

EDITORIAL TECHNOLOGY AND INNOVATION

Large language models and cognitive debt in plastic surgery practice

Jennifer Novo BSc, MD  0009-0005-3919-904X,^{1,a} Ishith Seth BSc, MD, PhD  0000-0001-5444-8925,^{2,3} Omar Shadid BSc, MD, PhD  0009-0005-3471-8656,² Yi Xie MD  0000-0002-9009-5837,² Richard J Ross MBBS, PhD, FRACS (Plast)  0009-0008-1068-3147,^{2,3} Warren M Rozen MBBS (Hons), PhD, FRACS (Plast)  0000-0002-4092-182X²⁻⁴

¹ School of Medicine, University of Notre Dame Sydney, Darlinghurst, New South Wales, AUSTRALIA. ² Department of Plastic and Reconstructive Surgery, Peninsula Health, Frankston, Victoria, AUSTRALIA. ³ Peninsula Clinical School, Monash University, Frankston, Victoria, AUSTRALIA. ⁴ Section Editor (Technology and Imaging), Australasian Journal of Plastic Surgery, St Leonards, New South Wales, AUSTRALIA.

Keywords: artificial intelligence, large language models, clinical decision-making, professionalism

Submitted: March 4, 2026 AEST Accepted: June 4, 2026 AEST

DOI: 10.34239/ajops.163047

Australasian Journal of Plastic Surgery

Vol 9, Issue 2, 2026

Large language models (LLMs) are increasingly being integrated into routine clinical workflows from operating theatres to hospital wards. Surgeons and other clinicians are now using tools such as ChatGPT not only to transcribe notes but to draft discharge summaries and operation reports, summarise case notes and even generate literature reviews in seconds. Trained on vast corpora of text, LLMs generate rapid and detailed responses that offer clear efficiency advantages in time-constrained environments. While reliance on technology to reduce workload is not new, the conversational and context-aware nature of LLMs allows them to generate outputs that appear complete and reasoned. Yet there is an unspoken risk: when we permit algorithms to speak on our behalf, are we also granting them the ability to think for us?

In plastic surgery, this risk may be subtle because the output can seem clinically sound before the underlying reasoning has been independently constructed. A registrar can use an LLM to draft a ward-round plan, operative report or follow-up letter that appears coherent and complete, and a consultant can sign it off quickly. Yet the cognitive work that normally builds judgement and detects case-to-case variation may not have occurred.

The real risk is not that LLMs write well, but that they may tempt us to stop thinking things through and start approving what ‘seems reasonable’. This shift, toward reduced scrutiny, aligns with established concerns about automation bias and unrecognised skill decay that occur when technological assistance is routinely trusted.^{1,2}

A recent research paper by Kosmyrna and colleagues at the Massachusetts Institute of Technology, alongside other electroencephalographic work, suggests that this risk might not be just a hypothetical concern.^{3,4} In a controlled experiment, participants who relied on ChatGPT to write essays showed reduced neural engagement and poorer memory for what they produced.³ The authors call this ‘cognitive debt’, a mental quieting that can persist even after the machine is turned off.³ This is mechanistic, preprint-level evidence, so should be read as a warning rather than proof of harm.³ Nonetheless, the signal warrants attention. When the LLM generates coherent output, the clinician’s role can shift from constructing the reasoning to editing it, and people may later struggle to explain or reproduce what was actually ‘theirs’.³ In a busy plastic surgery practice, that kind of drift could easily go unnoticed.

^a Corresponding author: Jennifer Novo BSc, MD; jenny.novo@gmail.com

Plastic surgery is particularly exposed to this risk because much of its decision-making occurs in the surgeon's mind in real time, not on a checklist. Plastic surgeons are trained to picture the reconstruction several steps ahead and to keep backup plans ready as the case unfolds. If venous outflow looks poor, they revise the anastomosis. If perfusion is borderline, they redesign or stage the reconstruction. An LLM can describe these steps neatly on the page, but it cannot supply the lived judgement that decides which step to take at the operating table.

The potential for cognitive offloading is pronounced in daily plastic surgery practice. Operation reports, consent summaries and perioperative notes can now be generated almost instantly from bullet points or voice prompts. When such practices become routine, surgeons risk displacing their own recall and clinical judgement, reviewing decisions they did not actively formulate and bypassing the reflection and variance detection, which Crebbin and colleagues identify as hallmarks of surgical expertise decision-making.⁵ What begins as documentation support risks evolving into cognitive substitution.

In plastic surgery, where small technical details can have significant downstream consequences, linguistic fluency can obscure clinically relevant omissions, including laterality, implant type and size, ischaemia time, recipient vessels, drain plans, anticoagulation instructions or the specific

intraoperative issue that should trigger early review.⁶ The danger is not that the LLM makes a mistake, but that a plausible narrative reduces the impulse to question what is missing or uncertain.²

The risk is amplified in training. This is because performance can improve while learning stalls. If an LLM generates the plan's structure, the explanation or the differential, the trainee's role shifts from construction to selection and editing. This is consistent with the concern that artificial intelligence (AI) assistance may hinder skill development or accelerate skill decay without the LLM user's awareness.² Surgical education already flags over-reliance on AI as a threat to independent decision-making. This matters most when the case deviates from expectations, which is precisely when judgement is required.⁷

Even among experienced surgeons, the 'illusion of oversight' is persuasive because the narrative sounds clinically plausible. Unlike dictation or templates that largely reflect the surgeon's own thinking, LLMs can introduce confident detail that was never verified. Existing studies on generative AI in plastic surgery have already highlighted ethical and safety concerns, including the risk that AI recommendations can lead surgeons to overlook their own clinical judgement.⁶ Over-reliance on LLMs is also linked with uncritical acceptance of incorrect or hallucinated outputs, which has implications for clinical documentation and patient-facing information.⁸

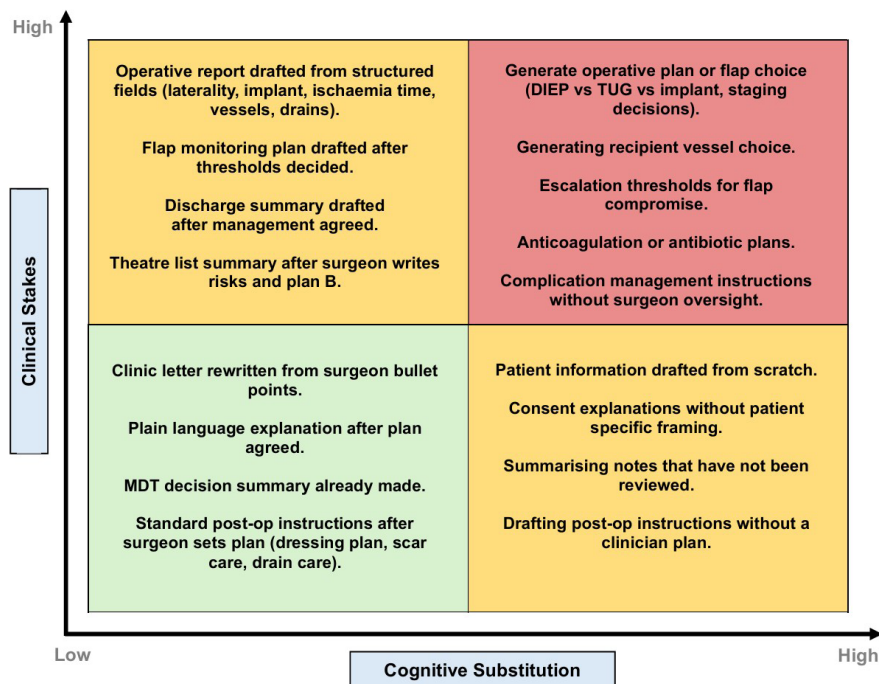


Fig 1. Cognitive debt risk map for large language model (LLM) use in plastic surgery. The map stratifies common examples of LLM use by clinical stakes and degree of cognitive substitution. The examples illustrate where LLM use primarily improves expression and organisation versus where it risks replacing cognitive work that underpins safe plastic surgery judgement.

The solution is not to avoid LLMs but to use them deliberately, in ways that protect the cognitive steps that matter (**Figure 1**). A simple rule is that the surgeon (or trainee) must first write the indication and operative plan in their own words, identify the key risks and state a realistic Plan B. Only then should the model be used to improve clarity, structure or patient-appropriate language. This preserves the cognitive ‘reps’ that underpin safe judgement and variance detection, rather than outsourcing them to a fluent draft.^{2,5}

Used well, LLMs can reduce administrative overhead and improve communication. Their best role in plastic surgery is to polish and organise what the clinician had already decided, not generate the reasoning that the clinician then merely signs off on.⁶ If the model can produce a plan faster than a trainee can explain it, that should be treated as a prompt to rebuild the mental model, not a reason to move on.² The cognitive process that safeguards patients cannot be delegated. In plastic surgery, passive fluency is never sufficient. AI must support, not replace, independent surgical judgement.⁶

Conflict of interest

The authors have no conflicts of interest to disclose.

Funding declaration

The authors received no financial support for the research, authorship and/or publication of this article.

References

- 1 Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;19(1):121–127. <https://doi.org/10.1136/amiajnl-2011-000089> PMID:21685142 PMCID:PMC3240751
- 2 Macnamara BN, Berber I, Çavuşoğlu MC, et al. Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cogn Res Princ Implic*. 2024;9(1):46. <https://doi.org/10.1186/s41235-024-00572-8> PMID:38992285 PMCID:PMC11239631
- 3 Kosmyna N, Hauptmann E, Yuan YT, et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv*. 2025;2506.08872. <https://doi.org/10.48550/arXiv:2506.08872>
- 4 Jiang T, Wu J, Leung SCH. The cognitive impacts of large language model interactions on problem solving and decision making using EEG analysis. *Front Comput Neurosci*. 2025;19:1556483. <https://doi.org/10.3389/fncom.2025.1556483> PMID:40741073 PMCID:PMC12307350
- 5 Crebbin W, Beasley SW, Watters DA. Clinical decision making: how surgeons do it. *ANZ J Surg*. 2013;83(6):422–428. <https://doi.org/10.1111/ans.12180> PMID:23638720
- 6 Dhawan R, Brooks KD, Shauly O, Shay D, Losken A. Ethical considerations for generative artificial intelligence in plastic surgery. *Plast Reconstr Surg Glob Open*. 2025;13(6):e6825. <https://doi.org/10.1097/GOX.0000000000006825> PMID:40469554 PMCID:PMC12133144
- 7 Varas J, Coronel BV, Villagrán I, et al. Innovations in surgical training: exploring the role of artificial intelligence and large language models (LLM). *Rev Col Bras Cir*. 2023;50:e20233605. <https://doi.org/10.1590/0100-6991e-20233605-en> PMID:37646729 PMCID:PMC10508667
- 8 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;11:28. <https://doi.org/10.1186/s40561-024-00316-7>



CC-BY-4. This is an open access article distributed under the Creative Commons Attribution 4.0 International License (CC-BY-4) which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.