

# Google AI Performs at 'Expert' Level on U.S. Medical Licensing Exam

— Latest version of medically tuned AI model achieved 85% accuracy, beating previous record

by [Michael DePeau-Wilson](#), Enterprise & Investigative Writer, MedPage Today

March 14, 2023



Recent accuracy scores from AI models taking a USMLE-style practice test. An earlier version of Med-PaLM also achieved 60%, the first ever passing score on the test, in December. Credit: Google Health AI

Google's medically focused generative artificial intelligence (AI) model achieved 85% accuracy on a U.S. Medical Licensing Examination (USMLE) practice test, the highest score ever recorded by an AI model, according to preliminary results shared by [the Google Health AI team](#).

The AI model, known as Med-PaLM 2 (Pathways Language Model), consistently performed at an "expert" physician level on the sample of USMLE-style practice questions, reported Alan Karthikesalingam, MD, PhD, a surgeon-scientist who leads the healthcare machine learning research group at Google Health in London, and co-authors.

Med-PaLM 2 answered both multiple choice and open-ended questions, provided written explanations for its answers, and evaluated its own responses. This result marks a notable improvement on previous AI models' attempts to reach near human accuracy and efficiency on a USMLE practice test, a benchmark that has been "a grand challenge" for this rapidly advancing technology, according to Karthikesalingam.

"If you look through the history of medicine, there have always been useful new tools that give clinicians what seemed like superpowers at the time," Karthikesalingam told *MedPage Today*.

"If AI can give caregivers back the gift of time, if AI can enable doctors and other caregivers to spend more time with their patients and bring time and humanity to medicine, and if it can increase accessibility and availability for people, that's our goal," he added.

The first version of Med-PaLM became the first AI model to achieve a passing score ( $\geq 60\%$  accuracy) on a similar USMLE-style practice test. Both versions of Med-PaLM were built by the Google Health AI team using large language models (LLMs) that were fine-tuned with increased capability and focus on medical information.

The preliminary results for Med-PaLM 2 were shared during an annual event today.

Vivek Natarajan, MCS, a research scientist at Google Health AI, suggested that the success behind this model comes from the technology advances available to the researchers, but also the specific medical expertise that helped the team determine exactly how to train the AI models.

"These models really learn quickly about the nuances of the safety of the medical domain and align itself very quickly," Natarajan told *MedPage Today*. "It's a combination of the very strong LLMs that we have at Google, the deep domain expertise ... as well as the pioneering techniques."

Despite the high marks for accuracy, the researchers noted that Med-PaLM 2, which was tested using 14 different criteria such as scientific accuracy and reasoning, still had significant limitations.

Medical News from Around the Web

---

[KAISER HEALTH NEWS](#)

Black Women Have Much at Stake in States Where Abortion Access May Vanish

[BBC](#)

Brain cancer research plea made by family of Ollie Gardiner

[CBS NEWS](#)

DNA identifies remains of Florida teen who disappeared in 1972. Detectives now say she may have been killed by a serial killer cop.

"These systems are certainly not perfect," Karthikesalingam said. "They will occasionally miss things. They will sometimes mention things that they shouldn't and vice versa. However, the potential to be a useful tool is clear."

He noted that the goal of this research has been to test the medical accuracy of these AI models to determine whether they can become tools that will complement clinicians and add value to healthcare systems.

With increasingly promising results from these tests, he believes that AI models, like Med-PaLM 2, can eventually reach the level of accuracy and consistency that would allow clinicians to use them in their daily practice to improve their patient care.

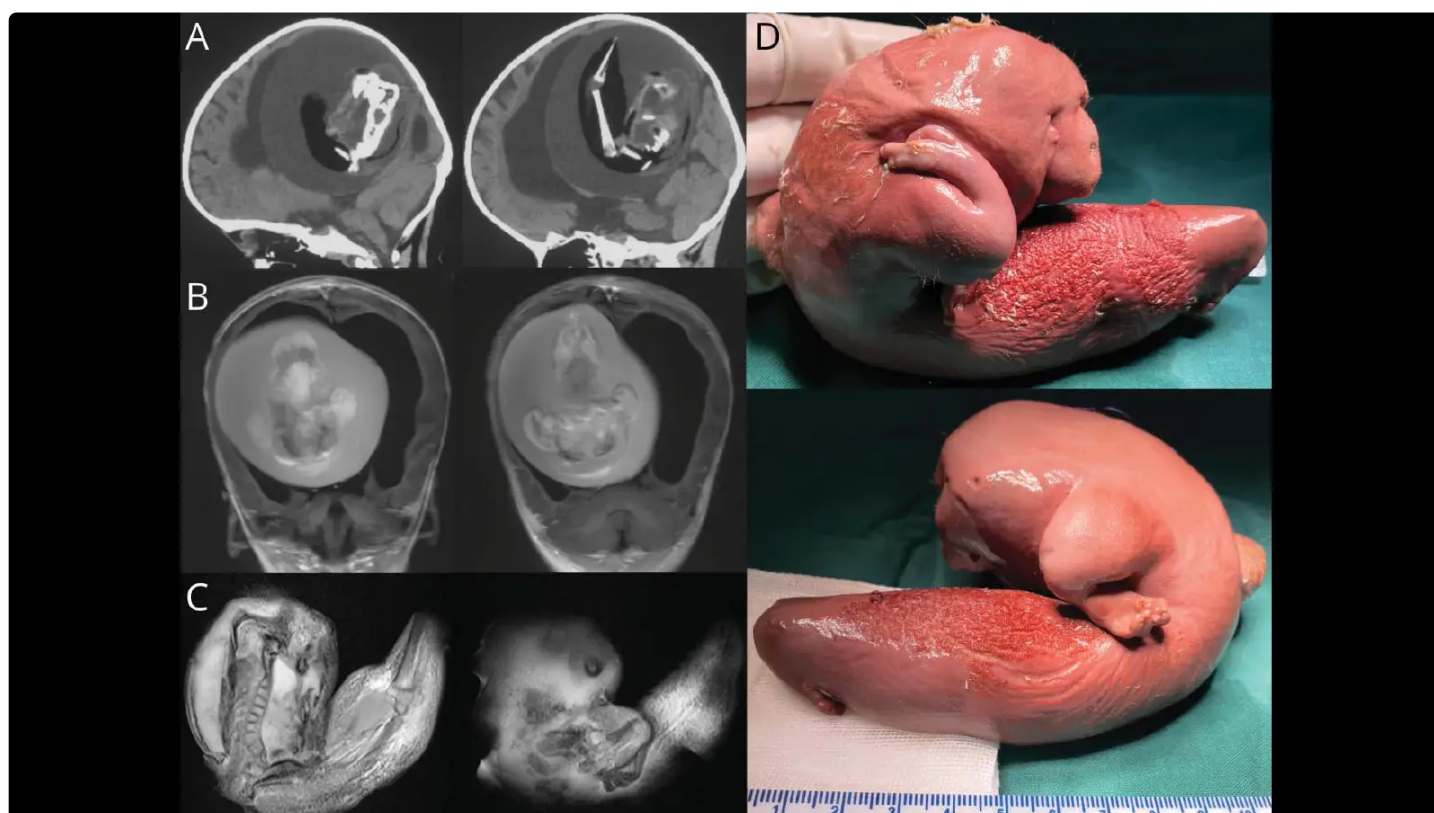
"This is shining a bright light toward a very hopeful and positive future in which these systems could become hopefully more cooperative and complementary tools that are better suited to workflows and help give clinicians superpowers," Karthikesalingam said.



[Michael DePeau-Wilson](#) is a reporter on MedPage Today's enterprise & investigative team. He covers psychiatry, long covid, and infectious diseases, among other relevant U.S. clinical news. Follow [Twitter](#)

4 Comments

## Recommended For You





**Oncology/Hematology**

Jimmy Carter's Melanoma; Post-Op Osimertinib Boosts OS; \$83K for Unproven Therapy

---

**Public Health & Policy**

Celeb Warns on Butt Shots; 27 Drugs Face Rebate Penalty; 'Shocking Amount of Misery'

---

## How Overturning Roe v. Wade Changed Match Day 2023

---

### Oncology/Hematology

Shortage of Prostate Cancer Drug; Win for DBT Mammo; Double Lung Transplant

---

### Infectious Disease

Evidence COVID Began in Raccoon Dog; Migraine Drug Recall; FDA Mum on Spring Booster

---

## Medical News From Around the Web

[KAISER HEALTH NEWS](#)

Black Women Have Much at Stake in States Where Abortion Access May Vanish

[BBC](#)

Brain cancer research plea made by family of Ollie Gardiner

[CBS NEWS](#)

DNA identifies remains of Florida teen who disappeared in 1972. Detectives now say she may have been killed by a serial killer cop.

[THE WASHINGTON POST](#)

How my dryer door gave me a concussion

[JOURNAL OF ADDICTIVE DISEASES](#)

Pain, smoking, and moderating effect of gender in a large, representative sample of Danish adults.

[THE JOURNAL OF ORTHOPAEDIC AND SPORTS PHYSICAL THERAPY](#)

All MCIDs Are Wrong, But Some May be Useful.