



OpenAI and Broadcom Unveil LLM-Optimized Intelligence Processor

June 24, 2026

Built from the ground up for current and future LLMs across the industry

Developed from design to production in nine months, accelerated by OpenAI's models

Will deliver performance per watt better than current state-of-the-art, based on early testing

Expands OpenAI's full-stack platform, from products to models and now to chips

To be deployed at gigawatt scale with data center partners, over multiple generations

SAN FRANCISCO and PALO ALTO, Calif., June 24, 2026 (GLOBE NEWSWIRE) -- OpenAI and Broadcom (NASDAQ: AVGO) today unveiled Jalapeño, OpenAI's first Intelligence Processor: an accelerator architected around OpenAI's vision for the future of LLM inference, and the first AI accelerator in a multi-generation compute platform the companies are building together to make advanced AI faster, more reliable, and more accessible to more people.

Jalapeño was delivered to OpenAI CEO Sam Altman and President Greg Brockman by Broadcom President and CEO Hock Tan and Semiconductor Solutions President Charlie Kawwas, marking an important step in OpenAI's strategy to build the full stack behind its models and products.

OpenAI designed the chip from scratch around its deep understanding of LLM fundamentals, informed by its roadmap of models, kernels, serving systems, and product needs, with partners Broadcom and Celestica, helping industrialize the platform through chip implementation, board, rack system integration, high-performance networking, and scalable production systems. Jalapeño is designed with flexibility to work with all LLMs guided by OpenAI's insights into the inference needs of current and future AI models across the industry. Engineering samples of the Jalapeño chip are running ML workloads in the lab at production target frequency and power, including GPT-5.3-Codex-Spark.

While OpenAI is still measuring final performance, early testing shows that Jalapeño will deliver performance per watt substantially better than current state-of-the-art. A detailed technical report on performance will be presented in the coming months. The architecture reduces data movement and balances compute, memory, and networking resources to achieve realized utilization much closer to theoretical peak performance. Broadcom's silicon implementation and networking technologies, including Tomahawk networking silicon, help bring the platform to large-scale production.

"The world is moving to a compute-powered economy," said Greg Brockman, President and Co-Founder, OpenAI. "Jalapeño is part of our long-term full-stack infrastructure strategy to make compute more abundant, resulting in AI which is faster, more reliable, more affordable for people and businesses, and can be used to solve more important problems. By designing more of the stack ourselves, we can serve more intelligence with greater efficiency and keep pushing advanced AI toward broader access."

"Jalapeño was designed from the ground up for LLM inference using detailed insights from our close collaboration with OpenAI researchers," said Richard Ho, who leads OpenAI's hardware program. "We optimized the architecture around the kernels, memory movement, networking, and serving patterns that matter most for frontier AI models. Based on early testing, Jalapeño will efficiently execute our most important workloads close to the hardware's theoretical limits."

"Our collaboration with OpenAI represents a fundamental commitment to scaling the physical infrastructure required for the next decade of AI," said Hock Tan, President and CEO, Broadcom. "This is just the beginning of a multi-generation roadmap. By co-developing our industry-leading silicon directly with OpenAI, we are enabling the deployment of gigawatt scale data centers with Microsoft and other partners beginning in 2026."

Designed to be the best inference platform for LLMs

Jalapeño is a blank-slate design for modern LLM inference, not a general-purpose accelerator adapted from earlier AI workloads. It is informed by the systems OpenAI runs every day across ChatGPT, Codex, the API, and future agentic products, while also being designed for current and future LLMs across the industry. The goal is to combine the power and throughput of today's leading AI accelerators with latency closer to the fastest specialized inference systems, making Jalapeño well suited for interactive LLM products at scale.

Nine-month tape-out, accelerated by OpenAI models

Jalapeño was co-developed from initial design to manufacturing tape-out in just nine months, and the custom AI accelerator program represents what may be the fastest ASIC development cycle ever achieved in high-performance advanced semiconductors. That speed reflects deep software-hardware co-development with OpenAI's engineering teams, Broadcom's silicon implementation expertise, and the use of OpenAI models to accelerate parts of the design and optimization process.

The same models served to users are helping improve the infrastructure used to run future models. If AI can help engineers design better chips faster, it can lower the cost of compute across the industry and help democratize access to advanced AI.

Building a multi-generation platform with partners

Jalapeño is the first step in a multi-generation compute platform designed for initial deployment by the end of 2026, and expanding in the years ahead, combining OpenAI-designed accelerators with Broadcom silicon implementation, networking, and connectivity technologies; and Celestica's board, rack and system expertise.

Additional Resource

[Read the OpenAI blog post](#)

About OpenAI

OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity.

About Broadcom

Broadcom Inc. (NASDAQ: AVGO) is a technology leader that designs, develops, and supplies semiconductors and infrastructure software for global organizations' complex, mission-critical needs. Broadcom combines long-term R&D investment with superb execution to deliver the best technology, at scale. Broadcom is a Delaware corporation headquartered in Palo Alto, CA. For more information, visit www.broadcom.com.

Press Contacts

OpenAI: press@openai.com

Broadcom: press.relations@broadcom.com

Cautionary Note Regarding Forward-Looking Statements

This announcement contains forward-looking statements (including within the meaning of Section 21E of the Securities Exchange Act of 1934, as amended, and Section 27A of the Securities Act of 1933, as amended) concerning Broadcom. These statements include, but are not limited to, statements regarding Broadcom's collaboration with OpenAI in delivering Jalapeño to OpenAI and Broadcom helping bring OpenAI's full-stack platform to large-scale production to enable the deployment of gigawatt scale datacenters. These forward-looking statements are based on current expectations and beliefs of Broadcom's management, current information available to Broadcom's management, and current market trends and market conditions, and involve risks and uncertainties that may cause actual results to differ materially from those contained in the forward-looking statements. Accordingly, undue reliance should not be placed on such statements.

Particular uncertainties that could materially affect future results include risks associated with: global political and economic conditions and uncertainty; government regulations, trade restrictions and trade tensions; fluctuations in the timing and volume of significant customer demand; ability to make successful investments in research and development and successfully expand Broadcom's business strategy or adopt Broadcom's business models; ability to continue winning business and the timing of such wins; dependence on contract manufacturing and outsourced supply chain; dependency on a limited number of suppliers; dependence on senior management and the ability to attract and retain qualified personnel; ability to protect against cybersecurity threats and a breach of security systems; ability to accurately estimate customers' demand and adjust the manufacturing and supply chain accordingly; ability to improve manufacturing capacity and quality; involvement in legal proceedings; quarterly and annual fluctuations in operating results; Broadcom's competitive performance; ability to maintain or improve gross margin; ability to protect Broadcom's intellectual property and the unpredictability of any associated litigation expenses; significant indebtedness and the need to generate sufficient cash flows to service and repay such debt; and other events and trends on a national, regional, industry-specific and global scale, including those of a political, economic, business, competitive and regulatory nature.

Broadcom's filings with the Securities and Exchange Commission (SEC) are available without charge at the SEC's website at <https://www.sec.gov> and include some important risk factors that may affect future results. Broadcom undertakes no intent or obligation to publicly update or revise the forward-looking statements made in this announcement, except as required by law.