Invited Commentary

AI IN SURGERY

Intelligent, Human-Centric Delivery Is Needed to Maximize AI

Carla M. Pugh, MD, PhD; Mitra Ann Alikhani, BSN, RN; Adam Rodman, MD, MPH

Artificial intelligence (AI)–augmented learning is here, and many believe it is superior to all past endeavors. AI's ability to deliver real-time, objective, standardized feedback is an unprecedented game changer. Giglio et al[1] investigated whether AI feedback accelerates surgical skill acquisition using a virtual reality simulator. Medical student participants (N = 87) were stratified into 3 groups: (1) computer-generated verbal feedback based on AI performance assessment; (2) human verbal feedback using an AI script identical to that of group 1; or (3) verbal feedback based on human interpretation and personalization of the AI performance assessment. The article emphasizes that group 3 outperformed the others; however, group 2 also outperformed group 1 in trial 5 and the final realistic task.

Related article

Beyond skill acquisition, student engagement improved by replacing synthetic, computer-generated speech (group 1) with a human voice (group 2), despite the words being identical. These results underscore that the method of interaction with an AI system is as important as the accuracy and automated performance of the system itself. Other studies, both in clinical diagnosis and radiological interpretation, show that AI outputs alone do not necessarily enhance human performance.[2,3] Put simply, delivery matters.

An area rich for further research is the natural depth and nuances of human feedback—timely eye contact mixed with verbal praise, varied tone, and content emphasis. This was not fully explored in the study by Giglio and colleagues, highlighting the need for mixed-methods research that captures instructors' tone and verbiage. Large language model systems are commonly optimized for positive engagement and sycophancy. Early evidence suggests that these factors decrease cognitive load and critical thinking in students and should be carefully considered as AI is integrated into medical curricula.[4] The concept of desirable difficulty has been identified as a key driver of durable learning.[5]

A quiet success of the study is its objective novice-to-expert scaling. Anchoring forces, path lengths, and speeds to expert benchmarks allow educators to ask, "How close to competent?" rather than merely "How fast?" The race for precision education requires objective, normative data; without this, both humans and models dispense subjective, qualitative feedback that risks perpetuating individual biases.[6] While AI cannot yet provide mentorship, emotional framing, or translation to the operating theater, standardized, real-time analytics cannot be delivered at scale by humans. Optimizing algorithms while ignoring point-of-care implementation yields only marginal gains. AI can facilitate learning, but cultivating human expertise still demands a human-centered interface—one that preserves friction, nuance, and the motivation to improve.

ARTICLE INFORMATION

Author Affiliations: Department of Surgery, Stanford University, Stanford, California (Pugh); Stanford University School of Medicine, Stanford, California (Alikhani); Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts (Rodman); Harvard Medical School, Boston, Massachusetts (Rodman).

Corresponding Author: Carla M. Pugh, MD, PhD, Department of Surgery, Stanford University, 300 Pasteur Dr, Grant, S-067, Stanford, CA 94305 (cpugh@stanford.edu).

Published Online: August 6, 2025. doi:10.1001/jamasurg.2025.2550

Conflict of Interest Disclosures: Dr Pugh reported serving on the product development advisory board for Johnson and Johnson. Dr Rodman reported personal fees from Google as a visiting researcher and grants from the Macy Foundation. No other disclosures were reported.

REFERENCES

1. Giglio B, Albeloushi A, Alhaj AK, et al. Artificial intelligence–augmented human instruction and surgical simulation performance: a randomized clinical trial. *JAMA Surg*. Published online August 6, 2025. doi:10.1001/jamasurg.2025.2564

2. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open*. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969

3. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. *Nat Med*. 2024;30(3):837-849. doi:10.1038/s41591-024-02850-w

4. Stadler M, Bannert M, Sailer M. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Comput Human Behav*. 2024;160:108386. doi:10.1016/j.chb.2024.108386

5. Nelson A, Eliasz KL. Desirable difficulty: theory and application of intentionally challenging learning. *Med Educ*. 2023;57(2):123-130. doi:10.1111/medu.14916

6. Desai SV, Burk-Rafel J, Lomis KD, et al. Precision education: the future of lifelong learning in medicine. *Acad Med*. 2024;99(4S)(suppl 1):S14-S20. doi:10.1097/ACM.0000000000005601