

Evaluation and mitigation of the limitations of large language models in clinical decision-making.

September 11, 2024

Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med. 2024;30(9):2613-2622. doi:10.1038/s41591-024-03097-1.

<https://psnet.ahrq.gov/issue/evaluation-and-mitigation-limitations-large-language-models-clinical-decision-making>

Researchers, clinicians, and other stakeholders are [hopeful](#) that integration of artificial intelligence and large language models (LLMs) can improve patient safety and reduce clinician [burden](#). This study used 2,400 real patient cases to test several LLM's ability to correctly diagnose common abdominal complaints. Each LLM performed significantly [worse](#) than physicians, did not follow treatment or diagnostic guidelines, could not interpret laboratory results, and often failed to follow instructions.