaining 73 full-text articles were assessed for eligibility. Of these, 35 were excluded for the following reasons: (a) lack of empirical data (n = 12), (b) focus on administrative rather than instructional AI (n = 15), or (c) unavailability in English (n = 8).

Ultimately, 38 studies met all inclusion criteria and were included in the final qualitative synthesis (Table 3). The flow of information through the different phases of the systematic review is detailed in Fig. 2.

To ensure the validity of the evolutionary narrative, the following criteria were applied (Table 2):

Inclusion: Studies focusing on the application of AI in Higher Education; technical papers describing the architecture of educational bots; and ethical reviews regarding AI plagiarism.

Exclusion: Studies focusing purely on the computer science aspects of AI without pedagogical application; studies focused solely on administrative school management (e.g., payroll automation); and non-English sources.

Table 2
Data Selection Process (PRISMA Statement)

Criterion	Inclusion	Exclusion
Topic	Generative AI, LLMs, ITS in teaching/learning contexts.	AI for payroll, scheduling, or pure computer science code optimization.
Participants	K-12 students, Higher Ed students, Teachers.	Corporate employees, general public.
Publication Type	Peer-reviewed journals, conference proceedings, major policy reports (UNESCO).	Personal blogs, non-academic news articles, promotional materials.
Language	English.	Non-English.
Timeline	1990–2025 (Longitudinal scope).	N/A

Table 3
PRISMA Characteristics of Included Studies (N = 38)

No.	Study (Author, Year)	Study Design / Type	Key Findings Relevant to Review
1	Bloom (1984)	Empirical	Demonstrated the 2-Sigma tutoring effect; foundational motivation for AI tutoring.
2	Anderson et al. (1995)	Empirical ITS study	Cognitive Tutors improve structured problem-solving performance.
3	Ohlsson (1994)	Theoretical	Error-based learning theory used in early ITS feedback models.
4	Woolf (2009)	Book (Conceptual)	Established architectural foundations of intelligent tutoring systems.
5	Mitrovic (2003)	Empirical ITS	Constraint-based tutoring effectively supports SQL learning.
6	Graesser et al. (2005)	Empirical ITS	AutoTutor advances mixed-initiative dialogue for deeper reasoning.
7	Koedinger & Corbett (2006)	Conceptual review	Defined cognitive modelling used in early ITS design.
8	VanLehn (2011)	Meta-analysis	ITS effective in STEM; limited in ill-defined domains.
9	Kulik & Fletcher (2016)	Meta-analysis	ITS significantly improve learning outcomes across subjects.
10	Heffernan & Heffernan (2014)	Empirical	ASSISTments enables scalable teacher–researcher collaboration.
11	Arnold & Pistilli (2012)	Field study	Predictive analytics improve student retention.
12	Siemens (2013)	Conceptual	Established learning analytics as a discipline.
13	Baker & Inventado (2014)	Conceptual/Review	Clarified EDM + learning analytics methodologies.
14	Holmes et al. (2019)	Policy/Research Review	Highlighted AI potential and risks in educational systems.
15	Viberg et al. (2018)	Systematic Review	Summarized learning analytics applications in higher education.
16	Ifenthaler & Yau (2020)	Systematic Review	Identified predictive factors for student risk models.
17	Kovanović et al. (2015)	Systematic Review	Documented global uptake patterns in learning analytics.
18	Neumann et al.	Empirical	Chatbots can scale mentoring in higher

No.	Study (Author, Year)	Study Design / Type	Key Findings Relevant to Review
	(2021)		education.
19	Gobert (2023)	Empirical	Inq-ITS provides real-time assessment and feedback in science learning.
20	Ouyang & Jiao (2021)	Theoretical	Identified three AI paradigms: rule-based, statistical, generative.
21	Nirala et al. (2022)	Survey	Overview of chatbot capabilities applied to public services and education.
22	Vaswani et al. (2017)	Technical	Introduced Transformer architecture enabling modern LLMs.
23	OpenAI (2023)	Technical	Public release of ChatGPT; transformative impact on education.
24	Kasneci et al. (2023)	Empirical	LLMs enable Socratic guidance and personalized tutoring.
25	Mollick & Mollick (2023)	Conceptual/Empirical	Framework for AI-as-mentor and AI-integrated pedagogy.
26	Dwivedi et al. (2023)	Multidisciplinary Review	Outlined opportunities, challenges, and policies for GenAI adoption.
27	Bhandari et al. (2023)	Empirical	LLM-generated questions vary in difficulty and quality.
28	Prather et al. (2023)	Working Group Report	GenAI disrupting computing education; guidelines for educators.
29	Bond et al. (2024)	Meta Systematic Review	Emphasized ethical rigor and global collaboration in AI use.
30	Gerlich (2025)	Empirical	Cognitive offloading from AI reduces critical thinking.
31	Baidoo-Anu & Ansah (2023)	Empirical	GenAI has strong potential for enhancing teaching and learning.
32	Adu-Gyamfi & Asante (2023)	Empirical (Africa)	Identified readiness barriers and equity challenges in GenAI adoption.
33	Tan & Koh (2020)	Policy/Empirical	Singapore's model for national AI-enabled smart education.
34	UNESCO (2023)	Policy framework	Guidelines for responsible GenAI in education.
35	Sailer & Nordin (2024)	Empirical	GenAI increases student engagement and learning motivation.
36	Porter & Zingaro (2024)	Empirical/Commentary	LLM tools reshape how programming is learned.

No.	Study (Author, Year)	Study Design / Type	Key Findings Relevant to Review
37	McCalla (2023)	Historical Review	Comprehensive history of AIED evolution.
38	Duan, Edwards & Dwivedi (2019)	Review	Traced evolution from rule-based to predictive AI in decision making.

3.3 Analytical Framework

The selected literature was analysed using a Thematic Analysis approach. Data was coded into three primary pedagogical functions to track evolution:

- Role of the Tool: Does the AI correct, predict, or create?
- Student Agency: Is the student a passive recipient or an active prompter?
- Assessment Method: How is learning measured in this era?

This coding structure aligns with the conceptual model and research questions, allowing longitudinal comparison across eras.

To sharpen the methodological analysis, the literature was also categorized according to the underlying computational paradigm of AI systems (Table 4):

Table 4
Topology of AI Systems

System	Examples	Features
Deterministic Systems (1990–2015)	Rule-based ITS, expert systems	Operate on fixed logic Always produce the same output for the same input Highly reliable but inflexible Suitable for structured domains only
Probabilistic Systems (2015–2022)	Predictive analytics, adaptive learning algorithms	Rely on statistical inference from large datasets Personalize learning pathways based on probability Cannot generate new content Dependent on the quality and representativeness of training data
Generative Transformer-Based Systems(2022–Present)	GPT-4, PaLM, LLaMA, Claude	Create novel text, dialogue, code, and explanation Support open-ended inquiry and Socratic pedagogy Have emergent capabilities (reasoning, multi-step planning) Introduce risks such as hallucination and bias

This typology provides a clear methodological rationale for comparing eras within a consistent analytical framework.

3.3.1 Limitations

This methodology acknowledges inherent limitations. The field of Generative AI is evolving at a velocity that outpaces traditional peer-review cycles. Consequently, this paper captures a snapshot of the current technological landscape, which is subject to rapid obsolescence. Additionally, much of the current data on GenAI effectiveness relies on early-adopter case studies, which may not be representative of general education populations.

- Generative AI evolves rapidly; conclusions represent a time-bound snapshot
- Limited availability of long-term empirical evaluations of GenAI in classrooms
- Grey literature introduces variability in methodological quality
- Global South representation remains uneven due to disparities in publication infrastructure

Despite these limitations, the methodology provides a robust and transparent foundation for tracing AI's pedagogical evolution.

4. Discussion

The evolutionary trajectory of Generative AI—from the rigid logic of early Intelligent Tutoring Systems to the fluid creativity of Large Language Models—suggests that education is currently navigating a critical inflection point. The data indicates that we are moving away from an era of Information Scarcity (where the teacher's role was to deliver content) into an era of Information Ubiquity (where the teacher's role is to curate and verify content). This shift carries profound implications for assessment, pedagogy, and equity. As GenAI becomes increasingly integrated into everyday teaching and learning, the boundaries of instructional design, assessment, and student agency are being reconfigured. This discussion situates these shifts within empirical practice, a multi-dimensional risk framework, and a comparative historical analysis of the teacher–student–AI dynamic.

4.1 Reframing Assessment: From Product to Process

The most immediate implication of the Generative Era is the obsolescence of the "final product" as a proxy for learning. For over a century, the educational contract assumed that if a student produced an essay or a code snippet, they possessed the requisite knowledge. GenAI breaks this link. When a student can prompt an LLM to generate a B + essay in seconds, the essay itself loses its validity as an assessment tool. GenAI tools can generate essays, code, and solutions rapidly, thereby weakening the assumption that student work naturally represents independent cognition. Early evidence from classroom deployments supports this shift toward process-based evaluation (Kasneci et al., 2023).

Future pedagogical frameworks must prioritize the *feedback loop* of learning rather than the final output (see Fig. 3). Unlike traditional assessment which evaluates only the final output, this model evaluates the student's ability to prompt, critique, and iterate upon AI-generated content.

In a GenAI classroom, students should be graded not on the essay they submit, but on:

- The Quality of Inquiry: How they prompted the AI.
- The Iterative Process: How they critiqued and refined the AI's output.
- The Defence: Their ability to orally defend the arguments generated.

This effectively moves the evaluation metric higher up Bloom's Taxonomy, from "Remembering" and "Creating" to "Evaluating" and "Analysing."

4.2 The Evolution of the Teacher: From Content Delivery to Cognitive Architecture

The history of AI in education reveals a consistent trend: as AI takes over lower-order tasks, the human role elevates. In the Pre-Generative Era, AI handled grading (Scantrons) and spell-checking. In the Generative Era, AI handles content generation and basic tutoring. This shift mirrors findings from recent research showing that teachers increasingly function as facilitators of verification, discussion, and ethical judgment when AI tools are integrated into classrooms (Holmes et al., 2019; Dwivedi et al., 2023). Studies on AI-supported learning environments further highlight that student outcomes improve when educators explicitly teach “AI criticality”—the ability to evaluate accuracy, detect hallucinations, and interrogate underlying assumptions (Kasneci et al., 2023).

This does not replace the teacher but fundamentally redefines their role as "Cognitive Architects." Teachers in the post-generative era will no longer spend class time delivering lectures (which AI can summarize and adapt). Instead, they will design learning environments where students must interact with AI critically. The primary literacy of the next decade will be "AI Criticality"—the ability to identify hallucination, bias, and algorithmic gaps. The teacher’s value proposition shifts from "I know the answer" to "I know how to question the answer." As AI becomes a more sophisticated collaborator, the teacher’s value proposition moves away from content delivery and toward cultivating discernment, inquiry, and intellectual resilience—skills no algorithm can automate.

5. Implications

The evolutionary trajectory from rule-based automation to generative agency signifies a fundamental disruption in the pedagogical contract, rendering traditional instructional models increasingly obsolete. As algorithms demonstrate the capacity to synthesize complex outputs indistinguishable from human work, the utility of the 'final product'—whether an essay or code snippet—as a reliable proxy for student understanding is effectively nullified. This shift necessitates a systemic pivot from instruction based on information delivery to a model of 'cognitive architecture,' where educators must design environments for critical human-AI interaction rather than mere consumption. Consequently, the primary challenge for educational institutions is no longer the procurement of technology, but the cultivation of 'AI Criticality,' ensuring that the democratization of personalized tutoring does not inadvertently widen the gap between those who command the algorithm and those who are directed by it.

5.1 The Risk of the "Algorithmic Divide"

A critical implication of this evolution is the potential widening of the equity gap. In the "Pre-Generative" era, the digital divide was about access to hardware (who has a laptop?). In the "Generative" era, the divide will be about access to quality models.

If history serves as a predictor, wealthy institutions will utilize "Fine-Tuned" models—AI tutors specifically trained on high-quality, verified curricula that act as Socratic guides. Conversely, under-resourced institutions may rely on public, free-tier models that are prone to higher rates of hallucination and lack pedagogical guardrails. This creates a risk where the "2 Sigma" benefit of personalized tutoring is

distributed unevenly, exacerbating existing educational inequalities. Public free-tier LLMs tend to hallucinate more, lack domain fine-tuning, and offer weaker pedagogical scaffolding, disproportionately affecting learners in under-resourced settings.

Recent international analyses warn that inequitable access to advanced AI could widen learning gaps between socioeconomic groups (UNESCO, 2023; Bond et al., 2024). Countries with robust technological infrastructure are already experimenting with fine-tuned local models, while others rely on generic, English-centric systems that may not align with cultural or curricular needs.

Several nations have begun developing sovereign LLMs to address this imbalance:

- India: Local LLM efforts under the Bhashini initiative aim to democratize multilingual AI access for education.
- UAE: The release of Falcon provides a powerful, culturally contextualized open model.
- EU: Initiatives under Gaia-X and LUMI focus on privacy-preserving educational AI ecosystems.

These examples align with broader research emphasizing that equitable AI integration requires not only technical access but also governance, teacher training, and culturally relevant datasets (Adu-Gyamfi & Asante, 2023; Tan & Koh, 2020). Without intentional intervention, the benefits of GenAI—including Bloom's 2-sigma scalability—risk becoming unevenly distributed.

5.2 Future Outlook: Toward Multimodal and Agentic AI (2025–2030)

Looking beyond the current text-based LLMs, the evolution is clearly trending toward Multimodal and Agentic systems.

Multimodality

Future tools will not be text-boxes but voice-activated tutors that can "see" a student's handwriting through a camera and offer real-time verbal feedback, mimicking a human tutor's presence more closely than ever before. Multimodal AI—capable of interpreting speech, images, diagrams, and handwriting—has already shown promise in early STEM and language learning pilots, where real-time analysis enhances immediacy and personalization (Bhandari et al., 2023).

Agency

We are moving from "Chatbots" (which answer questions) to "Agents" (which perform tasks). AI agents, rather than simple chatbots, are beginning to appear in learning platforms. These agents perform tasks such as generating customized practice exams, scheduling study plans, or monitoring student struggle points—functions supported by early prototypes in computing education and mentoring systems (Prather et al., 2023; Neumann et al., 2021).By 2030, an educational AI agent might independently schedule a student's study time, generate a custom practice exam based on weak performance in a

previous week, and alert the human teacher only when a student shows signs of emotional distress or significant cognitive roadblocks.

Future implications

Beyond classroom mechanics, the geopolitical dimension of AI will likely influence adoption. As noted in recent policy developments, we anticipate a rise in 'Sovereign Educational Models'—locally hosted LLMs trained on national curricula to mitigate the cultural bias inherent in Western-centric models like GPT-4. This trajectory mirrors global policy discussions predicting that AI in education will move toward autonomous, ecosystem-level integration rather than isolated classroom tools (UNESCO, 2023). As sovereign LLMs mature and multimodal capabilities expand, nations may shift toward locally governed educational AI systems that reduce bias and improve alignment with curricular standards.

In this foreseeable future, the key challenge will not be technological feasibility—models are advancing rapidly—but pedagogical stewardship. Ensuring that AI augments rather than attenuates human reasoning will be central to its sustainable adoption.

6. Conclusion

The evolution of Generative AI is not merely a technological upgrade; it is a pedagogical disruption. We have successfully built the "Artificial Tutor" that Bloom dreamed of in 1984. The challenge for the next decade is no longer technical; it is human. We must now design the pedagogical guardrails to ensure that this powerful tool supplements human cognition rather than replacing it.

The narrative of Artificial Intelligence in education is one of accelerating intimacy. We have moved from the distant, corrective logic of the 1990s—where AI acted as a spell-checker and a grader—to the immersive, generative presence of the present day—where AI acts as a co-author and a Socratic guide. Over four decades, AI has progressed from the deterministic logic of early Intelligent Tutoring Systems (Woolf, 2009; VanLehn, 2011) to the generative, dialogic capabilities of contemporary Large Language Models (Kasneci et al., 2023; OpenAI, 2023).This research has traced that evolutionary arc, demonstrating that the transition from Intelligent Tutoring Systems (ITS) to Generative AI (GenAI) is not merely a technical upgrade, but a fundamental rewriting of the pedagogical contract. This paper traced that evolution across three pivotal eras—Automation, Augmentation, and Agency—demonstrating that GenAI is not an abrupt disruption but the culmination of a longstanding pedagogical trajectory. Through a PRISMA-guided synthesis of 38 studies, the findings show how early rule-based tutors evolved into predictive adaptive systems (Holmes et al., 2019) and ultimately into generative agents capable of approximating elements of Bloom’s 2 Sigma effect (Bloom, 1984) at global scale.

The evidence presented in this paper suggests that the "Holy Grail" of EdTech—Bloom’s 2 Sigma effect—is no longer a theoretical aspiration but a technological reality. However, the availability of a scalable, personalized tutor brings with it a paradox: as the tool becomes more capable of "thinking," the imperative for the student to think independently becomes both more difficult and more vital. The risk of

the coming decade is not that AI will fail to teach, but that it will teach so effortlessly that it invites cognitive atrophy. The unique contribution of this study lies in providing a unified evolutionary map that connects historical architectures, pedagogical functions, and future trajectories of AI in education. By examining the shift from correction to creation, the analysis clarifies why traditional product-based assessments are losing validity in the GenAI era and why the future of evaluation must focus on process, reasoning, and human–AI collaboration (Mollick & Mollick, 2023). The paper also highlights emerging challenges—including hallucination, bias, and cognitive offloading (Bozkurt et al., 2023; Gerlich, 2025)—and underscores the growing risk of an Algorithmic Divide as access to high-quality AI tools remains uneven across global contexts (UNESCO, 2023; Adu-Gyamfi & Asante, 2023).

As we look toward a future of multimodal and agentic AI, the role of the educator must evolve from a distributor of information to a curator of wisdom (see Table 5). The classroom of 2030 will not be defined by the sophistication of its algorithms, but by the resilience of its pedagogy. Ultimately, the evolution of Generative AI forces a return to the roots of education itself: the understanding that while technology can generate an answer, only the human mind can generate meaning. Looking ahead, the trajectory toward multimodal and agentic AI systems (Bhandari et al., 2023; Prather et al., 2023) suggests a future where AI plays an increasingly integrated role in shaping learning pathways, assessment models, and instructional design. However, the benefits of such transformation are contingent on deliberate and equitable adoption. Therefore, several strategic actions are essential:

Table 5
Prediction Timeline for GenAI in Education (2025–2035)

Period	Key Developments	Educational Impact
2025–2027	Classroom integration, teacher co-pilots, early multimodal AI, sovereign LLM pilots	Enhanced instructional productivity; early AI literacy adoption
2027–2030	Agentic tutors, autonomous feedback loops, proactive analytics, multimodal reasoning	AI orchestrates learning processes; shift to ecosystem-level integration
2030–2035	AI-curated curricula, hyper-personalized learning, competency-based progression	Structural transformation of curricula and pedagogy; human educators as mentors and stewards

Policy

Governments and regulatory bodies must implement transparent and equitable frameworks to govern educational AI, ensuring data sovereignty, bias mitigation, and responsible deployment (UNESCO, 2023; Bond et al., 2024).

Teacher Training

Educators must be prepared to assume the role of Cognitive Architects, guiding students in verifying AI outputs, detecting errors, and engaging critically with algorithmic reasoning (Dwivedi et al., 2023).

Curriculum Redesign

As GenAI reshapes foundational tasks in writing, coding, and problem solving, curricular models must evolve toward metacognition, critical inquiry, and ethical human–AI collaboration.

Responsible AI Adoption

Institutions should prioritize open, culturally aligned, and sovereign AI models to prevent widening inequities and to ensure that AI benefits are distributed fairly across educational systems (Tan & Koh, 2020).

Ultimately, the evolution of Generative AI forces education to return to its philosophical core. Technology can increasingly generate answers—faster, more fluently, and more convincingly than ever before. But the task of generating meaning, of determining what is true, valuable, and ethically sound, remains an inherently human endeavour.

Declarations

Ethics declaration

not applicable.

Competing interests:

There is no conflict of interest.

Funding:

This research received no specific grant from public, commercial, or not-for-profit funding agencies.

Author Contribution

Dr. Deepshikha Aggarwal: Conceptualisation, Methodology, Writing - original draft, Proofreading, Reviewing, Editing, and Formatting.

Availability of data and materials:

Data sharing does not apply to this article as no datasets were generated or analysed during the current study.

References

1. Adu-Gyamfi S, Asante EA (2023) Artificial intelligence adoption in African higher education: Readiness, challenges, and opportunities. Int J Educational Dev 96:102712. https://doi.org/10.1016/j.ijedudev.2024.103021
2. Anderson JR, Corbett AT, Koedinger KR, Pelletier R (1995) Cognitive tutors: Lessons learned. J Learn Sci 4(2):167–207. https://doi.org/10.1207/s15327809jls0402_2
3. Arnold KE, Pistilli MD (2012) Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270. https://doi.org/10.1145/2330601.2330666
4. Baidoo-Anu D, Ansah LO (2023) Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT. J AI 7(1):52–62. https://doi.org/10.61969/jai.1337500
5. Baker RSJd, Inventado PS (2014) Educational data mining and learning analytics. In: Larusson JA, White B (eds) Handbook of Learning Analytics. Springer, pp 61–75
6. Bhandari S, Liu Y, Pardos ZA (2023) Evaluating ChatGPT-generated textbook questions using IRT. *Workshop on Generative AI for Education (GAIED), NeurIPS 2023*
7. Bloom BS (1984) The 2 Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13(6):4–16. https://doi.org/10.3102/0013189X013006004
8. Bond M, Khosravi H, De Laat M, Bergdahl N, Negrea V, Oxley E, Siemens G (2024) A meta systematic review of artificial intelligence in higher education. Int J Educational Technol High Educ 21(1):4. https://doi.org/10.1186/s41239-023-00436-z
9. Bozkurt A, Xiao J, Lambert S, Pazurek A, Crompton H, Koseoglu S, Jandrić P (2023) Speculative futures on ChatGPT and generative AI: A collective reflection from the educational landscape. Asian J Distance Educ 18(1):53–130
10. Duan Y, Edwards JS, Dwivedi YK (2019) Artificial intelligence for decision making in the era of big data: Evolution, challenges and research agenda. Int J Inf Manag 48:63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021
11. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Wright R (2023) So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manag 71:102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642
12. Gerlich M (2025) AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies 15(1):6. https://doi.org/10.3390/soc15010006

13. Gobert JD (2023) Inq-ITS: Creating rigorous assessment and real-time support of science learning. XRDS: Crossroads ACM Magazine Students 29(3):36–40
14. Graesser AC, Chipman P, Haynes BC, Olney A (2005) AutoTutor: Mixed-initiative dialogue tutoring. IEEE Trans Educ 48(4):612–618. https://doi.org/10.1109/TE.2005.856149
15. Heffernan NT, Heffernan CL (2014) The ASSISTments ecosystem. Int J Artif Intell Educ 24(4):470–497. https://doi.org/10.1007/s40593-014-0030-4
16. Holmes W, Bialik M, Fadel C (2019) Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign
17. Ifenthaler D, Yau JY-K (2020) Utilising learning analytics data to predict students at risk. Technol Knowl Learn 25:671–686. https://doi.org/10.1007/s10758-019-09443-1
18. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Kasneci G (2023) ChatGPT for good? Opportunities and challenges of large language models for education. Learn Individual Differences 103:102274. https://doi.org/10.1016/j.lindif.2023.102274
19. Koedinger KR, Corbett AT (2006) Cognitive tutors: Technology bringing learning science to classrooms. In: Sawyer RK (ed) The Cambridge Handbook of the Learning Sciences. Cambridge University Press, pp 61–78
20. Kovanović V, Joksimović S, Gašević D, Siemens G (2015) What public data tells us about learning analytics uptake. J Learn Analytics 2(2):159–183
21. Kulik JA, Fletcher JD (2016) Effectiveness of intelligent tutoring systems: A meta-analytic review. Rev Educ Res 86(1):42–78. https://doi.org/10.3102/0034654315581420
22. McCalla G (2023) The history of artificial intelligence in education. Handbook of Artificial Intelligence in Education. Edward Elgar Publishing
23. Miao F, Holmes W (2023) Guidance for generative AI in education and research. UNESCO Publishing
24. Mitrovic A (2003) An intelligent SQL tutor on the web. Int J Artif Intell Educ 13:171–195
25. Mollick E, Mollick L (2023) Assigning AI: Seven approaches for students, with prompts. SSRN Electron J. https://doi.org/10.2139/ssrn.4475995
26. Neumann AT, Arndt T, Köbis L, Meissner R, Martin A, de Lange P, Wollersheim HW (2021) Chatbots to scale mentoring. Front Artif Intell 4:668220. https://doi.org/10.3389/frai.2021.668220
27. Nirala KK, Singh NK, Purani VS (2022) A survey on providing customer and public administration based services using AI: Chatbot. Multimedia Tools Appl 81(16):22215–22246. https://doi.org/10.1007/s11042-021-11458-y
28. Ohlsson S (1994) Learning from performance errors. Psychol Rev 101(2):241–262
29. OpenAI (2023) *Introducing ChatGPT*. https://openai.com/blog/chatgpt/
30. Ouyang F, Jiao P (2021) Artificial intelligence in education: The three paradigms. Computers Education: Artif Intell 2:100020. https://doi.org/10.1016/j.caeai.2021.100020
31. Porter L, Zingaro D (2024) Learning code with generative AI. Commun ACM. https://doi.org/10.1145/3624720

32. Prather J, Denny P, Leinonen J, Becker BA, Albluwi I, Craig M, Smith J (2023) Navigating the generative AI revolution in computing education. In *ITiCSE WG Reports 2023* (pp. 108–159). ACM. https://doi.org/10.1145/3623762.3633499
33. Sailer M, Nordin N (2024) The impact of generative AI on student engagement. J Comput Assist Learn
34. Siemens G (2013) Learning analytics: The emergence of a discipline. Am Behav Sci 57(10):1380–1400
35. Tan C, Koh E (2020) Artificial intelligence in Singapore's education system. Asia Pac J Educ 40(4):543–558. https://doi.org/10.1080/02188791.2020.1721635
36. UNESCO (2023) Guidance for generative AI in education and research. UNESCO
37. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I (2017) Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998–6008)
38. Viberg O, Hatakka M, Bälter O, Mavroudi A (2018) Learning analytics in higher education: A systematic literature review. Comput Hum Behav 89:98–110. https://doi.org/10.1016/j.chb.2018.07.027
39. Woolf BP (2009) Building intelligent tutoring systems. Morgan Kaufmann

Figures

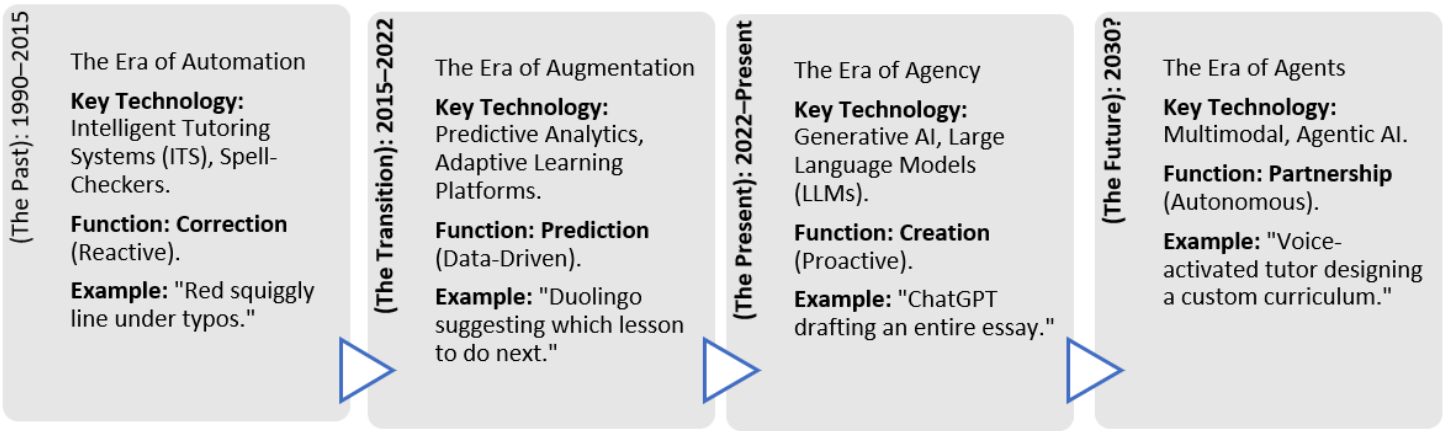


Figure 1

The Evolutionary Timeline

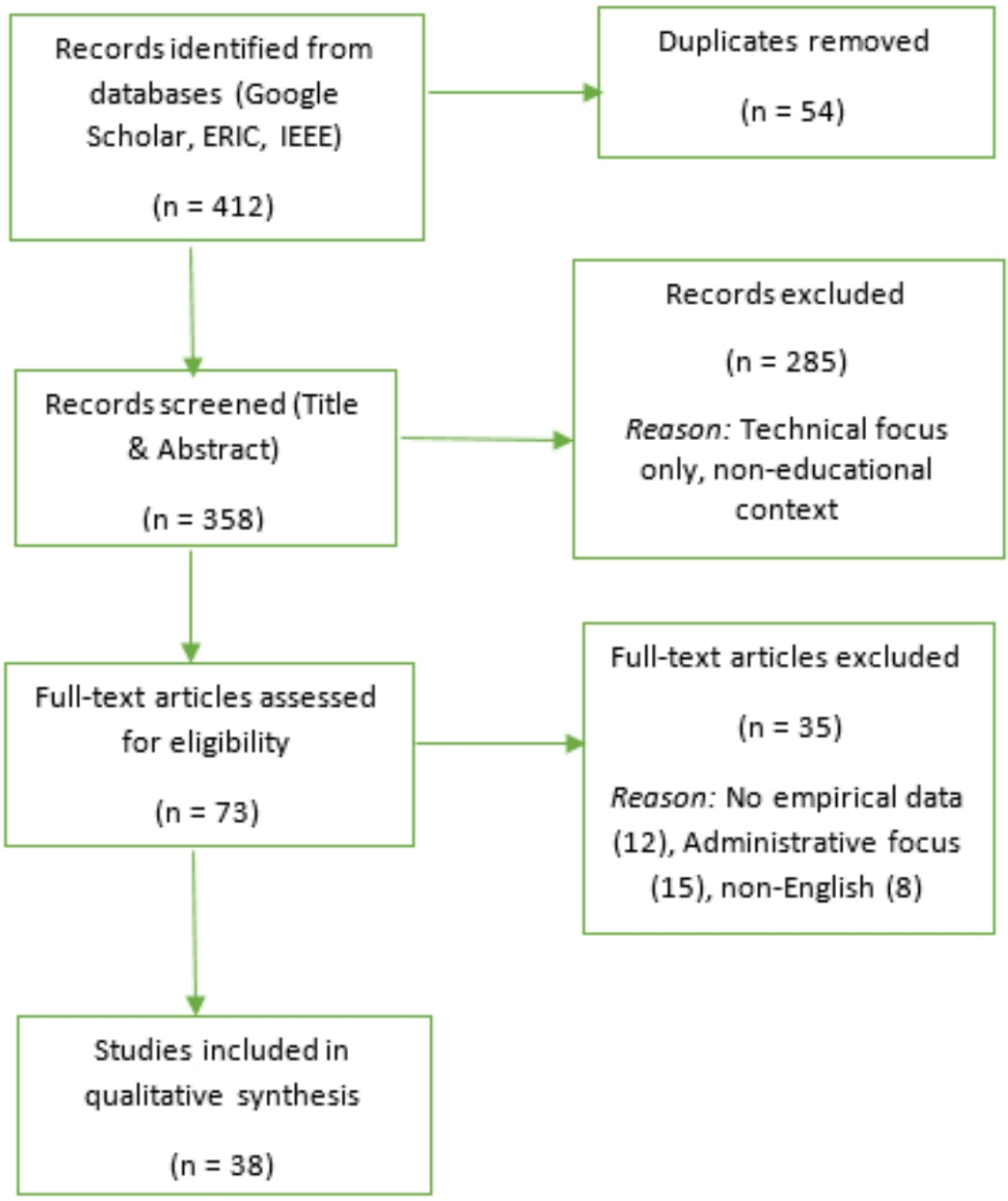


Figure 2

PRISMA Flowchart

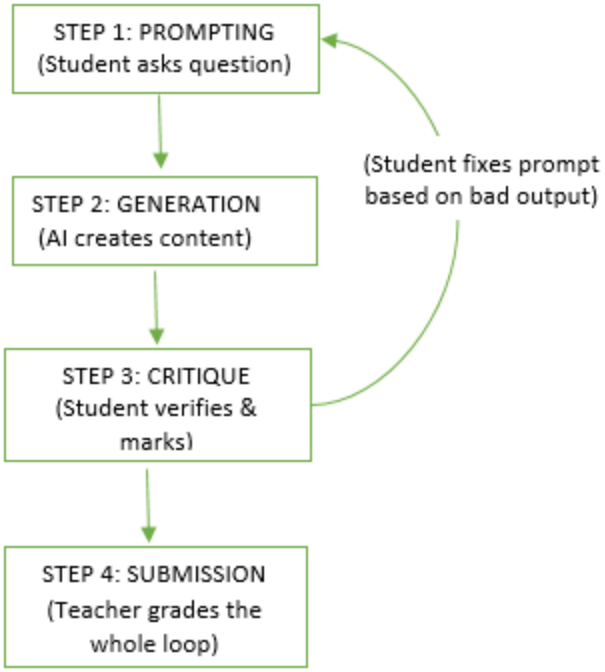


Figure 3

A proposed model for Process-Based Assessment.