Children's Hospital of Philadelphia PolicyLab

The complexity paradox of generalist AI agents in medicine: exploring GPT-5 for real-world multimodal medical diagnosis

Visit Article

The next generation of medical AI systems that integrate language understanding, web data retrieval, and multi-step reasoning are demonstrating increasing promise as potential diagnostic aids in technical evaluations. However, their real world diagnostic performance, explainability, and reliability remain unclear. We conducted a systematic quantitative analysis of ten GPT-5-based configurations on 161 clinician validated real world cases. Each model was assessed for diagnostic accuracy, reasoning quality, retrieval accuracy, evidence integration, and robustness to low-quality inputs. We found important limitations in these AI systems, including image misinterpretation, inconsistent web search that sometimes degraded performance, and limited transparency in agentic workflows. Yet the models also showed strengths, including effective use of patient history, clear stepwise reasoning, plausible differentials, improved specificity with explicit reasoning, and relevant evidence retrieval. To support clinical deployment and future model development, we recommend prioritizing reasoning-optimized AI models, improving information retrieval mechanisms for reflection and error correction, strengthening visual-textual integration, and establishing practical standards for image quality and clarity. Overall, reliable real world multimodal diagnostic performance depends more on transparent reasoning and high-quality inputs than on aggressive retrieval or complex agentic workflows. Persistent gaps in evidence integration and visual reasoning remain major barriers to safe routine clinical use.

Journal:

npj Digital Medicine

Authors: Shaohui Zhang, Jiarong Qian, Kai Zhang, Zhiling Yan, Michael A. Catalano, Adam Kiridly, Omar Toubat, Gregory Tasian, Quanzheng Li, Yonghui Wu, Lei Xing, Wei Liu, Hua Xu, Lifang He, Xiang Li, Zhiyong Lu, Lichao Sun, Jing Huang