

e-mentor

DWUMIESIĘCZNIK SZKOŁY GŁÓWNEJ HANDLOWEJ W WARSZAWIE
WSPÓŁWYDAWCA: FUNDACJA PROMOCJI I AKREDYTACJ KIERUNKÓW EKONOMICZNYCH

2025, nr 4 (111)



Kujundziski, A. P., Ademi, N. & Rahmani, D. (2025). AI chatbots in education – the future of teaching and learning. *e-mentor*, 4(111), 4–12. <https://www.doi.org/10.15219/em111.1724>

Aleksandra
Porjazoska
Kujundziski

AI Chatbots in Education – the Future of Teaching and Learning

Neslihan
Ademi

Damir
Rahmani

Abstract

AI chatbots in education can offer substantial benefits for both students and teachers by monitoring learners' progress, adapting to individual learning pace, supporting instruction and enhancing teachers' pedagogical practice. This study aims to outline key strategies and practical guidance for the responsible integration of chatbots into higher education. An initial evidence synthesis was conducted using a systematic–narrative hybrid review of the literature, a simplified PRISMA-style flow diagram and thematic coding of the narrative synthesis; 36 peer-reviewed articles were identified. The review makes an original contribution by mapping emerging pedagogical, technological and ethical dimensions of chatbot adoption that remain under-examined in existing secondary studies. Evolving roles of AI chatbots in learning design, student engagement, interaction and assessment, along with opportunities to strengthen their educational value, are identified. The study also offers practical implications by outlining evidence-informed strategies for instructors, curriculum designers and institutions to support the effective and responsible implementation of AI chatbots. Future work will extend the search to additional databases and examine discipline-specific applications of AI chatbots.

Keywords: chatbots, AI chatbots, education, higher education, teaching and learning


Introduction


Cutting-edge artificial-intelligence (AI) tools, including chatbots, are reshaping many aspects of life by supporting task completion, improving decision-making, and enhancing social interaction in both personal and professional contexts. Challenges facing the education sector by the COVID-19 pandemic accelerated the adoption of AI in teaching and learning. Since then, the integration of AI into educational settings has continued to expand (Ifelebuegu et al., 2023). Market analyses suggest that over 987 million people worldwide use AI chatbots, and forecasts by Research and Markets (2024) indicate that the global chatbot market will reach nearly \$46.64 billion by 2029, confirming their increasing adoption across industries.

In education, the use of chatbots has increased markedly since November 2022, when OpenAI released its generative AI (GenAI) tool, ChatGPT (Razak et al., 2023). Prior to ChatGPT, AI had been used in education for many years, primarily within learning-management systems to support administrative tasks. The emergence of GenAI and large language models (LLMs) such as ChatGPT, Google Bard, Apple Siri, and IBM Watson has begun to reshape traditional pedagogical approaches (Davar et al., 2025; Gill et al., 2024; Kujundziski & Bojadjiev, 2025). Educators and institutions have experimented with chatbots not only for administrative automation but also to support learning, formative assessment, and writing and idea generation (Davar et al., 2025; Ifelebuegu et al., 2023).

At the same time, advances in machine learning and access to larger training datasets have improved chatbot performance in tasks such as summarising texts, answering student queries, and offering language-practice activities – often beyond the capabilities of earlier models such as BERT and XLNet (Davar et al., 2025). The deployment of GPT models in education can help address a key limitation of online educational platforms: delays in providing students with timely responses (Ngo et al., 2024). Such

Aleksandra Porjazoska Kujundziski, International Balkan University, North Macedonia,  <https://orcid.org/0000-0001-8177-3328>

Neslihan Ademi, International Balkan University, North Macedonia,  <https://orcid.org/0000-0002-9510-8182>

Damir Rahmani, International Balkan University, North Macedonia,  <https://orcid.org/0000-0001-7570-0026>

systems are used across a range of domains, including student support (e.g., admissions and administrative procedures), personalised learning activities tailored to learners' needs and preferences (e.g., improving speaking and writing skills), and academic assistance with tasks such as idea generation and grammar correction, sometimes functioning as 'digital supervisors' (Krumsvik et al., 2025).

The application of chatbots in teaching and learning has attracted substantial scholarly attention (McGrath et al., 2024; Mishra, 2024). This work has focused mainly on their accuracy and the benefits they may offer to students (Abdallah et al., 2024; Wang et al., 2023) and teachers (Mishra, 2024), as well as the drawbacks they may introduce into the educational process (Birenbaum, 2023; Williams, 2023). Building on this literature, our study examines strategies for integrating chatbots effectively into higher education.

Background

Chatbots are computer programs designed to simulate written and/or spoken conversation, enabling users to interact on specific topics via text or voice. They are also referred to by various names, including 'virtual assistants', 'conversational agents', 'smart personal assistants', and 'dialogue systems'. Early chatbots achieved human-like exchanges largely through keyword matching or relatively simple natural language processing (NLP) techniques (Caldarini et al., 2022; Frangoudes et al., 2021; Pergantis et al., 2025). Subsequent advances in machine learning (ML), NLP, and generative AI have enabled more sophisticated applications (Davar et al., 2025; Pergantis et al., 2025), beginning with IBM Watson in 2006. This trajectory continued in 2011 with Apple's voice-activated personal assistant Siri, although Siri is not strictly a chatbot. In 2018, Google's Duplex demonstrated the ability to manage complex, real-time interactions (Davar et al., 2025).

Released by OpenAI in 2022, ChatGPT is a prominent LLM application based on the Generative Pre-trained Transformer (GPT) architecture, specifically GPT-3.5. Owing to its ability to generate human-like text and diverse content, GPT-3.5 and subsequent versions – GPT-4 and GPT-4o ('omni'), released in March 2023 and May 2024 respectively – have gained significant popularity (Kooli, 2023; Plevris et al., 2023). Google developed its own chatbot, Bard, in 2022, with performance comparable to OpenAI's ChatGPT.

The LLM developed by the AI start-up DeepSeek has attracted attention for its competitiveness, improved performance, and lower operating costs compared with ChatGPT. The DeepSeek R1 model can understand and generate text in both Chinese and English.

LLMs now handle a wide range of tasks and are increasingly used in teaching and learning (Leite, 2024). Pre-trained on vast datasets, contemporary generative AI models are capable of understanding context and producing coherent, contextually relevant responses

to user queries, reflecting advances in deep learning technology (Kooli, 2023).

AI chatbots play a multipurpose role in education. First, they can automate administrative tasks. Available at any time, AI-based chatbots can provide students with immediate responses and support, thereby improving institutional efficiency (Saleh et al., 2025). Second, through interaction with users, chatbots can map learners' patterns and pace and adjust learning materials to their capabilities. This personalises learning experiences, boosts engagement, and increases motivation (Labadze et al., 2023; Srinivasan et al., 2023). Moreover, chatbots can optimise resource allocation and serve as a critical link in the modern educational landscape (Abdallah et al., 2024). To foster an unbiased and engaging educational environment for all students, education systems need collaboration among educators, policymakers, and chatbot developers to establish principles for the responsible use of these tools, aligned with national strategies for AI implementation (Davar et al., 2025; Labadze et al., 2023).

Despite these benefits, the widespread use of AI chatbots in education also raises concerns and challenges, including academic integrity, plagiarism, and the risk of incorrect or biased outputs (often referred to as 'hallucinations'). Many publications discuss factors involved in integrating emerging AI chatbot technologies into educational settings and identify associated challenges (Annamalai et al., 2023). However, most studies focus on a single aspect of chatbot implementation in academia or on isolated instructional outcomes (e.g., educational methodologies, chatbot accuracy, academic perceptions of integration, or ethical challenges), and there remains a lack of coherent synthesis bringing together pedagogical, methodological, and ethical insights emerging alongside the rapid development of generative AI. Accordingly, this study adopts a comprehensive approach and aims to synthesise a broader set of conceptual patterns to clarify how chatbots are reshaping learning design, student engagement, interaction, and assessment practices in higher education, thereby making an original contribution. By examining current trends, conceptual developments, and implementation challenges, and by identifying the evolving roles and educational functions of AI chatbots in contemporary digital learning environments, the study also contributes to theory. In addition, it offers practical implications by outlining evidence-informed strategies and directions for the integration of chatbots by instructors, curriculum designers, and higher education institutions.

Research question: What strategies support the seamless and successful integration of AI-powered chatbots into academic curricula?

Methodology

This study undertakes an initial evidence synthesis using a systematic–narrative hybrid literature review (SNHLR), combining a structured search strategy with

interpretive narrative synthesis. This design is appropriate for emerging and rapidly developing areas such as AI chatbots in higher education, where the evidence base is methodologically diverse (Greenhalgh et al., 2018; Snyder, 2019; Turnbull et al., 2023). In this context, a full systematic review – requiring exhaustive multi-database coverage and formal risk-of-bias appraisal – was neither feasible nor well aligned with the exploratory and integrative aims of the review (Turnbull et al., 2023).

The evidence synthesis comprised peer-reviewed publications investigating the integration, pedagogical applications and implications of AI-powered chatbots in education.

The literature search was conducted in the complementary databases Web of Science (WoS), Google Scholar and Semantic Scholar. As institutional access to Scopus – often considered important for comprehensive systematic coverage – was unavailable, the study was positioned as an initial evidence synthesis supported by the SNHLR design.

Comparable Boolean search strings were used across the three databases, with syntax adapted to each platform to support transparency and reproducibility (Table 1, Appendix).

A PRISMA-style flow diagram (Coupe et al., 2019) is presented in Figure 1 (Appendix). Restricting results to publications from 2021 to 2025, the searches returned 2,487 records (WoS $n = 775$; Google Scholar $n = 1,230$; Semantic Scholar $n = 482$). After removing 1,059 duplicates, 1,428 records proceeded to title and abstract screening, guided by predefined inclusion and exclusion criteria (Table 2, Appendix). At this stage, 1,253 records were excluded because they were not related to education, did not involve AI chatbots, focused solely on technical model development without educational application, were not peer reviewed, or were not available as full-text articles in English. A total of 175 full-text articles were then assessed for eligibility.

Full-text eligibility was assessed in relation to the research question and the inclusion criteria. Eligible studies addressed AI chatbots, conversational agents or generative AI systems (e.g. ChatGPT, Bard, Gemini) in a higher-education context; provided empirical evidence, a conceptual model or pedagogical analysis relating to teaching, learning, assessment, or student/teacher perceptions; and were peer-reviewed publications available in full text.

Of the 175 full-text articles, 139 were excluded for the reasons summarised in Table 3. Exclusions included purely technical studies focusing on chatbot architecture or algorithm design (where this could not be determined at title/abstract stage); studies outside higher education; papers that were substantively irrelevant despite keyword matches; and papers with insufficient methodological or theoretical depth. The remaining 36 studies were taken forward for thematic synthesis.

Bibliographic data and thematic content were extracted from each included study (Table 4, Appendix).

Using thematic coding within a narrative analytic synthesis, the authors identified recurring patterns and emergent themes, including pedagogical integration and instructional design; student learning and engagement outcomes; educator perceptions and institutional readiness; and the ethical, motivational and evaluative dimensions of chatbot adoption. The coding framework and procedure (including Table 5) are provided in the Appendix. The coding was intended to support narrative synthesis rather than to produce quantitative reliability coefficients, consistent with qualitative interpretive approaches.

Content analysis across themes provided an integrated understanding of how AI chatbots contribute to, challenge and reshape educational practices. To enhance the reliability of the review and reduce potential bias, the two reviewers (A.P.K. and N.A.) analysed the papers independently. Disagreements were discussed and resolved by consensus.

Given the study's emphasis on an initial, structured evidence synthesis and the methodological diversity of the included studies, formal risk-of-bias tools (e.g. the Cochrane checklist or the Mixed Methods Appraisal Tool, MMAT) were not applied. This is a limitation of the review and reinforces its positioning as an initial evidence synthesis rather than a full systematic review. Nevertheless, quality was supported through clearly specified inclusion criteria, transparent screening procedures and an evidence-informed synthesis of findings (Wiboolyasarini et al., 2025).

The Integration of AI Chatbots in Education

The four emergent themes together provide a comprehensive framework for addressing the central research question: how best to integrate AI-powered chatbots into academic curricula. Each theme, synthesised through thematic content analysis, captures a distinct strategic dimension of integration – pedagogical design; learning impact, monitoring and evaluation; stakeholder readiness; and ethical and governance considerations. Taken together, these dimensions address the chatbots' roles and intended uses, pedagogical design principles, technology selection, implementation, and mechanisms for ongoing monitoring and evaluation (see Figure 2 in the Appendix). Although AI chatbots can benefit all educational stakeholders, their incorporation into education presents substantial challenges relating to reliability, accuracy, and ethical implications, which require careful scrutiny if their benefits are to be maximised (Groothuisen et al., 2024; Ngo et al., 2024).

Identifying Chatbot Roles and Principles for Pedagogical Design

Identifying a chatbot's primary role and functions is the first step towards integrating AI chatbots into education. Depending on the intended use, a chatbot may provide administrative (service-oriented) support, act as a teaching- or tutoring-oriented platform, or

support research and development by assisting with information retrieval and guidance (Baglivo et al., 2023; Moral-Sánchez et al., 2023; Okonkwo & Ade-Ibijola, 2021).

Service-oriented chatbots – designed primarily for administrative assistance – such as Ask L.U. (Abbas et al., 2022), respond to routine queries about timetables, grades, enrolment, fees, campus services, and admissions, providing efficient and timely support to users (Labrague & Sabei, 2025). They may also be used to orient students to university life, as in the cases of Lisa and Differ (Al-Abdullatif et al., 2023). Such chatbots can be integrated into an existing learning management system (LMS) (e.g., Moodle) or deployed via popular messaging applications such as WhatsApp and Twitter (Kumar, 2021), Telegram (Merelo et al., 2022), or Facebook Messenger; examples include Dina, Ask Holly, and Lola (Roca et al., 2024). This integration enables the automation of routine tasks, including responses to frequently asked questions. Performance is commonly assessed through user questionnaires or by the proportion of satisfactory responses generated, and reported results are generally encouraging (Abbas et al., 2022).

LLM-based systems that support adaptive and personalised learning can provide tailored tutoring by aligning learning materials with individual needs, learning styles, and proficiency levels, while also offering prompt feedback (Al-Abdullatif, 2023; Calonge et al., 2023; Chang et al., 2023; Lai, 2024; Liu & Reinders, 2025; Mageira et al., 2022). However, their use may reduce student–teacher interaction, encourage overreliance on chatbot responses, and expose learners to hallucinated outputs. Seamless integration in higher education therefore requires clear decisions about the chatbot’s role and the pedagogical goals it serves, alongside an assessment of ethical and operational risks and strategies to minimise or mitigate them.

Although the reviewed studies highlight the instructional value of chatbots, broad generalisation is not warranted because chatbots are conceptualised differently across contexts. Several authors (Calonge et al., 2023; Chang et al., 2023; Lai, 2024; Liu & Reinders, 2025) describe chatbots as instructional partners that provide adaptive tutoring and feedback. By contrast, studies focusing on administrative or advisory applications (Abbas et al., 2022; Merelo et al., 2022; Roca et al., 2024) imply different pedagogical expectations, contributing to inconsistency across settings (Okonkwo & Ade-Ibijola, 2021).

Experimental studies (Essel et al., 2022; Mageira et al., 2022) tend to offer stronger evidence than survey-based research (Awad & Moosa, 2024; Ilieva et al., 2023), which captures perceptions rather than directly measuring chatbot performance. Multiple studies (Al-Abdullatif, 2023; Calonge et al., 2023; Chang et al., 2023) report that adaptive systems can facilitate and enhance learning. Achieving such benefits may require instructors to shift from traditional, exam-focused approaches towards learning-oriented pedagogies that promote critical thinking (Chang et al., 2023).

Some researchers report improved learning performance and critical thinking when students interact with chatbots using well-designed prompts (Chang et al., 2023; Lai, 2024). Conversely, overreliance on AI may reduce human-to-human interaction (Moral-Sánchez et al., 2023) and limit learners’ critical reasoning (Husain, 2024). Taken together, these findings suggest that successful implementation depends not only on technical capability, but also on alignment between chatbot operation, pedagogical objectives, and instructor oversight.

Teaching- and learning-oriented chatbots can function as intelligent tutors. They deliver educational materials and adaptive support, helping students learn specific subjects through explanations, formative assessment tracking, and immediate, customised feedback (Abbas et al., 2022; Calonge et al., 2023; Essel et al., 2022; Ilieva et al., 2023; Mageira et al., 2022). AI chatbots can support students’ self-regulated learning (Chang et al., 2023; Essel et al., 2022) and self-directed learning (Al-Abdullatif, 2023; Chang et al., 2023). They may also enhance higher-order cognitive skills (e.g., critical thinking and problem-solving) (Frangoudes et al., 2021; Moral-Sánchez et al., 2023) and metacognitive skills, as reported for the chatbot Bashayer (Al-Abdullatif et al., 2023).

Chatbots can also surface learning-analytics information – such as time spent on a page, click history, assignment deadlines, and overall progress – enabling students to take greater control of their learning and allowing instructors to encourage active learning and make informed judgements about progress (Chang et al., 2023; Lai, 2024). Because these functions involve sensitive data, there are risks relating to privacy and non-compliance with data protection regulations (Ifelebuegu et al., 2023; Ilieva et al., 2023; Kumar, 2021). The use of anonymised datasets and adherence to institutional data-governance protocols can help mitigate these risks.

Findings on the development of students’ self-regulated learning are mixed (Chang et al., 2023; Hwang & Chang, 2023). Where improvements are observed, they appear to stem from chatbot designs that use structured prompts, guided questioning, and individualised feedback (Al-Abdullatif, 2023; Chang et al., 2023; Essel et al., 2022). When feedback is overly generic, students may treat chatbots as answer generators, resulting in shallow engagement (Husain, 2024). Students also often struggle to identify errors in AI-generated content, which may indicate limited critical scrutiny. Similarly, integrating chatbots within active-learning pedagogies, such as problem-based learning – often alongside gamification – can improve engagement, motivation, and learning outcomes, but may also create risks of overdependence on gameplay and a loss of learning focus (Kumar, 2021; Lin & Chang, 2023). Potential mitigations include monitoring depth of learning through engagement analytics and limiting chatbot features that are not aligned with learning goals. Overall, the literature suggests that chatbot success depends on well-designed instruction and

reflective learning activities (Calonge et al., 2023), as well as students' capacity to evaluate and improve AI-generated content (Essel et al., 2022).

Regardless of the purpose for which a chatbot is used – whether to improve teaching efficiency and reduce instructors' workload (Abbas et al., 2022; Calonge et al., 2023; Essel et al., 2022; Ilieva et al., 2023; Mageira et al., 2022) or to develop particular learner skills (e.g., language learning and communication skills) (Hwang & Chang, 2023; Kim et al., 2021) – the literature recognises a range of challenges. These include plagiarism, bias in teaching materials, threats to assessment validity, language bias, contextual inaccuracies, and cultural misinterpretations. Comparable risks arise when chatbots aggregate learning materials from multiple sources and provide simultaneous access to learners across geographical locations and time zones. At the same time, such systems can support collaborative learning and offer distinctive learning experiences (Al-Abdullatif, 2023; Lai, 2024; Liu & Reinders, 2025). To mitigate these issues, the literature recommends clear disclosure policies, hybrid human–AI evaluation workflows, and plagiarism-resistant assessment formats.

Practical Implementation Factors and Transferable Implications for Curriculum Design

As the literature indicates, some of the most significant practical implementation challenges associated with integrating chatbots in education concern compatibility with existing platforms, access to chatbot-development tools, instructor readiness, digital literacy, and curriculum alignment (Al-Abdullatif et al., 2023; Essel et al., 2022; Kumar, 2021; Labadze et al., 2023; Merelo et al., 2022). In practice, this can create tensions between innovation and academic integrity: whilst chatbots may enhance formative learning, their use in summative contexts remains problematic.

In terms of curriculum design, the main implications relate to pedagogical redesign (Abbas et al., 2022; Awad & Moosa, 2024; Calonge et al., 2023), the adaptation of learning environments (Chang et al., 2023; Lai, 2024), assessment practices (Essel et al., 2022), skills development and AI literacy (Labadze et al., 2023; Liu & Reinders, 2025), and staff capacity building (Labadze et al., 2023). Building on the preceding discussion of pedagogy, curricula should adopt a blended (hybrid) model in which AI provides personalised guidance, whilst instructors retain pedagogical control (Abbas et al., 2022; Calonge et al., 2023; Chang et al., 2023; Lai, 2024; Liu & Reinders, 2025). This approach can scaffold self-regulated learning (SRL) and student autonomy (Al-Abdullatif, 2023; Lai, 2024; Liu & Reinders, 2025), as illustrated by Lin and Chang's (2023) 'CHAT-ACTS' framework. Provided that these principles are followed, chatbots can be framed as complementary instructional partners rather than replacements for teachers (Husain, 2024).

Assessment practices are another critical factor in ensuring coherent chatbot integration. It has been

recognised that traditional assessment methods based on the recall of memorised facts are insufficient in AI-supported education to capture the competencies required in today's labour market. One strategy to address this gap, suggested by several authors (Abbas et al., 2022; Essel et al., 2022), is to adopt Assessment for Learning (AfL) principles. Studies indicate that shifting from paper examinations to oral examinations, group projects, portfolios, and case studies – and incorporating instructor-designed reflective components (e.g., guided self-assessment prompts, comparative analysis tasks, or ethics-focused questions) – can enable students to evaluate AI-generated content critically in terms of reliability, bias, and ethical implications (Calonge et al., 2023; Lai, 2024).

Chatbot integration is also more challenging when AI literacy among both instructors and students is low. Instructor readiness depends on understanding basic AI concepts, including how models are trained and how information is retrieved, alongside the ability to appraise AI-generated content for accuracy, bias, reliability, and ethical issues (Labadze et al., 2023; Liu & Reinders, 2025). By contrast, readily accessible no-code platforms such as DialogFlow, Flow XO, and Botsify allow educators to design and customise chatbots without extensive programming knowledge (Essel et al., 2022), thereby lowering the technical barrier to integration. Kumar (2021), for example, describes the chatbot QMT212, built using Textit and integrated with Telegram, Twitter, Facebook Messenger, and SMS.

However, no-code tools improve accessibility rather than guaranteeing pedagogical readiness. This highlights the need for hands-on training for instructors on chatbots' capabilities and limitations, potential applications, customisation, and best practices for integrating these tools into teaching methodologies (Essel et al., 2022; Labadze et al., 2023), alongside professional development programmes focused on designing AI-inclusive syllabi, developing blended teaching strategies, and establishing ethical guidelines for data use. Such professional development is necessary to retain educators' central role in a rapidly evolving technological environment (Al-Abdullatif, 2023; Labadze et al., 2023).

As previously noted, the implementation of chatbots in educational practice has faced challenges relating to the accuracy of the outputs they generate, as well as limited user familiarity with the technology. Accordingly, pre-implementation analysis and evaluation of chatbot accuracy and efficiency is an important element of integration. This may involve objective metrics derived from analytics (e.g., the proportion of correct answers provided by the chatbot, or time to task completion), whilst subjective measures rely on users' perceptions of chatbot accuracy (Frangoudes et al., 2021; Navas et al., 2024).

Beyond effectiveness and accuracy, recent studies have highlighted implications relating to transparency and ethics (Awad & Moosa, 2024; Krumsvik et al., 2025; Razak et al., 2023), the reliability of chatbots

across domains (DaFonte et al., 2025; Navas et al., 2024), and broader societal impact (Gill et al., 2024). Evaluations of AI models (e.g., ChatGPT, Gemini, and Claude) suggest that reliability is task- and context-dependent, limiting transferability across domains (Husain, 2024; Lai, 2024); for instance, performance in medical knowledge does not necessarily translate into statistical reasoning or critical thinking. As shown by Navas et al. (2024), both versions of ChatGPT (GPT-3.5 and GPT-4.0) demonstrated lower accuracy in solving statistical problems than Bing. DaFonte et al. (2025) similarly found Claude 3.5 Sonnet and ChatGPT o3-mini to be more reliable in medical-surgery tests than Gemini 2.0 Flash and ChatGPT-4o mini, but to underperform in critical thinking and decision-making. Different cultural contexts and infrastructural constraints introduce additional barriers to transferability (Al-Abdullatif, 2023; Gill et al., 2024; Liu & Reinders, 2025). For example, models trained predominantly on English-language data often exhibit linguistic and cultural biases when deployed in non-English-speaking regions or culturally diverse learning environments. Moreover, limited readiness, alongside geopolitical and legal constraints (such as internet censorship), can materially affect the feasibility and transferability of AI deployment (Davar et al., 2025). Overall, the literature suggests that effective incorporation in educational contexts requires robust implementation mechanisms and monitoring frameworks, as well as policies that support contextual flexibility whilst maintaining shared international principles of fairness, transparency, and accountability in the educational use of AI (DaFonte et al., 2025; Frangoudes et al., 2021; Navas et al., 2024).

Ethical Aspects, Policies and Risk Governance

Establishing robust, transparent and human-centred frameworks for policy, curriculum and risk governance is essential for the effective integration of AI chatbots and LLMs across diverse educational settings. As the curriculum implications have already been discussed, this section focuses on ethical and policy challenges. AI-powered systems can access and process large volumes of personal data, raising concerns about privacy and data security (Ifelebuegu et al., 2023; Ilieva et al., 2023; Kumar, 2021). This highlights the urgent need to develop detailed policies, guidelines and protocols that comply with the General Data Protection Regulation (GDPR) to mitigate these risks (Williams, 2023).

AI systems may also be trained on datasets that include unethical content, gender stereotypes or other forms of discrimination, which can result in biased models (Calonge et al., 2023). Although AI systems are increasingly widespread, their underlying mechanisms and decision-making processes often remain opaque; in this sense, they function as ‘black-box’ models. Consequently, users may not have access to clear explanations of how a system processes their data, which can increase scepticism towards AI chatbots (Ifelebuegu et al., 2023).

An ethical framework is therefore required to ensure the responsible and transparent use of AI chatbots in research and education. Academic dishonesty and the broader ethical implications of AI technologies are additional risks associated with deploying AI chatbots in education. The use of chatbot technologies for teaching, learning and research – together with the capacity of AI-generated content to circumvent conventional plagiarism-detection tools (Sobaih & Abuelnasr, 2025) – raises questions about authorship, academic integrity, plagiarism and cheating (Mariyono & Hidayatullahet, 2025; Raptopoulou, 2025). Collectively, these publications underline the need for universities and national education agencies to develop a robust legal framework, including clear guidance and data-protection standards aligned with the GDPR, to support the ethical use of AI chatbots. Moreover, cooperation between institutions across regions, in line with UNESCO’s (n.d.) recommendations on AI in education and the OECD’s (n.d.) AI principles could help to develop governance models that accommodate cultural and legal variation.

On the basis of the literature reviewed, several practical implications emerge for institutions, instructors and learning designers. Instructors should treat AI chatbots as supportive tools and prioritise low-risk uses, particularly in large classes or blended-learning environments, where they can provide automatically generated explanations and just-in-time feedback. Effective implementation also requires alignment with the curriculum and the avoidance of ad hoc deployment. In addition, professional development and training programmes that build AI literacy can help staff address academic integrity issues associated with generative AI. Finally, educational institutions should establish governance arrangements that set data-protection standards, ethical-use protocols and rules for transparency.

Conclusions

This study employs a systematic literature review to identify strategies and to propose a framework for the successful integration of AI tools – particularly AI chatbots – within education. An initial evidence synthesis was undertaken using a systematic–narrative hybrid review of three complementary databases. Boolean search strings, predefined inclusion and exclusion criteria, and thematic coding within a narrative analytic synthesis were used to identify 36 peer-reviewed articles published between 2021 and 2025. The synthesis highlighted themes relating to the pedagogical integration and instructional design of AI chatbots, learning outcomes, AI literacy and institutional readiness, and the ethical and legal compliance of educational chatbots. Effective integration depends on comprehensive, human-centred institutional policies and compliance with regulations such as the GDPR, in order to mitigate academic dishonesty and algorithmic bias and to ensure adequate transparency regarding chatbot operation.

Despite providing a comprehensive overview of strategies for the seamless, unbiased and transparent integration of chatbots in education, this review has several limitations. Although three databases were consulted, the study relied on only one major indexed database (Web of Science), as restricted access prevented the inclusion of Scopus. Consequently, the authors focused on an initial evidence synthesis rather than a full systematic literature review. Moreover, given the methodological heterogeneity of the included studies, formal risk-of-bias assessment tools were not applied, which constitutes a further limitation. Future research should validate these findings using additional databases and standardised quality appraisal frameworks. Further studies could also examine the application of AI chatbots in disciplines that remain under-represented in this literature, such as economics and administrative sciences.

The appendix is available in the online version of the journal.

References

- Abbas, N., Whitfield, J., Atwell, E., Bowman, H., Pickard, T., & Walker, A. (2022). Online chat and chatbots to enhance mature student engagement in higher education. *International Journal of Lifelong Education*, 41(3), 308–326. <https://doi.org/10.1080/02601370.2022.2066213>
- Abdallah, A. K., Alkaabi, A. M., Mehjar, D. A., & Aradat, Z. A. (2024). Chatbots in classrooms: Tailoring education and boosting engagement. In A. Abdallah, A. Alkaabi, & R. Al-Riyami (Eds.), *Cutting-Edge Innovations in Teaching, Leadership, Technology, and Assessment* (pp. 166–181). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-0880-6.ch012>
- Al-Abdullatif, A. M. (2023). Modeling students' perceptions of chatbots in learning: Integrating technology acceptance with the value-based adoption model. *Education Sciences*, 13(11), 1151. <https://doi.org/10.3390/educsci13111151>
- Al-Abdullatif, A. M., Al-Dokhny, A. A., & Drwish, A. M. (2023). Implementing the Bashayer chatbot in Saudi higher education: measuring the influence on students' motivation and learning strategies. *Frontiers in Psychology*, 14, 1129070. <https://doi.org/10.3389/fpsyg.2023.1129070>
- Annamalai, N., Rashid, R. A., Munir Hashmi, U., Mohamed, M., Harb Alqaryouti, M., & Eddin Sadeq, A. (2023). Using chatbots for English language learning in higher education. *Computers and Education: Artificial Intelligence*, 5, 100153. <https://doi.org/10.1016/j.caeai.2023.100153>
- Awad, W., & Moosa, J. (2024). Implications of AI chatbots in education: Challenges and solution. *Journal of Statistics Applications and Probability*, 13(2), 611–622. <https://doi.org/10.18576/JSAP/130203>
- Baglivo, F., de Angelis, L., Casigliani, V., Arzilli, G., Privitera, G. P., & Rizzo, C. (2023). Exploring the possible use of AI chatbots in public health education: Feasibility study. *JMIR Medical Education*, 9, e51421. <https://doi.org/10.2196/51421>
- Birenbaum, M. (2023). The chatbots' challenge to education: Disruption or destruction? *Education Sciences*, 13(7), 711. <https://doi.org/10.3390/educsci13070711>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, 13(1), 41. <https://doi.org/10.3390/info13010041>
- Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching*, 6(2), 346–357. <https://doi.org/10.37074/jalt.2023.6.2.22>
- Chang, D. H., Lin, M. P. C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using ai chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, 15(17), 12921. <https://doi.org/10.3390/su151712921>
- Coupe, N., Peters, S., Rhodes, S., & Cotterill, S. (2019). The effect of commitment-making on weight loss and behaviour change in adults with obesity/overweight: A systematic review. *BMC Public Health*, 19(1), 816. <https://doi.org/10.1186/s12889-019-7185-3>
- DaFonte, N., Cadiente, A., Implicito, C., Becker, N., & Surick, B. (2025). Does AI have utility in medical student surgical education? A comparative analysis of chatbots in answering standardized surgical multiple-choice questions. *Global Surgical Education – Journal of the Association for Surgical Education*, 4, 61. <https://doi.org/10.1007/s44186-025-00369-3>
- Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI chatbots in education: Challenges and opportunities. *Information*, 16(3), 235. <https://doi.org/10.3390/info16030235>
- Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, 19, 57. <https://doi.org/10.1186/s41239-022-00362-6>
- Frangoudes, F., Hadjjaros, M., Schiza, E. C., Matsangidou, M., Tsivitanidou, O., & Neokleous, K. (2021). An overview of the use of chatbots in medical and healthcare education. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and collaboration technologies. Games and virtual environments for learning* (pp. 170–184). 8th International Conference, LCT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Proceedings, Part II. Springer. https://doi.org/10.1007/978-3-030-77943-6_11
- Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19–23. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- Greenhalgh, T., Thorne, S., & Malterud, K. (2018). Time to challenge the spurious hierarchy of systematic over narrative reviews? *European Journal of Clinical Investigation*, 48(6), e12931. <https://doi.org/10.1111/eci.12931>
- Groothuisen, S., van den Beemt, A., Remmers, J. C., & van Meeuwen, L. W. (2024). AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning. *Computers and Education: Artificial Intelligence*, 7, 100290. <https://doi.org/10.1016/j.caeai.2024.100290>
- Husain, A. J. A. (2024). Potentials of ChatGPT in computer programming: Insights from programming

AI Chatbots in Education – the Future of Teaching...

- instructors. *Journal of Information Technology Education: Research*, 23, 002. <https://doi.org/10.28945/5240>
- Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099–4112. <https://doi.org/10.1080/10494820.2021.1952615>
- Ifelebuegu, A. O., Kulume, P., & Cherukut, P. (2023). Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching*, 6(2), 332–345. <https://doi.org/10.37074/jalt.2023.6.2.29>
- Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M., & Angelov, D. (2023). Effects of generative chatbots in higher education. *Information*, 14(9), 492. <https://doi.org/10.3390/info14090492>
- Kim, H. S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics*, 21, 712–734. <https://doi.org/10.15738/kjell.21..202108.712>
- Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15(7), 5614. <https://doi.org/10.3390/su15075614>
- Krumsvik, R. J. (2025). Chatbots and academic writing for doctoral students. *Education and Information Technologies*, 30(7), 9427–9461. <https://doi.org/10.1007/s10639-024-13177-x>
- Kujundziski, A. P. & Bojadjiev, J. (2025). Artificial Intelligence in education: Transforming learning landscapes. In M. Stevkovska, M. Klemenchich, & N. K. Ulutaş (Eds.), *Reimagining Intelligent Computer-Assisted Language Education* (pp. 1-54). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-4310-4.ch001>
- Kumar, J. A. (2021). Educational chatbots for project-based learning: investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education*, 18, 65. <https://doi.org/10.1186/s41239-021-00302-w>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Labrague, L. J., & Sabei, S. (2025). Integration of AI-powered chatbots in nursing education: A scoping review of their utilization, outcomes, and challenges. *Teaching and Learning in Nursing*, 20(1), e285–e293. <https://doi.org/10.1016/j.teln.2024.11.010>
- Lai, J. W. (2024). Adapting self-regulated learning in an age of generative artificial intelligence chatbots. *Future Internet*, 16(6), 218. <https://doi.org/10.3390/fi16060218>
- Leite, B. S. (2024). Generative Artificial Intelligence in chemistry teaching: ChatGPT, Gemini, and Copilot's content responses. *Journal of Applied Learning and Teaching*, 7(2), 190–204. <https://doi.org/10.37074/jalt.2024.7.2.13>
- Lin, M. P.-C., & Chang, D. (2023). CHAT-ACTS: A pedagogical framework for personalized chatbot to enhance active learning and self-regulated learning. *Computers and Education: Artificial Intelligence*, 5, 100167. <https://doi.org/10.1016/j.caeai.2023.100167>
- Liu, M., & Reinders, H. (2025). Do AI chatbots impact motivation? Insights from a preliminary longitudinal study. *System*, 128, 103544. <https://doi.org/10.1016/j.system.2024.103544>
- Magreira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, 12(7), 3239. <https://doi.org/10.3390/app12073239>
- Mariyono, D., & Alif Hidayatullah, A. N. (2025). Navigating the Moral Maze: Ethical challenges and opportunities of generative chatbots in global higher education. *Applied Computational Intelligence and Soft Computing*, 8584141. <https://doi.org/10.1155/acis/8584141>
- McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2024). Generative AI chatbots in higher education: a review of an emerging research area. *Higher Education*, 89, 1533–1549. <https://doi.org/10.1007/s10734-024-01288-w>
- Merelo, J. J., Castillo, P. A., Mora, A. M., Barranco, F., Abbas, N., Guillén, A., & Tsivitanidou, O. (2022). Exploring the role of chatbots and messaging applications in higher education: A teacher's perspective. In P. Zaphiris, & A. Ioannou, (Eds), *Learning and Collaboration Technologies. Novel Technological Environments. HCII 2022. Lecture Notes in Computer Science*, vol. 13329 (pp. 205–223). Springer. https://doi.org/10.1007/978-3-031-05675-8_16
- Mishra, R. (2024). Redefining education through Artificial Intelligence: An in-depth analysis of faculty knowledge dimensions and AI chatbots integration in enhancing teaching effectiveness in higher education institutions. *Pakistan Journal of Life and Social Sciences (PjLSS)*, 22(2), 20150-20160. <https://doi.org/10.57239/pjlls-2024-22.2.001476>
- Moral-Sánchez, S. N., Rey, F. J. R., & Cebrián-De-la-Serna, M. (2023). Analysis of artificial intelligence chatbots and satisfaction for learning in mathematics education. *International Journal of Educational Research and Innovation*, 20, 1-14. <https://doi.org/10.46661/ijeri.8196>
- Navas, G., Navas-Reascos, G., Navas-Reascos, G. E., & Proaño-Orellana, J. (2024). Exploring the effectiveness of advanced chatbots in educational settings: A mixed-methods study in statistics. *Applied Sciences*, 14(19), 8984. <https://doi.org/10.3390/app14198984>
- Ngo, T. T. A., An, G. K., Nguyen, P. T., & Tran, T. T. (2024). Unlocking educational potential: exploring students' satisfaction and sustainable engagement with ChatGPT using the ECM Model. *Journal of Information Technology Education: Research*, 23, 21. <https://doi.org/10.28945/5344>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1). <https://doi.org/10.1177/1609406917733847>
- OECD. (n.d.). *OECD AI principles overview*. OECD.AI. Retrieved November 1, 2025, from <https://oecd.ai/en/ai-principles>
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. <https://doi.org/10.1016/j.caeai.2021.100033>
- Pergantis, P., Bamicha, V., Skianis, C., & Drigas, A. (2025). AI chatbots and cognitive control: enhancing executive functions through chatbot interactions: A systematic review. *Brain Sciences*, 15(1), 47. <https://doi.org/10.3390/brainsci15010047>
- Plevris, V., Papazafeiropoulos, G., & Jiménez Rios, A. (2023). Chatbots put to the test in math and logic problems: A Comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google Bard. *AI*, 4(4), 949–969. <https://doi.org/10.3390/ai4040048>
- Raptopoulou, A. (2025). ChatGPT in higher education: Supporting academic literacy through ChatGPT-based activities. *European Journal of Education*, 60(2), e70131. <https://doi.org/10.1111/ejed.70131>

Razak, N. I. A., Yusoff, M. F. M., & Rahmat, R. W. O. K. (2023). ChatGPT review: A sophisticated chatbot models in medical & health-related teaching and learning. *Malaysian Journal of Medicine and Health Sciences*, 19(12), 98–108. <https://doi.org/10.47836/mjmhs.19.s12.12>

Research and Markets. (2024, October 28). *AI Chatbot Analysis Report 2024: Market Projected to Reach \$46.641 Billion by 2029, at a CAGR of 24.53%, Driven by Increasing Demand for Automated Customer Service Solutions and Operational Efficiency*. <https://www.globenewswire.com/news-release/2024/10/28/2969865/28124/en/AI-Chatbot-Analysis-Report-2024-Market-Projected-to-Reach-46-641-Billion-by-2029-at-a-CAGR-of-24-53-Driven-by-Increasing-Demand-for-Automated-Customer-Service-Solutions-and-Operati.html>

Roca, M. D. la, Chan, M. M., Garcia-Cabot, A., Garcia-Lopez, E., & Amado-Salvatierra, H. (2024). The impact of a chatbot working as an assistant in a course for supporting student learning and engagement. *Computer Applications in Engineering Education*, 32(5), e22750. <https://doi.org/10.1002/cae.22750>

Saleh, Z. T., Rababa, M., Elshatarat, R. A., Alharbi, M., Alhumaidi, B. N., Al-Za'areer, M. S., Jarrad, R. A., al Niarat, T. F., Almagharbeh, W. T., Al-Sayaghi, K. M., & Fadila, D. E. S. (2025). Exploring faculty perceptions and concerns regarding artificial intelligence Chatbots in nursing education: potential benefits and limitations. *BMC Nursing*, 24, 440. <https://doi.org/10.1186/s12912-025-03082-0>

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>

Sobaih, A. E. E., & Abuelnasr, A. (2025). Battle of AI chatbots: Graduate students' perceptions of ChatGPT

versus Gemini for learning purposes in Egyptian higher education. *Journal of Applied Learning and Teaching*, 8(1), 128–142. <https://doi.org/10.37074/jalt.2025.8.1.7>

Srinivasan, M., Venugopal, A., Venkatesan, L., & Kumar, R. (2024). Navigating the pedagogical landscape: Exploring the implications of AI and chatbots in nursing education. *JMIR Nursing*, 7(1), e52105. <https://doi.org/10.2196/52105>

Turnbull, D., Chugh, R., & Luck, J. (2023). Systematic-narrative hybrid literature review: A strategy for integrating a concise methodology into a manuscript. *Social Sciences and Humanities Open*, 7(1), 100381. <https://doi.org/10.1016/j.ssaho.2022.100381>

UNESCO. (n.d.). *Artificial intelligence in education*. Retrieved November 1, 2025, from <https://www.unesco.org/en/digital-education/artificial-intelligence>

Wang, T., Lund, B. D., Marengo, A., Pagano, A., Manuru, N. R., Teel, Z. A., & Pange, J. (2023). Exploring the potential impact of Artificial Intelligence (AI) on international students in higher education: generative ai, chatbots, analytics, and international student success. *Applied Sciences*, 13(11), 6716. <https://doi.org/10.3390/app13116716>

Wiboolyasarini, W., Wiboolyasarini, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand*, 14, 100224. <https://doi.org/10.1016/j.amper.2025.100224>

Williams, R. T. (2023). The ethical implications of using generative chatbots in higher education. *Frontiers in Education*, 8, 1331607. <https://doi.org/10.3389/educ.2023.1331607>

Aleksandra Porjazoska Kujundziski, full professor, completed her BSc and MSc studies in 1996 and 1999, respectively, at the University “Ss. Cyril & Methodius”, Faculty of Technology and Metallurgy, Skopje, R. North Macedonia, in the area of chemical engineering- polymer engineering. She finished her PhD studies in technical sciences (new materials – polymers) at the same faculty in 2006. Since 2010, Porjazoska Kujundziski has been working at the International Balkan University in Skopje, in the Department of Industrial Engineering. She has participated in several scientific research projects and national and bilateral projects with Bulgaria and Türkiye. Her research interests include enhancing teaching effectiveness in higher education through the application of new technologies, biodegradable polymers and materials (polymer systems), polymer nanomaterials for medical applications, and mathematical modelling of processes. She is the author of many scientific papers published in journals or presented at conferences.

Neslihan Ademi is a Teaching Assistant in the Department of Computer Engineering at the International Balkan University, Skopje. She holds a PhD in Computer Science and Engineering from St. Cyril and Methodius University, North Macedonia. She received her Master of Science degree in Electronic and Computer Education and her Bachelor of Science degree in Computer Systems Education from Gazi University, Türkiye. Her research interests include artificial intelligence, data science, educational systems, adaptive learning systems, learning analytics, and educational data mining.

Damir Rahmani is a computer engineer and AI researcher who completed both his Bachelor's (2019) and Master's (2021) degrees in Computer Engineering at International Balkan University with perfect merit-based scholarships. His thesis work produced “The Capsule,” an AI-powered HR chatbot, leading to multiple publications including journal articles in e-Mentor and conference papers at IEEE TELFOR on blockchain and Web 3.0 applications. Currently a Teaching Assistant at his alma mater, he teaches programming and mathematics courses while coordinating Erasmus+ projects on AI education. His research spans federated learning, AI in finance, and educational technologies, with regular presentations at European conferences and editorial work for the International Journal of Technical and Natural Sciences.