

Article

AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges

Diana-Margarita Córdova-Esparza 

Facultad de Informática, Universidad Autónoma de Querétaro, Av. de las Ciencias S/N, Juriquilla, Querétaro 76230, Mexico; diana.cordova@uaq.mx

Abstract: Recent advances in large language models (LLMs) have triggered rapid growth in AI-powered educational agents, yet researchers and practitioners still lack a consolidated view of how these systems are engineered and validated. To address this gap, we conducted a systematic literature review of 82 peer-reviewed and industry studies published from January 2023 to February 2025. Using a four-phase protocol, we extracted and coded them along six groups: technical and pedagogical frameworks, tutoring systems, assessment and feedback, curriculum design, personalization, and ethical considerations. Synthesizing these findings, we propose design principles that link technical choices to instructional goals and outline safeguards for privacy, fairness, and academic integrity. Across all domains, the evidence converges on a key insight: hybrid human–AI workflows, in which teachers curate and moderate LLM output, outperform fully autonomous tutors by combining scalable automation with pedagogical expertise. Limitations in the current literature, including short study horizons, small-sample experiments, and a bias toward positive findings, temper the generalizability of reported gains, highlighting the need for rigorous, long-term evaluations.

Keywords: human–AI interaction; LLM agents; AI and education; generative AI



Academic Editors: Jui-Long Hung, Xu Du and Juan Yang

Received: 30 March 2025

Revised: 21 May 2025

Accepted: 29 May 2025

Published: 31 May 2025

Citation: Córdova-Esparza, D.-M. AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. *Information* **2025**, *16*, 469. <https://doi.org/10.3390/info16060469>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial Intelligence (AI) has rapidly advanced in recent years, with technologies such as large language models (LLMs) playing a key role in a variety of domains, including education. These models, exemplified by GPT from OpenAI, enable the creation of intelligent educational agents that can understand, generate, and interact through natural language. By offering instant feedback, personalized instruction, and context-aware dialogue, LLM-based AI agents hold significant promise for transforming educational practices and learner experiences [1,2].

An AI agent is generally defined as a software entity capable of perceiving and interpreting its environment, making decisions based on these interpretations, and executing actions that help achieve specific objectives. Compared to conventional software programs, AI agents are distinguished by properties such as operating without continuous human oversight, responding adaptively to real-time changes, proactively anticipating needs, learning from past experiences, and communicating effectively with human users and other agents. These features align with established frameworks in the AI literature [3,4], and become more powerful when combined with the advanced language understanding and generation capabilities offered by LLMs [5,6].

The integration of AI agents into education has fueled innovations that aim to address key pedagogical challenges such as personalization, motivation, and scalability [7]. By

adapting content and feedback to the specific needs, abilities, and preferences of a learner, AI agents facilitate individualized learning pathways [1,2]. Natural language dialogue enables these agents to provide continuous guidance and clarify misconceptions, cultivating a more interactive and student-centered environment [4]. Data analytics embedded in agent technologies further inform instructional decisions by helping educators identify knowledge gaps and adjust their teaching strategies accordingly [2,8]. As a result, students can remain more engaged and motivated, especially when they receive timely and relevant feedback on their progress [5]. The advent of LLMs has expanded these capabilities, offering human-like conversational fluency and adaptive support at scale [6].

Despite this rapid growth, the field remains fragmented. To date, most studies focus on single applications—such as automated grading or dialogue tutoring—without offering a consolidated view of how design choices (e.g., retrieval strategies, multi-agent debate, prompt engineering) map onto pedagogical goals (e.g., personalization, metacognition, motivation). In addition, empirical evaluations are often short-term, use small samples, and employ various metrics, making it difficult to draw general conclusions about which architectures and strategies reliably improve learning outcomes. Finally, ethical considerations, such as privacy, fairness, and academic integrity, are addressed only sporadically, leaving practitioners without clear guidelines for responsible deployment.

To address these gaps, this article presents a systematic review of the literature of 82 peer-reviewed industry studies from January 2023 to February 2025. We extracted and coded discrete design decisions along several pedagogical intentions, yielding the following.

1. An evidence-based taxonomy that links specific technical choices (e.g., RAG, fine-tuning, multi-agent systems) to instructional objectives (e.g., scaffolding, assessment, engagement).
2. A comparative analysis of reported efficacy—quantifying, for example, how retrieval grounding reduces factual errors or how debate architectures boost accuracy on ill-structured tasks.
3. A set of design principles and ethical safeguards (privacy controls, bias audits, integrity guardrails) to guide practitioners in building reliable, equitable educational agents.

By synthesizing these findings, we aim to provide educators, developers, and policy makers with a clear map of current capabilities, validated benefits, and outstanding needs, thus supporting a more rigorous, ethical and impactful use of AI-powered educational agents.

In recent years, various strategies have emerged to enhance the effectiveness and pedagogical alignment of AI-powered tutoring systems [1]. Approaches such as Retrieval-Augmented Generation [9], the fine-tuning of models [10], the deployment of multiagent configurations, careful prompt engineering, the integration of established pedagogical strategies, and leveraging in-context learning capabilities offer distinct advantages, contributing to improved accuracy, relevance, and instructional quality [11]. Table 1 summarizes these complementary strategies, illustrating their respective roles in strengthening AI-driven educational interactions.

The subsections that follow unpack each strategy in turn and show how these approaches can be selectively combined to deliver tutoring agents that are both technically robust and pedagogically sound.

Table 1. Overview of common strategies for enhancing AI tutoring systems, highlighting their unique contributions to improving educational effectiveness and pedagogical alignment.

Strategies for Enhancing AI Tutoring Systems	Description
Retrieval-Augmented Generation	Connects AI with knowledge bases to ensure accurate, contextually relevant responses.
Prompt Engineering	Shapes AI behavior using specific instructions and guardrails.
In-Context Learning	Adapts the model’s behavior on-the-fly by leveraging examples or instructions contained within the current prompt.
Few-Shot Prompting	Steers the model with a handful of input–output demonstrations, eliciting task-specific, well-structured responses.
Dynamic Role Setting	Assigns (and can switch) clear personas—e.g., tutor, peer, coach—to align tone, depth, and feedback with each learner’s needs.
Fine-Tuning	Adapts AI models to specialize in domain-specific tasks, improving performance.
Multi-Agents Systems	Uses multiple AI agents to teach or critique each other, enhancing accuracy.
Pedagogical Strategies	Integrates educational theories to support student learning and reflection.

1.1. Retrieval-Augmented Generation (RAG)

Retrieval-augmented generation (RAG) integrates external knowledge directly into large language models (LLMs) to ensure contextually accurate and authoritative responses [9,12]. In education, it is used to build tutoring systems that leverage educational materials such as textbooks and lecture notes to obtain accurate, pedagogically aligned answers [5,13]. RAG involves two stages: retrieval and generation. The retrieval component finds relevant educational content, which is then used in prompts to guide LLMs in generating responses grounded in verified materials. This reduces inaccuracies and maintains consistency with instructional content. When students ask questions, RAG retrieves relevant excerpts from official documents, helping LLMs produce answers aligned with curriculum and educational goals [13]. By referencing verified sources, RAG enhances response accuracy and learners’ trust in these systems. Figure 1 shows the flow of interactions within a RAG-based educational agent, highlighting the relationship between the student, the retrieval, and the language model.

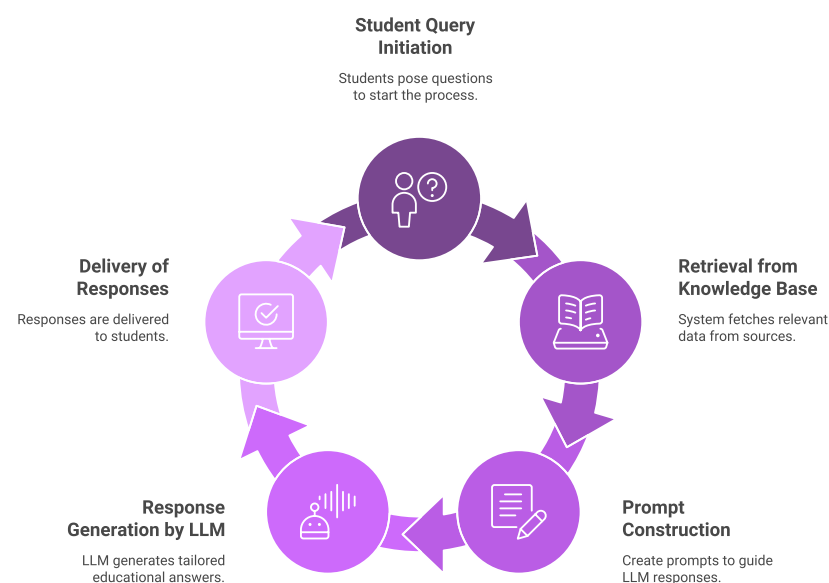


Figure 1. Illustration of the retrieval-augmented generation (RAG) pipeline in educational contexts, demonstrating how student queries initiate a circular flow involving retrieval from a knowledge base, prompt construction, response generation by an LLM, and delivery of accurate, curriculum-aligned responses back to the student.

1.2. Prompt Engineering

Prompt engineering involves crafting input instructions to guide AI behavior toward educational outcomes [14,15]. Designers use it to enforce instructional strategies, ensuring that AI responses align with educational goals. For example, they may direct AI to use guiding questions rather than direct answers, following the Socratic method [16]. Prompts can withhold complete solutions, gradually revealing information as hints. This technique, used in open source tutors such as CodeHelp with a GPT-4 backend, enhances problem solving without providing direct answers [17]. Effective prompt engineering also addresses ethics by promoting autonomy, preventing dishonesty, and maintaining pedagogical standards, allowing designers to shape AI for meaningful learning.

1.3. In-Context Learning

In-context learning (ICL) is a capability of large language models where the model learns to perform a task or adapt its behavior based solely on the information provided within the current prompt or context window, without requiring updates to its underlying parameters [14,18]. In essence, the LLM conditions its output on the patterns, examples, or instructions presented immediately before the target query. For educational agents, ICL allows for rapid, on-the-fly adaptation. For example, by including examples of Socratic dialogue or specific types of feedback within the prompt, the agent can adopt that pedagogical style for the subsequent interaction [19]. This strategy enables the dynamic personalization of the agent's responses based on the immediate conversational history or explicit examples, offering a flexible alternative or complement to more resource-intensive fine-tuning for achieving specific instructional behaviors.

1.4. Few-Shot Prompting

Few-shot prompting is a specific application of in-context learning where the prompt provided to the LLM includes a small number of demonstration examples (input–output pairs) illustrating the desired task or response format before presenting the actual query [14]. For example, a prompt might show 2–3 examples of a student question followed by an ideal hint before asking the LLM to generate a hint for a new question. These examples effectively teach the model the desired pattern or behavior within the context window. In educational agent design, few-shot prompting is valuable for guiding the LLM to produce outputs that adhere to specific pedagogical structures (e.g., step-by-step problem solving guidance, targeted feedback formats) or handle tasks where simple instructions (zero-shot prompting) might be insufficient or ambiguous [20,21]. It generally yields better task performance than zero-shot prompting by providing concrete examples of the expected output.

1.5. Dynamic Role Setting

Dynamic role setting involves explicitly assigning a persona or role to the LLM within the prompt to shape its interaction style, tone, and focus [22]. This is a specific technique within prompt engineering where instructions like “Act as a patient tutor explaining calculus concepts to a beginner” or “You are a peer debating the causes of the French Revolution” guide the LLM's behavior. In educational settings, the assignment of roles allows designers to create agents that embody specific pedagogical functions, such as a facilitator, a Socratic questioner, a domain expert, or even a simulated tutee for learning-by-teaching scenarios [23]. The dynamic aspect implies that these roles can potentially be adjusted during an interaction based on the learner's progress, needs, or the specific learning activity [19]. This technique enhances the agent's alignment with desired pedagogical goals and can make interactions feel more natural and contextually appropriate for the learner.

1.6. Fine-Tuning

Educational applications often use general-purpose LLMs such as GPT-4, but specialized contexts may need models fine-tuned on specific datasets [24]. Fine-tuning adapts a general LLM for better performance on focused tasks using specialized data. This is particularly useful in areas that require precise domain knowledge, such as grammar correction, language translation, code debugging, and explaining complex concepts. Fine-tuning improves the accuracy and contextual relevance of the model, aligning its language, style, and terminology with the educational domain [10]. Fine-tuned models excel at providing detailed feedback on student essays or explaining technical concepts in fields such as engineering or medicine [25]. However, fine-tuning requires considerable resources, such as computational power, quality data, and expertise [10,26]. It also requires maintaining and validating model performance over time, necessitating robust MLOps (machine learning operations), the set of practices that combines machine learning, software engineering, and operational management to streamline and scale model development, deployment, monitoring, and maintenance [27]. Hence, educators and developers must evaluate whether the performance gains justify the investment.

Figure 2 illustrates the iterative fine-tuning process, emphasizing the key stages involved in adapting a general large language model to specific educational contexts.

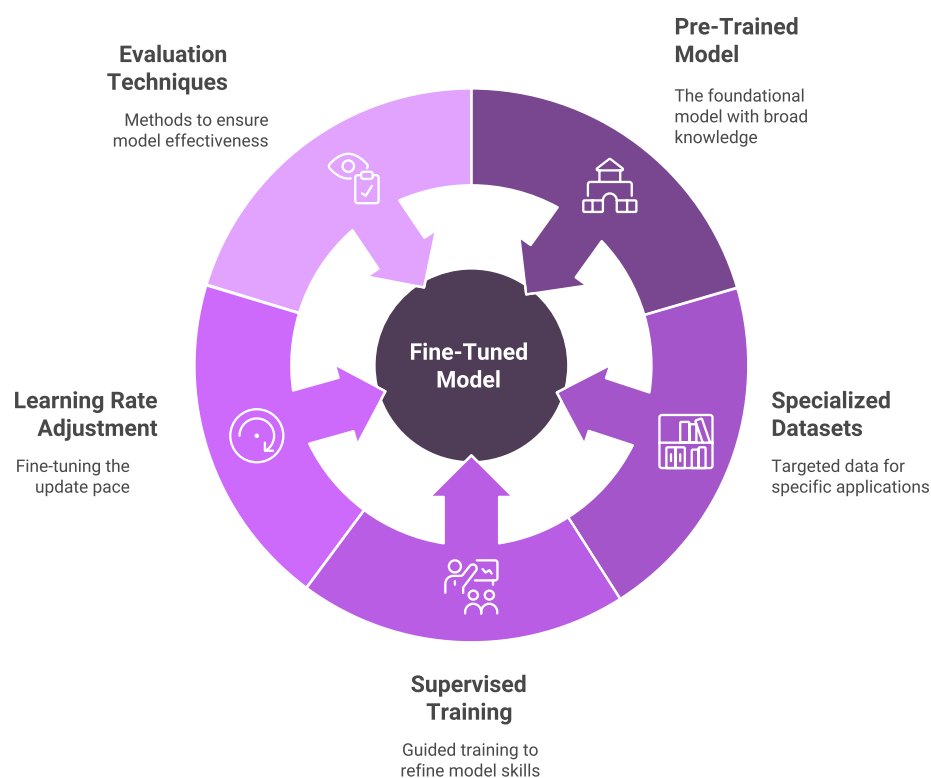


Figure 2. The fine-tuning workflow, depicting how a pretrained language model is adapted through specialized datasets, supervised training, careful learning rate adjustments, and rigorous evaluation techniques to achieve enhanced performance for educational applications.

1.7. Multi-Agent Systems

Researchers have used multi-agent systems, involving multiple LLM agents, to improve response accuracy by facilitating interactions like dialogues and critiques among agents [28–30]. These systems enhance the agents' ability to analyze and correct their reasoning, leading to more reliable answers. In education, agents often deliberate on complex problems before responding. For example, the Multi-Agent System for Conditional

Mining (MACM) uses structured interactions among agents to solve complex math problems, significantly boosting accuracy on difficult tasks in the MATH dataset from 54.68% to 76.73% [31]. Additionally, multi-agent setups enrich educational interactions by simulating group discussions and presenting various perspectives. Generative agents [32,33] demonstrate how different roles (students, teachers, parents) engage in classroom-style dialogue, promoting critical thinking and exposing learners to diverse viewpoints [34]. Despite these benefits, multi-agent systems introduce complexity, increasing computational costs, and challenges in coordination and consistency [35–37]. Thus, designers must weigh the benefits of improved accuracy and enriched discussions against the additional resources needed.

1.8. Pedagogical Strategies

On the educational theory side, LLM-driven agents frequently adopt strategies grounded in well-established pedagogical frameworks, such as instructional scaffolding, Socratic questioning, and the zone of proximal development (ZPD) [38,39]. Scaffolding involves structuring learning experiences by initially providing extensive guidance and gradually reducing assistance, known as fading, as students develop competence. Within a tutoring context, an agent can initially provide detailed explanations, illustrative examples, or step-by-step solutions, progressively shifting to brief hints or prompts to encourage independent problem solving as the student's skills advance [40].

Another key pedagogical technique widely implemented in agentic educational systems is Socratic questioning. Here, rather than directly providing answers, the educational agent poses open-ended, thought-provoking questions designed to stimulate critical thinking, deeper understanding, and self-directed learning. Such questioning encourages students to articulate their reasoning, justify their choices, or reconsider assumptions, actively involving them in the learning process and fostering robust conceptual mastery [41,42].

In addition to these approaches, educational agents increasingly incorporate metacognitive prompts, which explicitly encourage students to reflect upon their own cognitive processes. These prompts guide learners to monitor their understanding, evaluate their problem-solving strategies, and set self-directed learning goals [43]. This metacognitive engagement helps cultivate the self-awareness and adaptability of students, improving not only their immediate learning outcomes but also their capacity for lifelong learning [44,45].

Ultimately, by embedding these diverse pedagogical strategies, LLM-based agents can deliver personalized, adaptive, and cognitively engaging learning experiences, closely emulating the nuanced, reflective, and supportive interactions characteristic of human tutoring [46–48].

2. Methodology

We conducted a review of the literature on academic advances in the use of LLM-based AI-powered educational agents, aiming to gather evidence on the main opportunities, innovations, and challenges in the teaching–learning process. The procedure for selecting articles involved four phases [49]: planning, selection, information extraction, and execution, each with specific steps for implementation, as illustrated in Figure 3.

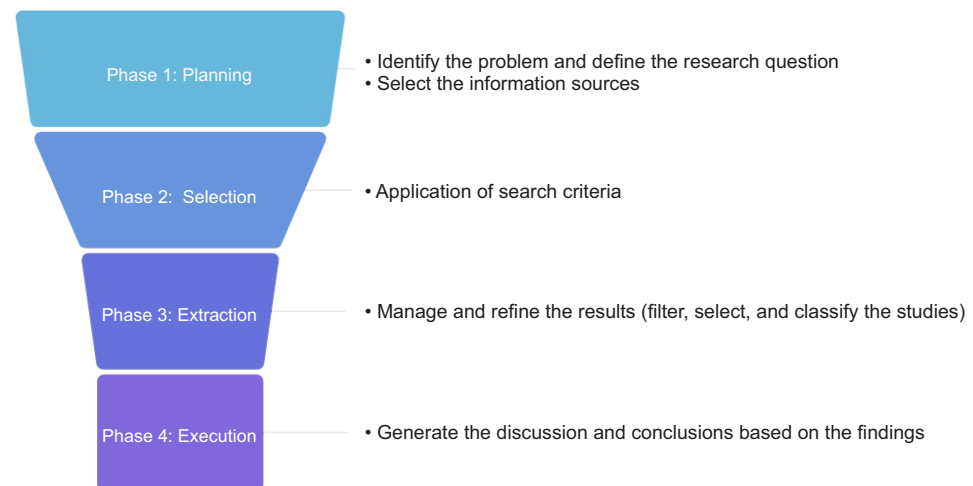


Figure 3. Phases of the systematic review, inspired by [49].

2.1. Phase 1: Planning

In this phase, we established the following research questions to guide the analysis.

- How can agents based on large language models (LLMs) enhance personalization, real-time feedback, and natural interaction in educational settings?
- What strategies and agentic tools can maximize the effectiveness of adaptive teaching and tutoring systems?
- What ethical challenges do educators face in the context of AI use?

The literature search was performed in the specialized Scopus database and the Google Scholar search engine. The search period was set from 2023 to 2025, given the emergence of ChatGPT (based on GPT-3.5), which frames the development of intelligent agents based on LLMs. Table 2 shows the search strings used to retrieve articles in the Google Scholar and Scopus databases.

Table 2. Search strings used for retrieving articles and theses from Google Scholar and Scopus databases.

Academic Database	Keywords
Google Scholar	(intitle:"large language model*" OR intitle:"LLM*") OR intitle:"agent*" AND intitle:"education"
Scopus	TITLE-ABS-KEY (("Large Language Models*" OR "LLM*") AND ("agent*") AND ("education")) AND PUBYEAR > 2022 AND PUBYEAR < 2026

2.2. Phase 2: Selection

To determine the articles for analysis, the following inclusion and exclusion criteria were applied.

Inclusion criteria:

- Articles published between 2023 and 2025.
- Articles written in English.
- Articles whose topics matched the search terms and research questions.
- Empirical primary studies (e.g., experimental, quasi-experimental, survey-based) that reported original data on LLM-based agents in education.

Exclusion criteria:

- Articles published before 2023, to focus on recent advancements.

- Articles not available in English.
- Articles not directly related to the search terms aligned with the research questions.
- Other academic works, including theses, technical reports, and briefs.

2.3. Phase 3: Extraction

The methodology used for this systematic review was based on the PRISMA Statement [50] (shown in Figure 4), aiming to identify and analyze the use and impact of LLM-based agents in education.

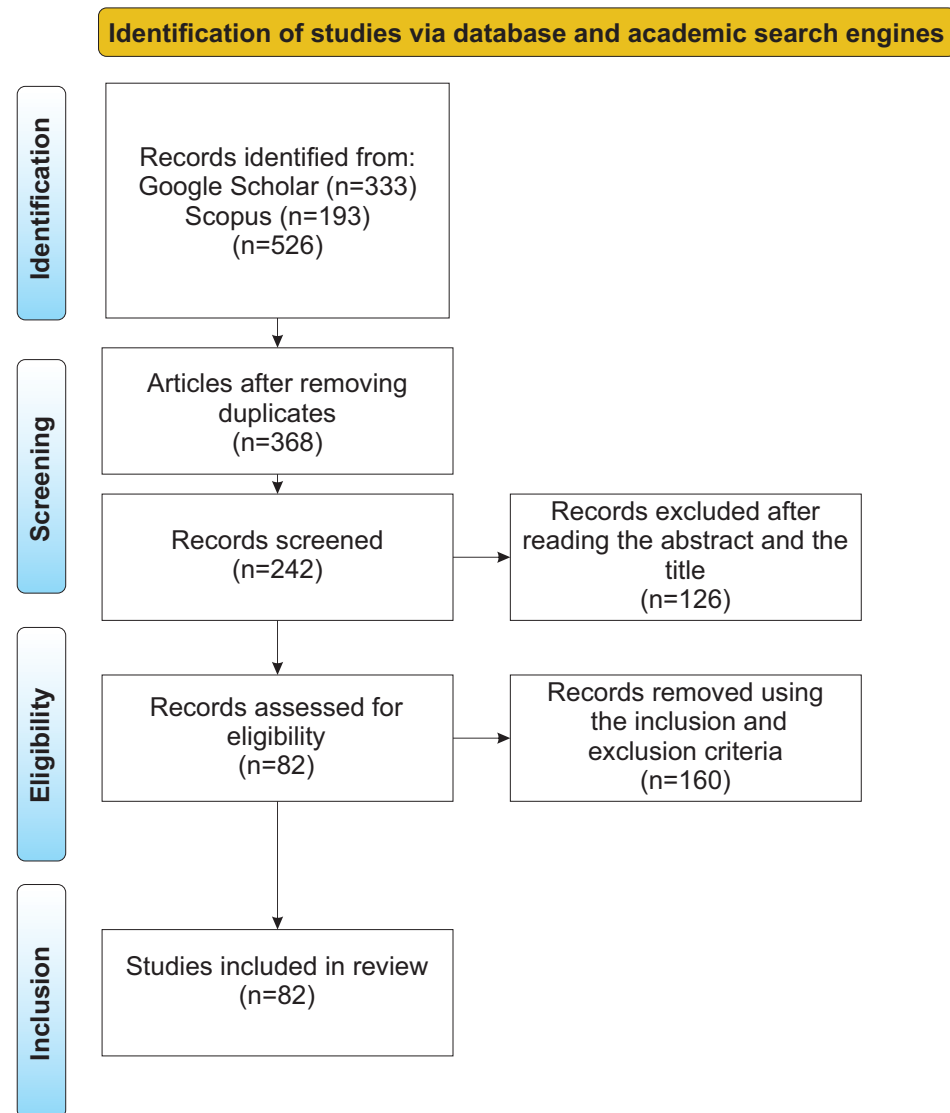


Figure 4. Process of selecting scientific articles for review.

Once the articles were extracted, they were analyzed and six groups were created (see Table 3), where we categorized each work. The findings of each group are described in Section 3.

1. Technical and Pedagogical Frameworks
2. LLM-Powered Tutoring Systems and Virtual Assistants
3. Assessment and Feedback with LLM Agents
4. Curriculum Design and Content Creation
5. Personalized and Engaging Learning Experiences
6. Ethical Considerations and Challenges

Table 3. Articles analyzed and classified in the systematic review. The table includes category, publication year, authors, and research title.

Category	Publication Year	Authors	Research Title
1. Technical and Pedagogical Frameworks	2025	Yusuf et al. [11]	Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art.
	2025	Kim et al. [51]	Assessing the Current Limitations of Large Language Models in Advancing Health Care Education.
	2025	Nikolovski et al. [52]	Advancing AI in Higher Education: A Comparative Study of Large Language Model-Based Agents for Exam Question Generation, Improvement, and Evaluation.
	2025	Li et al. [53]	ARCHED: A Human-Centered Framework for Transparent, Responsible, and Collaborative AI-Assisted Instructional Design.
	2025	Zhang et al. [54]	EduPlanner: LLM-Based Multi-Agent Systems for Customized and Intelligent Instructional Design.
	2024	Yu et al. [55]	From mooc to maic: Reshaping online teaching and learning through llm-driven agents.
	2024	Shen et al. [56]	Large Language Models for Education: A Survey and Outlook.
	2024	Wang et al. [57]	Enhancement of the performance of large 849 language models in diabetes education through retrieval-augmented generation: comparative study.
	2023	Liffiton et al. [17]	Codehelp: Using large language models with guardrails for scalable support in programming classes.
	2023	Eggmann et al. [58]	Implications of large language models such as ChatGPT for dental medicine.
2. LLM-Powered Tutoring Systems and Virtual Assistants	2025	Gao et al. [59]	Agent4Edu: Generating Learner Response Data by Generative Agents for Intelligent Education Systems.
	2025	Wang et al. [60]	CyberMentor: AI Powered Learning Tool Platform to Address Diverse Student Needs in Cybersecurity Education.
	2024	Jin et al. [23]	Teach ai how to code: Using large language models as teachable agents for programming education.
	2024	Marouf et al. [61]	Enhancing Education with Artificial Intelligence: The Role of Intelligent Tutoring Systems.
	2024	Alsafari et al. [62]	Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants.
	2024	Neira-Maldonado et al. [63]	Intelligent educational agent for education support using long language models through Langchain.
	2024	Maiti and Goel [64]	How Do Students Interact with an LLM-powered Virtual Teaching Assistant in Different Educational Settings?.
	2024	Molina et al. [65]	Leveraging LLM Tutoring Systems for Non-Native English Speakers in Introductory CS Courses.
	2024	Neumann et al. [66]	An llm-driven chatbot in higher education for databases and information systems.
	2024	Pan et al. [67]	ELLMA-T: an Embodied LLM-agent for Supporting English Language Learning in Social VR.
	2024	Gold and Geng [68]	Tracking the Evolution of Student Interactions with an LLM-Powered Tutor.
	2024	Gold and Geng [69]	On the Helpfulness of a Zero-Shot Socratic Tutor.
	2024	Henley et al. [70]	CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs.
	2024	Nagarajan [71]	AI-Enabled E-Learning Systems: A Systematic Literature
	2023	Sahlman et al. [5]	Revolutionizing Learning with GenAI.
	2023	OpenAI [6]	Filling Crucial Language Learning Gaps: Duolingo and GPT-4.

Table 3. Cont.

Category	Publication Year	Authors	Research Title
3. Assessment and Feedback with LLM Agents	2025	Yeung et al. [72]	A Zero-Shot LLM Framework for Automatic Assignment Grading in Higher Education.
	2024	Lagakis and Demetriadis [73]	EvaAI: A Multi-agent Framework Leveraging Large Language Models for Enhanced Automated Grading.
	2024	Pardos and Bhandari [74]	ChatGPT-Generated Help Produces Learning Gains Equivalent to Human Tutor-Authored Help on Mathematics Skills.
	2024	Doughty et al. [75]	ChatGPT for good? On opportunities and challenges of large language models for education.
	2023	Chan et al. [76]	ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate.
	2023	Pankiewicz and Baker [77]	Large Language Models (GPT) for Automating Feedback on Programming Assignments.
	2023	Kasneci et al. [1]	ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate.
	2023	Lex [78]	Introducing Q-Chat, the World's First AI Tutor Built with OpenAI's ChatGPT.
4. Curriculum Design and Content Creation	2025	Tan et al. [19]	ELF: Educational LLM Framework of Improving and Evaluating AI Generated Content for Classroom Teaching.
	2025	Mustofa et al. [79]	Integration of Artificial Intelligence (ChatGPT) into Science Teaching and Learning.
	2025	Durak et al. [80]	Comparison of Human-Written Versus AI-Generated Text in Discussions at Educational Settings: Investigating Features for ChatGPT, Gemini and BingAI.
	2024	Fleischhauer and Friedrich [81]	Factors determining the efficacy of AI-generated word problems for content-specific math language courses in higher education.
	2024	Luo et al. [82]	BPE: Exploring the Prompt Framework of Physics Exercises Generated from Bloom's Taxonomy in LLM.
	2024	Hu et al. [83]	Teaching Plan Generation and Evaluation with GPT-4: Unleashing the 908 Potential of LLM in Instructional Design.
	2024	Lin and Sun [84]	A Smart English Learning Curriculum Generation Mobile Platform based on Word Root Extension using Artificial Intelligence and LLM (Large Language Model.)
	2024	Robinson [85]	Harnessing AI for Structured Learning: The Case for Objective-Driven Design in E-Learning.
	2024	Amirjalili et al. [86]	Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature.
	2023	Denny et al. [87]	Can we trust AI-generated educational content? comparative analysis of human and AI-generated learning resources.
	2023	C. et al. [88]	Inculcating Digital Contents, Technological Tools: Agents for Enhancing Business Education Curriculum for Skill Acquisition in Nigerian Universities.

Table 3. Cont.

Category	Publication Year	Authors	Research Title
5. Personalized and Engaging Learning Experiences	2025	Hao et al. [21]	Student Engagement in Collaborative Learning with AI Agents in an LLM-Empowered Learning Environment: A Cluster Analysis.
	2025	Li et al. [89]	TutorLLM: Customizing Learning Recommendations with Knowledge Tracing and Retrieval-Augmented Generation.
	2025	Wang et al. [90]	LearnMate: Enhancing Online Education with LLM-Powered Personalized Learning Plans and Support.
	2025	Joe and Sebastin [91]	Human-Ai Collaboration In Education Using Leveraging Deep Learning And Natural Language Processing To Enhance Personalized Learning Systems.
	2024	Damasceno et al. [92]	DALverse: Assistive Technology for Inclusion of People with Disabilities in Distance Education through a Metaverse-Based Environment.
	2024	Choshi [93]	Addressing Challenges in Undergraduate Community Health Nursing Clinical: Kolb’s Experiential Learning Theory.
	2024	Choi [94]	The Effect of Early Childhood Science Education Class Based on Kolb’s Experiential Learning Theory on The Pre-service Early Childhood Teacher’s Science Knowledge, Science Attitude, and Science Teaching Efficacy.
	2024	Haritha and Rao [95]	A Holistic Approach to Professional Development: Integrating Kolb’s Experiential Learning Theory for Soft Skills Mastery.
	2024	Sajja et al. [96]	Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education.
	2024	Gunawan and Wiputra [97]	Exploring the Impact of Generative AI on Personalized Learning in Higher Education.
	2024	Muhammad and Orji [98]	Revolutionizing Education In The Digital Era: The Role Of Ai In Promoting Inclusivity, Equality, And Ethical Innovation.
	2024	Bhatnagar and Sharma [99]	Using AI to Create More Equitable and Inclusive Higher Education Learning Systems.
	2024	Qushwa and Onia [100]	AI Innovation in Education: Realizing Personalized Learning in the Digital Age.
	2023	Lawrence et al. [101]	How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process.
	2023	Echeverría et al. [102]	Designing Hybrid Human–AI Orchestration Tools for Individual and Collaborative Activities: A Technology Probe Study.
6. Ethical Considerations and Challenges	2023	Kwon and Shim [103]	The study on categorizing scaffolding types of AI chatbots for primary English learning-oriented assessment.
	2025	Rowsell [104]	Nine in 10 UK Undergraduates Now Using AI in Assessments – Survey.
	2025	Microsoft Presidio [105] Chu et al. [106]	Presidio: Data Protection and De-identification SDK. Llm agents for education: Advances and applications.

Table 3. Cont.

Category	Publication Year	Authors	Research Title
	2025	Kumar et al. [107]	Detecting and Mitigating Bias in LLMs through Knowledge Graph-Augmented Training.
	2025	Xu et al. [108]	Mitigating Social Bias in Large Language Models: A Multi-Objective Approach Within a Multi-Agent Framework.
	2024	Dey et al. [109]	Better to Ask in English: Evaluation of Large Language Models on English, Low-Resource and Cross-Lingual Settings.
	2024	Yong et al. [110]	Low-Resource Languages Jailbreak GPT-4.
	2024	Hussain [111]	Forming Policies for Ethical AI Use in Academic Writing.
	2024	Cheng et al. [112]	Reinforcement Learning from Multi-role Debates as Feedback for Bias Mitigation in LLMs.
	2024	Echterhoff et al. [113]	Cognitive bias in decision-making with LLMs.
	2024	Raza et al. [114]	Mbias: Mitigating bias in large language models while retaining context.
	2024	Xie et al. [115]	Differentially private synthetic data via foundation model apis 2: Text.
	2024	Sun et al. [116]	Empowering Users in Digital Privacy Management through Interactive LLM-Based Agents.
	2024	Grobler [117]	Exploring Essential Concepts In The Automation Of Student Plagiarism Management-A Case Study.
	2024	Davies [118]	Artificial Intelligent Branching Simulations (AIBS) in Critical Care Neonatal Nursing.
	2023	Khan Academy [119]	Introducing Khanmigo: An AI-Powered Guide Powered by GPT-4.
	2023	OpenAI [120]	ChatGPT Outage: Incident Report.
	2023	Zou et al. [121]	Universal and Transferable Adversarial Attacks on Aligned Language Models.
	2023	Muscanell and Robert [122]	EDUCAUSE QuickPoll Results: Did ChatGPT Write This Report?
	2023	Williams [123]	Turnitin Says One in 10 University Essays Are Partly AI-Written.
	2023	Kirchenbauer et al. [124]	A Watermark for Large Language Models.

3. Results

We created six broad categories of research findings that highlight the intersection between AI technologies and educational practices: technical and pedagogical frameworks, tutoring systems and virtual assistants, assessment and feedback, curriculum design and content creation, personalized learning experiences, and ethical considerations and challenges. In this section, we discuss each of these categories, beginning with the technical and pedagogical frameworks, which outline key strategies and conceptual foundations for effectively embedding large language models within educational contexts.

3.1. Technical and Pedagogical Frameworks

Implementing AI agents in education requires a careful synthesis of advanced technologies and solid pedagogical principles. Simply deploying a raw large language model (LLM) in a classroom, without additional support or oversight, can lead to challenges such as misinformation, superficial answer-giving, or misalignment with curricular goals [51,58]. Consequently, many systems integrate supplementary mechanisms, such as retrieval modules, multi-agent self-checking, guardrails, and pedagogical strategies, on top of a core LLM to ensure reliability and educational value [57].

Yusuf et al. [11] presented a framework highlighting the dual role of conversational agents (CAs) in education. On the one hand, CAs offer pedagogical applications such as providing on-demand answers (FAQ), fostering motivation, automating tasks like scheduling (administration), facilitating dialogue (communication), promoting self-reflection (metacognitive skills), providing role-playing for deeper learning (simulation), generating assessments (evaluations), and offering personalized guidance (mentorship). On the other hand, they exhibit technological functions like providing NLP for human-like text, using real-time data for authoritative responses (external data sources), integrating with social media for accessibility (social interfaces), using machine learning for user-adaptive behavior, developing custom features for educational needs (custom development), and enhancing interactions through feedback and collaboration tools (communication support). They also personalize recommendations and feedback based on learner progress.

We adapted the thematic coding of the studies reported in [11] into two tables. Table 4 shows the seven key pedagogical intents of conversational agents in higher education and their learning support. Table 5 details the six technological aspects that underpin these functions.

Table 4. Pedagogical purposes and learning intents addressed by conversational agents.

Pedagogical Function	Description
FAQ and knowledge base	Provides on-demand answers and feedback drawn from a curated knowledge repository.
Motivation and inspiration	Supplies emotional or cognitive cues that boost learners' motivation, engagement, and self-efficacy.
Administration and management	Automates routine admin and communication tasks (e.g., onboarding, scheduling), easing staff workload.
Reflective and metacognitive skills	Promotes analytical, critical, divergent, and creative thinking to deepen self-reflection and learning.
Simulation and experience	Creates safe simulated contexts (including affective interactions) for experiential practice.
Assessments	Generates formative or summative feedback and feed-forward guidance on learners' work.
Mentorship and coaching	Monitors learner progress and delivers personalized guidance, recommendations, or instruction.

Table 5. Technological functions and deployment features of conversational agents.

Technological Facet	Description
Natural language processing	Native or API-based NLP modules that parse and generate human language for understanding user intent and producing responses.
Connection to external data sources	Ability to query SQL databases, knowledge libraries, or other repositories in real time, enriching dialogue with up-to-date information.
Interface with social media/IM	Agent delivered through messaging platforms (e.g., WhatsApp, Facebook Messenger), using the chat UI as the main interaction channel.
Personalized development	Custom or otherwise unspecified codebases tailored to a specific use-case when off-the-shelf tools are insufficient.
Communication support	Multimodal interfaces such as voice bots, virtual humans, or avatars that add speech, audio, or embodied presentation layers.
Personalization functions	Machine-learning augmentations—eye-tracking, emotion or gaze detection, biometrics—that adapt dialogue flow to each user.

3.1.1. Integrating Technical and Pedagogical Layers

To be successful, LLM-based educational agents must intertwine these pedagogical applications with robust technical functionalities. For example, a retrieval-augmented agent (technical layer) that offers metacognitive prompts (pedagogical layer) can help learners develop reflective problem solving strategies while relying on verified course materials. Similarly, a multi-agent setup (technical) that engages in a debate or peer-review process (pedagogical) can detect and correct mistakes before delivering final answers to students.

Researchers have increasingly emphasized such blended approaches. Liffiton et al. [17] introduced guardrails and moderated dialogue flows to ensure that an AI-powered coding tutor provided hints rather than full solutions, reinforcing student learning rather than shortcutting it. Shen et al. [56] and other studies likewise demonstrate the importance of combining instructional design strategies with advanced model architectures or prompt engineering techniques.

3.1.2. Design Principles for LLM-Based Educational Agents

Several design principles have emerged for developers and educators implementing LLM-based agents. First, alignment with learning objectives is critical: the agent's conversational flow and knowledge sources should be explicitly mapped to course goals and curriculum standards, ensuring that automated feedback or on-demand answers remain accurate and relevant [52,53]. Furthermore, adaptive complexity, the agent's ability to dynamically adjust the difficulty of the questions or the depth of explanations based on student progress, is particularly valuable in diverse classrooms or self-paced learning environments [54,55]. Encouraging active learning and exploration is another essential principle: rather than providing direct answers, LLM-based tutors should guide students with probing questions, scaffolded hints, or reflective prompts to promote deeper engagement and self-discovery [53,125]. Ensuring technical robustness and reliability through mechanisms such as retrieval augmentation, multi-agent debate, or external verifier programs can significantly reduce hallucinations and improve factual correctness, especially when high stakes are involved (e.g., final assessments or critical concepts). Finally, ethical and safe interaction must be prioritized: privacy controls, bias audits, and content moderation are paramount; systems should transparently communicate if and when AI is uncertain, guiding learners to consult human instructors or trusted resources as necessary.

3.2. LLM-Powered Tutoring Systems and Virtual Assistants

Intelligent tutoring systems (ITSs) have long been studied in educational technology [61,126,127], but modern LLMs have dramatically increased their conversational skills due to their ability to understand and generate natural language responses using advanced natural language processing (NLP) and deep learning techniques.

In the educational domain, chatbots rely heavily on unstructured data sources such as textbooks, presentations, lecture transcripts, research papers, and more. This presented a significant challenge for natural language processing (NLP). Advanced language models offer a solution to overcome these challenges in educational chatbots utilizing methods such as RAG, providing access to a broader range of educational content while integrating AI models for both processes [62]. This combination leverages their strengths to provide more comprehensive and reliable educational responses. It also improves contextual awareness, as responses can be enriched with retrieved information, allowing for more in-depth and contextually rich educational interactions. Frameworks such as LangChain simplify the creation of applications that leverage advanced language models.

Neira-Maldonado et al. [63] developed an Intelligent Educational Agent (IEA) to enhance university students' learning. Integrating LangChain and OpenAI's GPT-3.5,

the architecture involved students, teachers, and tutors. The IEA effectively generated multiple-choice tests, academic plans, and valuable resource recommendations, supporting personalized and engaging education.

LLM-based tutors can answer questions, explain concepts, and guide problem-solving. The TeachYou system [23], featuring the AlgoBo chatbot, enhances learning-by-teaching (LBT) with a Reflect–Respond prompt pipeline. This simulates learning and feedback for programming instruction. Its mode-switching allows the agent to ask questions, promoting reflection and deeper algorithmic understanding. Key findings show that it increases knowledge building in learner–agent interactions. Interacting with the agent was intuitive and similar to human teaching. However, participants noted the need for better resources to support self-reflection. Limitations include insufficient exploration of diverse types of knowledge and a lack of direct learning measures. Future educational implementations could produce more robust effectiveness data.

Jill Watson, a virtual teaching assistant powered by LLMs [64], generates contextually relevant and coherent responses to students' questions using a RAG approach. It retrieves relevant information and generates responses aligned with questions. Integrated into Canvas and Blackboard, Jill Watson is easily accessible in learning environments. It supports autonomous learning and critical thinking, enhances question complexity and depth, and boosts cognitive engagement, adapting to course dynamics.

The study by Molina et al. [65] examined the use of large language models (LLMs) as tutors for non-native English-speaking students (NNESs) in introductory computer science courses, focusing on language and terminology challenges. The CodeHelp system, utilizing OpenAI's GPT-3.5, aims to improve programming course learning. Both native and non-native students value CodeHelp's accessibility, conversational style, and guidance. NNESs find it particularly helpful for communicating without perfect English, though some still struggle with technical terms, mixing English with their native language.

NNESs used the system like native speakers but less frequently. LLM-based tutors like CodeHelp can significantly aid computer science education by providing personalized support, bridging language gaps, and boosting student confidence. The paper highlights the importance of improving system visibility and student training to enhance usage and effectiveness.

The study from Neumann et al. [66] examined MoodleBot, an LLM chatbot in computer science classes, focusing on self-regulated learning and help-seeking support. MoodleBot uses GPT models with LangChain and LlamaIndex. The research addresses student acceptance and the alignment of chatbot responses with course content, showing an 88% accuracy in course support.

The work developed by Pan et al. [67] describes ELLMA-T, an embodied conversational agent that uses GPT-4 within social virtual reality (VR) environments to facilitate English-as-a-second-language learning. The system is designed to simulate realistic conversations in situational contexts, providing learners with opportunities for speech practice, language proficiency assessment, and personalized feedback.

The work of Gao et al. [59] introduced Agent4Edu, a personalized learning simulator that uses an LLM to simulate the response data of the learner and the problem solving behaviors of the learners. This simulator includes specific modules for learning profiling, memory, and action, designed for personalized learning scenarios.

The research by Wang et al. [60] focused on CyberMentor, an AI-driven platform for non-traditional cybersecurity students. It uses agentic workflow and LLM with RAG for precise information retrieval, enhancing accessibility and personalization. CyberMentor proves its worth by supporting knowledge acquisition, career preparation, and skill guidance, validated by LangChain's evaluation.

At Boston University, a GPT-4 powered "Socratic Tutor" was used in an introductory data science course [68,69]. The AI, equipped with homework and solutions, guided students with Socratic questions instead of direct answers. This interaction provided hints and clarifications, resulting in a helpfulness rating of 4.0/5, similar to human teaching assistants (4.2/5). It proved valuable for clarifying questions, code syntax, and problem solving steps while rarely giving full solutions and maintaining sound pedagogy. The study showed that with thoughtful prompt design, an LLM tutor can effectively aid instruction and maintain academic integrity.

Henley et al. [70] introduced CodeAid, an LLM-based helper, to support ~700 coding students. CodeAid explains the code, suggests error fixes, and gives examples, with safeguards to prevent giving direct solutions. In a semester-long study, many students used it extensively. Analysis showed that AI can enhance student support, but needs a balance between help and dependence. Students valued on-demand help as office hours, while instructors stressed ethical and pedagogical guidelines (e.g., hints over answers). The studies conclude that with proper methods, LLM tutors can enhance class assistance.

In relation to industry advances in LLM-based education agents, Khan Academy Khanmigo functions as a virtual tutor in all subjects, helping students with math problems, writing exercises, coding practice, and more. Early pilots show Khanmigo engaging learners in interactive dialogue and even role-playing as historical figures to make learning more immersive [5].

LLM-based tutors have enhanced language learning. Duolingo incorporates GPT-4 [6] to offer an AI conversation partner and explanation tool. With Duolingo Max, users engage in Role Play chats for realistic practice and receive grammar rule breakdowns. These features address previous gaps like open-ended speaking and detailed error feedback. Duolingo reports that GPT-4's abilities enable fluent, accurate AI experiences unlike earlier models, which struggled with complex dialogues and grammar explanations. This shows that LLM tutors enhance language learning by providing immersive practice and instant explanations, serving as a 24/7 language coach.

Virtual tutors extend beyond academics into professional training, with AI coaches increasingly guiding employees through complex tasks and onboarding. Companies like Oracle, IBM, and Accenture use AI-driven systems for personalized learning, on-demand training, and real-time, role-specific coaching [128]. In banking, gamified platforms let staff practice scenarios and access resources anytime [129]. Reviews show that these AI systems adapt to job roles and performance data, providing targeted support for growth and lifelong learning [71]. Although still evolving, this technology shows the wide applicability of LLM-based tutoring in educational and professional settings.

3.3. Assessment and Feedback with LLM Agents

LLM-based agents are also being applied to assessment, grading, and feedback tasks in education. One major opportunity is using LLMs to automatically grade assignments or provide feedback, which could save instructors significant time and provide students with instant guidance. Researchers have started to evaluate LLMs as automated graders for both programming tasks and written responses. The results so far are encouraging, though with important caveats. Table 6 summarizes the leading approaches, representative systems, and key findings in LLM-based assessment and feedback.

A recent study by Yeung et al. [72] introduced a Zero-Shot LLM-Based Automated Assignment Grading (AAG) system that leverages prompt engineering for grading and personalized feedback. Tested in higher education, it assessed both quantitative and explanatory text answers. The system consistently scored student work and provided tailored feedback on strengths and areas for improvement. Student surveys indicated

that this approach greatly enhanced their motivation, understanding, and preparedness compared to traditional grading with delayed feedback. Immediate, detailed AI feedback made students feel more informed and motivated to learn.

ChatEval [76] is a multi-agent framework where LLMs debate and evaluate responses, enhancing accuracy and alignment with human evaluations. Compared to single LLM methods, ChatEval's performance is measured by precision and correlation with human scores. A qualitative analysis explores agent contributions and dynamics like debate depth, perspective diversity, and consensus building. ChatEval represents an advanced evaluation tool and a notable advancement in the development of communicative agents for NLG tasks.

The study by Lagakis and Demetriadis [73] introduced EvaAI, a multi-agent system using LLMs and the AutoGen framework to automate grading in MOOCs, regardless of the topic. This system mimics human scoring, incorporates tutor feedback, and addresses scalability and personalization challenges in educational assessment. EvaAI provides precise grading with minimal initial data, outperforming single-agent systems. It uses a two-tier architecture: a Reverse Proxy Agent centralizes coordination by assigning expert graders based on subject, complexity, and grading rubrics; then, specialized agents in configurations like Grader–Reviewer or Multiple Graders collaboratively assess assignments across subjects.

The research by Pankiewicz and Baker [77] examined GPT models as coding tutors, finding that they can identify errors and suggest code style improvements, similar to human assistants. These systems require careful course-specific validation, as LLMs may overlook certain logical errors or edge cases. Researchers advise using LLMs to support grading instead of relying on them entirely.

For written assignments and essays, LLMs have shown mixed but improved results as graders. Early experiments using GPT-3.5 to score student essays or short answers found a moderate correlation with human grades. More recent work with GPT-4 suggests a closer alignment with human grade standards, especially if AI is given a well-crafted rubric as a prompt [1].

LLM graders match human accuracy but excel in efficiency. A study on AI-generated hints [74] in math education had GPT-3.5 create step-by-step algebra hints as feedback. Students receiving GPT hints improved as much as those with human hints, showing no significant test score difference. Yet, GPT was 20 times more efficient, highlighting the ability to rapidly provide personalized feedback for all students, unlike human graders.

Despite positive outcomes, challenges remain in fully adopting AI for grading. A key issue is the accuracy and fairness of LLM grading. Pardos [74] reported a 32% error rate in GPT's feedback for certain math topics, with some errors being significant. Relying solely on LLMs could lead to incorrect feedback or scores. Researchers suggest error mitigation strategies such as self-consistency checks, where AI generates and cross-verifies multiple solutions, and properly framing AI output to encourage critical acceptance by students. Current recommendations advocate using LLM-based grading to assist instructors by providing initial grades or comments for them to review and adjust, improving efficiency without losing oversight.

In addition to grading, LLM agents help in content creation for assessments, saving teachers and designers time. They can generate questions and complete multiple-choice tests. Doughty et al. [75] found that GPT-4 generated questions are comparable in quality to human-made ones. Educational tools like Quizlet use AI to automatically generate customized practice tests and flashcards, combining their databases with LLMs [78].

Table 6. Analysis of LLM-based assessment and feedback methods.

Aspect	Description	Example	Key Findings
Automated Grading	Zero-shot grading of student submissions via prompt engineering, without additional model training.	AAG system prompts for scoring quantitative and open-ended answers [72]	High consistency with human grades; boosted student motivation and preparedness [72,74]
Multi-Agent Evaluation	Multiple LLMs debate or peer-review each other's assessments to reduce single-model bias.	ChatEval's debate framework for open-ended responses [76]	Improved correlation with human evaluations and lower error rates [76]
Adaptive Feedback Generation	Real-time, personalized hints and comments tailored to each student's answer.	ChatGPT generated step-by-step algebra hints [74]	Equivalent learning gains to human-written hints, at 20× faster throughput [74]
Automated Essay Scoring	Rubric-guided essay evaluation using LLMs with in-prompt rubrics.	GPT-4 grading essays against instructor-provided rubrics [1]	Moderate-to-high agreement with human raters when given clear rubrics [74]
Question Generation	AI-driven creation of multiple-choice and practice questions for formative use.	GPT-4 MCQ generation in programming courses [75,78]	Expert ratings found AI-generated items on par with human-authored questions [75]

3.4. Curriculum Design and Content Creation

Beyond tutoring interactions and grading, LLM-based agents are being applied to curriculum design, lesson planning, and content creation [84,88]. Their contribution spans a continuum of instructional tasks. Figure 5 situates the studies reviewed in this section along that continuum, from minimal AI assistance, where the model merely suggests resources that teachers curate, to extensive automation, in which entire textbooks are generated under human oversight.

In higher education, artificial intelligence is increasingly being explored to assist in course material development. Professors have used LLMs such as ChatGPT to generate illustrative examples for lectures or create variations of homework problems, providing students with additional practice opportunities [81,82]. With careful editing and oversight, these AI-generated materials can enrich curricula and diversify learning resources [19,87]. For example, recent frameworks like ELF (Educational LLM Framework) have been developed to semi-automatically generate and refine entire sets of teaching materials—including study guides, dialogical content, and structured instructional modules—by leveraging LLMs in combination with expert teacher input and systematic evaluation. Although quality control remains essential, these approaches point to more efficient and scalable content-authoring workflows in education [19].

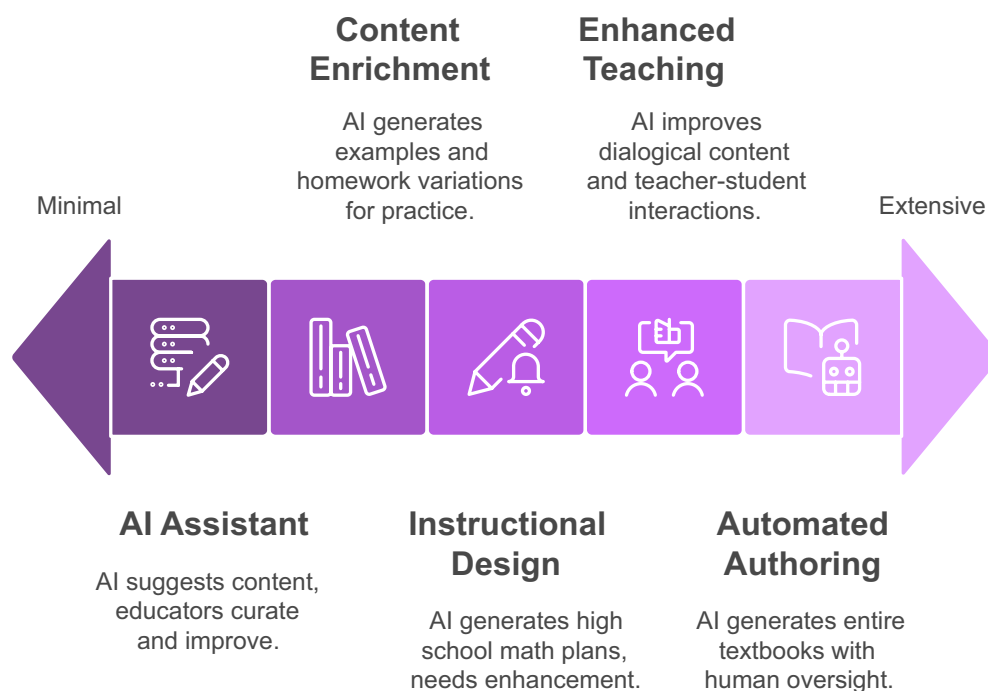


Figure 5. Spectrum of AI involvement in curriculum design and content creation. Moving left to right, the level of automation increases from AI assistant to automated authoring.

The work by Hu et al. [83] examined LLMs in instructional design, identifying strengths and weaknesses. Using GPT-4, they created high school math teaching plans based on mathematical problem chains and prompts. These plans, rated above six on an eight-point scale, showed practical usability. However, weaknesses were found in interdisciplinary design, mathematical culture integration, and addressing individual differences, indicating a need to improve instructional design components for diverse educational needs.

Although AI content generation is a powerful tool, human oversight remains essential. Educators are strongly advised to review and verify any AI-created lesson plans, quiz questions, or study materials for accuracy, appropriateness, and alignment with learning goals [79,85]. Studies have shown that while LLM-generated educational content can be useful as supplementary material, it often lacks the depth, contextual understanding, and pedagogical subtlety found in teacher-created resources [80,86,87]. For example, comparative analyses reveal that AI-generated explanations for math or programming problems are not always as effective or accurate as those written by experienced educators, and may miss important instructional cues or context [79,86,87]. As a result, best practices increasingly point toward a human–AI collaboration model for curriculum design: AI drafts or suggests content, and a skilled educator curates, edits, and improves it to ensure educational quality and relevance.

3.5. Personalized and Engaging Learning Experiences

LLM-based agents show potential in enabling personalized learning and boosting student engagement at scale [1]. Personalization in educational technology seeks to adjust instruction to individual differences [4,130] by tailoring content, pace, and interaction to the needs and preferences of each learner, a task difficult for teachers in large classrooms. LLMs provide new methods to achieve this through dynamic content generation and advanced conversational interaction [11].

One of the most robust findings in educational psychology is the large effect size of timely and specific feedback on achievement [131]. Modern LLM pipelines can now automate that loop: the system analyzes the work of the student, retrieves examples of common misconceptions, and generates hints that contain errors, all within a single conversational turn [89]. Meta-analytic evidence shows that intelligent tutoring systems delivering such adaptive feedback yield mean learning gains of $d = 0.66$ over business-as-usual controls [132].

In terms of affective and motivational scaffolds, engagement is not purely cognitive; LLM chatbots can detect sentiment in the learner's utterances and respond with encouragement or reframing strategies [103,133]. For example, research on AI chatbots for primary English learning found that affective scaffolding—such as regulating frustration and offering encouragement—not only helped learners self-correct errors but also improved their confidence and persistence in speaking tasks. Such findings echo the theory of self-determination, where autonomy-supportive feedback maintains intrinsic motivation and fosters deeper engagement and well-being [134,135].

Agents based on LLMs can be combined with vision or simulation engines to create multimodal and experiential learning environments. For example, platforms such as DALverse integrate assistive technology and metaverse-based environments to provide students with interactive multimodal tasks, such as manipulating virtual objects or exploring scientific phenomena in mixed reality, and have been shown to increase engagement and retention in distance education [92]. This approach aligns with Kolb's experiential learning theory, which posits that learning is deepened through cycles of concrete experience, reflective observation, abstract conceptualization, and active experimentation. Recent studies confirm that applying Kolb's framework in science and professional education improves knowledge, attitudes, and teaching effectiveness [93–95,136]. Furthermore, early research on virtual and puzzle-based platforms for engineering education demonstrates that multimodal tutors and interactive environments can significantly improve the retention and mastery of spatial concepts compared to traditional methods [136].

The Artificial Intelligence-Enabled Intelligent Assistant (AIIA) [96] uses AI for personalized adaptive learning in higher education. The AIIA framework uses NLP to build an interactive platform that meets individual needs and reduces the cognitive load. It adapts communication styles, offers empathetic responses, and includes Q&A, flashcards, quizzes, and summaries. The authors provided evidence that AI-enabled virtual teaching assistants (VTAs) improve academic performance while noting challenges such as academic integrity and the necessary policies for AI in assessments.

AI advancements, particularly LLMs, offer opportunities to enhance online education by moving beyond generic models. These adapt to individual preferences, allowing personalized learning. Wang et al. [90] proposed LearnMate, based on GPT-4, to create personalized learning plans based on goals, preferences, and progress. It offers personalized guidelines, assistance with course selection, and real-time help, enhancing understanding and engagement. The authors highlighted LLMs' potential for education and plan to expand LearnMate's functionality, interactivity, and adaptability for diverse contexts such as adult education and college support.

Hao et al. [21] explored Chinese university students' interactions with AI in collaborative LLM-assisted settings. After analyzing 110 students, they identified interaction patterns as active questioners, responsive navigators, or silent listeners. Few-shot LLM prompting and epistemic network analysis help in understanding these patterns. The study emphasizes the design of adaptable AI systems for diverse student needs to increase engagement and performance and highlights the need for targeted interventions to promote self-regulation and active participation.

Although AI-driven personalization offers the promise of more tailored learning experiences, it can also risk amplifying existing inequities or undermining teacher agency if not thoughtfully implemented [97–99]. Researchers are increasingly calling for a hybrid "AI-in-the-loop" or co-orchestration model, where teachers curate and adapt the recommendations generated by AI agents to ensure alignment with curricular goals and address the socioemotional needs of students [101,102]. Studies involving classroom technology probes and teacher-centered design processes show that such shared control between teachers and AI supports both equity and professional autonomy, allowing teachers to dynamically balance individual and collaborative learning while maintaining oversight over instructional decisions. Initial field trials in secondary schools suggest that this co-orchestration approach can improve learning outcomes without diminishing the essential role of teachers in guiding and supporting students [102].

In general, the capacity of LLMs to personalize and facilitate active participation represents a potential key advantage over standardized educational approaches. However, realizing this potential effectively probably requires a hybrid model in which human teachers remain indispensable to orchestrate learning experiences, provide higher-order guidance, and offer crucial socio-emotional support, while AI agents handle more routine aspects of personalization and interaction [91,100].

3.6. Ethical Considerations and Challenges

Building AI tutors that are powerful and trustworthy requires more than technical prowess; it demands compliance with internationally recognized ethical frameworks and transparent acknowledgment of real-world failure modes. In this section, we link each challenge to the UNESCO Recommendation on the Ethics of Artificial Intelligence [137] and the Ethics Guidelines for Trustworthy AI issued by the European Commission [138] (EC) and then illustrate the problem with concrete, education-specific cases and outline mitigation strategies in line with those guidelines.

3.6.1. Accuracy and Hallucination

LLMs occasionally generate authoritative but false statements, a phenomenon known as hallucination. In controlled classroom trials, Pardos and Bhandari [74] found that 32% of GPT-generated algebra hints contained substantive errors, and several students reproduced those errors in subsequent homework submissions. This threatens the UNESCO principles of safety and accuracy and violates the EC requirement for explainability.

A way to mitigate this is to use RAG to keep the answers grounded in vetted course materials. For example, a separate verifier agent rejects outputs with a low grounding score. In production systems such as Khan Academy Khanmigo, the tutor also displays inline citations and double-checks its answer when confidence is low [119].

3.6.2. Bias and Fairness

LLMs mainly trained on English data often underperform in low-resource languages, reproduce cultural biases, and suggest learning paths misaligned with minority contexts. For example, Dey et al. [109] reports that the macro-F1 of GPT-4 falls from ~86% in English to an average of ~64% in Bangla, Hindi, and Urdu, an accuracy gap large enough to distort the formative assessment. Yong et al. [110] further show that prompts in several African and indigenous languages can bypass GPT-4 safety filters, indicating additional equity and security risks. These behaviors contrast with the EC guidelines for the principles of diversity, nondiscrimination, fairness, and the call for inclusiveness.

Approaches to mitigating these problems are to (i) curate balanced domain corpora before fine-tuning [106,107], (ii) run bias audits using frameworks like Reinforcement Learning from Multi-role Debates (RLDF) to simulate student interactions and detect

disparities [112,113], and, for high-stakes educational applications, (iii) deploy multi-agent fairness checkers like MOMA or MBIAS to dynamically intervene in biased reasoning patterns while preserving pedagogical utility [108,114].

3.6.3. Privacy and Data Protection

A March 2023 glitch in the ChatGPT service briefly exposed the conversation titles and snippets of other users, illustrating how dialogue histories can leak even in closed SaaS deployments [120]. Beyond such bugs, Zou et al. [121] showed that carefully crafted prompt injection attacks can coerce GPT-4-class models into revealing private context, e.g., hidden system instructions, making educational deployments particularly vulnerable. These incidents violate the spirit of the Family Educational Rights and Privacy Act (FERPA) in the US [139] and run counter to the General Data Protection Regulation (GDPR) in Europe.

Techniques to mitigate these problems include encrypting learner data at rest and setting log rotation for less than 30 days, according to the FERPA cloud service guidelines [139]. Additionally, deploying the model within an on-premise cluster or Virtual Private Cloud (VPC), a logically isolated network infrastructure, provides stronger data location control [140]. Applying differentially private fine-tuning [115] and exposure-auditing techniques [116], together with a real-time Personally Identifiable Information (PII) scrubber (a tool designed to detect and remove or anonymize sensitive personal data) further reduces re-identification risk [105].

3.6.4. Academic Integrity

Large language models sharply reduce the cost of plagiarism and unauthorized assistance [111]. A recent Higher Education Policy Institute (HEPI) poll of 1041 undergraduates in the UK found that 88% had used generative AI tools in an assessment during the previous academic year and 25% admitted to incorporating AI-generated text into the version they submitted for grading [104]. Institutional data echo that trend: an EDUCAUSE QuickPoll of 278 US colleges reported that faculty believe that 23% of students are already handing in unedited AI output, while another 29% edit the text before submission [122]. Detection services are also signaling scale; in the first three months of Turnitin's AI writing detector, 10.3% of 65 million papers were marked as containing at least 20% AI-generated prose [123].

These figures clash with the UNESCO recommendation on the Ethics of AI principle of accountability and responsibility [137] and the Ethics Guidelines of the European Commission for Trustworthy AI requirements for technical robustness [138]. Institutions are therefore redesigning the course to make copy-and-paste cheating less efficient by applying the following techniques:

- Assessment redesign. Tasks that require personal reflection, process logs, authentic data, or oral defense reduce the value of recycled AI content and align with research on the authentic evaluation and mitigation of contract cheating [117,118].
- Socratic tutoring guardrails. Open-source systems such as CodeHelp refuse to reveal full solutions and instead guide students through hints and questions [17]. Similar "Socratic bots" have been deployed in data science courses with positive student ratings and minimal leakage of answers.
- Provenance technologies. Statistical watermarking of LLM output [124] and transcript-level provenance logs gives instructors post hoc tools to prove AI overuse.

4. Discussion

This survey aimed to map how large language model (LLM) agents are being designed, deployed, and studied across tutoring, assessment, curriculum design, and learner engagement contexts. Using a structured literature review protocol, we selected 82 sources

of peer review and industry (2023–2025) and coded them in six facets: technical and pedagogical frameworks, tutoring systems, assessment and feedback, curriculum design, personalization, and ethical considerations. Across these studies, five themes recurred: (i) retrieval grounding dramatically reduces hallucination; (ii) prompt-engineering guardrails preserve academic integrity; (iii) multi-agent debate boosts accuracy on ill-structured tasks; (iv) affective scaffolds raise persistence; and (v) co-orchestration with teachers mitigates equity risks. Taken together, the evidence suggests that hybrid human–AI workflows, rather than fully autonomous tutors, yield the most robust learning gains.

Our review conforms to the four-phase systematic protocol shown in Figure 3. The PRISMA 2020 statement recommends transparency in search and selection. We reported (i) all query strings, (ii) inclusion/exclusion criteria, and (iii) a PRISMA flow diagram (Figure 4).

This review is bounded by four constraints. First, the brisk release cycle of LLMs means that a snapshot of 2025 may age rapidly. Second, limiting the corpus to English-language publications probably excludes breakthroughs reported in low-resource settings. Third, most empirical studies we analyzed lasted fewer than eight weeks and drew on modest samples, weakening the external validity of their learning-gain estimates. Finally, the literature skews toward positive findings; null or negative results are scarce so that publication bias can overstate overall effectiveness.

Our research directions outline several key areas for future study. First, cross-lingual equity focuses on developing and auditing LLM tutors in underrepresented languages, using synthetic data augmentation in tandem with participatory design tailored to local contexts. Second, multimodal reasoning aims to combine vision or simulation engines with text-based agents to enhance support in STEM domains that involve diagrams, code execution, or virtual laboratories. Third, governance toolkits are needed to operationalize bias audits, provenance logging, and privacy-preserving fine-tuning pipelines to help institutions comply with regulations like GDPR and FERPA. Lastly, teacher co-orchestration analytics emphasizes designing dashboards that present AI recommendations, confidence scores, and student affect signals to enable educators to intervene strategically.

These avenues echo the call of UNESCO for AI systems that enhance human capacities while protecting fundamental rights [137] and chart a path toward the ethically aligned and evidence-based adoption of AI-powered educational agents.

5. Conclusions

The integration of AI-driven intelligent agents, especially those powered by large language models (LLMs), represents a transformative shift in educational technology, offering compelling opportunities for personalized instruction, increased engagement, and more effective assessment. As this paper has highlighted, advances in retrieval-augmented generation, prompt engineering, fine-tuning, multi-agent systems, and pedagogical strategy alignment have significantly enhanced the capabilities and reliability of agent tutoring systems. These advancements have broadened the scope of applications, from personalized tutoring and adaptive assessments to dynamic curriculum design, reshaping the educational environment.

However, as educators, technologists, and policymakers increasingly adopt AI agents, they must carefully navigate critical ethical considerations and practical challenges. Issues such as accuracy, privacy, fairness, academic integrity, and the evolving role of human educators must be proactively addressed through transparent and responsible AI system design, ongoing monitoring, and comprehensive regulatory frameworks. Interestingly, recent advances suggest feasible pathways to overcome these challenges and promote the responsible and beneficial use of AI in education.

In practice, the most effective approach combines the strengths of AI, scalability, instant feedback, and adaptive personalization with the subtle, empathetic, and ethically grounded roles of human educators. Rather than replacing teachers, AI technologies should serve as powerful companions, enabling educators to focus on higher-order mentoring and fostering critical thinking, creativity, and social-emotional learning. The key to successful integration lies in sustained collaboration between AI developers, educational researchers, policymakers, and teachers themselves.

Looking ahead, future research must continue to refine the capabilities of AI agents, explore novel pedagogical applications, rigorously assess educational outcomes, and ensure ethical and equitable access to these technologies. As LLM-based agents become more sophisticated, the goal remains clear: to enhance learning experiences, empower students, and equip educators with tools that meaningfully address individual learning needs at scale. By thoughtfully navigating these opportunities and challenges, we can leverage AI agents to build more engaging, effective, and equitable educational environments for learners around the world.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We thank the Ministry of Science, Humanities, Technology and Innovation (SECIHTI) for its support through the National System of Researchers (SNII).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]
2. Jeon, J.; Lee, S. Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Educ. Inf. Technol.* **2023**, *28*, 15873–15892. [CrossRef]
3. Woolf, B.P. *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*; Morgan Kaufmann: San Francisco, CA, USA, 2010.
4. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educ. Psychol.* **2011**, *46*, 197–221. [CrossRef]
5. Sahlman, W.A.; Ciechanover, A.M.; Grandjean, E. *Khanmigo: Revolutionizing Learning with GenAI*; Harvard Business School Publishing: Boston, MA, USA, Case No. 9-824-059, November 2023; revised April 2024. Available online: <https://www.hbs.edu/faculty/Pages/item.aspx?num=64929> (accessed on 30 May 2025).
6. OpenAI. Filling Crucial Language Learning Gaps: Duolingo and GPT-4. 2023. Available online: <https://openai.com/index/duolingo/> (accessed on 5 May 2025).
7. Bates, T. *Teaching in a Digital Age: Guidelines for Designing Teaching and Learning*; Tony Bates Associates Ltd.: Vancouver, BC, Canada, 2019.
8. Lee, S.J.; Kwon, K. A systematic review of AI education in K-12 classrooms from 2018 to 2023: Topics, strategies, and learning outcomes. *Comput. Educ. Artif. Intell.* **2024**, *6*, 100211. [CrossRef]
9. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.t.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9459–9474.
10. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730–27744.
11. Yusuf, H.; Money, A.; Daylamani-Zad, D. Pedagogical AI Conversational Agents in Higher Education: A Conceptual Framework and Survey of the State of the Art. *Educ. Technol. Res. Dev.* **2025**, *73*, 815–874. [CrossRef]

12. Guu, K.; Hashimoto, T.; Oren, Y.; Liang, P. REALM: Retrieval-Augmented Language Model Pre-Training. In Proceedings of the 37th International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020; pp. 3929–3938.
13. Swacha, J.; Gracel, M. Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications. *Appl. Sci.* **2025**, *15*, 4234. [[CrossRef](#)]
14. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
15. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24824–24837.
16. Gregorcic, B.; Polverini, G.; Sarlah, A. ChatGPT as a Tool for Honing Teachers’ Socratic Dialogue Skills. *Phys. Educ.* **2024**, *59*, 045005. [[CrossRef](#)]
17. Liffiton, M.; Sheese, B.; Savelka, J.; Denny, P. CodeHelp: Using Large Language Models with Guardrails for Scalable Support in Programming Classes. In Proceedings of the 23rd International Conference on Computing Education Research (Koli Calling ’23), Koli, Finland, 13–18 November 2023; Volume 3631830, ACM International Conference Proceeding Series; pp. 1–11. [[CrossRef](#)]
18. Dong, Q.; Li, L.; Dai, D.; Zheng, C.; Wu, Z.; Chang, B.; Sun, X.; Xu, J.; Sui, Z. A Survey on In-Context Learning. *arXiv* **2022**, arXiv:2301.00234.
19. Tan, K.; Yao, J.; Pang, T.; Fan, C.; Song, Y. ELF: Educational LLM Framework of Improving and Evaluating AI Generated Content for Classroom Teaching. *ACM J. Data Inf. Qual.* **2025**, *16*, 12:1–12:25. [[CrossRef](#)]
20. Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; Neubig, G. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* **2023**, *55*, 1–35. [[CrossRef](#)]
21. Hao, Z.; Jiang, J.; Yu, J.; Liu, Z.; Zhang, Y. Student Engagement in Collaborative Learning with AI Agents in an LLM-Empowered Learning Environment: A Cluster Analysis. *arXiv* **2025**, arXiv:2503.01694.
22. Wei, J.; Lauren, J.; Ma, T.; Chi, M. Larger Language Models Do In-Context Learning Differently. *arXiv* **2023**, arXiv:2303.03846.
23. Jin, H.; Lee, S.; Shin, H.; Kim, J. Teach ai how to code: Using large language models as teachable agents for programming education. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–28.
24. Fan, Y.; Jiang, F.; Li, P.; Li, H. GrammarGPT: Exploring Open-Source LLMs for Native Chinese Grammatical Error Correction with Supervised Fine-Tuning. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Foshan, China, 12–15 October 2023; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2023; Volume 15022, pp. 69–80.
25. Singhal, K.; Azizi, S.; Tu, T.; Mahdavi, S.S.; Wei, J.; Chung, H.W.; Scales, N.; Tanwani, A.K.; Cole-Lewis, H.; Pfohl, S.; et al. Large language models encode clinical knowledge. *Nature* **2023**, *620*, 172–180. [[CrossRef](#)]
26. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A Survey of Large Language Models. *arXiv* **2023**, arXiv:2303.18223.
27. Treveil, M.; Papadopoulos, M.; Karampatziakis, N.; Bradley, R.; Clarkson, B. *Introducing MLOps: How to Scale Machine Learning in the Enterprise*; O’Reilly Media, Inc.: Sebastopol, CA, USA, 2020.
28. Wang, Z.; Moriyama, S.; Wang, W.Y.; Gangopadhyay, B.; Takamatsu, S. Talk Structurally, Act Hierarchically: A Collaborative Framework for LLM Multi-Agent Systems. *arXiv* **2025**, arXiv:2502.11098.
29. Zhang, G.; Yue, Y.; Li, Z.; Yun, S.; Wan, G.; Wang, K.; Cheng, D.; Yu, J.X.; Chen, T. Cut the Crap: An Economical Communication Pipeline for LLM-based Multi-Agent Systems. *arXiv* **2024**, arXiv:2410.02506.
30. Yuksel, K.A.; Sawaf, H. A Multi-AI Agent System for Autonomous Optimization of Agentic AI Solutions via Iterative Refinement and LLM-Driven Feedback Loops. *arXiv* **2024**, arXiv:2412.17149.
31. Hendrycks, D.; Burns, C.; Kadavath, S.; Arora, A.; Basart, S.; Tang, E.; Song, D.; Steinhardt, J. Measuring mathematical problem solving with the math dataset. *arXiv* **2021**, arXiv:2103.03874.
32. Park, J.S.; O’Brien, J.; Cai, C.J.; Morris, M.R.; Liang, P.; Bernstein, M.S. Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual Acm Symposium on User Interface Software and Technology, San Francisco, CA, USA, 29 October–1 November 2023; pp. 1–22.
33. Neuberger, S.; Eckhaus, N.; Berger, U.; Taubenfeld, A.; Stanovsky, G.; Goldstein, A. SAUCE: Synchronous and Asynchronous User-Customizable Environment for Multi-Agent LLM Interaction. *arXiv* **2024**, arXiv:2411.03397.
34. Li, Q.; Xie, Y.; Chakravarty, S.; Lee, D. EduMAS: A Novel LLM-Powered Multi-Agent Framework for Educational Support. In Proceedings of the 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, 15–18 December 2024; pp. 8309–8316.
35. Cemri, M.; Pan, M.Z.; Yang, S.; Agrawal, L.A.; Chopra, B.; Tiwari, R.; Keutzer, K.; Parameswaran, A.; Klein, D.; Ramchandran, K.; et al. Why Do Multi-Agent LLM Systems Fail? *arXiv* **2025**, arXiv:2503.13657.
36. Helmi, T. Modeling Response Consistency in Multi-Agent LLM Systems: A Comparative Analysis of Shared and Separate Context Approaches. *arXiv* **2025**, arXiv:2504.07303.

37. Nooyens, R.; Bardakci, T.; Beyazit, M.; Demeyer, S. Test Amplification for REST APIs via Single and Multi-Agent LLM Systems. *arXiv* **2025**, arXiv:2504.08113.
38. Liu, Z.; Yin, S.X.; Lee, C.; Chen, N.F. Scaffolding language learning via multi-modal tutoring systems with pedagogical instructions. In Proceedings of the 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, 25–27 June 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1258–1265.
39. Sonkar, S.; Liu, N.; Mallick, D.B.; Baraniuk, R.G. CLASS: A Design Framework for building Intelligent Tutoring Systems based on Learning Science principles. *arXiv* **2023**, arXiv:2305.13272.
40. Pietersen, C.; Smit, R. Transforming embedded systems education: The potential of large language models. *South. J. Eng. Educ.* **2024**, *3*, 62–83. [\[CrossRef\]](#)
41. Favero, L.; Pérez-Ortiz, J.A.; Käser, T.; Oliver, N. Enhancing Critical Thinking in Education by means of a Socratic Chatbot. *arXiv* **2024**, arXiv:2409.05511.
42. Xie, X.; Yang, X.; Cui, R. A Large Language Model-Based System for Socratic Inquiry: Fostering Deep Learning and Memory Consolidation. In Proceedings of the 2025 14th International Conference on Educational and Information Technology (ICEIT), Guangzhou, China, 14–16 March 2025; pp. 52–57.
43. Chen, M.; Wu, L.; Liu, Z.; Ma, X. The Impact of Metacognitive Strategy-Supported Intelligent Agents on the Quality of Collaborative Learning from the Perspective of the Community of Inquiry. In Proceedings of the 2024 4th International Conference on Educational Technology (ICET), Wuhan, China, 13–15 September 2024; pp. 11–17.
44. Rehan, M.; John, R.; Nazli, K. Developing Self-Regulated Learning: The Impact of Metacognitive Strategies at Undergraduate Level. *Crit. Rev. Soc. Sci. Stud.* **2025**, *3*, 3402–3411. Available online: <https://thecrsss.com/index.php/Journal/article/view/402> (accessed on 30 May 2025).
45. Synekop, O.; Lavrysh, Y.; Lukianenko, V.; Ogienko, O.; Lytovchenko, I.; Stavyska, I.; Halatsyn, K.; Vadaska, S. Development of Students' Metacognitive Skills by Means of Educational Technologies in ESP Instruction at University. *BRAIN. Broad Res. Artif. Intell. Neurosci.* **2023**, *14*, 128–142. [\[CrossRef\]](#)
46. Sarfarajl, G.K. Intelligent Tutoring System Enhancing Learning with Conversational AI: A Review. *Int. J. Sci. Res. Eng. Manag.* **2025**, *9*, 1–9. [\[CrossRef\]](#)
47. Katsarou, E.; Wild, F.; Sougari, A.M.; Chatzipanagiotou, P. A Systematic Review of Voice-based Intelligent Virtual Agents in EFL Education. *Int. J. Emerg. Technol. Learn.* **2023**, *18*, 65–85. [\[CrossRef\]](#)
48. Gopi, S.; Sreekanth, D.; Dehbozorgi, N. Enhancing Engineering Education Through LLM-Driven Adaptive Quiz Generation: A RAG-Based Approach. In Proceedings of the 2024 IEEE Frontiers in Education Conference (FIE), Washington, DC, USA, 13–16 October 2024; pp. 1–8.
49. Grijalva, P.K.; Cornejo, G.E.; Gómez, R.R.; Real, K.P.; Fernández, A. Herramientas colaborativas para revisiones sistemáticas. *Rev. Espac.* **2019**, *40*, 1–10.
50. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* **2010**, *8*, 336–341. [\[CrossRef\]](#)
51. Kim, J.; Vajravelu, B.N. Assessing the Current Limitations of Large Language Models in Advancing Health Care Education. *JMIR Form. Res.* **2025**, *9*, e51319. [\[CrossRef\]](#)
52. Nikolovski, V.; Trajanov, D.; Chorbev, I. Advancing AI in Higher Education: A Comparative Study of Large Language Model-Based Agents for Exam Question Generation, Improvement, and Evaluation. *Algorithms* **2025**, *18*, 144. [\[CrossRef\]](#)
53. Li, H.; Fang, Y.; Zhang, S.; Lee, S.M.; Wang, Y.; Trexler, M.; Botelho, A.F. ARCHED: A Human-Centered Framework for Transparent, Responsible, and Collaborative AI-Assisted Instructional Design. *arXiv* **2025**, arXiv:2503.08931.
54. Zhang, X.; Zhang, C.; Sun, J.; Xiao, J.; Yang, Y.; Luo, Y. EduPlanner: LLM-Based Multi-Agent Systems for Customized and Intelligent Instructional Design. *IEEE Trans. Learn. Technol.* **2025**, *18*, 416–427. [\[CrossRef\]](#)
55. Yu, J.; Zhang, Z.; Zhang-li, D.; Tu, S.; Hao, Z.; Li, R.M.; Li, H.; Wang, Y.; Li, H.; Gong, L.; et al. From MOOC to MAIC: Reshaping Online Teaching and Learning through LLM-driven Agents. *arXiv* **2024**, arXiv:2409.03512.
56. Shen, W.; Xu, T.; Li, H.; Zhang, C.; Liang, J.; Tang, J.; Yu, P.S.; Wen, Q. Large Language Models for Education: A Survey and Outlook. *arXiv* **2024**, arXiv:2403.18105.
57. Wang, D.; Liang, J.; Ye, J.; Li, J.; Li, J.; Zhang, Q.; Hu, Q.; Pan, C.; Wang, D.; Liu, Z.; et al. Enhancement of the performance of large language models in diabetes education through retrieval-augmented generation: Comparative study. *J. Med. Internet Res.* **2024**, *26*, e58041. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Eggmann, F.; Weiger, R.; Zitzmann, N.U.; Blatz, M.B. Implications of large language models such as ChatGPT for dental medicine. *J. Esthet. Restor. Dent.* **2023**, *35*, 1098–1102. [\[CrossRef\]](#)
59. Gao, W.; Liu, Q.; Yue, L.; Yao, F.; Lv, R.; Zhang, Z.; Wang, H.; Huang, Z. Agent4Edu: Generating Learner Response Data by Generative Agents for Intelligent Education Systems. *arXiv* **2025**, arXiv:2501.10332. [\[CrossRef\]](#)
60. Wang, T.; Zhou, N.; Chen, Z. CyberMentor: AI Powered Learning Tool Platform to Address Diverse Student Needs in Cybersecurity Education. *arXiv* **2025**, arXiv:2501.09709.

61. Marouf, A.; Al-Dahdooh, R.; Ghali, M.J.A.; Mahdi, A.O.; Abunasser, B.S.; Abu-Naser, S.S. Enhancing Education with Artificial Intelligence: The Role of Intelligent Tutoring Systems. *Int. J. Eng. Inf. Syst. (IJEAIS)* **2024**, *8*, 10–16.
62. Alsafari, B.; Atwell, E.; Walker, A.; Callaghan, M. Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants. *Nat. Lang. Process. J.* **2024**, *8*, 100101. [CrossRef]
63. Neira-Maldonado, P.; Quisi-Peralta, D.; Salgado-Guerrero, J.; Murillo-Valarezo, J.; Cárdenas-Arichábala, T.; Galan-Mena, J.; Pulla-Sanchez, D. Intelligent educational agent for education support using long language models through Langchain. In Proceedings of the International Conference on Information Technology & Systems, Temuco, Chile, 8–10 February 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 258–268.
64. Maiti, P.; Goel, A.K. How Do Students Interact with an LLM-Powered Virtual Teaching Assistant in Different Educational Settings? *arXiv* **2024**, arXiv:2407.17429.
65. Molina, I.V.; Montalvo, A.; Ochoa, B.; Denny, P.; Porter, L. Leveraging LLM Tutoring Systems for Non-Native English Speakers in Introductory CS Courses. *arXiv* **2024**, arXiv:2411.02725.
66. Neumann, A.T.; Yin, Y.; Sowe, S.K.; Decker, S.J.; Jarke, M. An LLM-Driven Chatbot in Higher Education for Databases and Information Systems. *IEEE Trans. Educ.* **2024**, *68*, 103–116. [CrossRef]
67. Pan, M.; Kitson, A.; Wan, H.; Prpa, M. ELLMA-T: An Embodied LLM-agent for Supporting English Language Learning in Social VR. *arXiv* **2024**, arXiv:2410.02406.
68. Gold, K.; Geng, S. Tracking the Evolution of Student Interactions with an LLM-Powered Tutor. In Proceedings of the 14th International Learning Analytics & Knowledge Conference (LAK '24), Kyoto, Japan, 18–22 March 2024; ACM: New York, NY, USA, 2024; pp. 200–210.
69. Gold, K.; Geng, S. On the Helpfulness of a Zero-Shot Socratic Tutor. In Proceedings of the AIET 2024: 5th International Conference on AI in Education Technology, Barcelona, Spain, 29–31 July 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 45–56.
70. Henley, A.Z.; Kazemitabaar, M.; Ye, R.; Wang, X.; Henley, A.Z.; Denny, P.; Craig, M.; Grossman, T. CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs. In Proceedings of the CHI 2024, Honolulu, HI, USA, 11–16 May 2024; ACM: New York, NY, USA, 2024; pp. 1–20.
71. Nagarajan, S. AI-Enabled E-Learning Systems: A Systematic Literature. *Int. J. Recent Trends Multidiscip. Res.* **2024**, *4*, 23–28. [CrossRef]
72. Yeung, C.; Yu, J.; Cheung, K.C.; Wong, T.W.; Chan, C.M.; Wong, K.C.; Fujii, K. A Zero-Shot LLM Framework for Automatic Assignment Grading in Higher Education. *arXiv* **2025**, arXiv:2501.14305.
73. Lagakis, P.; Demetriadis, S. EvaAI: A Multi-agent Framework Leveraging Large Language Models for Enhanced Automated Grading. In Proceedings of the International Conference on Intelligent Tutoring Systems, Thessaloniki, Greece, 10–13 June 2024; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2024; Volume 14798, pp. 378–385.
74. Pardos, Z.A.; Bhandari, S. ChatGPT-Generated Help Produces Learning Gains Equivalent to Human Tutor-Authored Help on Mathematics Skills. *PLoS ONE* **2024**, *19*, e0304013. [CrossRef]
75. Doughty, J.; Wan, Z.; Bompelli, A.; Qayum, J.; Wang, T.; Sakr, M. A Comparative Study of AI-generated (GPT-4) and Human-crafted MCQs in Programming Education. In Proceedings of the 26th Australasian Computing Education Conference, Sydney, Australia, 29 January–2 February 2024; ACM: New York, NY, USA, 2024; pp. 114–123.
76. Chan, C.M.; Chen, W.; Su, Y.; Yu, J.; Xue, W.; Zhang, S.; Fu, J.; Liu, Z. ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate. *arXiv* **2023**, arXiv:2308.07201.
77. Pankiewicz, M.; Baker, R.S. Large Language Models (GPT) for Automating Feedback on Programming Assignments. In Proceedings of the 31st International Conference on Computers in Education, Matsue, Japan, 4–8 December 2023; Asia-Pacific Society for Computers in Education: Taoyuan City, Taiwan, 2023; Volume I, pp. 68–77. [CrossRef]
78. Lex. Introducing Q-Chat, the World's First AI Tutor Built with OpenAI's ChatGPT. Quizlet HQ Blog Post. 2023. Available online: <https://quizlet.com/blog/meet-q-chat> (accessed on 5 May 2025).
79. Mustofa, H.A.; Kola, A.J.; Owusu-Darko, I. Integration of Artificial Intelligence (ChatGPT) into Science Teaching and Learning. *Int. J. Ethnoscience Technol. Educ.* **2025**, *2*, 108–128. [CrossRef]
80. Durak, H.Y.; Eğin, F.; Onan, A. A Comparison of Human-Written Versus AI-Generated Text in Discussions at Educational Settings: Investigating Features for ChatGPT, Gemini and BingAI. *Eur. J. Educ.* **2025**, *60*, e70014. [CrossRef]
81. Fleischhauer, K.; Friedrich, K. Factors determining the efficacy of AI-generated word problems for content-specific math language courses in higher education. *Scripta Manent* **2024**, *19*, 4–24. Available online: <https://journals.uni-lj.si/scriptamanent/article/view/21366> (accessed on 30 May 2025). [CrossRef]
82. Luo, H.; Weng, J.; Lu, X.; Yu, Z. BPE: Exploring the Prompt Framework of Physics Exercises Generated from Bloom's Taxonomy in LLM. In Proceedings of the 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA), Hangzhou, China, 18–20 October 2024; pp. 26–29.
83. Hu, B.; Zheng, L.; Zhu, J.; Ding, L.; Wang, Y.; Gu, X. Teaching Plan Generation and Evaluation with GPT-4: Unleashing the Potential of LLM in Instructional Design. *IEEE Trans. Learn. Technol.* **2024**, *17*, 1471–1485. [CrossRef]

84. Lin, Z.; Sun, Y. A Smart English Learning Curriculum Generation Mobile Platform based on Word Root Extension using Artificial Intelligence and LLM (Large Language Model). *CS IT* **2024**, *14*, 87–96. [\[CrossRef\]](#)
85. Robinson, B. Harnessing AI for Structured Learning: The Case for Objective-Driven Design in E-Learning. *Int. J. E-Learn.* **2024**, *23*, 457–470. [\[CrossRef\]](#)
86. Amirjalili, F.; Neysani, M.; Nikbakht, A. Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Front. Educ.* **2024**, *9*, 1–11.
87. Denny, P.; Khosravi, H.; Hellas, A.; Leinonen, J.; Sarsa, S. Can we trust AI-generated educational content? Comparative analysis of human and AI-generated learning resources. *arXiv* **2023**, arXiv:2306.10509.
88. Ufondu, C.C.; Udeh, V.C.; Amobi, C.M. Inculcating Digital Contents, Technological Tools: Agents for Enhancing Business Education Curriculum for Skill Acquisition in Nigerian Universities. *Online J. TVET Pract.* **2023**, *8*, 108–115.
89. Li, Z.; Yazdanpanah, V.; Wang, J.; Gu, W.; Shi, L.; Cristea, A.I.; Kiden, S.; Stein, S. TutorLLM: Customizing Learning Recommendations with Knowledge Tracing and Retrieval-Augmented Generation. *arXiv* **2025**, arXiv:2502.15709.
90. Wang, X.J.; Lee, C.; Mutlu, B. LearnMate: Enhancing Online Education with LLM-Powered Personalized Learning Plans and Support. *arXiv* **2025**, arXiv:2503.13340.
91. Joe, A.; Sebastin, S. Human-AI Collaboration in Education using Leveraging Deep Learning and Natural Language Processing to Enhance Personalized Learning Systems. *ICTACT J. Soft Comput.* **2025**, *15*, 3695–3703.
92. Damasceno, A.R.P.; Silva, L.C.; Barros, E.M.; Oliveira, F.C.M.B. DALverse: Assistive Technology for Inclusion of People with Disabilities in Distance Education through a Metaverse-Based Environment. In Proceedings of the 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), Nicosia, Cyprus, 1–4 July 2024; pp. 142–146.
93. Choshi, M. Addressing Challenges in Undergraduate Community Health Nursing Clinical: Kolb’s Experiential Learning Theory. *J. Nurs. Educ.* **2024**, *63*, 1–4. [\[CrossRef\]](#)
94. Choi, Y. *The Effect of Early Childhood Science Education Class Based on Kolb’s Experiential Learning Theory on The Pre-Service Early Childhood Teacher’s Science Knowledge, Science Attitude, and Science Teaching Efficacy*; Educational Research Institute: Republic of Korea, 2024.
95. Haritha, D.; Rao, D. A Holistic Approach to Professional Development: Integrating Kolb’s Experiential Learning Theory for Soft Skills Mastery. *J. Eng. Educ. Transform.* **2024**, *37*, 415–424.
96. Sajja, R.; Sermet, Y.; Cikmaz, M.; Cwiertny, D.; Demir, I. Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information* **2024**, *15*, 596. [\[CrossRef\]](#)
97. Gunawan, A.; Wiputra, R. Exploring the Impact of Generative AI on Personalized Learning in Higher Education. In Proceedings of the 2024 IEEE 12th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 7 December 2024; pp. 280–285.
98. Muhammad, F.; Orji, C.M. Revolutionizing Education In The Digital Era: The Role Of Ai In Promoting Inclusivity, Equality, And Ethical Innovation. *Am. J. Soc. Sci. Educ. Innov.* **2024**, *6*, 151–157. [\[CrossRef\]](#)
99. Bhatnagar, R.; Sharma, D.K. Using AI to Create More Equitable and Inclusive Higher Education Learning Systems. *Commun. Appl. Nonlinear Anal.* **2024**, *32*, 25–44. [\[CrossRef\]](#)
100. Qushwa, F.G.; Onia, S.I. AI Innovation in Education: Realizing Personalized Learning in the Digital Age. *Educ. J. Ilm. Pendidik.* **2024**, *2*, 178–187. [\[CrossRef\]](#)
101. Lawrence, L.; Echeverría, V.; Yang, K.B.; Aleven, V.; Rummel, N. How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *Br. J. Educ. Technol.* **2023**, *55*, 823–844. [\[CrossRef\]](#)
102. Echeverría, V.; Yang, K.B.; Lawrence, L.; Rummel, N.; Aleven, V. Designing Hybrid Human–AI Orchestration Tools for Individual and Collaborative Activities: A Technology Probe Study. *IEEE Trans. Learn. Technol.* **2023**, *16*, 191–205. [\[CrossRef\]](#)
103. Kwon, H.; Shim, K. The study on categorizing scaffolding types of AI chatbots for primary English learning-oriented assessment. *Korea Assoc. Prim. Engl. Educ.* **2023**, *29*, 33–51.
104. Rowsell, J. Nine in 10 UK Undergraduates Now Using AI in Assessments—Survey, 2025. *Times Higher Education*, 26 February 2025. Available online: <https://www.timeshighereducation.com/news/nine-10-uk-undergraduates-now-using-ai-assessments-survey> (accessed on 2 May 2025).
105. Microsoft Presidio. Presidio: Data Protection and De-identification SDK. 2025. Available online: <https://microsoft.github.io/presidio/> (accessed on 2 May 2025).
106. Chu, Z.; Wang, S.; Xie, J.; Zhu, T.; Yan, Y.; Ye, J.; Zhong, A.; Hu, X.; Liang, J.; Yu, P.S.; et al. Llm agents for education: Advances and applications. *arXiv* **2025**, arXiv:2503.11733.
107. Kumar, R.; Kumar, H.; Shalini, K. Detecting and Mitigating Bias in LLMs through Knowledge Graph-Augmented Training. In Proceedings of the 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE), Nite, India, 6–7 February 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 608–613.
108. Xu, Z.; Chen, W.; Tang, Y.; Li, X.; Hu, C.; Chu, Z.; Ren, K.; Zheng, Z.; Lu, Z. Mitigating Social Bias in Large Language Models: A Multi-Objective Approach Within a Multi-Agent Framework. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 25579–25587.

109. Dey, K.; Tarannum, P.; Hasan, M.A.; Razzak, I.; Naseem, U. Better to Ask in English: Evaluation of Large Language Models on English, Low-Resource and Cross-Lingual Settings. *arXiv* **2024**, arXiv:2410.13153.
110. Yong, Z.X.; Menghini, C.; Bach, S.H. Low-Resource Languages Jailbreak GPT-4. *arXiv* **2024**, arXiv:2310.02446.
111. Hussain, Z.D. Forming Policies for Ethical AI Use in Academic Writing. In Proceedings of the SLIIT International Conference on Advancements in Sciences and Humanities (SICASH 2024), Sri Jayawardenepura Kotte, Sri Lanka, 4 December 2024.
112. Cheng, R.; Ma, H.; Cao, S.; Li, J.; Pei, A.; Wang, Z.; Ji, P.; Wang, H.; Huo, J. Reinforcement Learning from Multi-role Debates as Feedback for Bias Mitigation in LLMs. *arXiv* **2024**, arXiv:2404.10160.
113. Echterhoff, J.; Liu, Y.; Alessa, A.; McAuley, J.; He, Z. Cognitive bias in decision-making with LLMs. *arXiv* **2024**, arXiv:2403.00811.
114. Raza, S.; Raval, A.; Chatrath, V. Mbias: Mitigating bias in large language models while retaining context. *arXiv* **2024**, arXiv:2405.11290.
115. Xie, C.; Lin, Z.; Backurs, A.; Gopi, S.; Yu, D.; Inan, H.A.; Nori, H.; Jiang, H.; Zhang, H.; Lee, Y.T.; et al. Differentially private synthetic data via foundation model apis 2: Text. *arXiv* **2024**, arXiv:2403.01749.
116. Sun, B.; Zhou, Y.; Jiang, H. Empowering Users in Digital Privacy Management through Interactive LLM-Based Agents. *arXiv* **2024**, arXiv:2410.11906.
117. Grobler, M. Exploring Essential Concepts In The Automation Of Student Plagiarism Management-A Case Study. *J. Law Sustain. Dev.* **2024**, *12*, e4002. [CrossRef]
118. Davies, T. Artificial Intelligent Branching Simulations (AIBS) in Critical Care Neonatal Nursing; Pecha Kucha presentation, University of Melbourne, Melbourne, Australia, November 2024. Available online: [https://findanexpert.unimelb.edu.au/scholarlywork/1946498-artificial-intelligent-branching-simulations-\(aibs\)-in-critical-care-neonatal-nursing](https://findanexpert.unimelb.edu.au/scholarlywork/1946498-artificial-intelligent-branching-simulations-(aibs)-in-critical-care-neonatal-nursing) (accessed on 30 May 2025).
119. Khan Academy. Introducing Khanmigo: An AI-Powered Guide Powered by GPT-4. 2023. Available online: <https://blog.khanacademy.org/khanmigo> (accessed on 1 May 2025).
120. OpenAI. 20 March 2023 ChatGPT Outage: Incident Report. 2023. Available online: <https://openai.com/blog/march-20-chatgpt-outage> (accessed on 1 May 2025).
121. Zou, A.; Wang, Z.; Carlini, N.; Nasr, M.; Kolter, J.Z.; Fredrikson, M. Universal and Transferable Adversarial Attacks on Aligned Language Models. *arXiv* **2023**, arXiv:2307.15043.
122. Muscanell, N.; Robert, J. EDUCAUSE QuickPoll Results: Did ChatGPT Write This Report? *EDUCAUSE Review*, 14 February 2023. Available online: <https://er.educause.edu/articles/2023/2/educause-quickpoll-results-did-chatgpt-write-this-report> (accessed on 30 May 2025).
123. Williams, T. Turnitin Says One in 10 University Essays Are Partly AI-Written. *Times Higher Education*, 25 July 2023. Available online: <https://www.timeshighereducation.com/news/turnitin-says-one-10-university-essays-are-partly-ai-written> (accessed on 2 May 2025).
124. Kirchenbauer, J.; Geiping, J.; Wen, Y.; Katz, J.; Miers, I.; Goldstein, T. A Watermark for Large Language Models. In Proceedings of the 40th International Conference on Machine Learning, Hawaii, HI, USA, 23–29 July 2023; pp. 16918–16941. [CrossRef]
125. Maia, D.; dos Santos, S.C. Monitoring Students' Professional Competencies in PBL: A Proposal Founded on Constructive Alignment and Supported by AI Technologies. In Proceedings of the 2022 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 8–11 October 2022; pp. 1–8.
126. Nye, B.D.; Graesser, A.C.; Hu, X. AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *Int. J. Artif. Intell. Educ.* **2014**, *24*, 427–469. [CrossRef]
127. Roll, I.; Wylie, R. Evolution and revolution in Artificial Intelligence in Education. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 582–599. [CrossRef]
128. Pavitra, D.K.H.; Agnihotri, D.A. Artificial Intelligence in Corporate Learning and Development: Current Trends and Future Possibilities. In Proceedings of the 2023 Second International Conference on Smart Technologies for Smart Nation (SmartTechCon), Singapore, 18–19 August 2023; pp. 688–693.
129. Fauziyah, U.; Kaburuan, E.R.; Wang, G.; Aqsha. Gamification for Employee Training Platform in Banking Industries. In Proceedings of the 2019 International Conference on Information Management and Technology (ICIMTech), Jakarta/Bali, Indonesia, 19–20 August 2019; Volume 1, pp. 503–508.
130. Koedinger, K.R.; Corbett, A. Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom. In *The Cambridge Handbook of the Learning Sciences*; Sawyer, R.K., Ed.; Cambridge University Press: Cambridge, UK, 2006; pp. 61–77.
131. Hattie, J.; Timperley, H. The power of feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [CrossRef]
132. Kulik, J.A.; Fletcher, J.D. Effectiveness of intelligent tutoring systems: A meta-analytic review. *Rev. Educ. Res.* **2016**, *86*, 42–78. [CrossRef]
133. Lin, H.C.K. Employing AI Chatbot-Based Tutoring Systems and STEAM 6e Scaffold to Implement Cooperative Learning of Maker Education. *Impact* **2022**, *5*, 14–16.

134. Neufeld, A. Autonomy-supportive teaching in medicine: From motivational theory to educational practice. *MedEdPublish* **2021**, *10*, 117. [CrossRef]
135. Ryan, R.M.; Deci, E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **2000**, *55*, 68. [CrossRef]
136. Viswanathan, V.K.; Mirza, S.H.; Nayak, C.R.; Calhoun, M. Evaluation of a Puzzle-based Virtual Platform for Improving Spatial Visualization Skills in Engineering Freshmen. In Proceedings of the 2020 ASEE Annual Conference & Exposition, Virtual, 22–26 June 2020; Paper No. 34602; American Society for Engineering Education: Washington, DC, USA, 2020. [CrossRef]
137. UNESCO. Recommendation on the Ethics of Artificial Intelligence. 2021. Available online: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (accessed on 1 May 2025).
138. European Commission High-Level Expert Group on AI. Ethics Guidelines for Trustworthy Artificial Intelligence. 2019. Available online: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed on 1 May 2025).
139. U.S. Department of Education. Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices. 2014. Available online: <https://studentprivacy.ed.gov/resources/protecting-student-privacy-while-using-online-educational-services-requirements-and-best> (accessed on 1 May 2025).
140. Jansen, W.; Grance, T. *Guidelines on Security and Privacy in Public Cloud Computing*; Technical Report NIST SP 800-144; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2011. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.