Beyond Autonomy

A Plastic Surgeon's Responsibility in the Face of AI-Driven Misinformation

Pranav Rajaram, BS,[a] Megan Lane, MD,[b] Nazanin Andalibi, PhD,[c] Oliver L. Haimson, PhD,[c] Rachel C. Hooper, MD,[b] and Hannes Prescher, MD[b]

A promising development in medical practice is a shift from a "doctor knows best" model toward patient autonomy rooted in shared decisions, transparent risk, and respect for values.[1,2] Increasing Internet access, social media communities, and more recently generative large language models (LLMs) such as ChatGPT (OpenAI, San Francisco, CA) may have helped push patient autonomy forward by providing near-instant access to clinical information from trusted sources, but can also relay harmful medical misinformation. LLMs' fluent style can mask factual hallucinations and transmit the same gaps that already trouble online medical content.[1] As a result, medical visits increasingly begin with fact-checking claims, shifting time away from substantive decision making toward correction. We argue that without deliberate safeguards for autonomy, increasing patient reliance on LLMs can pull clinical encounters back toward paternalism.

CASE STUDY

A 50-year-old woman with diabetes presented to the emergency department after an outpatient MRI suggested early osteomyelitis of a toe. Before arrival, her partner entered the radiology impression into an LLM, which recommended urgent surgical evaluation with possible toe amputation and an infectious diseases consultation. In the emergency department, she was hemodynamically stable with a small distal second toe wound and mild cellulitis. The surgical team advised that debridement could be scheduled when operating room resources were available, but the family pressed for the LLM's plan. A mildly elevated creatinine prompted an additional LLM recommendation for a nephrology consultation, which the internal medicine team explained was unnecessary because the abnormality could be managed without subspecialist input.

Over the next 48 hours, plastic surgery and inpatient medicine provided unified bedside counseling. Daily 30-minute conversations began with a concrete review of symptoms, examination findings, imaging, and laboratory trends, with plain language explanations of osteomyelitis progression. Teams shared notes and prepared brief written summaries to avoid mixed messages and to model collaborative decision making. By the third encounter, the couple agreed to scheduled debridement, which proceeded uneventfully without clinical deterioration. The episode required several additional clinician hours but also illustrates that consistent, evidence-linked counseling can recalibrate expectations. As trust developed, the couple's questions shifted from the timing of amputation to postoperative wound care and antibiotic duration, reflecting a transition from crisis-driven demands to constructive engagement with the treatment plan.

ACCURATE INFORMATION UNDERPINS PATIENT AUTONOMY

Beauchamp and Childress[2] define autonomy as the ability of patients to make voluntary decisions once they receive reliable data. That premise underlies informed consent and falters when common sources are unreliable. Surveys of plastic surgery content on YouTube, TikTok, and Instagram show that 40%–70% omit complication rates, realistic recovery timelines, or total costs.[3–9] LLMs that learn from such material can inherit these gaps, and their fluent prose may confer unwarranted authority. In a randomized evaluator-blinded study of public patient questions, lay raters judged ChatGPT responses as higher quality and more empathic than physicians' and noted substantially greater length (211 vs 52 words); however, the clinical accuracy was not independently verified.[10]

Unlike ranked web search, LLMs return a single answer without transparent sources. They synthesize content, blending accurate and fabricated details ("hallucinations"), and deliver information in a fluent, validating, and confident tone that promotes anchoring and over-trust. Because platform incentives favor retention over verification, accuracy is uncertain, and these dynamics can mislead patients and erode clinicians' epistemic authority.

Received September 12, 2025, and accepted for publication, after revision December 5, 2025.
From the [a]School of Medicine, [b]Section of Plastic and Reconstructive Surgery, and [c]School of Information, University of Michigan, Ann Arbor, MI.
Pranav Rajaram: 0000-0003-1822-4750
Conflicts of interest and sources of funding: none declared.
No funding was received at any point during the ideation or preparation of this manuscript.
Reprints: Pranav Rajaram, BS, School of Medicine, University of Michigan, 1135 Catherine Street, Ann Arbor, MI 48105. E-mail: rbpranav546@gmail.com.

ISSN: 1536-3708/25/0000–0000
DOI: 10.1097/SAP.0000000000004608

PROFESSIONAL STEWARDSHIP PROTECTS PATIENT AUTONOMY

Correcting LLM errors is not a return to paternalism. In light of historical abuses that prompted autonomy safeguards (such as Tuskegee and the subsequent Belmont report),[11] the priority is information stewardship: clinicians should appraise AI-derived claims, direct patients to credible sources, and document the rationale for accepted or rejected advice while preserving patient agency. Correction grounds decisions in case-specific evidence and considers system-level risks, akin to antibiotic stewardship, aligning individual care with public health. When clinicians withdraw from guidance, some patients report feeling abandoned to decisions they do not feel qualified to make.[12,13]

Patients consult LLMs for varied reasons, and preferences for shared decision making differ by context and population.[14] Access barriers, convenience, and a desire for anonymity also drive online information seeking, and trust in clinicians has fluctuated since the pandemic.[15] Clinically, this means pairing validation with guardrails: ask what prompted the AI chatbot query, elicit what matters most to the patient, use teach-back to correct inaccuracies, and co-create an evidence-based plan.[16,17] Stewardship also requires explicit professional responsibility. Regardless of how persuasive an AI recommendation appears, the surgeon remains the final safeguard and medical authority; clinicians should scrutinize external guidance, correct errors, and document the rationale for accepted or rejected advice.[18]

EXPANDING STEWARDSHIP THROUGH RETRIEVAL-AUGMENTED GENERATION (RAG)

Individual clinician counseling cannot match the scale of online misinformation. LLMs can amplify erroneous information through hallucinations, producing confident but inaccurate outputs.[19] Patients need evidence-linked education at scale that delivers timely, consistent answers with transparent sources outside of provider visits without shifting the vetting burden to clinicians.

Rather than general-purpose LLMs, professional societies and academic centers could pair domain-restricted systems with RAG grounded in peer-reviewed libraries.[20] In such workflows, the model cites retrieved documents, abstains or escalates when confidence is low, and allows direct inspection of sources. RAG has reduced, though not eliminated, hallucinations while preserving conversational flow.[20] These tools must be built with guardrails, including a narrow clinical scope, approved source gating, transparency when retrieval is weak, clear uncertainty language, human handoff pathways, and ongoing evaluation of accuracy, faithfulness to sources, patient comprehension, and clinician time saved. Framed as information stewardship at scale, a responsible RAG approach meets patients where they seek answers while giving clinicians a verifiable, low-friction way to authenticate advice.[19,20]

WHY PLASTIC SURGERY IS A GOOD STARTING POINT

Plastic and reconstructive surgery (PRS) is a sensible first setting to test source-gated RAG. PRS has standardized care pathways, heavy exposure to image-driven misinformation, and frequent preference-sensitive choices, which together create a narrow but high-risk information domain. The same approach would fit other preference-sensitive specialties. Because LLMs are input-driven and cannot fill in missing clinical details on their own, even domain-restricted systems can mislead when inputs are incomplete or out of scope. Pilot programs could test efficacy by using evidence-linked RAG across the care continuum, show verifiable sources, adapt to literacy and language needs, and abstain or escalate to clinicians when confidence or evidence is low.

CLINICAL IMPORTANCE

Information stewardship is patient care. Clear, evidence-linked explanations lower anxiety and support informed, values-concordant choices.[21,22] For clinicians, this is high-value work that should be taught, practiced, and documented on par with informed consent so that, with institutional support and scalable RAG tools, patient autonomy is grounded in trustworthy information rather than fluent but erroneous LLM outputs. In complex encounters, ethics consultation can slow the discussion, align decisions with shared values, and has been associated with less lingering clinician moral distress and stronger team cohesion.[23,24] Integrating these reflective touchpoints into perioperative workflows helps ensure that redirecting patients from harmful advice is experienced as accompaniment rather than gatekeeping.

RESEARCH DIRECTIONS

Future work should scrutinize retrieval-augmented generation (RAG) rigorously. Early studies suggest that RAG curbs hallucinations by grounding answers in source documents, but whether it preserves essential safety details and avoids promotional bias remains uncertain.[20,25] A recent review highlights the lack of harmonized evaluation frameworks for health care RAG and calls for uniform metrics for assessing generated outputs.[26] Building on this, linguistic analyses should test for omission of high-risk warnings, promotional bias, response variability, and sociodemographic or linguistic skew that alters recommendations across otherwise identical scenarios.[27] Parallel eye-tracking studies could identify which portions of RAG-generated materials attract attention and link gaze patterns to comprehension and subsequent utilization.[28]

Despite progress on accuracy, RAG adoption remains unknown. Will clinicians and patients actually use such a tool in their routine workflow, and if so, how do we implement so it becomes as habitual as Google or ChatGPT? One pragmatic path is to use the Consolidated Framework for Implementation Research (CFIR) to map local barriers and facilitators, co-design with end users, and embed the tool at existing touchpoints such as previsit portals and consent modules with iterative refinement until it is easy, useful, and safe.[29]

NEXT STEPS

The next steps in integrating LLMs into the medical workflow for both patients and physicians rest on supervised integration rather than resistance or deflection. Institutions can pilot RAG LLMs within secure portals, offer residency workshops on counseling strategies for AI-primed patients, and recognize information stewardship efforts in care quality metrics. By placing digital health literacy in the forefront, creating an implementation structure, and building a shared agenda, the field of plastic surgery can model how to safeguard autonomy without surrendering professional responsibility.

Patient Consent

This case report has been fully anonymized, and no identifiable information is included. Although formal consent was not required, the patient provided permission for inclusion in the manuscript.

REFERENCES

1. Alber DA, Yang Z, Alyakin A, et al. Medical large language models are vulnerable to data-poisoning attacks. *Nat Med*. 2025;31:618–626.
2. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 7th ed. New York: Oxford University Press; 2013:459.
3. Maldonado J D, Arriagada I C, Conejero R A, et al. Social media and plastic surgery, the good, the bad, and where are we going? *Aesthetic Plast Surg*. 2025;49:5224–5236.
4. Montemurro P, Porcnik A, Hedén P, et al. The influence of social media and easily accessible online information on the aesthetic plastic surgery practice: literature review and our own experience. *Aesthetic Plast Surg*. 2015;39:270–277.
5. Nogueira R, Eguchi M, Kasmirski J, et al. Machine learning, deep learning, artificial intelligence and aesthetic plastic surgery: a qualitative systematic review. *Aesthetic Plast Surg*. 2025;49:389–399.
6. Samur Erguven S, Topsakal KG, Aksoy M. YouTube™ as a source of parents' information for craniosynostosis. *Orthod Craniofac Res*. 2024;27(Suppl 1):141–149.

7. Morena N, Ben-Zvi L, Hayman V, et al. How reliable are post-mastectomy breast reconstruction videos on YouTube? *JCO*. 2023;41(16_suppl):11021.
8. Kwak D, Park JW, Won Y, et al. Quality and reliability evaluation of online videos on carpal tunnel syndrome: a YouTube video-based study. *BMJ Open*. 2022; 12:e059239.
9. Gray MC, Gemmiti A, Ata A, et al. Can you trust what you watch? An assessment of the quality of information in aesthetic surgery videos on YouTube. *Plast Reconstr Surg*. 2020;145:329e–336e.
10. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183:589–596.
11. Adashi EY, Walters LB, Menikoff JA. The Belmont report at 40: reckoning with time. *Am J Public Health*. 2018;108:1345–1348.
12. Emanuel EJ. Four models of the physician-patient relationship. *JAMA*. 1992; 267:2221.
13. Pilnick A. Reconsidering patient-centred care: authority, expertise and abandonment. *Health Expect*. 2023;26:1785–1788.
14. Chewning B, Bylund CL, Shah B, et al. Patient preferences for shared decisions: a systematic review. *Patient Educ Couns*. 2012;86:9–18.
15. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. *Healthcare (Basel)*. 2021;9:1740.
16. Montori VM, Ruissen MM, Hargraves IG, et al. Shared decision-making as a method of care. *BMJ Evid Based Med*. 2023;28:213–217.
17. Elwyn G, Frosch D, Thomson R, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med*. 2012;27:1361–1367.
18. Naik N, Hameed BMZ, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front Surg*. 2022;9:862322.
19. Özer M. Is artificial intelligence hallucinating? *Turk Psikiyatri Derg*. 2024;35: 333–335.
20. Ozmen BB, Mathur P. Evidence-based artificial intelligence: implementing retrieval-augmented generation models to enhance clinical decision support in plastic surgery. *J Plast Reconstr Aesthet Surg*. 2025;104:414–416.
21. Ng SX, Wang W, Shen Q, et al. The effectiveness of preoperative education interventions on improving perioperative outcomes of adult patients undergoing cardiac surgery: a systematic review and meta-analysis. *Eur J Cardiovasc Nurs*. 2022;21:521–536.
22. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2017;4:CD001431.
23. Amos V, Whitehead P, Epstein B. Moral distress consultation services: insights from consultants. *HEC Forum*. 2025;37:217–233.
24. Austin W. What is the role of ethics consultation in the moral habitability of health care environments? *AMA J Ethics*. 2017;19:595–600.
25. Zhang G, Xu Z, Jin Q, et al. Leveraging long context in retrieval augmented language models for medical question answering. *NPJ Digit Med*. 2025;8:239.
26. Amugongo LM, Mascheroni P, Brooks S, et al. Retrieval augmented generation for large language models in healthcare: a systematic review. *PLOS Digit Health*. 2025;4:e0000877.
27. Omar M, Soffer S, Agbareia R, et al. Sociodemographic biases in medical decision making by large language models. *Nat Med*. 2025;31:1873–1881.
28. Tahri Sqalli M, Aslonov B, Gafurov M, et al. Eye tracking technology in medical practice: a perspective on its diverse applications. *Front Med Technol*. 2023;5:1253001.
29. Damschroder LJ, Aron DC, Keith RE, et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*. 2009;4:50.