Mayo Clinic to use Cerebras chips for medical AI models

Partnership is a multimillion and multi-year deal

January 09, 2024 By: Charlotte Trueman Have your say

Cerebras Systems has announced a new multi-year partnership with Mayo Clinic which will see the two entities collaborate on the development of LLMs for medical applications.

Mayo Clinic is a nonprofit organization that provides one the largest integrated health systems across the US, funding research and education in the medical field.

The partnership will see the clinic use computing chips and systems provided by Cerebras to develop AI models using anonymized medical records and patient data, including diagnostics, treatments, outcomes, imaging, and molecular research.

– Getty Images

According to a report from Reuters, some of the models will be used to summarize important parts of lengthy medical records, while others will be trained to look for patterns in medical images or analyze genome data. However, it was noted that all medical decisions will still be made by doctors and not the new AI systems.

The models will eventually be made available on the Mayo Clinic Platform, a data network that is accessible to various healthcare systems in the US, Canada, Brazil, and Israel.

“It is an honor to collaborate with Mayo Clinic, the top-ranked hospital in the nation. With its recognized leadership in delivering medical outcomes, we are uniquely positioned to combine AI and medicine. The state-of-the-art AI models we are developing together will work alongside doctors to help with patient diagnosis, treatment planning, and outcome estimation,” said Andrew Feldman, CEO and co-founder of Cerebras.

“To create the first truly patient-centric healthcare AI, Mayo Clinic selected Cerebras for its proven experience in designing and training large-scale, domain-specific generative AI models,” the company wrote on LinkedIn. “Together, Cerebras and Mayo Clinic seek to combine AI and domain expertise to produce better patient outcomes.”

Neither organization disclosed how much the partnership cost but multiple outlets reported that Cerebras said it was a multimillion-dollar deal.

Cerebras develops giant wafer-scale chips. Its Wafer Scale Engine 2 is the world's largest semiconductor, with 2.6 trillion transistors. Built on TSMC 7nm, it has 850,000 'AI optimized' cores, 40GB of on-chip SRAM memory, 20 petabytes of memory bandwidth, and 220 petabits of aggregate fabric bandwidth. The WSE-2 chip is sold packaged with the Cerebras CS-2, a 15U box that also includes HPE’s SuperDome Flex.

This is not the first time Cerebras has made forays into the health tech sphere. In May 2022, biopharmaceutical company AbbVie said that it used a Cerebras CS-2 system to train biomedical natural language processing (NLP) models.

The six billion parameter model is able to translate and make libraries of biomedical literature searchable across 180 languages using large Transformer models such as BERT, BERT LARGE, and BioBERT.

Subscribe to The Compute, Storage & Networking Channel for regular news round-ups, market reports, and more.

Create an Account to Subscribe Now

More in HPC & Quantum

16 May 2025

02 Oct 2025

01 Oct 2025

Dutch quantum computing community launches Tuna-5 quantum system via the public cloud

Rigetti sells two Novera quantum computers to customers in Asia and the US

SRV and CSC sign agreement for development phase of Lumi 'AI Factory'

More in Investment / M&A / Financing

02 May 2025

Meta raises AI data center capex forecast to up to $72bn, blames Trump tariffs for increased cost

24 Oct 2025

Stack secures green financing for data center in Japan's capital, Tokyo

Episode **2026 Trends and Outlooks: Data centers in Asia**

Tags

Cerebras Systems | LLM | Cerebras | Mayo Clinic

Comments

Please log in or sign up to post and reply to comments.

Whitepapers

Navigating liquid cooling architectures for data centers with AI workloads

28 Oct 2025

Charting a course for compliance, carbon emission reductions, and efficiency with Low GWP Refrigerants

24 Oct 2025

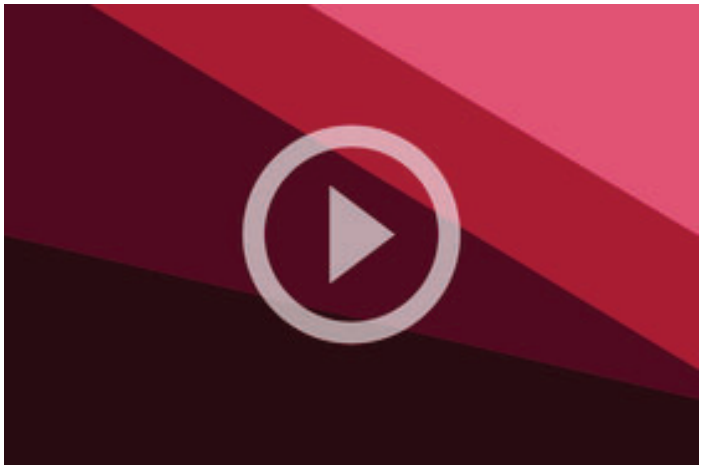
Guidance for RDHx deployment

24 Oct 2025

AI operations for IP and data center networking

22 Oct 2025

More



De-risking your next data center migration

*Email:

*Country:

Select...

Please select which newsletters or magazine you would like to receive:

- ☐ Daily Newsletter
- ☐ EMEA (Weekly)
- ☐ North America (Weekly)
- ☐ Asia Pacific (Weekly)
- ☐ MENA (Bi-Weekly)
- ☐ Skills & Workforce (Monthly)
- ☐ DCD>Magazine (Quarterly)

*I have read and agree to the Terms and Conditions and Privacy Policy

☐ Yes, I agree

Submit

Browse by Topic

Business Operations

- Construction & Site Selection
- Investment / M&A / Financing
- Standards & Regulations
- Workforce & Skills

Cooling

Critical Power

- Grid Level
- Rack Level

Connectivity

- Carrier Networks
- Networking
- Subsea Infrastructure

IT Hardware & Semiconductors

Sectors

- Cloud & Hyperscale
- Colocation & Wholesale
- Edge & IoT
- Enterprise
- Government & Military
- HPC & Quantum
- Telecoms & 5G

Browse by Region

Products

© 2025 DCD

Data Centre Dynamics Ltd (DCD), 32-38 Saffron Hill, London, EC1N 8FH
Email. **info@datacenterdynamics.com**

DCD is a subsidiary of **InfraXmedia**

More from InfraXmedia: SDxCentral - Data Center Nation