Toward Human-Centered Agentic AI in Education: A Systematic Review and Multi-Layer Integration Framework

Tutik Lestari[1], Supriadi Panggabean[2], Dana Indra Sensuse[3], Hammam Riza[4], Dito Eka Cahya[5]
Darunnajah University[12], Universitas Indonesia[2], Badan Riset dan Inovasi Nasional[45]
tutik.lestari@darunnajah.ac.id[1], supriadipanggabean@darunnajah.ac.id[2], dana@cs.ui.ac.id[3], hamm001@brin.go.id[4], dito001@brin.go.id[5]

Abstract

The rapid integration of agentic artificial intelligence (AI) in educational environments has generated growing interest in its potential to transform teaching and learning processes. This systematic literature review (SLR) synthesizes findings from 75 peer-reviewed studies published between 2025 and 2026, identified through Scopus, Web of Science, and IEEE Xplore databases, following PRISMA guidelines. The review examines the applications, dominant technologies, performance outcomes, and ethical challenges associated with agentic AI in education. Findings reveal that agentic AI systems characterized by autonomy, goal-directed behavior, and adaptability are increasingly deployed in personalized learning, adaptive assessment, and collaborative learning environments. Machine learning and natural language processing emerged as dominant enabling technologies. Despite notable improvements in student engagement and learning outcomes, significant challenges persist, including data privacy concerns, implementation costs, insufficient educator training, and ethical risks such as algorithmic bias. A novel multi-layer conceptual framework is proposed to guide the ethical and pedagogically grounded integration of agentic AI. Future research directions are outlined across short-term, medium-term, and long-term horizons, emphasizing interdisciplinary collaboration, scalable deployment, and governance.

Keywords: *agentic AI, artificial intelligence in education, systematic literature review, personalized learning, adaptive learning, machine learning, natural language processing, educational technology, PRISMA.*

Abstrak

Integrasi pesat kecerdasan buatan agenik (Agentic Artificial Intelligence/AI) dalam lingkungan pendidikan telah memunculkan minat yang semakin besar terhadap potensinya untuk mentransformasi proses pembelajaran dan pengajaran. Tinjauan literatur sistematis (Systematic Literature Review/SLR) ini mensintesis temuan dari 75 studi yang telah melalui proses peer-review dan dipublikasikan antara tahun 2025 hingga 2026, yang diidentifikasi melalui basis data Scopus, Web of Science, dan IEEE Xplore dengan mengikuti pedoman PRISMA. Tinjauan ini mengkaji aplikasi, teknologi dominan, hasil kinerja, serta tantangan etis yang terkait dengan penerapan AI agenik dalam pendidikan. Hasil penelitian menunjukkan bahwa sistem AI agenik—yang dicirikan oleh otonomi, perilaku berorientasi tujuan, dan kemampuan beradaptasi—semakin banyak diterapkan dalam pembelajaran personalisasi, asesmen adaptif, dan lingkungan pembelajaran kolaboratif.

Pembelajaran mesin (machine learning) dan pemrosesan bahasa alami (natural language processing) muncul sebagai teknologi utama yang mendukung implementasi tersebut. Meskipun terdapat peningkatan yang signifikan dalam keterlibatan peserta didik dan hasil pembelajaran, berbagai tantangan penting masih perlu diatasi, termasuk isu privasi data, tingginya biaya implementasi, kurangnya pelatihan bagi pendidik, serta risiko etis seperti bias algoritmik. Sebagai kontribusi utama, penelitian ini mengusulkan sebuah kerangka konseptual multilapis yang dirancang untuk mendukung integrasi AI agenik secara etis dan berlandaskan prinsip-prinsip pedagogis. Selain itu, arah penelitian masa depan diuraikan dalam perspektif jangka pendek, menengah, dan panjang, dengan menekankan pentingnya kolaborasi multidisipliner, implementasi yang dapat diskalakan, serta tata kelola yang efektif.

Kata Kunci: AI agenik, kecerdasan buatan dalam pendidikan, tinjauan literatur sistematis, pembelajaran personalisasi, pembelajaran adaptif, pembelajaran mesin, pemrosesan bahasa alami, teknologi pendidikan, PRISMA.

1. Introduction

The transformative role of artificial intelligence in shaping modern education has become a defining research theme over the past decade [1]. Among the most recent developments is the emergence of *agentic AI*—systems capable of autonomous operation, goal-directed behavior, and real-time adaptation to individual learner needs [2], [3]. Unlike conventional AI tools, which respond reactively to discrete inputs, agentic AI orchestrates multi-step instructional workflows, monitors learner progress, and reconfigures content and pacing without direct human intervention [4].

The growing deployment of agentic AI across K-12 schooling, higher education, corporate training, and lifelong learning platforms reflects an urgent need for comprehensive empirical synthesis [5]. Despite a proliferating body of individual studies, the field lacks a unified review that spans methodologies, application domains, performance outcomes, and ethical dimensions simultaneously [6], [7]. Fragmented evidence makes it difficult for practitioners and policymakers to draw evidence-based conclusions [8].

This systematic literature review (SLR) addresses that gap by synthesizing 75 peer-reviewed studies following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [9]. The review is guided by five research questions:

RQ1: What are the current trends and theoretical underpinnings of agentic AI in education?

RQ2: What frameworks, methods, and architectures support its implementation?

RQ3: In which application domains is agentic AI deployed, and through what approaches?

RQ4: What challenges and limitations characterize existing implementations?

RQ5: What research gaps and future opportunities exist in this field?

The remainder of this paper is organized as follows: Section II reviews the theoretical background; Section III details the methodology; Section IV presents results; Section V provides the discussion; Section VI proposes the conceptual framework; and Section VII outlines future research directions before concluding remarks in Section VIII.

2. Theoretical Foundation and Literature Synthesis

2.1 Evolution of AI in Education

The history of AI in education spans over four decades, beginning with early intelligent tutoring systems (ITS) such as SOPHIE and LISP Tutor [10]. These rule-based systems laid the conceptual groundwork for adaptive instruction by providing personalized feedback based on student responses [11]. The subsequent introduction of machine learning enabled more sophisticated modeling of learner behavior, culminating in the adaptive learning systems prevalent today [12], [13].

The term *agentic AI* reflects a qualitative leap beyond reactive systems. Agentic systems are characterized by persistent goals, memory across sessions, tool use, and the capacity to decompose complex tasks into sub-goals [14]. In educational contexts, such systems function as co-teachers, learning companions, and automated assessors [3], [15].

2.2 Personalized Learning and Adaptive Systems

Personalized learning—the customization of instructional content, pace, and modality to individual learner profiles—represents the most extensively researched application of agentic AI [1], [2]. Studies consistently report improvements in engagement and academic performance when AI-driven personalization is employed [4], [16]. Adaptive assessments, which dynamically calibrate difficulty to learner progress, constitute a closely related domain [25].

2.3 Ethical and Equity Dimensions

Scholarship has increasingly foregrounded the ethical dimensions of AI in education. Concerns center on data privacy, algorithmic bias, lack of transparency in decision-making, and the risk of exacerbating existing educational inequalities [5], [7], [8]. Holmes et al. [7] argue that poorly designed AI systems may systematically disadvantage already-marginalized learners, underscoring the need for equity-centered design principles. Parallel debates concern the appropriate redistribution of cognitive and instructional labor between human educators and AI agents [9].

3. Research Methodology

3.1 Search Strategy

A comprehensive Boolean search was executed across three databases: Scopus, Web of Science, and IEEE Xplore. Searches were limited to publications from 2025 to 2026. The Scopus query was structured as follows:

TITLE-ABS-KEY("agentic AI" OR "autonomous AI" OR "intelligent tutoring systems" OR "adaptive learning" OR "personalized education") AND TITLE-ABS-KEY("education" OR "learning" OR "teaching") AND (LIMIT-TO(PUBYEAR, 2025) OR LIMIT-TO(PUBYEAR, 2026))

Analogous strings were constructed for Web of Science using the TS= field tag and for IEEE Xplore using the publication year filter. Duplicate records were removed using reference management software prior to screening [9].

3.2 Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Publication type	Peer-reviewed articles	Conference abstracts, editorials
Year	2025–2026	Prior to 2025
Language	English	Non-English
Topic relevance	Agentic AI in educational contexts	Technical AI without educational focus

Table I. Inclusion and Exclusion Criteria.

3.3 *PRISMA Selection Process*

The PRISMA 2020 guidelines [9] structured the selection workflow across four phases. Database searching yielded 350 initial records; supplementary hand-searching contributed 20 additional references. After deduplication, 320 records remained. Title and abstract screening eliminated 180 records, leaving 140 for full-text review. Of these, 65 were excluded for failing to meet inclusion criteria, resulting in **75 studies** included in the final synthesis.

3.4 *Quality Assessment*

Study quality was appraised using three complementary tools: the Cochrane Risk of Bias Tool for randomized controlled trials, the Newcastle-Ottawa Scale for non-randomized studies, and the Joanna Briggs Institute (JBI) Critical Appraisal Tools for qualitative and mixed-methods research [17], [18], [19].

3.5 *Data Synthesis*

A hybrid synthesis approach was employed. Quantitative findings were pooled where sufficient methodological homogeneity existed; thematic analysis [20] was applied to qualitative data to surface cross-study patterns. Bibliometric analysis characterized publication trends and citation networks.

4. Results

4.1 Publication Trends and Geographic Distribution

Year	Publications (n)
2025	40
2026	35
Total	**75**

Table II. Annual distribution of included studies.

North America produced the largest share of publications (n=30, 40%), followed by Europe (n=25, 33.3%) and Asia (n=15, 20%), consistent with regional patterns reported in bibliometric analyses of AI-education literature [21]. No included studies originated from Africa or South America, signaling a geographic research gap warranting deliberate redress [8].

4.2 Thematic Clusters

Theme	n	%
Personalized Learning	25	33.3
Implementation Challenges	20	26.7
Ethical Considerations	15	20.0
Technological Frameworks	10	13.3
Performance Outcomes	5	6.7

Table III. Thematic distribution of included studies.

4.3 Dominant Methods, Frameworks, and Technologies

Case studies, surveys, and controlled experiments constituted the predominant methodological approaches (n=40) [22]. Machine learning (ML) and natural language processing (NLP) were the most frequently employed technologies (n=30), consistent with broader trends in AI-enhanced education research [1], [4], [23]. The ADDIE and SAMR instructional design models appeared in 20 studies as guiding frameworks, while cloud-based and hybrid system architectures were reported in 25 studies.

4.4 Application Domains

Domain	Studies (n)
K-12 Education	30
Higher Education	25
Corporate Training	10
Lifelong Learning	10

Table IV. Application domains of agentic AI.

4.5 Performance Outcomes

Student engagement was addressed in 40 studies; 75% reported positive effects attributable to real-time adaptive feedback and personalized content delivery [1],

[14]. Learning outcome improvements were reported in 25 of 35 relevant studies. Teacher satisfaction was assessed in 20 studies, with 15 reporting favorable perceptions of AI-augmented instruction [4]. Ethical concerns were flagged in 15 studies, with neutral or mixed impacts on adoption decisions [5], [7].

4.6 Challenges Identified

Challenge	Studies (n)
Data Privacy	25
Implementation Costs	20
Lack of Educator Training	15
Ethical Concerns (bias, transparency)	10

Table V. Challenges identified in the literature.

4. Discussion

4.1 Research Gaps

a. Addressing the Research Questions

With respect to **RQ1**, the review confirms that personalized learning remains the dominant theoretical and practical paradigm in agentic AI research [1], [2], [3]. The proliferation of generative AI tools—including large language models deployed as tutors—marks a qualitative shift from earlier adaptive systems toward true agentic behavior [4], [6].

Addressing **RQ2**, ML and NLP dominate the technological landscape, with cloud-based architectures enabling scalability [23], [24]. The ADDIE and SAMR frameworks provided pedagogical anchoring for AI integration, though the review notes a deficit of frameworks specifically theorized for agentic—as distinct from adaptive—systems.

On **RQ3**, K-12 and higher education account for 73% of application domains, while corporate training and lifelong learning remain underexplored [5], [6]. Collaborative AI-mediated learning platforms appear across all domains but are more mature in higher education [4].

Regarding **RQ4**, data privacy emerges as the most frequently cited challenge (n=25), consistent with findings in broader AIED literature [7], [8]. The redistribution of educator roles—particularly the risk that AI-mediated instruction erodes professional autonomy—represents a structural tension requiring governance attention [9].

Concerning **RQ5**, the geographic concentration of research in North America and Europe, the near-absence of studies from the Global South, and the limited focus on experiential and informal education contexts constitute the most salient research gaps [8], [21].

b. Strengths and Limitations of Existing Work

The reviewed studies demonstrate methodological diversity and increasing empirical rigor. Strengths include the adoption of controlled experimental designs and the use of validated psychometric instruments to measure engagement and satisfaction. However, weaknesses include reliance on short-duration pilots, convenience sampling, and the under-representation of qualitative and participatory approaches that could illuminate lived student experience [20], [22].

c. Emerging Trends

Three emergent trends warrant particular attention: (1) multimodal agentic systems that integrate text, voice, and gesture recognition for richer learner modeling [6]; (2) affective co-learning agents that monitor emotional states to adjust instructional tone and difficulty [15]; and (3) the deployment of generative AI as a conversational tutoring scaffold, with implications for both learning gains and academic integrity [7], [8].

4.2 Proposed Conceptual Framework

To address the integration challenges identified, this review proposes a novel **Multi-Layer Agentic AI Integration Framework (MLAIF)** comprising nine interdependent layers:

Layer	Function and Theoretical Basis
Input Layer	Captures multi-source data: student interactions, assessments, behavioral signals, and instructor inputs [14].
Data Layer	Processes, cleans, and stores collected data; enforces privacy-by-design principles [7].
Knowledge Layer	Embeds pedagogical theories (e.g., constructivism, Bloom's Taxonomy) as structured priors [11].
Processing Layer	Applies ML and NLP algorithms to detect learning patterns and generate predictive insights [23].
Intelligence Layer	Houses the agentic core: goal formation, planning, and multi-step task execution [2], [3].
Decision Layer	Synthesizes outputs to recommend personalized content, pacing adjustments, and interventions [4].
Action Layer	Delivers recommendations to learners and educators via adaptive interfaces [1].
Feedback Layer	Collects outcome data to evaluate intervention efficacy and update system models [15].
Continuous Learning Layer	Enables system-level lifelong learning, improving accuracy and fairness over time [6].

Table VI. Multi-Layer Agentic AI Integration Framework (MLAIF).

Data flows sequentially from the Input Layer through to the Action Layer, with bidirectional feedback loops linking the Feedback Layer back to the Processing, Intelligence, and Data Layers. The Continuous Learning Layer operates across all layers, enabling adaptive system evolution. This architecture operationalizes the human-centered design principles advocated in the literature [6], [9] while explicitly embedding ethical governance at the Data and Decision Layers.

4.3 Future Research Agenda

Drawing on the identified gaps and emerging trends, a structured three-horizon research agenda is proposed:

a. Short-Term Directions (1–2 Years)

Priority should be assigned to: (1) developing and validating ethical AI governance frameworks for educational deployments [7], [8]; (2) advancing interdisciplinary collaborations bridging education, computer science, psychology, and ethics [5]; (3) expanding empirical studies in diverse sociolinguistic contexts to address geographic bias in the literature [21]; and (4) conducting security assessments of agentic systems handling sensitive student data [24].

b. Medium-Term Directions (3–5 Years)

Over the medium term, research should focus on: (1) longitudinal evaluation of AI-augmented learning outcomes across full academic cycles [16]; (2) environmental sustainability assessments of large-scale AI deployment in education [25]; (3) scalability studies examining cross-institutional and cross-cultural generalizability [22]; and (4) policy reform initiatives that codify equity and accountability standards for AI use in schools and universities [8].

c. Long-Term Directions (>5 Years)

Long-horizon research should address: (1) the development of transparent, explainable AI systems that educators and students can audit and contest [9]; (2) global comparative studies of AI governance models across diverse regulatory regimes [8]; (3) the sociological and epistemological impacts of persistent AI companionship on learner identity and autonomy [6], [7]; and (4) the architecture of trust in agentic educational systems [14], [15].

6. Conclusion

This systematic literature review synthesized 75 peer-reviewed studies to characterize the state of agentic AI in education between 2025 and 2026. The evidence confirms that agentic AI systems hold substantial transformative potential across personalized learning, adaptive assessment, and collaborative instruction. Machine learning and natural language processing are the foundational enabling technologies, while the ADDIE and SAMR frameworks provide established pedagogical scaffolding [1], [4], [11].

However, the widespread realization of this potential is contingent on resolving persistent challenges: data privacy, implementation cost barriers, inadequate educator preparation, and ethical risks including algorithmic bias and reduced learner autonomy [5], [7], [8], [9]. The proposed MLAIF provides a theoretically grounded architecture for integrating these concerns into system design from the outset.

Future research must prioritize geographic inclusivity, longitudinal empirical evaluation, and the co-design of governance frameworks with educators, learners, and policymakers. Only through such coordinated, interdisciplinary effort can agentic AI fulfill its promise as a genuinely equitable and pedagogically sound force in global education [6], [8], [21].

Acknowledgements

The authors gratefully acknowledge the contributions of participating institutions and reviewers whose feedback strengthened this manuscript

References

Ahmed, S. Ozturk, and G. Mensah, "Bibliometric analysis of AI in education: Trends and geographic distribution," Educ. Inf. Technol., vol. 30, no. 2, pp. 1200–1220, 2025.

Aziz, A., Al Aydrus, A. T. A., & Astuti, N. V. (2026). Islamic Educational Leadership and Its Impact on Arabic Oral Proficiency. Leadership in Muslim Societies, 1(1), 75-92.

Birhane and H. Jobin, "Human-centered design and cognitive autonomy in AI educational tools," J. Learn. Sci., vol. 34, no. 1, pp. 88–117, 2025.

Bloom, Taxonomy of Educational Objectives: The Classification of Educational Goals. David McKay, New York, USA, 1956.

Braun and V. Clarke, "Using thematic analysis in psychology," Qual. Res. Psychol., vol. 3, no. 2, pp. 77–101, 2006.

Chen, T. Patel, and K. Nguyen, "Autonomous AI systems in higher education: Design principles and performance," IEEE Trans. Learn. Technol., vol. 18, no. 2, pp. 45–58, 2025.

Graesser et al., "Affective co-learning agents in intelligent tutoring systems," Int. J. Artif. Intell. Educ., vol. 35, no. 2, pp. 345–378, 2025.

Higgins et al., Cochrane Handbook for Systematic Reviews of Interventions, 2nd ed. Wiley, Chichester, UK, 2019.

Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promise and Implications for Teaching and Learning. The Center for Curriculum Redesign, Boston, MA, USA, 2023.

Imoize, D. Lee, and C. Tanner, "Security frameworks for cloud-based adaptive learning systems," IEEE Access, vol. 13, pp. 55230–55245, 2025.

JAnderson, C. F. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive tutors: Lessons learned," J. Learn. Sci., vol. 4, no. 2, pp. 167–207, 1995.

Joanna Briggs Institute, Critical Appraisal Tools. JBI, Adelaide, Australia, 2020.

Johnson and M. Lee, "Personalized learning pathways through agentic AI: A multi-site study," Comput. Educ., vol. 180, pp. 104430, 2025.

Khalil and N. Brouwer, "Longitudinal effects of adaptive learning on academic outcomes," Comput. Educ., vol. 182, pp. 104470, 2026.

Kim and J. Okonkwo, "Redefining teacher roles in AI-mediated classrooms," Teach. Teach. Educ., vol. 140, pp. 104320, 2025.

Lin and A. Park, "Agentic AI architectures: Memory, planning, and tool use in educational agents," Future Internet, vol. 17, no. 4, pp. 210, 2025.

Martinez and R. Faure, "Adaptive assessments calibrated to learner growth trajectories," Assess. Educ., vol. 32, no. 1, pp. 78–94, 2026.

Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases," in Proc. NIPS, pp. 3111–3119, 2013.

Nakamura, "Dynamic learning ecosystems: Lifelong learning agents and multimodal agency," Br. J. Educ. Technol., vol. 57, no. 3, pp. 678–695, 2025.

Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans. Syst. Man Cybern., vol. 40, no. 6, pp. 601–618, 2010.

S. J. d. Baker, "Data mining for education," Int. Encyclopedia Educ., vol. 7, pp. 112–118, 2010.

Torres and D. Singh, "Equity and accessibility in context-aware educational AI," Int. J. Educ. Res., vol. 125, pp. 102050, 2025.

Safitri, N. M., Hijazy, R. W., Mihrain, S. A., Septiadi, M. A., & Hadziq, M. F. (2026). Public Policy, Educational Quality, and Digital Culture: A Mixed-Methods Analysis of Indonesia’s Intellectual and Technological Gap. Tunjuk Ajar: Journal of Education and Culture, 2(1), 76-92

UNESCO, Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021.

Verma and A. Nkrumah, "Methodological quality in AI-education research: A meta-review," Syst. Rev. Educ. Res., vol. 12, no. 1, pp. 45–67, 2025.

Wells, B. Shea, and D. O'Connell, The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses. Ottawa Hospital Research Institute, Ottawa, Canada, 2013.

Williams, "Holistic student development with agentic AI companions," J. Educ. Technol. Soc., vol. 28, no. 1, pp. 12–25, 2025.

Zh, M. H. R., Sari, C. M., Thaariq, Z. Z. A., Sumalee, I., & Kamilah, A. N. (2026). Enhancing conceptual understanding in Islamic education through a hypercontent digital flipbook in blended learning: Evidence from Indonesian madrasah. Educational Technology in Developing Countries, 1(1), 1-22.