AI Infrastructure Investment

WHITEPAPER
JUNE 2025

CONTRIBUTORS
ALISON ADAMS, PHD

Since OpenAI launched its generative artificial intelligence application, ChatGPT, in late 2022, AI has catapulted to the forefront of technology companies in the US. The growing competition in AI development has driven a significant surge in investments in technological research and innovation. In the US, the so-called "Magnificent Seven" companies have each developed their own proprietary large language models (LLMs).

The training of LLMs is only the first step in the AI commercialization revolution. The development and deployment of AI demands substantial upgrading of technology infrastructure. For example, the commercial development of AI will require the development and production of advanced semiconductor chips, larger, faster and more deeply connected cloud computing capabilities, and new sources of power generation. While the training of LLMs was largely a project of concentrated computing resources, the deployment and commercialization of AI will require a substantial upgrade of current technology infrastructure.

This is the first in a series looking at AI investment and potential implications for economic growth in the US. This note is focused on the investment needed to meet AI-computing demand, from domestic production of advanced semiconductor chips to a massive expansion of server farms and their networks, as well as an increase in power generation capacity to power the servers and chip fabs.

This note examines how technology firms are investing in physical infrastructure to meet the compute power needed to monetize AI with commercialization efforts. Like the build-out of the US electrical grid or the internet, which involved substantial initial investment that subsequently drove economic growth, the investment in AI infrastructure should provide the undergirding for wide-spread adoption of Generative AI (GenAI) commercialization.

Key takeaways

- → The commercialization of AI requires significant investments in upgrading technology infrastructure, including advanced semiconductor chips, cloud computing capabilities, and power generation.
- → Innovative semiconductor development and the construction of specialized factories (i.e., "fabs") are pivotal in meeting AI computing demands, alongside enhanced data storage and network connectivity infrastructure.
- → AI-related upgrades involve tailored developments such as proprietary semiconductors, software integration, expanded server farms, and new sources of electricity to power these systems.
- → Like previous infrastructure revolutions, such as the creation of the electrical grid and the internet, AI investments are expected to drive widespread adoption and substantial economic growth over time.

What is AI infrastructure?

AI infrastructure investment involves a substantial upgrading of current compute capabilities.[1] But unlike a traditional brownfield investment where a company expands on a current productive asset, AI infrastructure investment requires a substantial investment in research and development to meet new levels of technological advancement. The front-loading of AI investment costs is largely being made in expectation of future demand from enterprises and consumers, which represent a necessary component for profitable commercialization of AI.

To upgrade AI computing capacity in existing server farms, AI related upgrades will require the customized development of new AI-semiconductors, software integrations, data center construction, network upgrades, and new sources of power (see Figure 1).

AI development and commercialization demands faster and more sophisticated semiconductors. While cutting-edge chips are designed by firms like Nvidia, they typically do not make the chips themselves. Rather, they outsource their production to a handful of companies that have the capabilities to manufacture these advanced chips at specialized factories, or foundries, known as "fabs". Once the chips have been produced and delivered, they require an immense amount of data storage (usually via the cloud), which of itself then requires connectivity and electricity.

FIGURE 1
Illustration of Private AI Related Infrastructure Investment Announcements in 2025

Source: Meketa survey of company announcements as of March 2025.

		Compute Resources		Data Storage & Management	Power Generation	
	2025 AI Related Plans Announced	Proprietary AI Semiconductor Development	AI Semiconductor Foundry (<7 nm chips)	Cloud Storage & High-Speed Networks	Renewables	Nuclear
Alphabet (Google)	$75B	Yes		Yes	Yes	Yes
Amazon	$100B	Yes		Yes	Yes	Yes
Apple	$500B (4 Years)	Yes		Yes	Yes	
Meta (Facebook)	$65B	Yes		Yes	Yes	Yes
Microsoft	$40B (+Stargate)	Yes		Yes	Yes	Yes
Nvidia	Stargate	Yes	With TSMC in AZ			
Tesla	Yes	Yes		Yes	Yes	Possible
Oracle	Stargate	Yes		Yes	Yes	
Intel	Yes	Yes		Yes		
Arm	Stargate	Yes				
AMD		Yes				
TSMC	$165B (+Stargate)	Yes			Possible	
OpenAI	Stargate	Yes		Yes	Yes	Yes
Project Stargate	$500B (4+ Years)	Yes		Yes	Yes	Yes

Advanced chip production

In 1990, the US manufactured about 40% of the world's semiconductors, but today Asia dominates global semiconductor supply, led by Taiwan's TSMC and followed by UMC, SMIC, and Samsung.[2] The global pandemic and supply chain disruptions of 2020 through 2022 exposed this dependency on semiconductors produced overseas.[3] With tensions between the US and China on the rise, particularly regarding the independence of Taiwan, there was a "growing fear that lagging behind China in such critical technologies will undermine US national security and economic competitiveness."[4] Hence, in 2022, the US Congress passed the 'Creating Helpful Incentives to Produce Semiconductors & Science Act' (i.e., the CHIPS & Science Act).[5]

The scale of private and public investment since the passage of the CHIPS and Science Act is quite impressive when compared to the modest $39 billion headline program. According to a White House press release on the Act's two-year anniversary, nearly $400 billion in private investments have been announced.[6] In the year following the passage of the CHIPS Act, the US semiconductor industry added 44,000 new jobs.[7] But in an economy as large as the US, the number of jobs created from the CHIPS act thus far is less than the average monthly job gain the US normally produces over the same period of time.[8]

The US Semiconductor Industry Association estimated that investment in foundry capacity due to the CHIPS act will create at least 115,000 additional jobs for technicians, engineers, and computer scientists by 2030.[9] But the semiconductor foundry jobs created are a small portion of the total number of information technology and computer science related jobs that are expected to materialize by 2032. In total, the US economy is expected to add 3.85 million new jobs in the IT and related sectors over the next eight years.[10]

While the CHIPS Act is funded by the government, its implementation depends on active partnership and coordination with private enterprise, including both US and foreign companies (see Figure 2). For example, TSMC recently announced plans to build foundry capacity in Arizona for its most advanced semiconductors. Amazon (AWS), Google and Meta are in close partnership with TSMC to develop their own proprietary AI semiconductors. Several others are following suit, such as Intel, Samsung, Micron Technology and Texas Instruments among others.[11] In 2025, Nvidia announced plans to manufacture American made AI supercomputers with its next generation Blackwell semiconductor to be manufactured in Arizona.[12] Nvidia is building over one million square feet of manufacturing space in Arizona and Texas to build and test its Blackwell Chips.[13] Through partnerships with TSMC, Foxconn, Wistron, Amkore and SPIL, Nvidia currently plans to invest approximately half a trillion dollars in the US to build and develop next generation semiconductors.[14] CHIPS & Science Act grants, loans, and tax incentives have been broadly distributed around the US.[15]

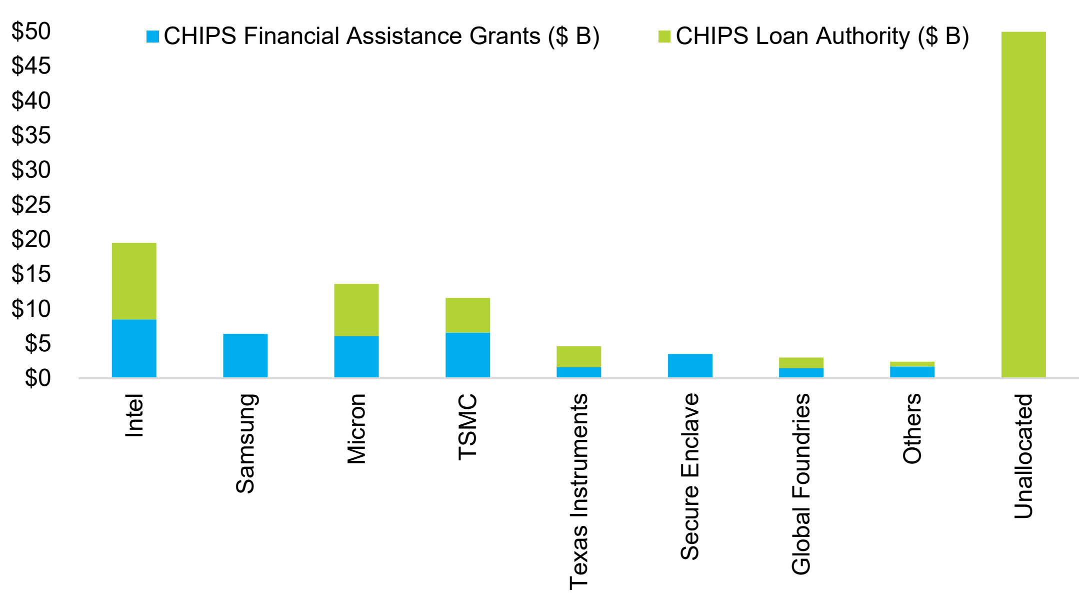

FIGURE 2

CHIPS Act Grants & Loans to Private Semiconductor Companies ($ B)

Source: CSIS, S. Shivakumar et al., "A World of Chips Acts: The Future of US-EU Semiconductor Collaboration," August 2024.

Data centers

Traditional data centers cannot train or support AI models due to the high resource needs of AI workloads.[16] Traditional data centers must be upgraded with higher density racks of graphics processing units (GPUs). GPUs deliver parallel processing capabilities needed to train and run AI applications.[17] Hyperscalers are currently in the process of upgrading CPU data centers adding GPUs to meet demand for AI compute. Moreover, investors and companies have already or are in the process of building dedicated AI data centers. The StarGate Project, a four-year half-trillion dollar AI collaboration, is one such effort to build new AI data centers. The Stargate Project is a collaboration between OpenAI, SoftBank, Oracle, and MGX.[18] Another is xAI's Colossus in Tennessee, a server farm that is expected to use 200,000 Nvidia GPUs.[19]

This is in addition to the traditional hyperscalers like Amazon, Google, and Microsoft, who have announced plans to continue to upgrade and build AI data centers. There are newcomers to the AI data center space that are providing AI-GPU cloud capacity called "neo-cloud" providers who rent access to their AI-GPUs. Perhaps the most visible is CoreWeave, which went public in 2025 with Nvidia as a key investor.[20] Private neo-cloud firms are investing in AI data centers as well.[21]

The need for AI capable data centers will go well beyond those enterprises who are engaged in the training, development, and commercialization of their own AI models. Large cloud providers and other corporations are building LLMs capable of natural language processing and understanding. These models must be trained on vast amounts of information using power-intensive processors. According to a recent study, leading data center players could spend as much as $1.8 trillion on AI data centers between 2024 and 2030.[22] Since the start of 2025, Amazon, Microsoft and Meta have all announced plans to build new data centers in the US.[23] These hyperscalers will most likely increase the average size of a US data center from 46 MW to 60 MW by 2028.[24]

Moreover, many enterprises are moving beyond the training of AI models and shifting to the inference phase of AI model development. Inference AI applies the AI trained model to new data or an application or service featuring the underlying AI model.[25] Open AI models such as LLaMA 3 and China's DeepSeek R1 are making commercial adoption of AI tools and agents more widespread.[26] The efficiency of DeepSeek's R1 might drive capex costs lower.[27]

Cheaper access to AI could mean more commercial demand on the one hand, but it could also present a risk of over-building AI data center capacity,[28] similar to what happened with fiberoptic cable during the early days of the internet.[29] If generative AI is to be widely adopted and integrated into computational activities, AI data center capacity must grow. According to Nvidia CEO Jensen Huang, AI GPU capacity could be entering a period of hyper Moore's Law progress during which he estimated a doubling of capacity every two years.[30] "Over the next ten years, our hope is that we could double or triple performance every year at scale."[31]

Power demand

After decades of very low demand growth, power demand in the US is forecast to grow at least 3% a year over the next decade (see Figure 3).[32] According to the US Department of Energy, "Domestic energy usage from data centers is expected to double or triple by 2028."[33] In 2023, data centers accounted for just 4.4% of total US power demand, but this is expected to increase to 12% of demand by 2028.[34] Generative AI is expected to account for approximately 35% of total data center demand by 2030.[35]

While traditional sources of power generation may be able to provide some of the incremental power needed by data centers, the US Department of Energy found that renewable energy sources could also be used to meet anticipated demand.[36] Sources of clean energy may scale easily for dedicated data center power needs.[37] Indeed, Google, Meta, Microsoft, Tesla, Amazon, Apple and Nvidia have all indicated a continued interest in clean energy sources for power generation.[38] For example, Microsoft has agreed to purchase power from the Three Mile Island nuclear reactor for the next twenty years to provide clean energy for its data centers that are running its AI applications.[39] Amazon, Google, and Meta have also announced plans to support the tripling of nuclear power production in the US by 2050.[40] In May 2025, President Trump signed an executive order designed to help fast track the construction of at least ten nuclear power plants.[41]

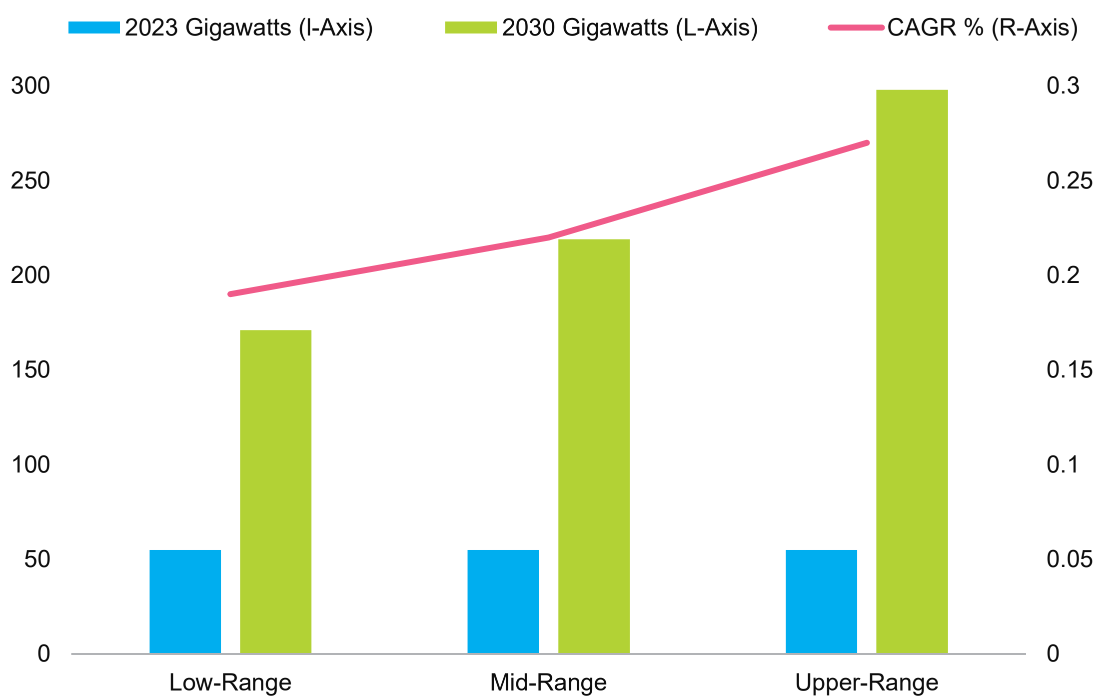

FIGURE 3
Global Data Center Power Demand Growth Estimate (GW)

Source: McKinsey Data Center Power Demand Model, October 29, 2024. Three scenarios showing the upper, low and midrange estimates of demand based on analysis of AI adoption trends, growth in shipments of different types of chips (application specific integrated circuits, graphics processing units, and associated power consumption; and typical compute, storage, and network needs of AI workloads. Demand is measure by power consumption to reflect the number of services a facility can house.

Whereas power was traditionally owned and operated by local utility companies, in recent years, many private equity and institutional investors have entered the space (see Figure 4). Thus, additional sources of capital are helping to meet the rising power generation needs.

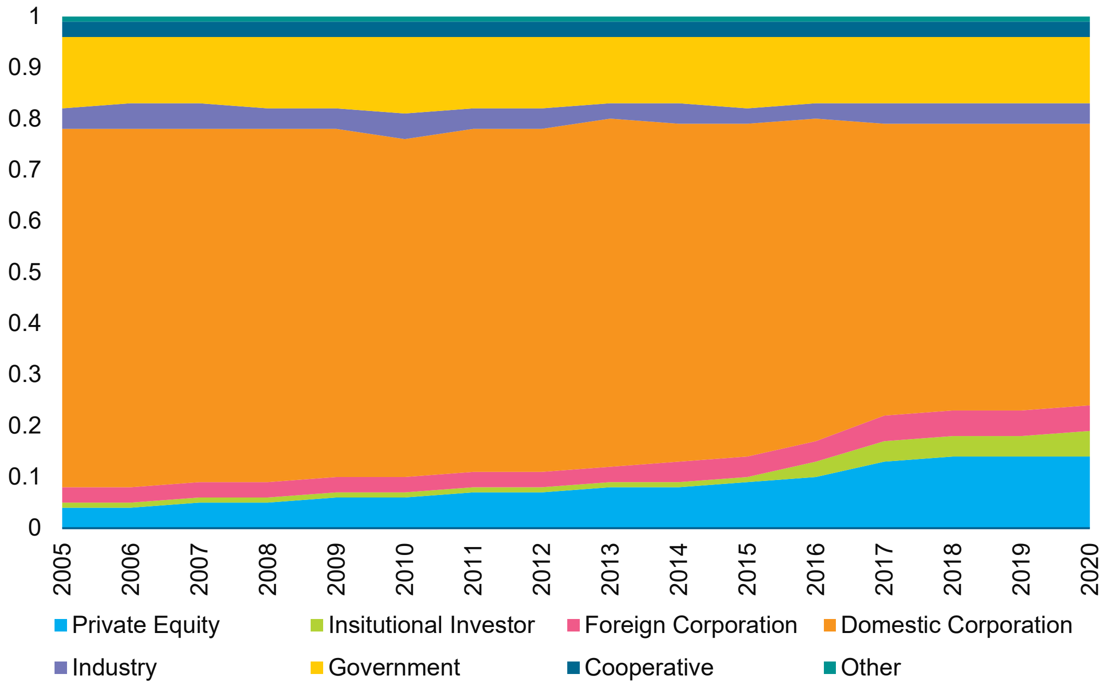

FIGURE 4
Institutional Investors & Private Equity Power Generation Ownership (% Monthly Power Generation in the US)

Source: NBER, J. D. Rauh et al., "The Shifting Finance of Electricity Generation," September 2024. Meketa illustration.

Connectivity, towers & satellites

The competition for AI advancement is not just confined to GPUs and data centers. Global telecom companies are looking to expand their digital wireless capacity. After years of rapid growth for mobile data capacity, demand growth had slowed in recent years.[42] For global telecom networks, the need to build-out sufficient digital wireless capacity to support ever smarter and faster iphones appears to have met an inflection point.[43]

The rapidly expanding capacity of generative AI and AI assistants will also grow demand for mobile digital capacity.[44] In recent years, much of the new demand for mobile digital video connectivity was limited to video traffic.[45] But increasingly powerful generative AI assistants will replace simple text-only AI applications like ChatGPT.[46] New generative AI assistants "are used for multi-media tasks, including audio, image, and video generation. They are highly effective at capturing long range dependencies and are particularly powerful for 'multimedia tasks such as text-to-image."[47] The promise of AI to deliver personalized and specialist services, data, and advertising could account for the largest AI-driven demand for data capacity in an indirect manner (see Figure 5).[48]

Simply upgrading existing 4G towers may not be enough to accommodate the increased demand for wireless digital capacity in the coming years.[49] "The explosion in data created by the increased use of AI is another reason why telecom firms around the world are continuing to invest in so-called 5G stand alone mobile networks."[50]

Stand-alone 5G networks have higher capacity and speed than upgraded or hybrid networks built for 4G spectrum.[51] Global telecom companies are looking to expand their wide-area-network (WAN) capabilities to integrate public and private local-area-networks (LANs) as well as cloud facilities like those being built by Amazon, Google, and Microsoft.[52] Enterprise AI demand for network access is also forecast to growth rapidly over the coming decade (see Figure 5).

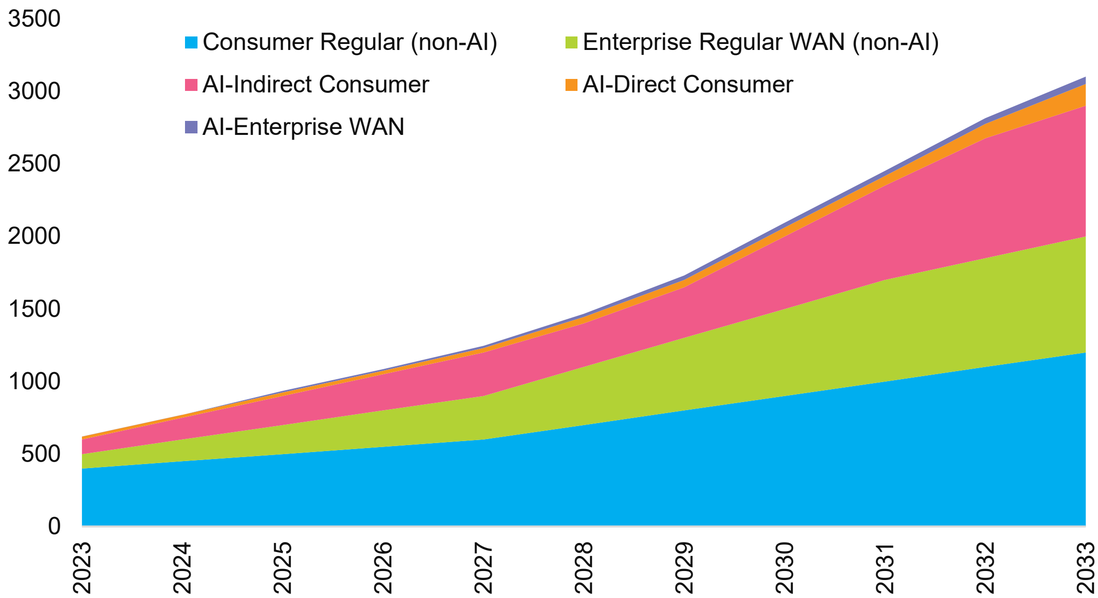

FIGURE 5

Global AI Wide Area Network (WAN) Demand Growth 2023 – 2033 (EB/M)

Source: Nokia, "Global Network Traffic Report," 2023. Wide Area Network (WANs) connect small local area networks (LANs) that include cellular service providers, public and private cellular towers and cloud storage facilities. Wireless communications depend on WANs infrastructure. EB/month is a measure of digital information storage per month where EB is an exabyte of data or one billion gigabytes.

The surge in AI-driven demand is ushering in a new era for digital infrastructure. With increasingly sophisticated generative AI applications and AI-enabled devices poised to reshape how we consume and create content, companies are preparing for a future of exponential data growth. This has the potential to present a host of unique investment opportunities. Upgrading existing networks will not suffice; stand-alone 5G networks and, eventually, 6G capabilities will be necessary to meet the demands of AI-enabled multimedia and enterprise connectivity.

Conclusion

The rapid evolution of artificial intelligence and its integration into global infrastructure heralds a transformative era for technology and society. From the unprecedented demands placed on data centers and wireless networks to the specialized advancements in semiconductor technology, AI is undeniably reshaping the way we consume, generate, and interact with information. These developments are not only pushing the boundaries of innovation but are also creating significant economic and investment opportunities. As telecommunication companies transition towards stand-alone 5G networks and even contemplate 6G, the need to accommodate AI-driven multimedia and enterprise connectivity becomes ever more pressing. Similarly, the advancements of GPUs illustrates the industry's commitment to meeting the computational demands of increasingly sophisticated AI applications.

The themes explored in this paper underscore a fundamental truth: the synergy between AI's capabilities and the supporting infrastructure is both a challenge and an opportunity. The promise of AI to deliver personalized services, complex multimedia applications, and advanced analytical solutions requires a robust framework of connectivity and computational power. Companies worldwide are investing heavily in upgrading existing systems, ensuring that they remain competitive in this rapidly changing landscape. As AI continues to evolve, the collaboration between innovative hardware development, expansive network infrastructures, and the growing demand for intelligent applications sets the stage for a future of unparalleled technological growth.

Ultimately, the trajectory of AI and its supporting infrastructure highlights the profound interconnectedness of technological innovation, economic strategy, and societal advancement. By addressing the challenges of exponential data growth and harnessing AI's potential, the global industry is poised to lead humanity into a future that thrives on creativity, efficiency, and connectivity. The next decade will be pivotal, as the world witnesses how these advancements redefine the boundaries of possibility.

Appendix

Why does AI need advanced semiconductors typically called GPUs?

The training of an AI large language model can take months and tens of millions of dollars. "This enormous computational power is delivered by computer chips that not only pack the maximum number of transistors – basic computational devices that can be switched between on (1) and off(0) states but also are tailor-made to efficiently perform specific calculations required by AI systems . . . trying to deliver the same AI application using older AI chips or general-purpose chips (CPUs) can cost tens to thousands of time more."[53]

As general purpose semiconductors (CPUs) have gotten smaller and faster, they have approached the atomic and engineering limits. Between the 1960s and the 2010s semiconductors have followed Moore's Law where they doubled their capacity and speed roughly every two years.[54] But today, the transistors are only a few atoms thick making smaller and faster transistors more difficult and costly.[55] "The economies of scale historically favoring general-purpose chips like central processing units (CPUs) have been upset by rising demand for specialized applications like AI and the slowing of Moore's Law-driven CPU improvements."[56] The CPU is a critical 'brain' of a computer often responsible for many procedure. Far from being replaced by GPUs, CPUs work in conjunction with GPUs.

AI chips include GPUs and field programmable gate arrays (FPGAs) and application-specified integrated circuits (ASICs).[57] "AI chips gain speed and efficiency (that is, they are able to complete more computations per unit of energy consumed) by incorporating huge numbers of smaller and smaller transistors, which run faster and consumer less energy." (see Figure 6) But AI chips also have the capacity to run calculations in parallel and even store an entire AI algorithm on one chip.[58]

Major US technology companies are currently in the process of developing and upgrading AI semiconductor chips. The advancement is perhaps most evident in the release of ever faster GPUs. For example, last summer Nvidia has released an even faster GPU than the H100: the Blackwell B200 MLPerf is estimated to be four times faster than the GPU (H100) that trained the first ChatGPT release in 2022.[59] In February 2025, Nvidia has confirmed that even faster GPUs are coming in 2025 and 2026.[60]

"While the standard cloud needs supreme flexibility and fungibility in compute storage, RAM, and networking, the GPU cloud needs far fewer options due to the relative homogeneity or workloads . . . and the H100 (Nvidia) is the optimal GPU for basically all modern use cases including LLM training and high volume LLM/diffusion inference."[61]

FIGURE 6

GPU & CPU Economics

Source: Semianalysis, D. Patel, et al, "DeepSeek Debates: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts," January 31, 2025.

	Unit	CPU (Traditional)	GPU (AI)
Server Capital Costs			
Upfront Server Capex	USD	$15,000	$350,000
Useful Life in Years	Years	6	6
Cost of Capital	%	13%	13%
Total Server Capital Costs Per Month	USD/month	$301	$7,026
Server Hosting Costs			
Electricity	USD/kWh	0.087	0.087
Utilization Rate	%	80%	80%
Power Usage Effectiveness	Ratio	1.25	1.25
Colocation Cost (rack space rental cost, infrastructure, install)	USD/kW/month	63.5	63.5
Total Hosting Cost pe kW per Month	USD/kW/mth	183.5	183.5
Server Power Consumption	W	1,200	10,200
Total Server Hosting Costs per Month	USD/month	$220	$1,872
Comprehensive Server Cost per Month	USD/month	$521	$8,898
Number of vCPU/GPUs per Server	vCPU/GPU	128	8
Capital Cost per Unit, per Hour	USD/hr	$0.003	$1.203
Hosting Cost per Unit, per Hour	USD/hr	$0.002	$0.321
Rental Cost per unit per Hour	USD/hr	$0.006	$1.524
Cost in Hosting vs Capex	%	73.1%	26.5%
W=watts; kW=kilowatts; kWh=Kilowatt-hours			
OEM pricing with networking costs allocated to server level not hyperscale			

Measuring an inference AI model's speed can be done in "time to first token" (TTFT). "This metric shows how long a user need to wait before seeing the model's output. This is the time it takes from submitting the query to receiving the first token."[62]

Notes

[1]"Compute" is industry shorthand for computing power or computational resources. It refers to the hardware and systems (like GPUs and data centers) that perform the calculations and data processing required for AI tasks.
[2]Major semiconductor companies include Taiwan Semiconductor Manufacturing Company (TSMC), United Microelectronics Corporation (UMC), Semiconductor Manufacturing International Corporation (SMIC) and Samsung.
[3]Source: Council on Foreign Relations, D. Roy, "The CHIPS Act: How US Microchip Factories Could Reshape the Economy," October 8, 2024.
[4]Source: Council on Foreign Relations, D. Roy, "The CHIPS Act: How US Microchip Factories Could Reshape the Economy," October 8, 2024.
[5]Source: Council on Foreign Relations, D. Roy, "The CHIPS Act: How US Microchip Factories Could Reshape the Economy," October 8, 2024.
[6]Source: The Whitehouse, "Two Years after the CHIPS and Science Act, Biden-Harris Administration Celebrates Historic Achievements in Bringing Semiconductor Supply Chains Home, Creating Jobs, Supporting Innovation, and Protecting National Security," August 9, 2024.
[7]Source: Semiconductor Industry Association, "Chipping Away: Assessing & Addressing the Labor Market Gap Facing the US Semiconductor Industry," 2023.
[8]Source: FRED as of August 2024. On average, the US economy added at least 200,000 jobs a month.
[9]Source: Semiconductor Industry Association, "Chipping Away: Assessing & Addressing the Labor Market Gap Facing the US Semiconductor Industry," 2023.
[10]Source: Semiconductor Industry Association, "Chipping Away: Assessing & Addressing the Labor Market Gap Facing the US Semiconductor Industry," 2023.
[11]Source: Tom's Hardware, A. Shilov, "US Semiconductor Renaissance: All the Upcoming Fabs," August 29, 2022.
[12]Source: Nvidia, "Nvidia to Manufacture American-Made AI supercomputers in US for First Time," April 14, 2025.
[13]Source: Nvidia, "Nvidia to Manufacture American-Made AI supercomputers in US for First Time," April 14, 2025.
[14]Source: Nvidia, "Nvidia to Manufacture American-Made AI supercomputers in US for First Time," April 14, 2025.
[15]Source: PIIE, G. Hufbauer et al, "Industrial Policy Through the CHIPS and Science Act: A Preliminary Report," January 2025. See also, US Treasury, "US Department of the Treasury Releases Final Rules to Strengthen US Semiconductor Industry, October 22, 2024.
[16]Source: Macquarie Data Centers, "What is an AI Data Center, and How Does It Work?" July 15, 2024.
[17]Source: Google, "What is a GPU and its role in AI?" 2025. CPUs or central processing units continue to operate in data centers and are often low cost and great for some AI workloads like inference.
[18]Source: OpenAI, "Announcing the Stargate Project," January 21, 2025. Project Stargate will also actively invest in AI related semiconductor foundries and equipment and has included Arm, Nvidia, Oracle, Microsoft, and OpenAI in designing and integrating software and semiconductor innovation and production.
[19]Source: Financial Times, "Elon Musk Plans to Expand Colossus AI Supercomputer Ten-Fold," December 3, 2024.
[20]Source: Financial Times, T. Kinder et al., "CoreWeaver Looks to Raise Up to $2.7Bn in IPO," March 20, 2025. CoreWeave has relied on significant investment, public and private debt, as well as public equity markets to raise sufficient funds to invest in AI GPU capacity.
[21]Source: Reuters, S. Nellis et al., "Behind $500 Billion AI Data Center Plan, US Startups Jockey with Tech Giants," January 23, 2025.
[22]Source: Boston Consulting Group, V. Lee et al., "Breaking Barriers to Data Center Growth," January 20, 2025.
[23]Source: Reuters, "Meta in talks for $200 B AI Data Center Project," February 25, 2025. Meta has not confirmed rumors of $200 B project. Mark Zuckerberg has publicly affirmed plans to invest $65 B in 2025.
[24]Source: Boston Consulting Group, V. Lee et al., "Breaking Barriers to Data Center Growth," January 20, 2025. S&P Global as of January 2025.
[25]Source: Nvidia, "What is the Big Difference Between Deep Learning Training and Inference?" April 22, 2016.
[26]Source: Semianalysis, D. Patel el at., "DeepSeek Debates: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts," January 31, 2025.
[27]Source: Tom's Hardware, S. Shivlov, "DeepSeek Might not be as disruptive as claimed, firm reportedly has 50,000 Nvidia GPUs and spent $1.6 billion on buildouts," February 1, 2025. Subsequent analysis of DeepSeek revealed that the true cost of training the AI platform was approximately $1.6 billion dollars and the developer owned 50,000 Nvidia H100 GPUs.
[28]Source: New York Times, M. Farrell, "Wall St. Is All In on AI Data Centers. But Are They the Next Bubble," June 2, 2025.
[29]Source: "AI to Drive 165% Increase in Data Center Power Demand by 2030," February 4, 2025.
[30]Source: ASML, Intel co-founder Gordon Moore predicted that the number of transistors on a microchip would double every two years in 1975. The increase in the number of transistors increased the computing power of microchips. Today a semiconductor chip may have up to tens of billions transistors on the most advanced chips.
[31]Source: Barrons, T. Kim, "Nvidia CEO Jensen Huang Predicts "Hyper Moore's Law' Pace for AI," November 7, 2024.
[32]Goldman Sachs Research forecasts global power demand from data centers will increase 50% by 2027 and by as much as 165% by the end of the decade (compared with 2023). Source: Goldman Sach, "AI to drive 165% increase in data center power demand by 2030," February 4, 2025.

Notes

[33]Source: US Department of Energy, "DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers," December 20, 2024.
[34]Source: US Department of Energy, "DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers," December 20, 2024.
[35]Source: Boston Consulting Group, V. Lee et al., "Breaking Barriers to Data Center Growth," January 20, 2025.
[36]Source: US Department of Energy, "DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers," December 20, 2024.
[37]Source: Boston Consulting Group, V. Lee et al., "Breaking Barriers to Data Center Growth," January 20, 2025.
[38]Source: Power Magazine, D. Proctor, "Generational Shift – Data Centers Bring Change to Energy Landscape," January 7, 2025. Data Center Knowledge, C. Tozzi, 7 Top Data Center Sustainability Strategies for 2025," February 19, 2025.
[39]Source: A. Salzman, "Microsoft Wants Nuclear Power So Bad for AI that Its Paying Double, Experts Say," September 25, 2024.
[40]Source: Financial Times, M. Moore et al., "Amazon, Google and Meta support tripling of nuclear capacity by 2025," March 12, 2025.
[41]Source: Barrons, A. Salzman, "Trump's Nuclear Orders Could Loosen Radiation Standards. DOGE's Key Role," May 29, 2025.
[42]Source: Fierce Network, J. Madden, "Data Demand is Slowing, but AI Will Spike Uplink Data," August 28, 2024.
[43]Source: Source: Ericsson, "Impact of GenAI on Mobile Network Traffic,"
[44]Source: Ericsson, "Impact of GenAI on Mobile Network Traffic," 2024. See also Nokia, "Global Network Traffic Report," 2023.
[45]Source: Fierce Network, J. Madden, "Data Demand is Slowing, but AI Will Spike Uplink Data," August 28, 2024.
[46]Source: BBC, M. Wall, "How Mobile Phone Networks Are Embracing AI," June 19, 2024.
[47]Source: Ericsson, "Impact of GenAI on Mobile Network Traffic," 2024.
[48]Source: McKinsey & Company, "AI Power: Expanding data center capacity to meet growing demand," October 29, 2024. McKinsey estimates that personalized genAI may account for over 40% percent of the total demand for AI data center capacity globally by 2030.
[49]Source: BBC, M. Wall, "How Mobile Phone Networks Are Embracing AI," June 19, 2024.
[50]Source: BBC, M. Wall, "How Mobile Phone Networks Are Embracing AI," June 19, 2024.
[51]Source: BBC, M. Wall, "How Mobile Phone Networks Are Embracing AI," June 19, 2024.
[52]Source: Nokia, "Global Network Traffic Report," 2023.
[53]Source: Center for Security & Emerging Technology, S. Khan et al., "AI Chips: What They Are and Why They Matter: An AI Chips Reference," April 2020.
[54]Source: Center for Security & Emerging Technology, S. Khan et al., "AI Chips: What They Are and Why They Matter: An AI Chips Reference," April 2020.
[55]Source: Center for Security & Emerging Technology, S. Khan et al., "AI Chips: What They Are and Why They Matter: An AI Chips Reference," April 2020.
[56]Source: Center for Security & Emerging Technology, S. Khan et al., "AI Chips: What They Are and Why They Matter: An AI Chips Reference," April 2020.
[57]Source: Center for Security & Emerging Technology, S. Khan et al., "AI Chips: What They Are and Why They Matter: An AI Chips Reference," April 2020.
[58]Source: Center for Security & Emerging Technology, S. Khan et al., "AI Chips: What They Are and Why They Matter: An AI Chips Reference," April 2020.
[59]Source: Tom's Hardware, "Nvidia publishes first Blackwell B200 MLPerf results: Up to 4X faster than its H100 predecessor, when using FP4," August 28, 2024.
[60]Source: Tom's Hardware, :Nvidia Confirms Blackwell Ultra and Vera Rubin GPUs are on Track for 2025 and 2026," February 27, 2025. Nvidia Blackwell 300-series (Blackwell Ultra) "unofficial performance uplift enabled by Nvidia's b300-series will be around 50%."
[61]Source: Semianalysis, D. Patel et al., "DeepSeek Debates: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts," January 31, 2025.
[62]Source: Nvidia as of March 10, 2025.

Important Information

This report (the "report") has been prepared for the sole benefit of the intended recipient (the "recipient").

Significant events may occur (or have occurred) after the date of this report, and it is not our function or responsibility to update this report. The information contained herein, including any opinions or recommendations, represents our good faith views as of the date of this report and is subject to change at any time. All investments involve risk, and there can be no guarantee that the strategies, tactics, and methods discussed here will be successful.

The information used to prepare this report may have been obtained from investment managers, custodians, and other external sources. Some of this report may have been produced with the assistance of artificial intelligence ("AI") technology. While we have exercised reasonable care in preparing this report, we cannot guarantee the accuracy, adequacy, validity, reliability, availability, or completeness of any information contained herein, whether obtained externally or produced by the AI.

The recipient should be aware that this report may include AI-generated content that may not have considered all risk factors. The recipient is advised to consult with their meketa advisor or another professional advisor before making any financial decisions or taking any action based on the content of this report. We believe the information to be factual and up to date but do not assume any responsibility for errors or omissions in the content produced. Under no circumstances shall we be liable for any special, direct, indirect, consequential, or incidental damages or any damages whatsoever, whether in an action of contract, negligence, or other tort, arising out of or in connection with the use of this content. It is important for the recipient to critically evaluate the information provided.

Certain information contained in this report may constitute "forward-looking statements," which can be identified by the use of terminology such as "may," "will," "should," "expect," "aim," "anticipate," "target," "project," "estimate," "intend," "continue," or "believe," or the negatives thereof or other variations thereon or comparable terminology. Any forward-looking statements, forecasts, projections, valuations, or results in this report are based upon current assumptions. Changes to any assumptions may have a material impact on forward-looking statements, forecasts, projections, valuations, or results. Actual results may therefore be materially different from any forecasts, projections, valuations, or results in this report.

Performance data contained herein represent past performance. Past performance is no guarantee of future results.