Cerebras AI Inference Wins Demo of the Year Award at TSMC North America Technology Symposium

December 5, 2025

World's Fastest AI Inference Selected by Global Community at TSMC's Technology Symposium



SUNNYVALE, Calif.--Cerebras Systems, makers of the fastest AI infrastructure, today announced that Cerebras AI Inference has been named Demo of the Year at 2025 TSMC North America Technology Symposium. Voted on by attendees from TSMC's customer and partners, the award recognizes the most compelling and impactful innovation demonstrated in the Innovation Zone at TSMC's annual Technology Symposium.

> "Wafer-scale computing was considered impossible for fifty years, and together with TSMC we proved it could be done," said Dhiraj Mallick, COO, Cerebras Systems.
> Share

"Wafer-scale computing was considered impossible for fifty years, and together with TSMC we proved it could be done," said Dhiraj Mallick, COO, Cerebras Systems. "Since that initial milestone, we've built an entire technology platform to run today's most important AI workloads more than 20x faster than GPUs, transforming a semiconductor breakthrough into a product breakthrough used around the world."

"At TSMC, we support all our customers of all sizes—from pioneering startups to established industry leaders—with industry-leading semiconductor manufacturing technologies and capacities, helping turn their transformative idea into realities," said Lucas Tsai, Vice President of Business Management, TSMC North America. "We are glad to work with industry innovators likes Cerebras to enable their semiconductor success and drive advancements in AI."

In 2019, Cerebras introduced the industry's first functional wafer-scale processor—a single-die chip 50 times larger than conventional processors—breaking a half-century of semiconductor assumptions through its partnership with TSMC. The Cerebras CS-3 extends this lineage and continues a scaling law unique to Cerebras.

A Showcase of Innovation and Partnership

Cerebras demonstrated CS-3 inference in TSMC North America Technology Symposium's Innovation Zone, a curated exhibition area highlighting breakthrough technologies from across TSMC's emerging customers. Cerebras AI Inference received the highest number of votes from attendees at the North America event, reflecting both the technical achievement and the excitement it generated among event attendees.

Cerebras AI Inference Leading the Industry

Cerebras AI Inference is now used across the world's most demanding environments. It is available through AWS, IBM, Hugging Face, and other cloud platforms. It supports cutting-edge national scientific research at U.S. Department of Energy laboratories and the Department of Defense, and global enterprises across healthcare, biotech, finance, and design have adopted Cerebras to accelerate their most complex AI workloads with real-time performance that GPUs cannot deliver.

Cerebras is also the fastest platform for AI coding—one of the fastest growing and most strategic AI verticals. It generates code more than 20 times faster than competing solutions.

Cerebras has been a pioneer in supporting open-source models from OpenAI, Meta, G42 and others, consistently achieving the fastest inference speeds as verified by independent benchmarking firm Artificial Analysis.

Cerebras now serves trillions of tokens per month across the Cerebras Cloud, on-premises deployments, and leading partner platforms.

For more information on Cerebras AI Inference, please visit www.cerebras.ai.

About Cerebras Systems

Cerebras Systems builds the fastest AI infrastructure in the world. We are a team of pioneering computer architects, computer scientists, AI researchers, and engineers of all types. We have come together to make AI blisteringly fast through innovation and invention because we believe that when AI is fast it will change the world. Our flagship technology, the Wafer Scale Engine 3 (WSE-3) is the world's largest and fastest AI processor. 56 times larger than the largest GPU, the WSE uses a fraction of the power per unit compute while delivering inference and training more than 20 times faster than the competition. Leading corporations, research institutes and governments on four continents chose Cerebras to run their AI workloads. Cerebras solutions are available on premise and in the cloud, for further information, visit cerebras.ai or follow us on

LinkedIn, X and/or Threads.

Contacts

Media Contact
PR@zmcommunications.com