

Review article

Artificial intelligence in plastic surgery: current applications, limitations, and ethical considerations

Tinh Nguyen Trung *Department of Surgery, Faculty of Medicine, College of Health Sciences, Nam Can Tho University, Can Tho, Vietnam*

Abstract

Artificial intelligence (AI) is increasingly being applied in plastic surgery, which is a specialty that relies heavily on visual assessment, technical precision, and individualized clinical judgment. To summarize current applications of AI in plastic surgery and evaluate the associated methodological limitations and ethical considerations relevant to clinical practice and education. A narrative literature review was performed using PubMed, Scopus, and Web of Science from database inception through January 2026. Studies were included if they addressed AI applications in plastic surgery practice and education, as well as the ethical domain thereof. Evidence was synthesized across thematic areas, including image-based systems, large language models (LLMs), methodological constraints, and regulatory considerations. Published studies primarily describe image-based systems supporting facial analysis, surgical planning, and simulation. However, recent work has explored the use of LLMs for patient information, clinical support, and resident education. Reported performance varies depending on the dataset characteristics, study design, and clinical context. Common limitations include selection bias, small or non-diverse datasets, limited external validation, inconsistent assessment standards, and reduced model transparency. The ethical concerns include bias, informed consent, disclosure of AI involvement, and accountability in academic use, particularly in an industry where outcomes are subjective and closely linked to patient expectations. Current evidence supports the cautious use of AI as a supplementary tool rather than an independent decision-making system. Furthermore, broader clinical integration will require prospective validation, standardized reporting, and clear ethical and regulatory guidance.

Keywords: Artificial intelligence, ethics, machine learning, plastic surgery, surgical education.

Artificial intelligence (AI) has become increasingly applied across medical specialties, influencing areas such as diagnosis, treatment planning, and medical education. ⁽¹⁻³⁾ In plastic surgery, where visual evaluation, technical execution, and individualized clinical judgment are central to care, these developments have garnered attention for AI's potential as a supportive tool in practice and training. ^(4,5)

Advances in machine learning, including convolutional neural networks for image analysis and large language models (LLMs) for text-based tasks, have enabled the automated processing of complex visual and written information. ⁽⁶⁻⁸⁾ Such developments are particularly relevant to plastic surgery, especially

in facial analysis and symmetry assessment. Furthermore, their use raises familiar concerns regarding reliability, bias, and the responsible application of facial recognition technologies. ⁽⁹⁾

Early studies in plastic surgery primarily focused on facial aesthetic assessment, surgical simulation, and outcome prediction. ^(4,10) However, recent work has evaluated their potential role as tools for resident education and patient-oriented communication, alongside broader efforts in medicine to explore AI-supported clinical decision systems. ⁽¹¹⁻¹⁴⁾ The emergence of LLMs has expanded these discussions beyond image-based systems to include generative tools capable of producing medical explanations, procedural summaries, and educational material. ^(12,15) As these tools have become more accessible, questions related to transparency, bias, authorship, and scientific integrity have received increasing attention. ⁽¹⁶⁻¹⁸⁾

Despite sustained academic interest, the routine integration of AI into standard plastic surgery practice remains limited. Much of the available literature is

***Correspondence to:** Tinh Nguyen Trung, Department of Surgery, Faculty of Medicine, College of Health Sciences, Nam Can Tho University, Can Tho, Vietnam.

E-mail: nttinh@nctu.edu.vn

Received: March 20, 2026

Revised: June 12, 2026

Accepted: August 28, 2026

exploratory and is often based on small datasets, retrospective designs, and limited external validation, patterns that reflect broader translational challenges in clinical AI research. ⁽¹⁹⁾ Moreover, ethical and regulatory structures have not consistently kept pace with technological development, which leaves issues of accountability, bias, and governance only partially addressed as these applications continue to expand. ⁽²⁰⁾

Although previous reviews have discussed AI in medicine more broadly, specialty-specific syntheses that combine clinical applications, educational implications, and ethical considerations within the plastic surgery domain remain relatively limited. ⁽¹⁾ In particular, the rapid uptake of LLMs and other generative tools warrants further examination, given their emerging influence on patient communication, scholarly practice, and the development of educational content. ^(12, 14, 18)

Therefore, this narrative review aims to summarize current applications of AI in plastic surgery, examine the methodological limitations, and discuss ethical and regulatory considerations relevant to clinical care and surgical education.

Methods

A narrative review methodology was selected to capture the breadth and heterogeneity of the reported applications of AI in plastic surgery. This approach was considered appropriate given the wide variation in AI models, outcome measures, and study designs currently described in the literature, which limits the feasibility of a systematic review or meta-analysis.

A literature search was performed using PubMed, Scopus, and Web of Science from database inception through January 2026. Search terms included combinations of “artificial intelligence,” “machine learning,” “deep learning,” “large language models,” “ChatGPT,” and “plastic surgery,” along with relevant subspecialty terms such as “facial plastic surgery,” “hand surgery,” “microsurgery,” and “aesthetic surgery.”

Studies were included if they addressed AI applications relevant to plastic surgery practice, education, or ethical considerations. Original research studies and high-quality review articles were eligible for inclusion, with particular emphasis placed on recent publications examining LLMs, AI-generated content, and specialty-specific ethical concerns. Articles that were not directly related to clinical or educational contexts in plastic surgery were excluded.

Included studies were organized into thematic domains (**Figure 1**): 1) clinical applications; 2) education and training; 3) limitations and performance concerns; and 4) ethical and regulatory considerations.

Priority was given to studies with clear methodology, clinical relevance, and transparent reporting of model development and validation; however, a formal risk-of-bias assessment was not performed due to heterogeneity in study designs.

Current applications and evidence synthesis

• **Image-based analysis and surgical planning (Figure 1)**

The most established and extensively studied applications of AI in plastic surgery include image-based analysis, particularly for facial aesthetic and reconstructive procedures. ^(4,10) Advances in computer vision and deep learning have enabled the development of algorithms capable of identifying facial landmarks, quantifying symmetry, and analyzing the proportional relationships that are central to aesthetic evaluation and surgical assessment. ⁽⁴⁾

These image-based systems have been applied in facial plastic surgery to support preoperative planning by enabling the automated assessment of anatomical features and visual simulation of potential surgical outcomes (**Table 1**). Such applications may assist surgeons by providing structured visual references during consultations and facilitating discussions regarding the potential surgical approaches and anticipated postoperative changes. ^(4,10) In addition, AI-generated visualizations have been investigated as adjunctive tools for patient counseling, particularly in procedures where the outcomes are closely linked to subjective aesthetic expectations and patient perception.

Despite these advances, most reported image-based AI models have been trained and evaluated on curated datasets collected under controlled conditions. These datasets often lack adequate representation across diverse ethnic backgrounds, age groups, and real-world imaging environments, thus raising concerns about their applicability to routine clinical practice. ^(9,19) Consequently, algorithmic performance may vary when applied outside the original training context, thereby limiting confidence in their generalizability across broader patient populations.

Furthermore, external validation of image-based AI systems in plastic surgery remains limited. Many studies rely on retrospective analyses and single-institution datasets, with relatively few assessments conducted in prospective or multi-center clinical

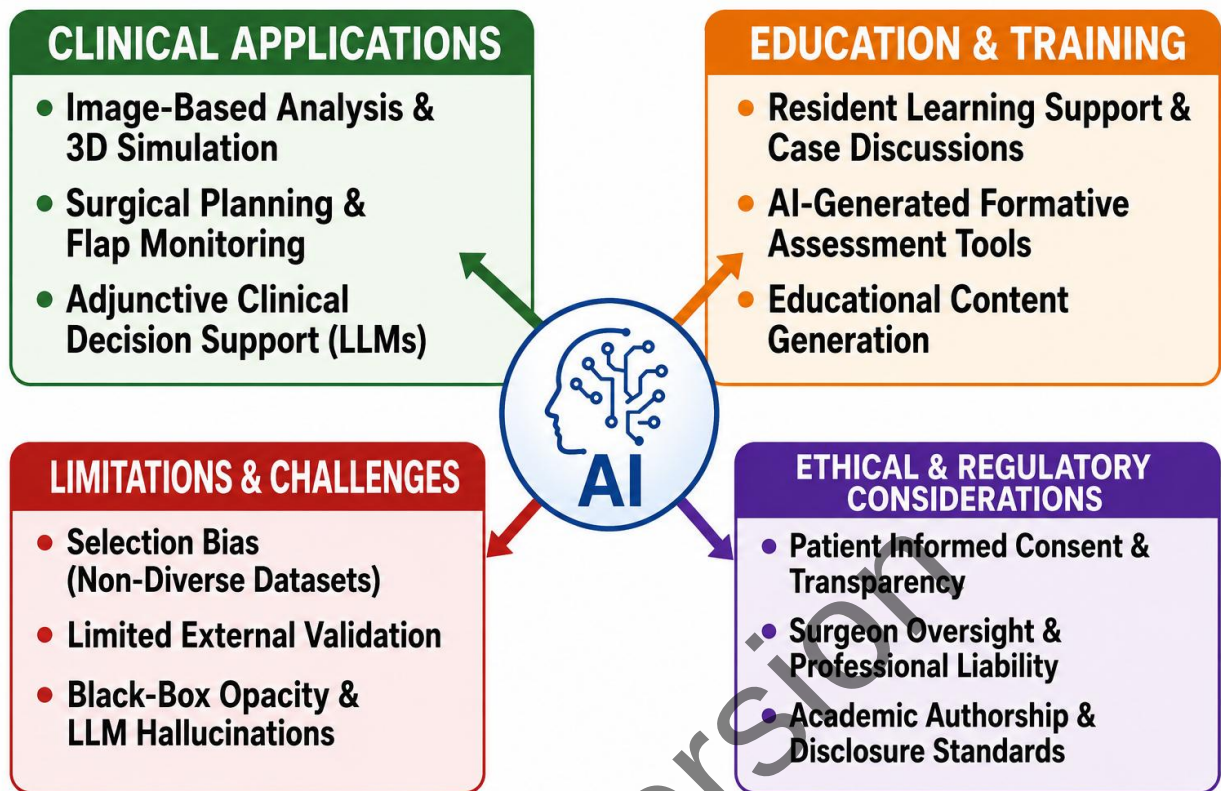


Figure 1. Key domains of AI integration in plastic surgery. Overview of current domains of artificial intelligence applications in plastic surgery, including clinical practice, education and training, limitations, and ethical and regulatory considerations. AI tools are intended to function as adjunctive resources under surgeon oversight rather than as replacements for clinical judgment. AI, Artificial intelligence.

settings.^(10,19) Variability in image acquisition, lighting conditions, and patient positioning may further affect the algorithmic performance in everyday practice. Collectively, these limitations emphasize the necessity of cautious interpretation of reported results and reinforce the view that image-based AI tools should be regarded as adjunctive aids rather than definitive decision-making instruments in surgical planning.

• LLMs in clinical decision support and patient counseling

The emergence of LLMs has broadened the scope of AI applications in plastic surgery beyond image-based analysis to involve text-based clinical support and patient communication.^(12,15) These models can respond to medical queries in natural language, provide summaries of procedural information, and offer general educational content, which has led to a growing interest in their use as supportive tools in clinical and educational settings.

Recent evaluations of ChatGPT and similar LLMs indicate that these systems can generate coherent, contextually relevant responses to frequently asked questions about plastic surgery, with particular attention to gender-affirming procedures.^(12,15) Such

capabilities suggest a potential role for LLMs in facilitating access to preliminary information and supporting patient engagement, which is particularly relevant in settings with limited consultation time or when patients seek information before attending an in-person evaluation.

However, multiple studies have identified important limitations in the accuracy and reliability of LLM-generated content. The responses produced by these models may be incomplete, overly generalized, or factually inaccurate, especially when addressing nuanced clinical scenarios that require individualized assessment, procedural judgment, or specialty-specific expertise.^(12,15) The tendency of LLMs to generate confident-sounding but incorrect statements raises notable concerns regarding their unsupervised use in clinical contexts.

Moreover, performance variability becomes more pronounced in specialty-specific decision-making tasks. In assessments of complex hand surgery scenarios, LLMs have demonstrated inconsistent accuracy and, in some cases, have generated recommendations that deviate from established clinical standards.⁽²¹⁾ These findings underscore the inherent limitations of models trained primarily on general medical or non-specialty-specific datasets.

Table 1. Summary of AI applications in plastic surgical diagnostics and pre-operative support.

Domain	Application area	Specific AI model / technology	Primary function & clinical utility	Key methodological limitations
Facial aesthetic & reconstructive surgery	Facial symmetry and landmark assessment	Convolutional neural networks (CNNs), deep learning	Automated detection of facial asymmetry; preoperative planning and quantitative outcome assessment	Datasets often lack ethnic diversity; sensitive to lighting and head positioning
Rhinoplasty & maxillofacial planning	3D visual outcome simulation	Computer vision algorithms, Generative adversarial networks	Generates predictive post-procedure 3D models to guide surgical execution and patient consultation	Risk of raising unrealistic patient expectations; predictions are probabilistic
Breast reconstruction & aesthetics	Tissue volume estimation & flap monitoring	Machine learning image classification models	Calculates breast volume; assists in flap choice; early detection of vascular compromise	Limited multicenter prospective validation; small single-institution sample sizes
Burns & wound care management	Wound area segmentation & burn depth evaluation	Deep learning computer vision	Automated boundary detection; differentiation between partial and full-thickness burns	Varies across skin pigmentation; require high-standard standardized lighting
Hand & microsurgery	Structural injury detection & diagnostic support	Fine-tuned LLMs, Computer vision	Assist in interpreting complex radiological/clinical presentations in hand trauma	High variability in LLM responses; potential for generating non-standard clinical advice

LLM, large language models

Taken together, current evidence suggests that while LLMs may serve as supplementary tools for information dissemination and educational support, their role in clinical decision-making must remain limited and subject to active clinician oversight.^(21,22) Reliance on LLM-generated outputs without expert review carries a risk of misinformation and inappropriate guidance, thus reinforcing that these systems should function as adjuncts rather than substitutes for professional judgment in plastic surgery practice.

• AI in surgical education and training

AI has also been evaluated as a supplementary tool in plastic surgery education and training, particularly for resident learning and academic support (**Table 2**).^(13, 14) Previously reported applications include the automated generation of educational materials, structured case discussions, and formative assessment resources that are designed to complement established teaching methods.

Evaluations of this AI-generated educational content suggest that such materials may provide accessible and timely information for trainees; however, the quality and reliability of this content remain inconsistent.⁽¹⁴⁾ Although AI systems can generate coherent summaries and discussion prompts, previous studies have identified inaccuracies, omissions, and limited depth, particularly when the content is produced without specialty-specific input or expert oversight. Furthermore, robust evidence demonstrating measurable improvements in learning outcomes or clinical competence remains limited.

LLMs have received particular attention in surgical education because their outputs often closely resemble conventional educational text.^(13, 18) This feature raises concerns related to academic integrity and transparency, as AI-generated material may be difficult to distinguish from human-authored content. Moreover, in training environments, residents may rely on AI-generated outputs without fully appreciating their limitations or potential inaccuracies.

Table 2. Comparison of AI modalities in surgical practice and education.

AI Modality	Primary applications	Key benefits	Main challenges /risks	Human oversight requirement
Image-based systems (CNNs, computer vision)	Cephalometric analysis, flap monitoring, surgical simulation	Objective landmark detection; precise quantitative assessment	Algorithmic bias; performance drop on unstandardized real-world images	Mandatory: Surgeon must verify all landmark and volumetric calculations
LLMs	Patient frequently asked questions generation, procedural summaries, resident study aids	Rapid processing of textual queries; accessible educational materials	“Hallucinations” (confident factual errors); lack of clinical intuition	Mandatory: Expert clinician review required before sharing outputs
Generative AI tools	Educational case scenarios, draft generation for academic writing	Streamlines material creation; offers interactive learning prompts	Risk of academic integrity violations; lack of reliable detection tools	Mandatory: Full disclosure of AI involvement; human verification of references

AI, artificial intelligence; CNN, Convolutional Neural Networks; LLM, large language models.

The incorporation of AI tools into educational workflows has introduced challenges related to authorship, attribution, and disclosure. Studies assessing the detectability of AI-generated text have revealed that neither human reviewers nor currently available automated tools can reliably identify such content, thereby highlighting limitations in the existing oversight mechanisms.^(16,18) These findings highlight the need for clear institutional guidance regarding the acceptable use, disclosure practices, and accountability in educational and scholarly contexts.

Collectively, available evidence supports a cautious approach to the use of AI in plastic surgery education. While AI-generated material may serve as supplemental learning aids, it should not replace structured curricula, faculty mentorship, or the development of critical appraisal skills that are central to surgical training.^(13, 14, 18) Ensuring that trainees understand the capabilities and limitations of AI tools will be essential for their responsible integration into educational practice.

Limitations and challenges

Despite growing interest and encouraging early findings, several limitations continue to constrain the current application of AI in plastic surgery. Among the most substantial challenges are the quality and representativeness of data used to train AI models. Many existing systems rely on datasets that are limited in size and lack adequate representation of diverse patient populations, which increases the risk of biased

outputs and uneven performance across demographic groups.^(9, 19)

The predominance of retrospective and single-institution study designs further limits confidence in the reported outcomes. Although such studies may demonstrate feasibility under controlled conditions, they provide limited insight into how these AI systems perform in routine clinical practice, where variability in imaging quality, clinical workflows, and patient characteristics is common.^(10, 19) The scarcity of prospective and multicenter validation studies, therefore, remains a major barrier to broader clinical adoption.

LLMs introduce additional challenges related to reliability and interpretability. These systems are known to generate responses that appear fluent and authoritative but may include factual inaccuracies or clinically inappropriate recommendations.^(12, 21) This limitation is particularly concerning in contexts that involve clinical decision support and patient education, where misleading information may directly influence patients' understanding or expectations.

Model opacity represents another important limitation. Many AI systems operate as so-called “black boxes,” producing outputs without transparent explanations as to how the conclusions are derived.⁽²²⁾ This lack of interpretability complicates error identification, limits clinician confidence, and introduces concerns about accountability when AI-assisted recommendations are incorporated into clinical workflows.

Finally, the rapid pace of AI development has outstripped the establishment of standardized evaluation frameworks. Inconsistent performance metrics, variable reporting standards, and limited external benchmarking impede meaningful comparisons across studies and evidence-based integration into practice.^(10, 22) Addressing these methodological and structural challenges will be essential before AI tools can be responsibly incorporated into routine plastic surgery care.

Ethical and regulatory considerations

Ethical considerations associated with integrating AI into plastic surgery pose distinct challenges, largely due to the subjective nature of the aesthetic outcomes and fundamental role of patient expectations in surgical decision-making.^(16, 17, 23) The use of AI-generated analyses, recommendations, or visual simulations may introduce additional complexity regarding interpretation, transparency, and the appropriate degree of reliance on algorithmic outputs. These ethical challenges are not unique to plastic surgery, but their consequences may be amplified in this specialty due to the subjective valuation of the outcomes.

In aesthetic and reconstructive procedures, AI-generated predictions or simulations may influence patients' understanding of the anticipated outcomes. Although such tools may support preoperative discussions, there is a risk that the algorithmic outputs could be interpreted as definitive or guaranteed results rather than probabilistic estimates.^(4, 20, 24) Therefore, surgeons retain the responsibility for contextualizing AI-assisted information and clearly communicating its limitations within the shared decision-making processes.

Algorithmic bias remains a crucial ethical concern in plastic surgery applications, particularly in facial analysis systems that may exhibit variable performance across skin tones, facial morphologies, and culturally influenced aesthetic norms.^(9, 17, 23, 25) Bias arising from non-representative training datasets or unclear development processes may unintentionally exacerbate disparities in care. Transparent reporting of dataset composition, validation strategies, and known limitations is thus essential to promote fairness and responsible clinical use.^(16, 23, 25)

Informed consent represents another important ethical challenge. Patients may not fully understand when or how AI tools are incorporated into their care, particularly when these AI-generated recommendations or visualizations influence clinical discussions.^(15, 20, 24)

The ethical integration of AI requires explicit disclosure of its role, scope, and limitations, along with clear communication that such systems function solely as adjunctive tools and do not replace professional clinical judgment.

Moreover, the increasing use of generative AI in academic writing and surgical education has raised concerns related to authorship, originality, and disclosure.^(16, 18, 26) Empirical studies have demonstrated that AI-generated text is not consistently identifiable by either expert reviewers or automated detection tools, thereby highlighting the limitations in current oversight mechanisms.^(16, 18, 26) These findings emphasize the need for clear journal and institutional policies that address disclosure requirements, accountability, and the responsible use of AI in scientific publishing.

From a regulatory perspective, the current frameworks remain underdeveloped to encompass the specific challenges of AI-assisted surgical practice. In addition, questions regarding liability, documentation, and responsibility for AI-informed decisions remain incompletely defined.^(16, 22, 23, 26)

It is paramount to emphasize that AI systems function strictly as supportive decision-aid tools. The ultimate legal, professional, and ethical responsibility for clinical evaluation, diagnostic interpretation, surgical decisions, and overall patient management resides entirely with human healthcare professionals, including surgeons, physicians, and medical staff. Algorithmic outputs cannot delegate or diminish professional accountability. Until clearer regulatory guidance is established, plastic surgeons should be cautious when adopting AI technologies, prioritizing patient safety, transparency, and ethical integrity in clinical implementation.

Future direction

Future research in AI and plastic surgery should prioritize enhancing the quality of evidence supporting its clinical application. Emphasis should be placed on prospective and multicenter studies that assess AI tools in real-world clinical environments, where variability in patient populations, imaging conditions, and surgical workflows may influence performance.^(19, 27) Such investigations are critical for determining whether findings from retrospective or proof-of-concept studies translate into clinically meaningful and reproducible outcomes.

The development of larger, more diverse, and carefully curated datasets represents another crucial direction for future work. Ensuring adequate representation across age groups, ethnic backgrounds, and surgical indications may help mitigate algorithmic bias and improve generalizability.^(9, 17, 27) In addition, the adoption of standardized reporting frameworks and validation protocols would enhance the transparency, reproducibility, and meaningful comparison across AI studies in plastic surgery.^(10, 19, 27)

Furthermore, specialty-specific model development may improve the clinical relevance of AI systems in complex or underrepresented domains. Tailoring model training to subspecialties, such as hand surgery, microsurgery, and gender-affirming care, may enhance the contextual understanding and reduce the risk of inappropriate recommendations observed with more broadly trained models.^(15, 27)

Regulatory considerations will also play an important role in shaping the future of AI integration into plastic surgery practice. Clearer guidance on approval pathways, liability, documentation, and disclosure requirements will be necessary before the broader clinical adoption can be pursued responsibly.^(16, 22, 28) Emerging regulatory and ethical frameworks increasingly emphasize the principles of human oversight, accountability, and transparency in medical AI, which align closely with the values of surgical practice.^(27, 28) Central to these principles is the explicit recognition that healthcare professionals retain sole responsibility for patient care pathways, ensuring that technology enhances rather than replaces clinician accountability.

Furthermore, educational initiatives should evolve alongside technological development. Incorporating formal instruction on AI capabilities, limitations, and ethical implications into surgical training programs may promote informed and vital engagement with these tools.^(13, 18, 28) By fostering AI literacy rather than uncritical reliance, training programs can better prepare future surgeons to integrate AI responsibly into clinical and academic practice.

Overall, the most appropriate role of AI in plastic surgery is likely to remain that of a surgeon-directed adjunct. Continued collaboration among clinicians, researchers, educators, and regulatory bodies will be crucial to advancing AI technologies that prioritize patient safety, ethical integrity, and professional responsibility.^(16, 22, 27, 28) Without such coordinated efforts, premature or unregulated adoption of AI tools risks reinforcing bias, eroding clinician trust, and undermining patient confidence.

Conclusion

AI presents opportunities to augment preoperative planning, education, and patient communication in plastic surgery. However, current evidence supports its use only as an adjunct to surgeon expertise and clinical judgment. A cautious, evidence-based approach that emphasizes transparency, patient safety, and professional responsibility will be essential as AI continues to evolve within the specialty.

Author contributions

TNT solely conceived and designed the study, performed the literature synthesis, drafted and critically revised the manuscript, and approved the final version for publication. TNT takes full responsibility for the integrity of the work and its academic content.

Acknowledgments

None.

Conflicts of interest statement

TNT has completed the International Committee of Medical Journal Editors Form for Disclosure of Potential Conflicts of Interest and declares no conflicts of interest.

Data sharing statement

The present review is based on the references cited. All data generated or analyzed during the present study are included in this published article and the citations herein. Further details, opinions, and interpretation are available from the corresponding author on reasonable request.

ORCID

Tinh Nguyen Trung  <https://orcid.org/0009-0006-2692-9725>

References

1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* 2022;28:31–8.
2. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical AI. *Nature* 2023;616:259–65.
3. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med* 2022;28:1773–84.
4. Bayar Muluk N. Artificial intelligence in facial plastic procedures. *Clin Otolaryngol* 2026;51:1–6.
5. Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med* 2018;1:54.

6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
7. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42: 60–88.
8. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst* 2020;33: 1877–901.
9. Raji ID, Gebru T, Mitchell M, Buolamwini J, Lee J, Denton E. Saving face: investigating the ethical concerns of facial recognition auditing. *Proc AAAI ACM Conf AI Ethics Soc* 2020;145–51.
10. Spoer DL, Onyewu C, Armstrong MB, et al. A systematic review of artificial intelligence applications in plastic surgery: looking to the future. *Plast Reconstr Surg Glob Open* 2022;10:e4608.
11. Chen JH, Asch SM. Machine learning and prediction in medicine-beyond the peak of inflated expectations. *N Engl J Med* 2017;376:2507–9.
12. Rothchild E, Jung G, Ricci JA. Evaluating ChatGPT's ability to address frequently asked questions in gender-affirmation surgery. *J Homosex* 2026;73: 142–53.
13. Luce EA, Le P, Pumilia A, Bian J, McCarthy J, Tidwell T, et al. The validity, accuracy, and comprehensiveness of artificial intelligence for generating resident educational material and discussion guides. *Ann Plast Surg* 2026;96:83–6.
14. Fine KS, Zona EE, O'Shea AW, Via EC, Attaluri PK, Wirth PJ, et al. Evaluating the detection accuracy of AI-generated content in plastic surgery: a comparative study of medical professionals and AI tools. *Plast Reconstr Surg* 2026;157:129e-36e.
15. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. *arXiv:2303.13375v2* [Preprint]. 2023 [cited 2026 Mar 2]: [35 p.]. Available from: <https://arxiv.org/abs/2303.13375>
16. Thorp HH. ChatGPT is fun, but not an author. *Science* 2023;379:313.
17. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
18. Else H. Abstracts written by ChatGPT fool scientists. *Nature* 2023;613:423.
19. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
20. Floridi L, Cows J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach* 2018;28:689–707.
21. Savran A. Performance and reliability of state-of-the-art LLMs in complex hand surgery scenarios. *J Orthop Surg* 2026;34:10225536261416605.
22. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA* 2018;320: 2199–200.
23. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. *Camb Q Healthc Ethics* 2020;29:567–80.
24. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018;268:70–6.
25. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proc Mach Learn Res* 2018;81:77–91.
26. Stokel-Walker C. AI-authored papers are flooding scientific journals. *Nature* 2023;613:620–1.
27. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378.
28. Price WN, Cohen IG. Privacy in the age of medical big data. *Nat Med* 2019;25:37–43.