Clinical Research

Applications of artificial intelligence in neurosurgical practice and residency education: a systematic review

Dewa Putu Wisnu Wardhana[1], Irawan Yusuf[2,3], Irwin Aras[3]

Check for updates

pISSN: 0853-1773 • eISSN: 2252-8083
https://doi.org/10.13181/mji.oa.268320
Med J Indones. 2026;35:186–94

Received: July 10, 2025
Accepted: April 13, 2026

Authors' affiliations:
[1]Department of Surgery, Faculty of Medicine, Universitas Udayana, Universitas Udayana Hospital, Bali, Indonesia, [2]Department of Physiology, Universitas Hasanuddin, Makassar, Indonesia, [3]Department of Medical Education, Universitas Hasanuddin, Makassar, Indonesia

Corresponding author:
Dewa Putu Wisnu Wardhana
Neurosurgery Subdivision, Department of Surgery, Faculty of Medicine, Universitas Udayana, Universitas Udayana Hospital, Jalan Rumah Sakit Unud, Jimbaran, South Kuta, Badung, Bali 80361, Indonesia
Tel/Fax: +62-361-222510/ +62-361-246656
E-mail: wisnuwardhana@me.com

ABSTRACT

BACKGROUND Artificial intelligence (AI) offers innovative and efficient training methods for neurosurgery residents, yet existing evidence remains fragmented, often limited to specific technological applications and failing to fully assess their impact on clinical competence. Therefore, this study aimed to map the current evidence on AI applications in neurosurgery and curriculum development.

METHODS This study was conducted following the preferred reporting items for systematic reviews and meta-analyses guidelines by searching for observational or experimental studies on AI in neurosurgery residency education (2020–2025) in PubMed, ScienceDirect, and Google Scholar. Keywords included variations of "training," "artificial intelligence," and "neurosurgery". After screening 719 initial records and assessing their eligibility, 10 studies were selected for narrative synthesis.

RESULTS A review of 10 studies reveals 2 key roles for AI in neurosurgical training: improving psychomotor skills through machine learning-integrated simulators and reinforcing theoretical knowledge through large language models (LLMs). Virtual reality (VR) simulators such as NeuroVR and virtual operative assistant provide objective metrics and personalized feedback, with assessment accuracy of up to 80%. LLMs such as ChatGPT show 50.4–84.8% accuracy in answering exam questions. However, simulators cannot yet capture all surgical competencies, and LLMs are still weak in processing image-based data.

CONCLUSIONS AI can improve neurosurgical training through personalized, objective learning tools, but AI simulators and LLMs remain supplements and cannot replace human expertise in clinical reasoning and complex image interpretation.

KEYWORDS artificial intelligence, education, internship and residency, medicine, neurosurgery, simulation training

The development of digital technology, particularly after the coronavirus disease 2019 pandemic, has accelerated the transformation of education and healthcare, including the increased use of artificial intelligence (AI).[1,2] AI can mimic human thought processes, such as reasoning and problem solving, and is increasingly accepted as a learning support tool.[3,4] In medical education, this technology is used to personalize learning materials, provide automatic feedback, and manage large-scale data. A cross-sectional survey using an online questionnaire revealed that 99.1% of students believed that AI would play an important role in healthcare services, and 82.1% of the respondents agreed that medical students should be trained using AI.[5]

In clinical medicine, AI improves the quality of education and healthcare services through its strong analytical capabilities, assisting in diagnosis, clinical

Copyright @ 2026 Authors. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are properly cited. For commercial use of this work, please see our terms at https://mji.ui.ac.id/journal/index.php/mji/copyright.

decision-making, and the prediction of treatment outcomes quickly and with precision.[4,6] In medical education, AI supports adaptive learning, data-driven simulations, and administrative efficiency, allowing educators to focus more on clinical interactions.[7] Its relevance is even greater in specialized fields such as neurosurgery, which require long-term intensive training and high technical skills but have limited practical experiences because of complex anatomy and the risk of technical errors.[8,9] Neurosurgery is a data-rich field that frequently uses medical imaging, surgical recordings, and robotic procedures. AI has a strong potential for recognizing complex patterns, analyzing unstructured data such as images and videos, and improving procedural precision, thereby accelerating the learning curve and enhancing training safety.[10–12] Furthermore, there is a lack of synthesis regarding the quality of evidence, effectiveness of skill transfer, and readiness for AI integration into the neurosurgical curricula. Therefore, this study aimed to collate the current evidence on AI applications in neurosurgery, identify existing knowledge gaps, and provide recommendations to guide future research and curriculum development.

METHODS

A systematic literature search considering studies related to using AI in neurosurgery was conducted. This systematic review adhered to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines and was registered with the prospective register of systematic reviews (CRD420251131073). The studies included in this review were articles published in online scientific databases, including PubMed, ScienceDirect, and Google Scholar.

The inclusion criteria used in the systematic review consisted of observational or experimental studies investigating the use of AI in neurosurgery resident education published from January 1, 2020, to January 1, 2025. Furthermore, the inclusion criteria focused on specific educational targets, categorized into three domains: psychomotor skills acquisition, theoretical knowledge enhancement, and clinical reasoning support. The exclusion criteria included irrelevant data, inaccessible full texts, and studies not written in English.

A literature search of relevant studies using controlled keywords from population/patient/problem, intervention, comparison, and outcome development was conducted to obtain keywords (Table 1). During journal search using search engines, researchers initially determined the keywords to be used along with their synonyms using Medical Subject Headings related to "training," "artificial intelligence," and "neurosurgery" in the article title, abstract, and keywords.

To identify relevant studies, abstracts and titles were reviewed by three independent reviewers (DPWW, IA, and IY). The same reviewers independently determined whether the full texts were relevant to the topic. Where there were differences in judgment, they were resolved through a discussion.

Data extraction was performed by three independent reviewers (DPWW, IA, and IY) using an agreed standardized form. Any disagreements were discussed or resolved through consensus.

Data collection and management were performed using Microsoft Office LTSC Professional Plus 2021 (Microsoft Corp., USA), and no automation tools were used. The extracted data included the author, year of publication, study design, AI model, outcomes, and limitations.

Data were analyzed using a narrative synthesis approach, allowing for a comprehensive exploration of various AI applications in the training of neurosurgery resident doctors without relying on quantifiable

Table 1. PICO framework and search strategy

Component	Description	Keywords/search terms
Population	Neurosurgical field	"Neurosurgery" OR "Neurosurgeries" OR "Neurosurgical"
Intervention	AI technology	"Artificial Intelligence" OR "AI" OR "Machine Intelligence" OR "Reasoning" OR "Computational Intelligence" OR "Computer Vision System"
Comparison	Traditional/standard methods	"Training" OR "Education"
Outcome	Psychomotor skills, theoretical knowledge, and clinical reasoning	Not specifically used as string keywords

AI=artificial intelligence; PICO=population/patient/problem, intervention, comparison, and outcome

Table 2. Risk of bias assessment summary

First author, year	Study design	Risk of bias level	Primary reason
Alkadri,[13] 2021; Li,[14] 2024	Observational/ simulator	Low	Use of consistent and objective simulator performance metrics.
Fazlollahi,[15] 2022	Randomized clinical trial	Low	Robust randomization, appropriate controls, and blinded outcome measurement.
Fazlollahi,[16] 2023	Retrospective cohort	Low	Transparent methodology and objective metric extraction from simulator data.
Chen,[17] 2023; Stengel,[18] 2024; Powers,[19] 2024	Comparative	Low	Use of standardized question banks and predefined reference standards.
Williams,[20] 2024; Bartoli,[21] 2024	Comparative	Moderate to high	Small convenience samples (n = 8 and n = 10), limiting generalizability.
Dabbas,[22] 2024	Comparative	High	Lack of blinding in evaluator scoring, introducing interpretation bias.

effect sizes. The studies were grouped based on characteristics such as publication year, research design, country, sample size, and AI model. Following a detailed description of the study findings in a manuscript, a cross-study synthesis was conducted to identify patterns, themes, differences, and correlations to comprehensively understand the potential of AI in neurosurgical education.

Methodological quality assessment

To evaluate the methodological quality and risk of bias of the included studies, we used standardized appraisal instruments tailored to specific study designs. Cohort and experimental studies were assessed using the Joanna Briggs Institute Critical Appraisal Checklist, and diagnostic accuracy was evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 tool (GitHub, USA). These tools enabled a systematic examination of the study design, conduct, and analysis to identify potential sources of bias. The assessment revealed a heterogeneous risk-of-bias profile across the selected studies, with studies demonstrating varying levels of methodological rigor depending on their design, sample size, and blinding protocols (Table 2).[13–22]

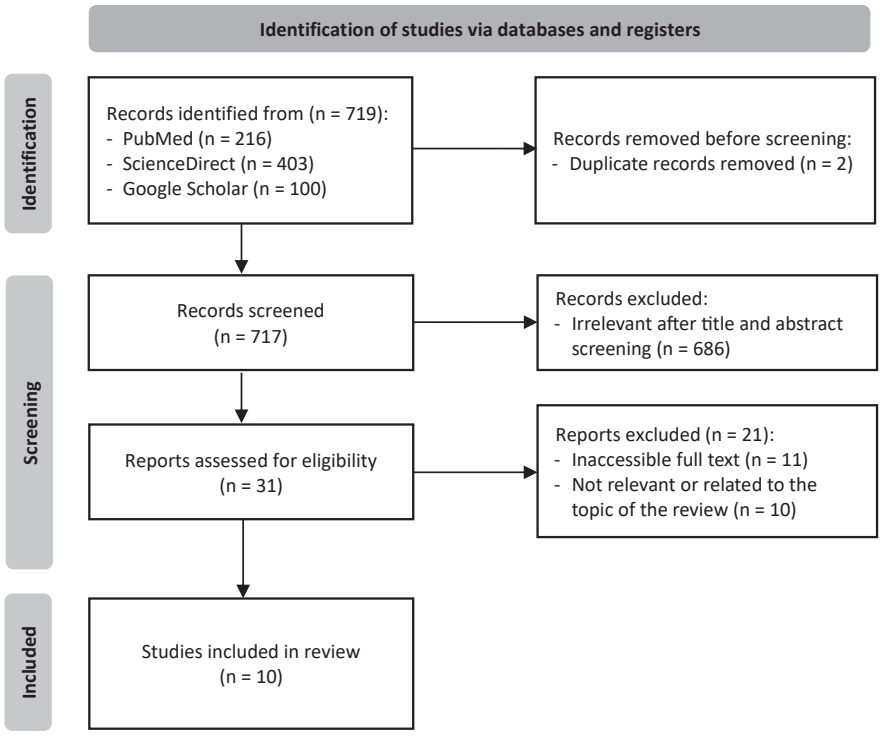


Figure 1. Flowchart of the search and study selection process

Table 3. Characteristics and key findings of the included studies on AI in neurosurgical education

First author, year	Study design	AI model	Outcomes	Limitations
Alkadri,[13] 2021	Observational study	Sim-Ortho simulator	ANN model trained on 9 selected surgical metrics, including motion, safety, and efficiency achieved a test accuracy of 80%.	The ACDF simulator used in this study does not include many of the complex interactions that occur in the performance of a patient's ACDF procedure.
Fazlollahi,[15] 2022	RCT	VOA (NeuroVR)	VOA can improve performance during training, better skill transfer, improve surgical performance metrics and the vast majority of participants (96%) reported that they prefer to learn from a combination of expert instruction and AI guidance.	VR simulation platforms fail to capture the full spectrum of competencies required in surgery, such as interdisciplinary teamwork.
Fazlollahi,[16] 2023	Cohort retrospective	VOA	The AI-enhanced simulation curriculum demonstrated significant unintended changes in 52 out-of-curriculum performance metrics that were not observed in the control cohort.	NA
Li,[14] 2024	Observational study	ML in VR	By using mRMR and SVM-RFE algorithms, 91 surgical performance metrics are reduced to 15 metrics making it more efficient than traditional training models.	NA
Chen,[17] 2023	Retrospective analysis	ChatGPT (GPT-4)	ChatGPT correctly answered 335 out of 509 questions (65.8%) on its first try. After 3 tries, ChatGPT's success rate increased to 383 questions (75.3%).	NA
Bartoli,[21] 2024	Comparative study	ChatGPT 3.5	ChatGPT scored 21.4 (25.5/50 based on 42 questions) and ranked 6th out of 11 candidates.	- ChatGPT cannot process queries containing images or schematic diagrams to interpret. - Formulating requests to ChatGPT requires an iterative process to ensure the questions and answers are correct.
Dabbas,[22] 2024	Comparative study	ChatGPT 3.5	Across 46 neurolocalization scenarios, ChatGPT 3.5 achieved an accuracy score of 84.8%, generating completely complete and largely correct responses. The average score was 4.5 out of 5 with an SD of 0.97.	This simplified case scenario may not fully capture the complexity of real-world neurological cases.
Stengel,[18] 2024	Comparative study	ChatGPT 3.5, Bing, and Bard	The Bard LLM achieved the highest accuracy with 62% of questions correct overall and 69% (excluding IB questions), outperforming human examinees at 59% and 59%, respectively. All LLMs scored highest on theory-based questions (excluding IB questions): ChatGPT: 79%; Bing: 83%; and Bard: 86%, significantly outperforming human examinees at 60%.	- Not taking the exam questions multiple times, which could theoretically improve the overall score of the AI model. - The questions used do not cover every aspect or topic within the broad field of neurosurgery.
Williams,[20] 2024	Prospective comparative study	ChatGPT 3.5	The average overall score for the ChatGPT interview was 41.7/60 (69.5% of overall) and ranked 7th out of 9 participants.	Lack of interpretation of radiological images, such as cranial or spinal imaging.
Powers,[19] 2024	Comparative study	ChatGPT	ChatGPT had an accuracy of 50.4% (1,068/2,120), with the highest and lowest accuracy in the pharmacology (81.2%, 13/16) and vascular (32.9%, 91/277) subcategories.	- The exclusion of imaging questions results in a difference in complexity from the actual board exam. - Smaller samples of evaluated questions hinder the analysis of the quality of education in specific subjects.

ACDF=anterior cervical discectomy and fusion; AI=artificial intelligence; ANN=artificial neural network; ChatGPT=chat generative pre-trained transformer; IB=image-based; LLM=large language model; ML=machine learning; mRMR=minimum redundancy maximum relevance; NA=not available; RCT=randomized clinical trial; SD=standard deviation; SVM-RFE=support vector machine-recursive feature elimination; VOA=virtual operative assistant; vr=virtual reality

RESULTS

Of 10 studies met the inclusion criteria, 4 were comparative studies, 2 were observational studies, 1 was randomized controlled trial, 1 was retrospective cohort study, 1 was retrospective analysis, and 1 was prospective comparative study. The systematic literature search and study selection process, conducted according to PRISMA guidelines, resulted in an initial 719 records, with the final selection summarized in the PRISMA flowchart (Figure 1).

A systematic review of 10 selected studies identified two primary applications of AI in neurosurgical residency training: the use of AI-integrated simulators for technical skill development and the utilization of AI platforms to support theoretical medical education. Data from these studies characterize the integration of AI as a tool for mapping educational evidence and identifying specific knowledge gaps within the current neurosurgical curricula (Table 3).

Training of surgical skills through virtual reality (VR) simulation

The results of the analyzed studies indicate that AI-integrated surgical simulators, such as NeuroVR, utilize performance data recorded at 50 Hz. The implementation of a virtual operative assistant (VOA) provides post-hoc feedback based on predetermined performance metrics. The trainee outcomes within the VR environment deviated from the expert benchmarks in two primary categories: safety and efficiency. Safety metrics demonstrated a reduction in healthy tissue removal and maximum instrument divergence distance. Contrastingly, the efficiency metrics decreased in dominant hand speed and acceleration, along with a reduced rate of tumor removal. The limitations identified in the VOA system include a lack of contextual components and time-specific directions. This study identified the development of an Intelligent Continuous Expertise Monitoring System (ICEMS) for real-time error projections and feedback.

Exam preparation and medical knowledge

The use of AI in education is reflected in the ability of large language models (LLMs), such as chat generative pre-trained transformer (ChatGPT), Bing, and Bard, to handle exam questions in neurology and neurosurgery. In a text-based neurolocalization study using 46 case scenarios, ChatGPT-3.5 achieved an accuracy score of 84.8% in generating "completely correct" and "mostly correct" responses. Although promising, the accuracy is not sufficiently reliable for routine patient care. Regarding the neurology board exam questions (from the commercial question bank BoardVitals), ChatGPT (likely version 3.5) answered 65.8% of the questions correctly on the first attempt and 75.3% on the third attempt. However, another study evaluating ChatGPT for neurosurgery board exam questions (Congress of Neurological Surgeons Self-Assessment in Neurological Surgery) reported an overall accuracy of 50.4%. This performance was lower than that of humans in most subcategories and was below the required pass rate for the exam (72% by 2023).

DISCUSSION

This systematic review of 10 studies identified two distinct but complementary AI roles in neurosurgical training: enhancement of psychomotor skills through machine learning (ML)-integrated simulation and augmentation of theoretical knowledge via LLMs. The principal finding is that while AI tutors in VR environments can objectively classify expertise and improve safety metrics, such as reducing healthy tissue removal, they may inadvertently compromise procedural efficiency. Conversely, while LLMs, such as ChatGPT, demonstrate potential in answering board-level text-based questions with accuracy ranging from 50.4–84.8%, their utility is currently limited by significant inconsistencies in handling complex clinical reasoning and image-based diagnostics. Collectively, the evidence suggests that AI is currently best utilized as a supplementary educational tool rather than to replace human instruction.

While traditional neurosurgical training relies on direct mentorship, this approach is subject to increasing constraints due to reduced resident work hours, stringent patient safety requirements, and ethical concerns regarding intraoperative learning.[23] Our review identifies that AI-driven simulation platforms address these systemic limitations by providing standardized, risk-free training environments that objectively quantify technical proficiency. Consequently, AI is not an alternative to but a supplementary framework to traditional mentorship, enabling skill acquisition independent of clinical caseload or faculty availability.[24]

VR simulations are increasingly used as surgical simulators, providing new benefits for

training and assessing neurosurgical skills.[13,14] Using AI through simulators will significantly impact neurosurgical education and training.[25,26] AI simulators provide objective and quantitative performance assessments.[15,26] AI systems generate measurable and reliable metrics from the analysis of performance data recorded during simulations. The ICEMS is an AI simulator that uses deep learning algorithms to predict skill levels based on performance metrics derived from simulator data.[15,16,26] This can be used by instructors and trainees to evaluate their technical performance and comprehensively understand skill gaps for individuals.[13]

Simulators can provide personalized and formative feedback to trainees. VOA systems use support vector machine algorithms to classify trainee performance and provide goal-oriented, metrics-based audiovisual feedback[15,16] specific to the learning needs of an individual, and it helps learners to identify areas for improvement and set realistic performance goals. Fazlollahi et al[15] demonstrated that metrics-based feedback from an AI tutor was more effective in improving surgical skills than remote expert instruction, consistent with the findings by Yilmaz et al,[27] who revealed that AI-guided feedback outperformed human instructors in a neurosurgical training study, resulting in significantly better performance outcomes for trainees. AI provides consistent and objective feedback, and through its ability to identify deficiencies, can contribute to more effective and efficient skill development of resident doctors compared with that of potentially subjective and variable human training.[27]

In addition to improving technical performance, using AI simulators can lead to unexpected learning outcomes.[16] A cohort study conducted by Fazlollahi et al[16] revealed that an AI-enhanced simulation curriculum led to significant changes in 52 performance metrics beyond the taught curriculum. These unexpected changes included significant improvements in procedural safety, such as reduced damage to healthy tissues and decreased motion and efficiency metrics, indicating the need for continuous evaluation and expert intervention to optimize educational objectives and ensure that unexpected outcomes are managed effectively.[16] Alkadri et al[13] conducted a data extraction study, resulting in 157 surgical performance metrics spanning three categories: motion, safety, and efficiency. An artificial neural network model trained on nine selected surgical metrics, covering all three categories, achieved a testing accuracy of 80%.[13] The findings by Li et al[14] were consistent with those of the two previous studies, reporting that ML-based training in VR, using the mRMR and SVM-RFE algorithms, generated 91 surgical performance metrics that were subsequently reduced to 15, making the training method more efficient than traditional approaches.

AI tutoring systems can provide automated feedback and assessments, reducing the need for intensive supervision by instructors. Fazlollahi et al[15] found that an AI intervention saved approximately 53 hours of expert supervision and assessment over 13 weeks, compared to that via an instructor-only group. This finding is consistent with those of Chawla et al[24] and Patel et al,[28] suggesting that AI simulators provide more efficient and cost-effective training. AI-powered simulators can deliver training that is not limited by the availability of operating room time, expert supervision, or physical location,[24] particularly in resource-constrained settings.[28] AI-enhanced simulation can replicate neurosurgical training, making high-quality practice opportunities more accessible to various trainees regardless of geographic or logistical constraints. Unlike traditional training, which relies on scheduled surgeries and the availability of supervising surgeons, AI simulators can be accessed anytime and anywhere, thus providing flexibility and convenience to trainees.

Previous studies have emphasized the potential of AI in identifying key performance metrics that differentiate skill levels.[13,14,16] By analyzing data from experts and trainees with varying levels of experience, ML algorithms can identify the factors that contribute the most to expert performance.[14,16] This information can be used to develop more focused and efficient training curricula for residents and establish measurable competency benchmarks for neurosurgical education.[16]

However, AI simulators have limitations. VR simulators may not capture all aspects of surgical competence, such as interdisciplinary teamwork and situational decision making. Therefore, it is necessary to develop systems that provide real-time feedback and combine the power of AI tutors with human instruction.[15,16]

LLMs, such as ChatGPT, have been identified as tools for personalizing neurosurgical education through the generation of customized learning materials and simulated patient encounters.[29,30] Our review of the included studies demonstrates that LLMs

can facilitate quick access to complex neurosurgical information, assist in clinical decision-making, such as determining adjuvant therapy in neuro-oncology.[31–33] Furthermore, the evidence revealed that these tools support surgical planning by evaluating medical records and recognizing patterns to assist in risk-benefit assessments for individual treatment plans.[34] In addition to its clinical applications, the literature suggests that AI can improve work efficiency by simplifying medical writing and data handling. However, the utility of LLMs remains primarily supplementary as their performance in important neurosurgical examinations and complex image-based diagnostics continues to exhibit significant variability.[35,36]

Considering the potential benefits of ChatGPT, it is often used in neurosurgery resident education. Previous studies have compared the accuracy of ChatGPT and neurosurgery residents in answering medical questions. In a study conducted by Bartoli et al,[21] ChatGPT 3.5 scored 21.4 out of 42 questions and ranked sixth out of 11 candidates, consistent with the findings of a study by Williams et al,[20] where the average overall score for ChatGPT 3.5 interviews was 41.7/60 (69.5% accuracy overall), ranking seventh out of nine participants. In another study, ChatGPT 3.5 achieved an accuracy score of 84.8% in generating completely complete and largely correct responses, out of 46 neurolocalization case scenarios.[22] Stengel et al[18] observed contrasting results, with ChatGPT exhibiting 69% accuracy compared to 59% for human participants on theory-based questions. These results align with ChatGPT's high accuracy of 81.2% in the field of pharmacology.[19] Collectively, these results demonstrate significant variability in ChatGPT's performance across different neurosurgical subspecialties and assessment formats, possibly reflecting inconsistencies in its ability to process complex clinical data.

However, a critical comparison of these studies revealed significant inconsistencies. ChatGPT has not surpassed the overall capabilities of residents, but has demonstrated its potential to contribute to improving education and other medical fields.[20,21] ChatGPT has weaknesses in image-based diagnostics.[17,18] ChatGPT performs poorly in classifying glioma types and has low accuracy in the vascular field,[19,33] suggesting that it has weaknesses in processing complex data and requires further elaboration. Contrastingly, Stengel et al[18] reported that LLMs outperformed humans in theory-based questions, while Powers et al[19] revealed that ChatGPT failed the American Board of Neurological Surgery primary exam with 50.4% accuracy. This inconsistency is attributed to the "exclusion bias" in some studies where image-based questions were removed, artificially inflating the text-based accuracy scores.

ChatGPT operates on ML principles, requiring continuous database refinement and improving accuracy with increasing data input.[17,37] Furthermore, to achieve good accuracy, the data provided to ChatGPT must be accurate and concise. This demonstrates that the quality of the data provided is more important than the quantity of data provided to determine the accuracy.[22] Furthermore, most studies still use ChatGPT 3.5, while ChatGPT 4 is now available, offering higher accuracy than ChatGPT 3.5. ChatGPT 4 addresses the weaknesses of ChatGPT 3.5, demonstrating 20% greater accuracy than that of ChatGPT 3.5 in imaging-based diagnosis.[38]

Although AI tools have potential, their results must be continuously reviewed.[39] Furthermore, the ethical implications of ChatGPT, legal responsibilities, data privacy, licensing, and regulations are a major concern. Integration with hospital personnel on the judicious use of AI is required to address ethical and validation challenges.[40] The above data show that ChatGPT has advantages and disadvantages when applied to neurosurgery resident education. Therefore, collaboration among supervising physicians, conventional teaching methods, and the use of ChatGPT is necessary to improve the quality of neurosurgery resident education.

The integration of AI tools has profound implications for patient safety and clinical standards. The ability of AI-enhanced simulators to enforce safety thresholds, such as limiting force application and maximizing healthy tissue preservation, potentially reduces iatrogenic injuries during practical procedures. However, reliance on LLMs for decision support is a public health risk if not rigorously supervised, given their documented propensity for factual inaccuracies.

Despite the promising initial results, this review identified a critical scientific gap: the lack of longitudinal data correlating AI-based training with long-term skill retention and operative outcomes in patients. Most of the included studies were cross-sectional and assessed immediate performance.

Future research must bridge the 'simulation-to-theatre' gap to validate the clinical efficacy of these educational interventions.

This systematic review had limitations. First, several comparative studies were based on small convenience samples (n = 10), which limits the statistical power and generalizability of the results to a broader population of trainees. Second, methodological quality assessments revealed a critical risk of bias in some studies. Third, the evaluation of LLMs often excludes image-based questions owing to technical constraints, artificially inflating the accuracy scores, and not reflecting the multimodal complexity of practical neurosurgical operations. Additionally, most studies employed cross-sectional designs that assessed immediate performance in simulated environments, leaving a gap in the evidence regarding long-term skill retention and the transferability of these skills to clinical patient outcomes. Finally, the review was restricted to English-language publications from 2020 to 2025, potentially introducing publication bias, and excluding relevant data from other languages or earlier timeframes.

Future research should prioritize longitudinal studies that bridge the "simulation-to-operating room" gap by correlating the performance of AI-based simulators with long-term surgical outcomes and patient safety metrics. Furthermore, to address the current diagnostic limitations of LLMs, prospective studies should focus on adopting multimodal AI architectures capable of integrating complex radiological imaging and real-time surgical video into clinical reasoning frameworks. Additionally, the development of next-generation simulators is necessary to incorporate non-technical competencies, such as interdisciplinary teamwork and situational decision-making, which are currently underrepresented in VR platforms. Finally, the establishment of standardized ethical guidelines and policy frameworks is crucial to ensuring data privacy and algorithmic transparency before AI can be fully institutionalized within global neurosurgery residency curricula.

In conclusion, this study underscores that AI is a transformative and supplementary tool in neurosurgical education, primarily through the dual application of ML-integrated simulators for psychomotor skill acquisition and LLMs for theoretical knowledge enhancement. Although AI-driven VR simulators provide objective and personalized feedback that improves safety metrics and procedural precision, they do not capture all surgical competencies, and current LLMs exhibit significant inconsistencies when handling complex clinical reasoning or image-based diagnostics. Ultimately, AI cannot replace clinical reasoning and human expertise provided by traditional mentorship, necessitating a collaborative approach between supervised instruction and AI tools. Future research must address the current scientific gap by focusing on longitudinal studies that correlate AI-based training with long-term operative outcomes and patient safety to ensure that these technologies are rigorously validated before full integration into neurosurgical residency curricula.

Conflict of Interest

The authors affirm no conflict of interest in this study.

Acknowledgment

None.

Funding Sources

None.

REFERENCES

1. Ambarita N, Nurrahmatullah MF. Impacts of artificial intelligence on student learning: a systematic literature review. J Varidika. 2024;36(1):13–30.
2. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MA, Dwivedi YK. A systematic literature review of artificial intelligence in the healthcare sector: benefits, challenges, methodologies, and functionalities. J Innov Knowl. 2023;8(1):100333.
3. Danilov GV, Shifrin MA, Kotik KV, Ishankulov TA, Orlov YN, Kulikov AS, et al. Artificial intelligence in neurosurgery: a systematic review using topic modeling. Part I: major research areas. Sovrem Tekhnologii Med. 2021;12(5):106–12.
4. Bandara WM, Senanayaka SG. Use of artificial intelligence in education: a systematic review. 2024 International Research Conference on Smart Computing and Systems Engineering (SCSE); 2024 April 04-04; Colombo, Sri Lanka. New York: Institute of Electrical and Electronics Engineers (IEEE); 2024. p. 15.
5. Budu, Aras I, Amalia A. [Standards of education for master of medical education and health professions]. Yogyakarta: Indonesian College of Health Profession Education; 2023. Indonesian.
6. Buabbas AJ, Miskin B, Alnaqi AA, Ayed AK, Shehab AA, Syed-Abdul S, et al. Investigating students' perceptions towards artificial intelligence in medical education. Healthcare (Basel). 2023;11(9):1298.
7. Salam A, Abhinesh N. Revolutionizing dermatology. IP Indian J Clin Exp Dermatol. 2024;10(2):107–12.
8. Yaeger KA, Munich SA, Byrne RW, Germano IM. Trends in United States neurosurgery residency education and training over the last decade (2009-2019). Neurosurg Focus. 2020;48(3):e6.
9. Ayuningtyas DS. [Implementation of workplace-based learning in residency education]. Jurnal Implementa Husada. 2024;5(2):77–84. Indonesian.
10. Iop A, El-Hajj VG, Gharios M, de Giorgio A, Monetti FM, Edström E, et al. Extended reality in neurosurgical education: a systematic review. Sensors. 2022;22(16):1–20.
11. Amru K, Hamid F, Afandi IN, Amalia A, Mannyu B, Nelwan BJ, et

al. The association of "first 1000 days of life" training program on communication skill and empathy of undergraduate medical students: a cross-sectional study. J Adv Med Educ Prof. 2023;11(4):222–9.
12. 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 Neurosurg. 2025;197:123809.
13. Alkadri S, Ledwos N, Mirchi N, Reich A, Yilmaz R, Driscoll M, et al. Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. Comput Biol Med. 2021;136:104770.
14. 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.
15. Fazlollahi AM, Bakhaidar M, Alsayegh A, Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. JAMA Netw Open. 2022;5(2):e2149008.
16. Fazlollahi AM, Yilmaz R, Winkler-Schwartz A, Mirchi N, Ledwos N, Bakhaidar M, et al. AI in surgical curriculum design and unintended outcomes for technical competencies in simulation training. JAMA Netw Open. 2023;6(9):e2334658.
17. Chen TC, Multala E, Kearns P, Delashaw J, Dumont A, Maraganore D, et al. Assessment of ChatGPT's performance on neurology written board examination questions. BMJ Neurol Open. 2023;5(2):e000530.
18. Stengel FC, Stienen MN, Ivanov M, Gandía-González ML, Raffa G, Ganau M, et al. Can AI pass the written European Board Examination in neurological surgery? - ethical and practical issues. Brain Spine. 2024;4:102765.
19. Powers AY, McCandless MG, Taussky P, Vega RA, Shutran MS, Moses ZB. Educational limitations of ChatGPT in neurosurgery board preparation. Cureus. 2024;16(4):e58639.
20. Williams SC, Starup-Hansen J, Funnell JP, Hanrahan JG, Valetopoulou A, Singh N, et al. Can ChatGPT outperform a neurosurgical trainee? a prospective comparative study. Br J Neurosurg. 2024; 39(5):627–36.
21. Bartoli A, May AT, Al-Awadhi A, Schaller K. Probing artificial intelligence in neurosurgical training: ChatGPT takes a neurosurgical residents written exam. Brain Spine. 2024;102715.
22. Dabbas WF, Odeibat YM, Alhazaimeh M, Hiasat MY, Alomari AA, Marji A, et al. Accuracy of ChatGPT in neurolocalization. Cureus. 2024;16(4):e59143.
23. Suri A, Sharma R, Singh R, Baby B, Dhanakshirur R, Dhingra R, et al. Neurosurgery simulation-based skills training: from mannequins and cadavers to virtual reality and artificial intelligence. Natl Board Exam J Med Sci. 2023;1(1):41–54.
24. Chawla S, Devi S, Calvachi P, Gormley WB, Rueda-Esteban R. Evaluation of simulation models in neurosurgical training according to face, content, and construct validity: a systematic review. Acta Neurochir (Wien). 2022;164(4):94766.
25. Piñera-Castro HJ, Borges-García C. Applications of artificial intelligence in neurosurgical education: a scoping review. Egypt J Neurol Psychiatry Neurosurg. 2025:61(83).
26. Park JJ, Tiefenbach J, Demetriades AK. The role of artificial intelligence in surgical simulation. Front Med Technol. 2022;4:1076755.
27. Yilmaz R, Bakhaidar M, Alsayegh A, Abou Hamdan N, 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.
28. Patel EA, Aydin A, Cearns M, Dasgupta P, Ahmed K. A systematic review of simulation-based training in neurosurgery, part 1: cranial neurosurgery. World Neurosurg. 2020;133:e850–73.
29. Wu Z, Li S, Zhao X. The application of ChatGPT in medical education: prospects and challenges. Int J Surg. 2025;111(1):1652-3.
30. Scherr R, Halaseh FF, Spina A, Andalib S, Rivera R. ChatGPT interactive medical simulations for early clinical education: case study. JMIR Med Educ. 2023;9(1):e49877.
31. Rubalcava NS, Guetter CR, Kapani N, Quiñones PM. How artificial intelligence is expected to transform surgical training [Internet]. Chicago: American College of Surgeons; 2023 [cited 2025 Dec 13];108(8). Available from: https://www.facs.org/for-medical-professionals/news-publications/news-and-articles/bulletin/2023/august-2023-volume-108-issue-8/how-artificial-intelligence-is-expected-to-transform-surgical-training/.
32. Kulkarni S, Seneviratne N, Baig MS, Khan AH. Artificial intelligence in medicine: where are we now? Acad Radiol. 2020;27(1):62–70.
33. Haemmerli J, Sveikata L, Nouri A, May A, Egervari K, Freyschlag C, et al. ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board? BMJ Health Care Inform. 2023;30(1):e100775.
34. Roman A, Al-Sharif L, AL Gharyani M. The expanding role of ChatGPT (chat-generative pre-trained transformer) in neurosurgery: a systematic review of literature and conceptual framework. Cureus. 2023;15(8):e43502.
35. He Y, Tang H, Wang D, Gu S, Ni G, Wu H. Will ChatGPT/GPT-4 be a lighthouse to guide spinal surgeons? Ann Biomed Eng. 2023;51(7):1362–5.
36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in Neural Information Processing Systems 30 [Internet]. 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec 4–9; Long Beach, California. San Diego: NeurIPS; 2017 [cited 2025 Dec 13]. p. 5999–6009. Available from: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.
37. Chintagunta B, Katariya N, Amatriain X, Kannan A. Medically aware GPT-3 as a data generator for medical dialogue summarization [Internet]. Proceedings of the Second Workshop on Natural Language Processing for Medical Conversations; 2020 Jul 10; Seattle, Association for Computational Linguistics; 2021 [cited 2025 Dec 13]. p. 66–76. Available from: https://aclanthology.org/2021.nlpmc-1.9/.
38. Ali R, Tang OY, Connolly ID, Fridley JS, Shin JH, Zadnik Sullivan PL, et al. Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral boards preparation question bank. Neurosurgery. 2023;93(5):1090–8.
39. Biswas SS. Role of Chat GPT in public health. Ann Biomed Eng. 2023;51(5):868–9.
40. Bečulić H, Begagić E, Skomorac R, Mašović A, Selimović E, Pojskić M. ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. Med Glas. 2024;21(1).