

Science study: OpenAI's o1 matched or beat physicians at diagnosis on real ER cases

Source copy kept by wheresthe.ai for the story "Science study: OpenAI's o1 matched or beat physicians at diagnosis on real ER cases" (<https://wheresthe.ai/d/us-science-o1-ed-diagnosis-study-2026-04/>).

Source: <https://www.newswise.com/articles/landmark-test-of-clinical-reasoning-finds-ai-outperformed>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 13705bcd398bb991ab0239de9e0e5533b7ee02d8bcb0110ae02d4d778427184f

Landmark Test of Clinical Reasoning Finds AI Outperformed | Newswise

Landmark Test of Clinical Reasoning Finds AI Outperformed

AI can pass the hardest exams medical school has to offer. But can it handle the real world's inherent messiness? Harvard Medical School and Beth Israel Deaconess Medical Center researchers sought to find out.

30-Apr-2026 at 2:00 PM EDT , by Harvard Medical School



Newswise — BOSTON - In one of the largest studies to compare artificial intelligence and physicians on a wide array of clinical reasoning tasks including real emergency department data, a team of physicians and computer scientists at Harvard Medical School and Beth Israel Deaconess Medical Center evaluated whether an AI system could do what physicians do every day: review a messy patient chart and use that information to determine diagnosis and next steps. In a new study published April 30, 2026 in Science, co-senior authors Arjun (Raj) Manrai, assistant professor of biomedical informatics at HMS and Adam Rodman, MD, MPH, a hospitalist and clinical researcher at BIDMC and team report that a large language model (LLM) outperformed physicians across many common clinical reasoning tasks including emergency room decisions, identifying likely diagnoses, and choosing next steps in management. The LLM's performance indicated that long-standing ways of testing medical AI may no longer capture current systems' performance, pointing to a possible turning point for the field. "We tested the AI model against virtually every benchmark, and it eclipsed both prior models and our physician baselines," said co-senior author Manrai. "However, this does not mean AI will necessarily improve care—how and where it should be deployed remain understudied, and we desperately need rigorous prospective trials to evaluate the impact of AI on clinical practice." "Models are increasingly capable," said Peter Brodeur, MD, MA, the study's co-first author. "We used to evaluate models with multiple-choice tests; now they are consistently scoring close to 100 percent and we can't track progress anymore because we're already at the ceiling." Incorporating standards first created in the 1950s to train and evaluate doctors, the researchers compared how an AI system performed against hundreds of clinicians. The comparisons included case study diagnostic challenges, reasoning exercises, and real emergency department cases. In one of their experiments, the investigators tasked the LLM with evaluating patients at various points in a standard emergency department setting, ranging from early triage to later admission decisions. At each stage, the model was given only the information available at that point — drawn directly from real-world electronic health records — and asked to generate likely diagnoses and suggest what should happen next. "To better understand real-world performance, we needed to test performance early in the patient course, when clinical data is sparse," said co-first author Thomas Buckley, Harvard Kenneth C. Griffin School of Arts and Science doctoral student, Dunleavy Fellow in HMS' AI in Medicine PhD program, and a member of Manrai's lab. Unlike many prior studies, the team did not smooth out the messiness of real-world care before testing the AI. The emergency department cases were presented exactly as they appeared in the electronic health record. "We didn't pre-process the data at all," Rodman said. "The model is literally just processing data as it exists in the health record." At the early decision points in the real-world emergency department cases, the model matched or exceeded attending physicians in diagnostic accuracy. That result surprised even the researchers. "I thought it was going to be a fun experiment but that it wouldn't work that well," Rodman said. "That was not at all what happened." The results make the case that medical AI is ready to be studied the same way as all new medical interventions — through carefully controlled clinical trials in real care settings. The researchers are clear that their results do not suggest that AI systems are ready to practice medicine autonomously, or that physicians can be removed from the diagnostic process. "A model might get the top diagnosis right but also suggest unnecessary testing that could expose a patient to harm," said Brodeur. "Humans should be the ultimate baseline when it comes to evaluating performance and safety." About Harvard Medical School Harvard Medical

School brings together the brightest minds in science and medicine to improve health and well-being for all. The school and its affiliated hospitals and research institutions are home to 12,000 faculty members and 1,600 medical and graduate students. Together, they function as a magnet, pulling together the best and most passionate researchers, clinicians, students, and changemakers in science, medicine, and health. About Beth Israel Deaconess Medical Center Beth Israel Deaconess Medical Center is a leading academic medical center, where extraordinary care is supported by high-quality education and research. BIDMC is a teaching affiliate of Harvard Medical School and consistently ranks as a national leader among independent hospitals in National Institutes of Health funding. BIDMC is the official hospital of the Boston Red Sox. Beth Israel Deaconess Medical Center is a part of Beth Israel Lahey Health, a healthcare system that brings together academic medical centers and teaching hospitals, community and specialty hospitals, more than 4,700 physicians and 39,000 employees in a shared mission to expand access to great care and advance the science and practice of medicine through groundbreaking research and education.

[Request an Expert](#)

[MEDIA CONTACT](#)

[Register for reporter access to contact details](#)

[CITATIONS](#)

[Science](#)

[TYPE OF ARTICLE](#)

[Research Results](#)

[SECTION](#)

[MEDICINE](#)

[CHANNELS](#)

[All Journal News](#) [Artificial Intelligence](#) [Clinical Trials](#) [Emergency Medicine](#) [Healthcare Technology](#) [Stories Getting News Coverage](#)

[KEYWORDS](#)

[LLM](#) [Artificial Intelligence \(AI\)](#) [AI Electronic Health Records](#)

[COMMENTS | COMMENTING POLICY](#)

×

[Connecting Research and Experts with Journalists](#)

[Unlock Your Access to Newswise Research News including Embargoed News and Expert Pitches](#)

[9 AM Daily Wire \(click to select custom wire\)](#) [0:00 Eastern Standard Time](#) [1:00 Eastern Standard Time](#) [2:00 Eastern Standard Time](#) [3:00 Eastern Standard Time](#) [4:00 Eastern Standard Time](#) [5:00 Eastern Standard Time](#) [6:00 Eastern Standard Time](#) [7:00 Eastern Standard Time](#) [8:00 Eastern Standard Time](#) [9:00 Eastern Standard Time](#) [10:00 Eastern Standard Time](#) [11:00 Eastern Standard Time](#) [12:00 Eastern Standard Time](#) [13:00 Eastern Standard Time](#) [14:00 Eastern Standard Time](#) [15:00 Eastern Standard Time](#) [16:00 Eastern Standard Time](#) [17:00 Eastern Standard Time](#) [18:00 Eastern Standard Time](#) [19:00 Eastern Standard Time](#) [20:00 Eastern Standard Time](#) [21:00 Eastern Standard Time](#) [22:00 Eastern Standard Time](#) [23:00 Eastern Standard Time](#)

Name:

Email:

Media Outlet:

[Continue Reading](#) Used only to deliver research news. Unsubscribe anytime.

[View All Latest News](#)

o

[Login](#)

[News](#)

[Latest News](#) [Research News](#) [Releases](#) [Research Alert](#) [News with Video/Audio](#) [Browse More Sections](#)

[Medicine](#) [Science](#) [Life](#) [Business](#) [Expert Pitches](#)

[Channels](#)

[Medicine](#)

[Biotech](#) [Cancer](#) [COVID](#) [Clinical Trials](#) [Diabetes](#) [Genetics](#) [Infectious Disease](#) [Neuro](#) [Obesity](#) [Women's Health](#) [View all Medicine Channels](#)

[Science](#)

[Chemistry](#) [Climate Science](#) [Dinosaurs](#) [DOE Science News](#) [Engineering](#) [Materials Science](#) [Physics](#) [Space](#) [View all Science Channels](#)

[Life](#)

[Behavioral Science](#) [Education](#) [Government](#) [History](#) [Politics](#) [Social Media](#) [View all Life Channels](#)

[Business](#)

[Economics](#) [Wall Street](#) [In the Workplace](#) [Women in Business](#) [View all Business Channels](#)

[Journal News](#)

[Cell \(journal\)](#) [JAMA](#) [Journal of Experimental Medicine](#) [Nature \(journal\)](#) [NEJM](#) [Neurology \(journal\)](#) [PLoS](#) [View all Journal Channels](#)

[Trends and Top Stories](#)

[Staff Picks](#) [Top Clipped Stories](#) [Top Hit Stories](#)

[By Location](#)

[Afghanistan News](#) [African News](#) [China News](#) [Cuba News](#) [Europe News](#) [Germany News](#) [Gulf of Mexico](#) [India/Pakistan News](#) [Iraq News](#) [View all by Location Channels](#)

[Meetings, Grants, and Events](#)

[Grant Funded News](#) [Medical Meetings](#) [Newswise Live - Event in Progress](#) [Newswise Live - Expert Spotlight](#) [Newswise Live - Events](#) [Scientific Meetings](#) [View all Meeting, Grants & Events Channels](#)

[RSS](#)

[Experts](#)

[Expert Pitch](#) [Expert Query](#) [Expert Directory](#)

[Journalists](#)

[Pricing](#)

[About](#)

[Member Services](#) [Accessibility Statement](#) [Newswise Live Invoice Lookup](#) [Services for Journalists](#) [Archived Wires](#) [Participating Institutions](#) [Media Subscribers](#) [Sample Effectiveness Reports](#) [Terms of Service](#) [Privacy Policy](#) [Our Staff](#) [Contact Newswise](#)

[Blog](#) [FAQ](#) [Help](#)

0.36327