

Ann. Acad. Med. Siles. (Online) 2026; DOI: 10.18794/aams/240194

Review

Artificial intelligence in neurosurgical education – current applications, opportunities and limitations: A narrative review

Mateusz Stępień¹, Marta Strauchman¹, Anna Siodłak¹, Marcin Setlak², Bartłomiej Błaszczuk²

¹Students' Scientific Club, Department of Neurosurgery, Faculty of Medical Sciences in Katowice,
Medical University of Silesia, Katowice, Poland

²Department of Neurosurgery, Faculty of Medical Sciences in Katowice,
Medical University of Silesia, Katowice, Poland

Address for correspondence:

Marta Strauchman
Studenckie Koło Naukowe
Klinika Neurochirurgii
Uniwersyteckie Centrum Kliniczne im. prof. K. Gibińskiego ŚUM
ul. Medyków 14, 40-752 Katowice
e-mail: poczta.martys@gmail.com

Received: 05.08.2026, Revised: 21.09.2026, Accepted: 23.09.2026, Published: October 2026

This is an open access article made available under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license, which defines the rules for its use. It is allowed to copy, alter, distribute and present the work for any purpose, even commercially, provided that appropriate credit is given to the author and that the user indicates whether the publication has been modified, and when processing or creating based on the work, you must share your work under the same license as the original. The full terms of this license are available at <https://creativecommons.org/licenses/by-sa/4.0/legalcode>.

© Copyright by Author(s)

Publisher: Medical University of Silesia, Katowice, Poland

ABSTRACT

The evolution of artificial intelligence (AI) and large language models (LLMs) is updating medical and neurosurgical education. This narrative review provides an updated synthesis of recent applications of AI in neurosurgical education, with particular emphasis on LLMs, knowledge assessment, simulation-based training, automated competency evaluation, and implementation risks. Through machine learning, deep learning, and virtual reality systems, AI enables personalized, adaptive learning pathways, 24/7 on-demand academic assistance, and objective, data-driven competency evaluations in risk-free environments. In neurosurgery, intelligent surgical simulators and automated 3D spatial planning tools optimize technical execution and track procedural metrics. Generative AI streamlines curriculum design, literature synthesis, and examination drafting. However, the adoption of foundation models introduces critical limitations and risks. These systems remain vulnerable to hallucinations, logic confabulations, and fabricated bibliographic citations, which pose risks in high-stakes clinical contexts. Excessive dependence on AI can cause cognitive fatigue, diminish independent critical thinking, and elevate risks related to academic plagiarism, algorithmic bias, and data privacy. To address these challenges, continuous human verification is required. AI should function as a supportive, transparent assistant of human clinical judgment, ensuring that technological innovation reinforces core values of medical practice and patient safety.

KEYWORDS

artificial intelligence, neurosurgical education, large language models, virtual reality simulation, personalized learning

Introduction

The field of artificial intelligence (AI) experienced a major turning point following the public release of OpenAI's ChatGPT in 2022 [1]. Within the surgical domain, the digitization of operating rooms and the generation of massive volumes of data have caused a major surge in clinical AI applications [2].

Large language models (LLMs) represent a highly advanced branch of natural language processing computer systems designed to simulate, process, and generate human conversation in written or spoken form. Rather than being trained for a single narrow task, modern LLMs are pretrained on billions of words and tokens derived from extensive internet text files, articles, books, and platforms [3,4]. Through a multi-staged training process, LLMs learn how words are used with each other in language and can apply these learned patterns to complete natural language processing tasks.

There is an ongoing discussion in the medical sector that digital competence is a mandatory requirement for lifelong learning in various professional fields [5]. Since future medical care will be increasingly upgraded by technologies like e-health services and telemedicine, it is recommended that health profession curricula directly integrate specific teachings on digital literacy to cultivate a workforce fully prepared to use these digital systems safely and effectively [6].

In neurosurgery, AI is being utilized to enhance preoperative diagnosis, optimize surgical candidate selection, guide intraoperative navigation, and refine postoperative management. Neurosurgical resident education can also benefit from AI integration. When paired with virtual reality (VR) surgical training simulators, AI can autonomously analyze tracking data from a trainee's technical execution – such as speed, pressure, accuracy, precision – and use it to create performance metrics. Through automated feedback, AI can significantly enhance the resident training and accelerate safe technical skill acquisition [2].

Recent literature syntheses have evaluated the integration of AI in neurosurgical education. The review by Harley et al. [7] focused on how AI and VR synergize to assess technical skill acquisition. Their analysis established that automated AI metrics reliably discriminate between expert neurosurgeons and trainees in VR environments. They also highlighted the efficacy of platforms like the Virtual Operative Assistant (VOA) in providing automated feedback, while critically identifying major gaps in the literature, including a severe underutilization of AI for assessing non-technical skills. The scoping review by Piñera-Castro and Borges-García [8] demonstrated that LLMs like GPT-4 surpassed earlier versions and junior trainees on text-based written board questions, but experienced severe accuracy drops when confronted with image-based or visuospatial queries. They also emphasized persistent technical bottlenecks, including black-box algorithmic opacity.

This narrative review differs from these previous studies by offering an updated synthesis of recent AI applications in neurosurgical education between 2024 and 2026. In particular, we address developments in advanced multimodal models, newer large language architectures, Retrieval-Augmented Generation (RAG) systems, and AI chatbots strictly grounded in verified educational materials. Furthermore, we explore modern automated competency assessment frameworks alongside newer AI applications in neurosurgical simulation, such as rapid patient-specific 3D anatomical model generation and real-time intraoperative guidance tools, as well as key implementation risks.

Methods

Relevant studies were identified through searches of electronic databases including PubMed, Scopus, Web of Science and Google Scholar, up to 30 June 2026, with focus on studies published

between 2024 and 2026. They were supplemented by manual screening of reference lists. The literature search was last updated on 3 August 2026.

Search terms included combinations such as: "Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Large Language Models" OR "LLMs" OR "LLM" OR "Generative AI" OR "Virtual Reality" OR "VR" OR "Augmented Reality" OR "Extended Reality" OR "Retrieval-Augmented Generation" OR "3D" OR "Virtual Operative Assistant" AND "Neurosurgery" OR "Neurosurgical" AND "Education" OR "Training" OR "Residency" OR "Resident" OR "Simulation" OR "Competency" OR "Assessment" OR "Skill Acquisition" OR "Performance Metrics".

The inclusion criteria for this review were as follows: papers focusing on the application, integration, evaluation, opportunities, or limitations of artificial intelligence, machine learning, deep learning, large language models, generative AI, virtual reality, augmented reality, 3D anatomical modeling, or intelligent tutoring systems explicitly within the context of neurosurgical education, resident training, knowledge acquisition and technical or non-technical competency assessment. When it comes to publication types, we included: original research studies, systematic reviews, scoping reviews, umbrella reviews, narrative reviews, and preprints from recognized repositories to capture the most recent advancements in fast-evolving AI domains. All included papers had to be written in English.

Exclusion criteria involved studies focusing solely on general clinical diagnostics, patient care outcomes, or pure intraoperative surgical navigation that did not feature a core educational, training, resident assessment, or pedagogical component. Non-academic commentaries, general news items, editorials, materials lacking structured data and articles published in languages other than English were also excluded.

The reference lists of all relevant articles were manually screened to identify any additional studies that matched the inclusion criteria. The screening of titles, abstracts, and full texts, along with the final article selection, was performed by the authors, with any differences resolved through discussion and consensus.

Fundamentals of artificial intelligence

AI is defined as an artificial system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation [9]. While early AI implementations relied heavily on expert systems and handcrafted, top-down "if-then" statements, these rigid structures consistently failed when applied to complex pattern recognition tasks like speech, language comprehension, or image classification [10,11].

Machine learning (ML) focuses specifically on the learning aspect of AI, describing the capacity of systems to automatically discover meaningful relationships, regularities, and hidden insights from

problem-specific training data [11,12]. Depending on the learning paradigm, such as supervised, unsupervised, or reinforcement learning, ML algorithms calibrate their parameters using target labels, intrinsic data structures, or environmental feedback to automatically generate underlying analytical models [10,11].

Deep learning is a concept within ML based on deep artificial neural networks containing an input layer, an output layer, and multiple deeply nested hidden layers. It performs automated feature learning, hierarchically assembling low-level abstract features into highly complex representations directly from raw, high-dimensional input data. The mathematical building blocks of deep learning consist of connected processing units known as artificial neurons, which represent generalized evolutions of early models such as the perceptron, transforming inputs through non-linear activation functions [11,12]. The massive complexity of these multi-layer networks introduces a prominent "black-box" problem, making their decision-making processes opaque and non-auditable unless explicitly augmented with explainable AI techniques [11].

LLMs represent the latest technical evolution of computational language modeling [13]. LLMs scale up parameter counts to tens or hundreds of billions and are pre-trained on massive multi-source text corpora via distributed parallel architectures across extensive GPU or TPU hardware infrastructures [14]. Architecturally, LLMs optimize training stability and capacity through specific design configurations. Scaling language models beyond a critical parameter threshold unlocks highly advanced, qualitative "emergent abilities" that are entirely absent in smaller networks. These primary emergent abilities include In-Context Learning, Instruction Following, and Step-by-Step Reasoning. To align with human intentions, models undergo post-training alignment utilizing text-level optimization. Retrieval-Augmented Generation enhances factual reliability by combining an LLM with external, curated information retrieved at the time of query. This allows responses to be grounded in verified knowledge sources [13,15]. We have created a scheme that illustrates the relationship between modern AI technologies (Figure 1).

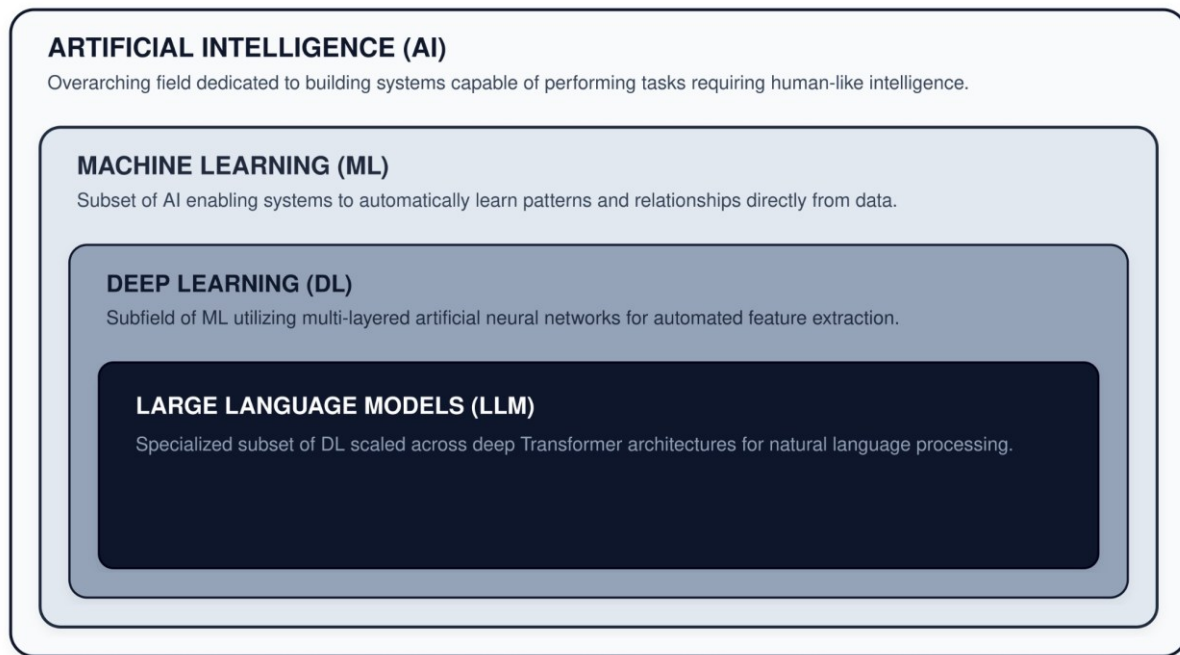


Fig. 1. Conceptual relationship between artificial intelligence, machine learning, deep learning, and large language models

Artificial intelligence in medical and surgical education: General context

Medical syllabus planning and course content generation can be significantly accelerated with the use of AI [16]. A study by Zhang et al. [17] demonstrated the feasibility of using ChatGPT 4.0 to generate ophthalmology clinical case scenarios that are highly relevant for undergraduate medical training. However, the same study revealed that raw AI-generated materials are highly susceptible to hallucinations and diagnostic logic errors, which were observed in 16.67% of the generated cases. Furthermore, a literature review by Kiyak and Emekli [18] indicated that specific prompt writing tactics allow educators to precisely modulate the tone, formatting, and structural parameters of the generated output. When properly validated by experts, these generative technologies can effectively relieve medical faculty from administrative work [16].

Creating high quality multiple-choice questions is a very time consuming task for medical lecturers [18]. A prospective cohort study conducted by Law et al. [19] demonstrated that ChatGPT-4o offers massive time efficiency, reducing the overall time required for question generation.

However, AI-generated questions were significantly easier. A systematic review by Artsi et al. [20] evaluated multiple studies and confirmed that while LLMs are capable of producing competent questions with acceptable discrimination indexes, they consistently present structural flaws.

The synthesis of vast literature by AI helps researchers manage information overload and support clinical decision-making [16,21]. A study conducted by Tang et al. [22] demonstrated that conventional automatic evaluation metrics fail to correlate with actual human-evaluated text quality. The evaluation revealed that zero-shot language models are highly extractive and heavily susceptible to generating factually inconsistent summaries and omitting critical PICO elements. In

contrast, a clinical reader study published by Van Veen et al. [21] demonstrated that when appropriate task-specific adaptation methods are applied, an adapted GPT-4 model can outperform medical experts.

A multi-cohort deployment study published by Thesen and Park [23] investigated an online teaching assistant called NeuroBot TA, which was strictly limited to instructor-curated course materials to eliminate internet hallucinations. The empirical usage data revealed a highly strategic, exam-focused adoption pattern among medical students, characterized by a massive 329.6% increase in conversation volume during the three days leading up to major exams and significant after-hours utilization post-5 PM for basic neuroanatomy and physiology clarifications. However, collected feedback also showed that the students were frustrated by AI's refusal to answer broader medical questions, leading to an overall system adoption rate of 31.4%. We have provided a summary of AI application and their potential benefits in the medical education (Table I).

Table I. Applications of artificial intelligence (AI) in medical education

Application	Potential benefits
Syllabus and Clinical Case Scenario Generation	• Accelerates course design and syllabus planning. • Rapidly generates realistic, highly relevant clinical case scenarios for medical training.
Multiple-Choice Question Drafting	• Allows time efficiency in exam preparation. • Enables automated scoring and rapid creation of simulated practice question repositories.
Literature Synthesis and Summarization	• Compresses vast literature to manage information overload. • Supports rapid evidence synthesis and clinical decision-making.
Virtual Teaching Assistants and AI Facilitators	• Provides 24/7 on-demand academic support. • Restricting models to instructor-curated course materials eliminates internet hallucinations

Artificial intelligence in neurosurgical education

Knowledge acquisition

LLMs can support personalized learning and simplify the acquisition of complex anatomical and surgical knowledge for neurosurgical trainees [24]. As highlighted in reviews by Moglia et al. [25] and Xu et al. [26], these tools assist residents in information retrieval, literature summarization, and building a structured knowledge base. During case-based preparations, trainees can utilize LLMs to compile surgical approach preferences, facilitate learning, and strengthen the overall coordination of clinical processes [24]. Furthermore, LLMs allow trainees to generate customized practice materials, flashcards, and summary guides from existing syllabus content to support self-directed study and continuous self-assessment [25,27]. However, literature consistently emphasizes that trainees must maintain critical oversight. Because general-purpose models can generate false information, duplicate responses, or provide contextually inappropriate claims without verifiable references, residents must rigorously evaluate AI-generated content against authoritative neurosurgical textbooks and clinical guidelines [24,26].

Separately from their use as personal study aids, LLMs are increasingly evaluated as test-takers on specialized neurosurgical written board examinations to benchmark their baseline clinical knowledge and reasoning capabilities [28]. Evaluative studies demonstrate both strong text-based capabilities and significant structural limitations. For instance, Çamlar et al. [24] tested six distinct LLMs on 599 board-style questions, showing that advanced models including ChatGPT-o1pro, ChatGPT-4, and the neurosurgery-specific AtlasGPT achieve near-expert accuracy on Single Best Answer-Recall of Knowledge, Single Best Answer-Interpretative Application of Knowledge, and True/False formats, yet suffer severe performance drops or total processing failures when faced with image-based anatomical queries. Similarly, a comprehensive evaluation by Vishwanath et al. [28] testing 28 LLMs on 2,904 questions from the Congress of Neurological Surgeons Self-Assessment in Neurosurgical Surgery (CNS-SANS) revealed that while six models achieved passing scores, their competence was highly fragile when exposed to a distraction framework containing clinically irrelevant but familiar terminology. When irrelevant distractor statements were embedded into the prompt, the accuracy across various model architectures was significantly reduced by as much as 20.4%. While foundation models like GPT-4 and Med-PaLM 2 have successfully passed broader licensing exams such as the USMLE, current evidence establishes that their reliability on specialized, image-dense neurosurgical board benchmarks remains constrained [25,27].

High performance by an LLM on standardized examinations does not inherently demonstrate its

educational efficacy, as benchmark scores reflect static knowledge retrieval rather than a tool's capacity to foster clinical reasoning or support effective pedagogical outcomes.

Surgical training

Surgical simulations powered by VR and AI deliver interactive, safe environments for mastering technically demanding interventions. While VR itself is not a form of artificial intelligence, AI serves as an analytical or generative layer integrated into these immersive environments. AI-enhanced neurosurgical simulation refers to the integration of ML algorithms, computer vision, or generative models with immersive hardware to automatically evaluate performance metrics, reconstruct patient-specific anatomy, or deliver real-time, objective feedback without reliance on a human preceptor. A review by Isikay et al. [29] emphasized that patient-specific 3D visualization and reality technologies fundamentally enhance spatial orientation, approach selection, and intraoperative navigation in complex anatomical regions like the skull base. A systematic review by Khalifa et al. [30] demonstrated that AI performance assessments, incorporating motion tracking, kinematics, force sensors, and computer vision, provide objective and reproducible skill evaluations. Furthermore, a study by Li et al. [31] utilized ML algorithms such as Support Vector Machines combined with feature selection to classify surgical skill levels with high accuracy. They also employed Shapley values to interpret individual psychomotor metrics for targeted training. Within neurosurgery, AI-enhanced intelligent tutoring systems have become valuable teaching and learning tools. A narrative review by Halman et al. [32] highlighted the VOA, developed by Mirchi et al. [33], as an explainable AI tool that achieved 92% accuracy in distinguishing experts from novices during simulated brain tumor resections. The system condenses complex simulator data into four core performance metrics: two safety indicators and two movement indicators. VOA implements a unique, two-step educational paradigm rooted in cognitive load theory, requiring trainees to master safety metrics before advancing to movement mechanics. The platform combines objective benchmarks with multi-modal feedback, incorporating textual guidance, auditory instructions mimicking a clinical mentor, and video demonstrations of expert execution. In a systematic review and meta-analysis, Hanna et al. [34] evaluated the efficacy of AI tutoring versus expert instruction, finding that AI-augmented tutoring systems had comparable or slightly superior technical skill acquisition (measured by OSATS scores) relative to human expert instruction. Similarly, in a randomized controlled trial by Yilmaz et al. [35], trainees using the real-time multifaceted AI feedback of the Intelligent Continuous Expertise Monitoring System achieved significantly better performance outcomes during practice repetitions compared to active control groups, while demonstrating OSATS ratings comparable to in-person expert instruction. Despite these performance gains, the integration of automated systems poses distinct cognitive

and behavioral challenges. A mixed-method review by Natali et al. [36] cautioned against AI-induced deskilling and upskilling inhibition. They describe the risk of expert atrophy resulting from habitual reliance on automated systems. Moreover, specific behavioral consequences have been documented. A randomized trial revealed that while AI feedback improved bimanual coordination, it also induced unintended behavioral adaptations, such as significantly slower resection rates and reduced dominant hand speed [32].

To enhance procedural realism, patient-specific 3D anatomical models provide digital representations of complex pathology, allowing trainees to visualize spatial relationships and evaluate approaches in a controlled environment. For spinal care, Ranabhat et al. [37] developed a fast automated pipeline combining VertDetect and TotalSegmentator that constructs high-fidelity patient-specific 3D models for spinal decompression from raw CT and MRI datasets in approximately 2.5 minutes, achieving excellent volumetric accuracy with a Dice Similarity Coefficient of 0.95 for vertebral bone.

Preoperative procedure planning using 3D spatial models enables surgical teams to evaluate approaches and anticipate intraoperative challenges. A retrospective study by Roh et al. [38] demonstrated that reviewing patient-specific anatomy inside a 3D VR environment significantly enhanced tumor-related structure detection and optimized surgical approaches for anterior skull base meningiomas compared to conventional 2D reviews. Furthermore, Saher et al. [39] noted that current planning procedures face critical limitations, including visual fatigue, motion sickness, and a lack of tactile haptic feedback in standard virtual interfaces.

Scholarly training and academic skills

The efficiency of generative chatbots in clinical literature searching faces structural limitations. This is demonstrated by an empirical evaluation from Gwon et al. [40], where ChatGPT-3.5 achieved a mere 0.08% relevance rate when tasked with a systematic search on Peyronie's disease, frequently inventing highly plausible but entirely non-existent "hallucinated" studies. Even search engines backed by web-browsing capabilities, such as Microsoft Bing AI, demonstrated a low absolute volume of relevant papers and regularly committed classification errors, such as misidentifying standard reviews as randomized controlled trials. To evaluate search efficiency a multitask assessment by Sollini et al. [41] compared six state-of-the-art LLMs against human researchers in conducting a literature review for a radioligand therapy. While certain high performing models processed screening tasks rapidly in under 30 minutes, human researchers proved distinctly superior in constructing a complete, accurate, and fully reliable systematic review.

Citation and reference organization remains a critical weakness for general-purpose LLMs. Early

data documented that ChatGPT achieved only a 10% accuracy rate for head and neck surgery references, though GPT-4 subsequently improved citation accuracy to 74.3% for otolaryngology topics. It is recommended to bypass standard conversational chatbots in favor of specialized literature tools like Perplexity, Elicit, or Consensus, which actively couple LLMs with verified scholarly databases to suppress reference fabrication [42]. While these specialized AI platforms effectively support initial literature discovery and reference organization, definitive literature searches must still be conducted in verified scholarly databases.

LLMs are highly effective at assisting with scientific writing and manuscript preparation. In a blinded, randomized controlled study, GPT-4 generated introduction sections for clinical trials that demonstrated significantly higher readability scores compared to human written text. LLMs are also heavily utilized to draft case reports, generate clinical summaries, and improve technical writing fluency for non-native English scholars [42]. Lieberum et al. [43] emphasized that because these networks generate text based on probabilistic word associations rather than factual comprehension, their outputs carry an ongoing risk of data contamination and misinformation. LLM applications must remain strictly confined to supportive tasks under direct human supervision, requiring thorough fact-checking and rigorous editing by the authors before submission.

Opportunities and advantages

Conversational AI tools, intelligent tutoring systems, and virtual patients function as adaptive teaching assistants that provide continuous academic support [44]. These systems create low-stakes learning environments where medical trainees can practice clinical reasoning, history taking, and diagnostic decision-making without fear of negative evaluation [45]. In medical education, interactive dialogue and real-time feedback enable learners to clarify complex pathophysiological concepts and rectify mistakes immediately during self-directed study [46,47]. Personalized learning is achieved as ML algorithms continuously analyze individual performance data to adapt educational pathways. These intelligent platforms track student progress, identify specific knowledge gaps, and dynamically adjust content difficulty and instructional pacing based on objective competency metrics [48]. Additionally, personalized AI systems can be tailored to address specific learning disabilities such as dyslexia or ADHD by providing alternative explanations and multimodal sensory formats, which may improve overall accessibility if properly designed [49].

The integration of conversational AI tools with mobile technologies accelerates medical knowledge retrieval by allowing trainees to access clinical guidelines and literature directly on personal devices [44]. This multi-device compatibility supports flexible, self-paced learning and facilitates

rapid information verification during fast-paced clinical workflows [47]. Retrieving key terms or clinical evidence through AI interfaces proves significantly faster than manual searches in traditional, paper-based textbooks [46,50].

AI-driven language assistance tools offer vital solutions to overcome linguistic barriers. Automated translation, speech recognition, and voice-capable language models allow bidirectional communication in multilingual environments [51]. This allows non-native English speakers to follow complex instructional materials, improve conversational fluency, and engage effectively in healthcare training [47,51].

Limitations and risks

Hallucinations in many models result primarily from autoregressive training objectives that prioritize token-likelihood optimization over validated accuracy, creating a high level of unwarranted certainty [52]. These errors can be structurally categorized as confabulations, where outputs are completely arbitrary and highly sensitive to irrelevant details like random seeds [53]. In educational and clinical settings, these inaccuracies present themselves as either internal hallucinations that directly contradict the provided context, or external hallucinations that introduce entirely ungrounded, fabricated claims [54]. LLMs operate as statistical mimics that remain fundamentally indifferent to empirical reality and objective truth [55]. Furthermore, the non-transparent nature of these algorithms makes it exceptionally difficult for users to trace, verify, or identify the logical reasoning behind a generated clinical or academic decision [56]. The rapid evolution of medical and scientific knowledge presents an ongoing challenge for static neural networks, leading to a high risk of model obsolescence [52]. When systems are trained on fixed or historical datasets, the presence of outdated data causes them to confidently cite disproven guidelines or distorted instructional concepts [54]. This temporal limitation can be partially mitigated by integrating RAG, external search capabilities, continuously updated databases, and specialized literature tools [56].

Excessive AI dependence impairs cognitive capabilities, directly leading to a measurable decline in students' critical thinking. This dependency is partially mediated by psychological mechanisms of cognitive fatigue, which arise from information overload and high metacognitive monitoring costs. While high information literacy can buffer the negative impact on critical thinking, it paradoxically amplifies mental exhaustion and subjective fatigue due to the heavy demands of continuous cross-referencing and verification [57]. Over-reliance encourages students to adopt mental shortcuts, promoting cognitive disengagement and an uncritical acceptance of flawed information [58]. Conversational interfaces also encourage quick, shallow interactions over deep reflection, potentially undermining long-term analytical skills [59].

The integration of AI into higher education creates a threat to academic integrity by facilitating data fabrication, and plagiarism [58]. Experimental evidence shows that both human educators and automated AI detection tools struggle to reliably distinguish between human-authored and AI-generated text. Current detection algorithms remain fundamentally unreliable, producing high rates of both false-positive and false-negative errors. False positives frequently misclassify legitimate student writing, while false negatives fail to identify AI edited outputs. Consequently, institutions cannot rely on automated detectors or authentic assessment design alone to safeguard academic standards [60].

Data intensive foundation models require massive datasets that expose protected health and student records to deep confidentiality and security threats. During their life cycle, these applications are highly susceptible to adversarial privacy attacks, including re-identification, membership inference, attribute disclosure, and model inversion [61]. The constant harvesting of personal inputs (such as names, prompts, and uploaded images) without explicit user consent effectively transforms educational aids into intrusive surveillance mechanisms [58].

The adoption of advanced AI frameworks and immersive technologies introduces severe equity concerns due to unequal access to subscription-based foundation models, VR infrastructure, and high quality surgical simulators. While well-funded academic centers can integrate AI tools and virtual simulation platforms, low-resource institutions face substantial financial and technological barriers. This widening digital divide risks intensifying existing educational disparities in medical and surgical training [56]. We prepared a table to summarize advantages and limitations of AI integration (Table II).

Table II. Advantages vs. limitations of artificial intelligence (AI)

Domain	Advantages and strengths	Limitations and risks
Information Retrieval and Accessibility	• Rapid, 24/7 on-demand retrieval of clinical definitions and literature. • Compresses vast biomedical evidence to support decision-making.	• Susceptible to hallucinations, confabulations, and logic errors. • Fixed training data leads to static memory and technological obsolescence.
	• Delivers adaptive, personalized learning pathways tailored to individual	• Over-reliance impairs higher-order critical thinking and causes

Pedagogy and Personalized Learning	anatomical and clinical knowledge gaps. • Creates a supportive, low-stakes environment encouraging help-seeking behaviors. • Automatically adjusts content difficulty, instructional pacing, and multimodal presentation formats.	cognitive fatigue. • Promotes cognitive laziness and reactive, surface-level responses.
Assessment and Faculty Support	• Reduces exam question drafting time. • Relieves faculty of administrative course planning and syllabus design. • Powers objective, automated skill benchmarking in VR simulations.	• AI-generated questions mostly test surface-level knowledge and often contain structural errors. • Facilitates academic shortcuts, data fabrication, and plagiarism threats.
Operational, Linguistic and Technical Infrastructure	• Breaks down linguistic barriers via instant neural translation and real-time speech-to-text captioning. • Powers realistic VR/AR simulation feedback loops and patient-specific 3D modeling.	• "Black box" neural architecture makes decision logic non-auditable. • Vulnerable to data privacy attacks, unauthorized harvesting, and security threats.

Ethical considerations

Generative models directly challenge traditional concepts of authorship and academic integrity [59]. Since detection software remains unreliable, maintaining academic integrity depends on clear, mandatory disclosure rather than technical enforcement [60]. Uncritical acceptance of automated literature synthesis risks embedding fabricated findings into the scientific record [58]. It is a danger with direct clinical ramifications, as evidenced by 91.8% of surveyed clinicians who encountered medical hallucinations and 84.7% who warned of potential patient harm [52]. When an algorithm produces an incorrect prediction or misleading guidance that results in direct harm, the boundaries of liability become blurred among developers, healthcare institutions, and

end-users [56]. While commercial deployment practices often attempt to bypass this by offloading the burden of verification onto downstream users. However, it must be emphasized that the ultimate responsibility for assessment, evaluation, and educational decision-making rests with educators and academic institutions, rather than leaving end-users to navigate algorithmic risks without structured support [55].

Future perspectives

Future studies on personalized education need to move past initial testing and gather solid proof that these systems actually work. Early pilot implementations such as LLM-based automated grading tested on small cohorts of 96 students successfully reduced evaluation times from hours to approximately 24 minutes [62]. However, these findings are limited by single-center designs. Khakpaki [63] highlights the need for solid methodological evaluation, while Izquierdo-Condoy et al. [62] emphasize that future investigations must conduct large-scale, longitudinal studies to assess whether adaptive learning pathways improve clinical competence. Sriram et al. [64] and Tsang and Dai [65] recommend multi-center randomized controlled trials to validate AI-personalized curricula across diverse student cohorts. While pilot trials show a reduction in procedural acquisition times, larger frameworks are needed to ensure equitable access and knowledge retention.

Integrating VR and AI demands a clear research strategy designed to overcome the boundaries of preliminary studies. Early mixed-reality evaluations report high user engagement and short-term skill gains. Chance [66] proposes a framework requiring that future research must transition to longitudinal studies to measure actual clinical error reduction rather than short-term simulation metrics. Mergen et al. [67] point out that current literature relies on a fragmented mix of inconsistent measurements, and they propose building standardized, validated evaluation tools like tailored Kirkpatrick models with the help of multidisciplinary experts. Additionally, Saher et al. [39] call for future studies to refine AI-driven VR models beyond preliminary testing, expand global accessibility, evaluate non-invasive brain-machine interface integrations, and test remote telesurgery training protocols.

To move intelligent surgical simulators from early laboratory concepts into real hospital use, they must undergo strict and thorough validation. While early prototypes show strong technical viability, Raghavan et al. [68] propose extensive multicenter trials with extended follow-up periods to evaluate their true generalizability. Sankar et al. [69] emphasize that current neurosurgical datasets suffer from severe imbalance. They argue that future research must explore federated learning architectures to diversify training data and rigorously test Explainable AI models in high stakes environments. Leal [70] highlights the need for multi-center validations of computer vision

and deep learning systems, alongside comprehensive economic and cost-effectiveness analyses of simulator implementation.

Future investigations into automated course creation and content generation must address the methodological vulnerabilities identified in early trials. Preliminary studies on generative AI highlight high rates of reference fabrication and content inaccuracies. Izquierdo-Condoy et al. [62] recommend the adoption of structured reporting frameworks such as the FACETS model to ensure transparency and cross-study comparability. Khakpaki [63] proposes a dedicated curriculum research centered on formal AI literacy, to prevent uncritical automation. It will allow educators to maintain active control when using AI to draft multiple-choice questions and syllabi. We present a concept of integration of these AI pedagogical tools (Figure 2).

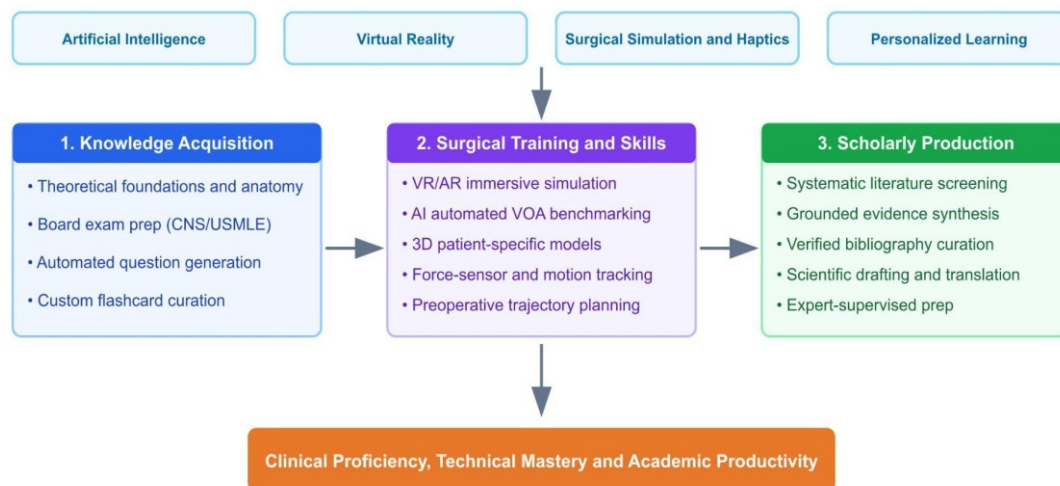


Fig. 2. Integrated ecosystem of AI-driven neurosurgical education

Limitations

As a narrative review, the literature identification and selection process was not rigorous and may be subject to selection bias. Due to the rapid evolution of AI technologies and LLMs, certain tools, architectures, and performance benchmarks discussed are subject to rapid technological out datedness. The scope of this review was restricted to English-language publications, potentially missing relevant local or non-English research.

Conclusions

AI has demonstrated potential to support selected components of neurosurgical education, particularly knowledge assessment, simulation-based feedback, and automated analysis of technical performance. The integration of AI with VR and intelligent surgical simulators provides objective competency evaluations that significantly enhance precision in a risk-free environment. However, because foundation models rely on probabilistic text predictions, they remain

vulnerable to hallucinations, logic errors, and fabricated citations. These flaws demand constant caution. That is why human verification is still required. Addressing risks like student over-reliance, algorithmic bias, and privacy threats calls for clear ethical guidelines and targeted AI literacy. Technology should enhance student learning and training, not replace human mentorship. It is essential to carefully balance between technological innovation and traditional education.

Use of AI tools statement

Google Gemini 3.5 Flash-Lite with Extended Thinking was used as an assistant tool for language editing and stylistic improvement of the manuscript text. It was also used in generating the SVG vector graphics code and layout formatting based on authors' conceptual design and content specifications for the figures. The literature selection, interpretation and final approval of the manuscript were performed by the authors.

Funding

This research received no external funding.

Conflict of interest

The authors declare no conflict of interest.

Authors' contribution

Study design – M. Setlak, B. Błaszczuk, M. Stępień, M. Strauchman, A. Siodłak

Data collection – M. Stępień, M. Strauchman, A. Siodłak, M. Setlak

Manuscript preparation – M. Stępień, M. Strauchman, A. Siodłak, M. Setlak

Literature research – M. Stępień, M. Strauchman, A. Siodłak, M. Setlak

Final approval of the version to be published – M. Setlak, B. Błaszczuk, M. Stępień, M. Strauchman, A. Siodłak

REFERENCES

1. Mesko B. The ChatGPT (Generative Artificial Intelligence) Revolution Has Made Artificial Intelligence Approachable for Medical Professionals. *J Med Internet Res.* 2023;25:e48392. doi: 10.2196/48392.
2. Windermere SA, Shah S, Hey G, McGrath K, Rahman M. Applications of Artificial Intelligence in Neurosurgery for Improving Outcomes Through Diagnostics, Predictive Tools, and Resident Education. *World Neurosurgery.* 2025;197:123809. doi: 10.1016/j.wneu.2025.123809.
3. Egli A. ChatGPT, GPT-4, and Other Large Language Models: The Next Revolution for Clinical Microbiology? *Clin Infect Dis.* 2023;77(9):1322–1328. doi: 10.1093/cid/ciad407.

4. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930–1940. doi: 10.1038/s41591-023-02448-8.
5. Mainz A, Nitsche J, Weirauch V, Meister S. Measuring the Digital Competence of Health Professionals: Scoping Review. *JMIR Med Educ*. 2024;10:e55737. doi: 10.2196/55737.
6. Campanozzi LL, Gibelli F, Bailo P, Nittari G, Sirignano A, Ricci G. The role of digital literacy in achieving health equity in the third millennium society: A literature review. *Front Public Health*. 2023;11:1109323. doi: 10.3389/fpubh.2023.1109323.
7. Harley JM, Tawakol T, Azher S, Quaiattini A, Del Maestro R. The role of artificial intelligence, performance metrics, and virtual reality in neurosurgical education: an umbrella review. *Glob Surg Educ J Assoc Surg Educ*. 2024;3(1):83. doi: 10.1007/s44186-024-00284-z.
8. Piñera-Castro HJ, Borges-García C. Applications of artificial intelligence in neurosurgical education: a scoping review. *Egypt J Neurol Psychiatry Neurosurg* 2025;61:83. doi: 10.1186/s41983-025-01015-x.
9. Haenlein M, Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *Calif Manag Rev*. 2019;61(4):5–14. doi: 10.1177/0008125619864925.
10. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol*. 2020;9(2):14. doi: 10.1167/tvst.9.2.14.
11. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electron Mark*. 2021;31(3):685–695. doi: 10.1007/s12525-021-00475-2.
12. Taye MM. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*. 2023;12(5):91. doi: 10.3390/computers12050091.
13. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A Survey of Large Language Models. *arXiv*. Preprint posted online March 18, 2026. doi: 10.48550/arXiv.2303.18223.
14. Kumar P. Large language models (LLMs): survey, technical frameworks, and future challenges. *Artif Intell Rev*. 2024;57(10):260. doi: 10.1007/s10462-024-10888-y.
15. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A Comprehensive Overview of Large Language Models. *arXiv*. Preprint posted online October 17, 2024. doi: 10.48550/arXiv.2307.06435.

16. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Med Educ.* 2023;9:e48291. doi: 10.2196/48291.
17. Zhang Q, Huang Z, Huang Y, Wang G, Zhang R, Yang J, et al. Generative AI in medical education: feasibility and educational value of LLM-generated clinical cases with MCQs. *BMC Med Educ.* 2025;25(1):1502. doi: 10.1186/s12909-025-08085-8.
18. Kiyak YS, Emekli E. ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: a literature review. *Postgrad Med J.* 2024;100(1189):858–865. doi: 10.1093/postmj/qgae065.
19. Law AK, So J, Lui CT, Choi YF, Cheung KH, Kei-Ching Hung K, et al. AI versus human-generated multiple-choice questions for medical education: a cohort study in a high-stakes examination. *BMC Med Educ.* 2025;25(1):208. doi: 10.1186/s12909-025-06796-6.
20. Artsi Y, Sorin V, Konen E, Glicksberg BS, Nadkarni G, Klang E. Large language models for generating medical examinations: systematic review. *BMC Med Educ.* 2024;24(1):354. doi: 10.1186/s12909-024-05239-y.
21. Van Veen D, Van Uden C, Blankemeier L, Delbrouck JB, Aali A, Bluethgen C, et al. Adapted large language models can outperform medical experts in clinical text summarization. *Nat Med.* 2024;30(4):1134–1142. doi: 10.1038/s41591-024-02855-5.
22. Tang L, Sun Z, Idnay B, Nestor JG, Soroush A, Elias PA, et al. Evaluating Large Language Models on Medical Evidence Summarization. *medRxiv [Preprint].* 2023;2023.04.22.23288967. doi: 10.1101/2023.04.22.23288967.
23. Thesen T, Park SH. A generative AI teaching assistant for personalized learning in medical education. *NPJ Digit Med.* 2025;8(1):627. doi: 10.1038/s41746-025-02022-1.
24. Çamlar M, Sevgi UT, Erol G, Karakaş F, Doğruel Y, Güngör A. Comparative performance of neurosurgery-specific, peer-reviewed versus general AI chatbots in bilingual board examinations: evaluating accuracy, consistency, and error minimization strategies. *Acta Neurochir.* 2025;167(1):241. doi: 10.1007/s00701-025-06628-y.
25. Moglia A, Georgiou K, Cerveri P, Mainardi L, Satava RM, Cuschieri A. Large language models in healthcare: from a systematic review on medical examinations to a comparative analysis on fundamentals of robotic surgery online test. *Artif Intell Rev.* 2024;57(9):231. doi: 10.1007/s10462-024-10849-5.

26. Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. *J Educ Eval Health Prof.* 2024;21:6. doi: 10.3352/jeehp.2024.21.6.
27. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: A scoping review. *iScience.* 2024;27(5):109713. doi: 10.1016/j.isci.2024.109713.
28. Vishwanath K, Alyakin A, Ghosh M, Lee JV, Alber DA, Sangwon KL, et al. Evaluating the performance and fragility of large language models on the self-assessment for neurological surgeons. *arXiv.* Preprint posted online 29 May 2025. doi: 10.48550/arXiv.2505.23477.
29. Isikay I, Cekic E, Baylarov B, Tunc O, Hanalioglu S. Narrative review of patient-specific 3D visualization and reality technologies in skull base neurosurgery: enhancements in surgical training, planning, and navigation. *Front Surg.* 2024;11:1427844. doi: 10.3389/fsurg.2024.1427844.
30. Khalifa A, Tahhan O, Albazooni M, Saeed M, Hamdi R, Stanners M, et al. Automated and Artificial Intelligence (AI)-Derived Performance Assessment in Surgical Simulation: A Systematic Review. *Cureus.* 2025;17(12):e100477. doi: 10.7759/cureus.100477.
31. Li F, Qin Z, Qian K, Liang S, Li C, Tai Y. Personalized assessment and training of neurosurgical skills in virtual reality: An interpretable machine learning approach. *Virtual Real Intell Hardw.* 2024;6(1):17–29. doi: 10.1016/j.vrih.2023.08.001.
32. Halman J, Tencer S, Siemiński M. Artificial Intelligence and Extended Reality in the Training of Vascular Surgeons: A Narrative Review. *Med Sci.* 2025;13(3):126. doi: 10.3390/medsci13030126.
33. Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS One.* 2020;15(2):e0229596. doi: 10.1371/journal.pone.0229596.
34. Hanna F, Gad MM, Helmy MBF, Eisa MM, Omran MWZ, Youssef M, et al. Artificial intelligence augmented tutoring vs expert instruction on learning simulated general surgical skills: a systematic review and meta-analysis. *BMC Med Educ.* 2026;26(1):952. doi: 10.1186/s12909-026-09606-9.
35. Yilmaz R, Bakhaidar M, Alsayegh A, Hamdan NA, Fazlollahi AM, Tee T, et al. Real-Time multifaceted artificial intelligence vs In-Person instruction in teaching surgical technical skills: a randomized controlled trial. *Sci Rep.* 2024;14(1):15130. doi: 10.1038/s41598-024-65716-8.

36. Natali C, Marconi L, Dias Duran LD, Cabitza F. AI-induced Deskilling in Medicine: A Mixed-Method Review and Research Agenda for Healthcare and Beyond. *Artif Intell Rev.* 2025;58(11):356. doi: 10.1007/s10462-025-11352-1.
37. Ranabhat RK, Ross TD, Jiao T, Larouche J, Finkelstein J, Hardisty M. Virtual-reality based patient-specific simulation of spine surgical procedures: A fast, highly automated and high-fidelity system for surgical education and planning. *arXiv. Preprint posted online April 29, 2026.* doi: 10.48550/arXiv.2604.26781.
38. Roh TH, Choi S, Moon JH, Kim EH, Kang SG, Chang JH. Extended Reality in Neurosurgery : Surgical Planning, Navigation, Education, and Patient Communication. *J Korean Neurosurg Soc.* 2026;69(3):349–360. doi: 10.3340/jkns.2026.0047.
39. Saher L, Asif A, Mehmood Z, Fatima SS, Nizam R. Virtual Reality in Neurosurgery: Advancing Training, Education, and Surgical Planning. *Asian J Neurosurg.* 2025;21(1):26–37. doi: 10.1055/s-0045-1813259.
40. Gwon YN, Kim JH, Chung HS, Jung EJ, Chun J, Lee S, et al. The Use of Generative AI for Scientific Literature Searches for Systematic Reviews: ChatGPT and Microsoft Bing AI Performance Evaluation. *JMIR Med Inform.* 2024;12:e51187. doi: 10.2196/51187.
41. Sollini M, Pini C, Lazar A, Gelardi F, Ninatti G, Bauckneht M, et al. Human researchers are superior to large language models in writing a medical systematic review in a comparative multitask assessment. *Sci Rep.* 2025;16(1):173. doi: 10.1038/s41598-025-28993-5.
42. Ahn S. The transformative impact of large language models on medical writing and publishing: current applications, challenges and future directions. *Korean J Physiol Pharmacol.* 2024;28(5):393–401. doi: 10.4196/kjpp.2024.28.5.393.
43. Lieberum JL, Toews M, Metzendorf MI, Heilmeyer F, Siemens W, Haverkamp C, et al. Large language models for conducting systematic reviews: on the rise, but not yet ready for use-a scoping review. *J Clin Epidemiol* 2025;181:111746 doi: 10.1016/j.jclinepi.2025.111746.
44. Akintola AS, Akintayo M, Kadri T, Oforgu CM, Michael M, Nwanna M. Adaptive AI Systems in Education: Real-Time Personalised Learning Pathways for Skill Development. *Artif Intell Mach Learn Data Sci* 2025;3(1):2489–2494. doi: 10.51219/JAIMLD/Akinyemi-Sadeeq-Akintola/534.
45. Vrdoljak J, Boban Z, Vilović M, Kumrić M, Božić J. A Review of Large Language Models in Medical Education, Clinical Decision Support, and Healthcare Administration. *Healthcare.* 2025;13(6):603. doi: 10.3390/healthcare13060603.

46. Yarlagadda KC. AI in Education: Personalized Learning and Intelligent Tutoring Systems. *EJCSIT*. 2025;13(32):15–27. doi: 10.37745/ejcsit.2013/vol13n321527.
47. Michalczak M, Zgoda W, Michalczak J, Żądło A, Nasser A, Tokarek T. Artificial Intelligence in Medical Education: A Narrative Review. *AI*. 2025;6(12):322. doi: 10.3390/ai6120322.
48. Hariyanto, Kristianingsih FXD, Maharani R. Artificial intelligence in adaptive education: a systematic review of techniques for personalized learning. *Discov Educ*. 2025;4(1):458. doi: 10.1007/s44217-025-00908-6.
49. Strielkowski W, Grebennikova V, Lisovskiy A, Rakhimova G, Vasileva T. AI -driven adaptive learning for sustainable educational transformation. *Sustainable Development*. 2025;33(2):1921–1947. doi: 10.1002/sd.3221.
50. Sami A, Tanveer F, Sajwani K, Kiran N, Javed MA, Ozsahin DU, et al. Medical students' attitudes toward AI in education: perception, effectiveness, and its credibility. *BMC Med Educ*. 2025;25(1):82. doi: 10.1186/s12909-025-06704-y.
51. Fitas R. Inclusive education with AI: supporting special needs and tackling language barriers. *AI Ethics*. 2025;5(6):5729–5757. doi: 10.1007/s43681-025-00824-3.
52. Kim Y, Jeong H, Chen S, Li SS, Park C, Lu M, et al. Medical Hallucinations in Foundation Models and Their Impact on Healthcare. *arXiv*. Preprint posted online November 2, 2025. doi: 10.48550/arXiv.2503.05777.
53. Farquhar S, Kossen J, Kuhn L, Gal Y. Detecting hallucinations in large language models using semantic entropy. *Nature*. 2024;630(8017):625–630. doi: 10.1038/s41586-024-07421-0.
54. Danyaro KU, Abdullahi S, Abdallah AS, Chiroma H. Hallucinations in Large Language Models for Education: Challenges and Mitigation. *IJTL*. 2025;4(6):13–19. doi: 10.22161/ijtle.4.6.2.
55. Bélisle-Pipon JC. Why we need to be careful with LLMs in medicine. *Front Med*. 2024;11:1495582. doi: 10.3389/fmed.2024.1495582.
56. Li X, Yan X, Lai H. The ethical challenges in the integration of artificial intelligence and large language models in medical education: A scoping review. *PLoS One*. 2025;20(10):e0333411. doi: 10.1371/journal.pone.0333411.
57. Tian J, Zhang R. Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. *Acta Psychol*. 2025;260:105725. doi: 10.1016/j.actpsy.2025.105725.

58. Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn Environ.* 2024;11(1):28. doi: 10.1186/s40561-024-00316-7.
59. Williams RT. The ethical implications of using generative chatbots in higher education. *Front Educ.* 2024;8:1331607. doi: 10.3389/feduc.2023.1331607.
60. Kofinas AK, Tsay CHH, Pike D. The impact of generative AI on academic integrity of authentic assessments within a higher education context. *Brit J Educ Technol.* 2025;56(6):2522–2549. doi: 10.1111/bjet.13585.
61. Chen Y, Esmailzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. *J Med Internet Res.* 2024;26:e53008. doi: 10.2196/53008.
62. Izquierdo-Condoy JS, Arias-Intriago M, Montero Corrales L, Ortiz-Prado E. Artificial Intelligence in Medical Education: Transformative Potential, Current Applications, and Future Implications. *JMIR Med Educ.* 2026;12:e77127-e77127. doi: 10.2196/77127.
63. Khakpaki A. Advancements in artificial intelligence transforming medical education: a comprehensive overview. *Med Educ Online.* 2025;30(1):2542807. doi: 10.1080/10872981.2025.2542807.
64. Sriram A, Ramachandran K, Krishnamoorthy S. Artificial Intelligence in Medical Education: Transforming Learning and Practice. *Cureus.* 2025;17(3):e80852. doi: 10.7759/cureus.80852.
65. Tsang MO, Dai SCK. Artificial Intelligence–Driven Personalised Surgical Curricula for Medical Students. *Premier Journal of Science* 2025;14:100189. doi: 10.70389/PJS.100189.
66. Chance EA. The combined impact of AI and VR on interdisciplinary learning and patient safety in healthcare education: a narrative review. *BMC Med Educ.* 2025;25(1):1039. doi: 10.1186/s12909-025-07589-7.
67. Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education – a scoping review. *BMC Med Educ.* 2024;24(1):788. doi: 10.1186/s12909-024-05777-5.
68. Raghavan N, Patel PC, Limon D, Morris MX, Kempenich JW, Rajesh A. Artificial Intelligence in Surgical Education: A 2025 Update on Adaptive Training, Feedback, and Competency-Based Education. *Am Surg.* 2026;92(3):675–686. doi: 10.1177/00031348251397597.

69. Sankar S, Rabbi KA, Crosby RF, Rajhi YA, Karunaratne V, Rathnayake M, et al. Artificial Intelligence in Neurosurgical Training: Applications, Challenges, and Future Directions. *Asian J Neurosurg*. Published online June 30, 2026. doi: 10.1055/s-0046-1824448.

70. Leal JA. Artificial Intelligence in Orthopaedic and Trauma Surgery Education: Applications, Ethics, and Future Perspectives. *J Am Acad Orthop Surg Glob Res Rev*. 2025;9(9):e25.00174. doi: 10.5435/JAAOSGlobal-D-25-00174.