# OpenAI’s $300bn AI hardware expansion redefines compute and energy scale



OpenAI is leading a $300 billion expansion in AI hardware infrastructure, creating a tightly integrated ecosystem that links chip suppliers, financiers and energy providers. The initiative includes multi-year contracts with AMD and Broadcom to deliver 16 gigawatts of new compute capacity between 2026 and 2029—comparable to the electricity use of some small nations.

AMD will provide 6 gigawatts of Instinct GPUs and has issued equity warrants to OpenAI tied to performance milestones, deepening the financial link between supplier value and AI scaling. Broadcom will deliver 10 gigawatts of custom silicon and rack systems, co-designed for OpenAI’s AI workloads. This collaboration builds on the Stargate infrastructure project with Oracle and SoftBank, which includes five new US data centre sites. Stargate alone accounts for 7 gigawatts of planned capacity, with total investment across the initiative expected to exceed $400 billion by the end of 2025.

The approach reflects a circular AI economy, where equity incentives, capital flows and purchase agreements bind model operators to vendors. Nvidia’s 7 percent stake in CoreWeave and the latter’s $22.4 billion contract book with OpenAI exemplify this model, as does Nvidia’s reported $100 billion in chip orders driven by OpenAI demand.

Execution risks remain. Energy access and utilisation efficiency will be key to financial returns. Goldman Sachs projects global data centre electricity demand will rise 165% by 2030, while McKinsey warns that US centres alone could consume 14% of national power by decade’s end. Long-term power deals and onsite generation will be essential for sustainable scaling. Regulatory concerns also persist—though the UK Competition and Markets Authority declined to probe the Microsoft-OpenAI partnership in March 2025, future equity-linked supply deals may attract renewed scrutiny.

On the hardware front, Broadcom’s custom silicon, built with OpenAI and fabricated by TSMC, is set to boost efficiency for inference workloads from 2026 to 2029. OpenAI is also working with Arm to develop a server-grade CPU optimised for this ecosystem. These advances are expected to reduce compute costs and improve scalability.

AMD is progressing its Helios rack-scale AI platform, unveiled at the 2025 Open Compute Project Global Summit. Helios integrates EPYC CPUs, MI450 GPUs and high-speed networking, supporting 72 GPUs per rack and offering 1.4 exaFLOPS of FP8 performance—50% more memory than Nvidia’s comparable systems. Oracle plans to deploy 50,000 MI450 GPUs from Q3 2026, reinforcing AMD’s presence in AI-first data centres. A landmark $40 billion acquisition of Aligned Data Centers by a consortium including Microsoft, Nvidia and BlackRock further signals deepening private sector investment. Aligned controls over 5 gigawatts of capacity across the US and Latin America.

The next two to three years will be pivotal. As deployments go live and power agreements close, OpenAI will focus on converting purchase commitments into usable compute and revenue. If enterprise usage scales in line with infrastructure buildout, the financial structure behind this AI ecosystem could evolve into a sustainable compute economy rather than a source of vendor risk.

OpenAI’s infrastructure drive blends hardware innovation with financial and energy strategy, laying a foundation for AI leadership. For the UK, this underscores the need to develop integrated frameworks that align innovation with regulation and sustainability—ensuring global competitiveness in the era of scalable AI.

Created by [Amplify](https://www.hbmadvisory.com/amplify): AI-augmented, human-curated content.

## Bibliography

1. <https://bitcoinethereumnews.com/tech/the-feedback-loop-rewiring-wall-street-finance/?utm_source=rss&utm_medium=rss&utm_campaign=the-feedback-loop-rewiring-wall-street-finance> - Please view link - unable to able to access data
2. <https://www.reuters.com/business/oracle-offer-cloud-services-using-amds-upcoming-ai-chips-2025-10-14/> - Oracle has announced a collaboration with AMD to offer cloud services powered by AMD's upcoming MI450 artificial intelligence chips, aiming to meet the growing demand for AI infrastructure supporting tools like ChatGPT. The initial deployment will include 50,000 MI450 processors in Q3 2026, with plans for further expansion in 2027 and beyond. This partnership strengthens AMD’s client portfolio and enhances Oracle’s processor capabilities amid increasing competition for compute power. AMD shares rose over 3% following the news, despite broader market concerns related to U.S.-China trade tensions.
3. <https://www.reuters.com/legal/transactional/blackrock-nvidia-buy-aligned-data-centers-40-billion-deal-2025-10-15/> - An investor consortium including BlackRock, Microsoft, Nvidia, Abu Dhabi’s MGX, Elon Musk’s xAI, the Kuwait Investment Authority, and Singapore's Temasek is acquiring Aligned Data Centers for $40 billion. The purchase of the U.S.-based company from Australian Macquarie Asset Management marks the first major deal for the AI Infrastructure Partnership, established to secure the infrastructure necessary to meet surging demand for artificial intelligence applications. Aligned Data Centers operates over 5 gigawatts of capacity across nearly 80 centers in the U.S. and Latin America.
4. <https://www.tomshardware.com/pc-components/cpus/openai-arm-partner-on-custom-cpu-for-broadcom-chip> - OpenAI is collaborating with Arm, a SoftBank-owned company, to develop a custom server-class CPU that will support its in-house AI accelerator, co-developed with Broadcom. This CPU, intended for next-generation AI infrastructure, could be one of Arm's largest ventures into the data center space. OpenAI's accelerator, designed specifically for inference workloads, is expected to enter production in late 2026, supporting up to 10 gigawatts of compute capacity by 2029. Fabricated by TSMC, the Broadcom-built chip has been in development for 18 months.
5. <https://www.wired.com/story/openai-oracle-softbank-data-center-stargate-us/> - OpenAI, Oracle, and SoftBank are expanding their Stargate project with five new AI data center sites in the U.S. These additions bring Stargate to nearly 7 gigawatts of planned capacity and over $400 billion in investment over the next three years, positioning the project to secure the full $500 billion, 10-gigawatt commitment by the end of 2025. The new sites are located in Shackelford County, Texas; Doña Ana County, New Mexico; and an upcoming site in the Midwest, with additional expansions near Abilene, Texas, and in Lordstown, Ohio, and Milam County, Texas.
6. <https://www.tomshardware.com/tech-industry/amd-debuts-helios-rack-scale-ai-hardware-platform-at-ocp-global-summit-2025-promises-easier-serviceability-and-50-percent-more-memory-than-nvidias-vera-rubin> - At the 2025 Open Compute Project Global Summit, AMD introduced its Helios rack-scale AI hardware platform, purpose-built for the Meta OpenRack Wide form factor. Aimed at AI-first data center deployments, Helios is AMD's strategic answer to Nvidia’s high-performance rack solutions like the GB300 NVL72. This platform integrates AMD EPYC CPUs, Instinct MI450 GPUs, Pensando networking, and the ROCm software stack for optimized performance and efficiency. Helios supports up to 72 MI450 GPUs per rack, each with access to 432GB of HBM4 memory, totaling 31TB of memory and offering 1.4 exaFLOPS of FP8 performance. AMD claims this provides 50% more memory than Nvidia's upcoming Vera Rubin platform.
7. <https://www.reuters.com/business/oracle-offer-cloud-services-using-amds-upcoming-ai-chips-2025-10-14/> - Oracle has announced a collaboration with AMD to offer cloud services powered by AMD's upcoming MI450 artificial intelligence chips, aiming to meet the growing demand for AI infrastructure supporting tools like ChatGPT. The initial deployment will include 50,000 MI450 processors in Q3 2026, with plans for further expansion in 2027 and beyond. This partnership strengthens AMD’s client portfolio and enhances Oracle’s processor capabilities amid increasing competition for compute power. AMD shares rose over 3% following the news, despite broader market concerns related to U.S.-China trade tensions.