State of the Global AI Industry
for Mr. Huang

Friedrich Bruck
School of Business, University of Texas at Arlington
10/10/2025
2019 words

Executive Summary:

The AI and technology market is expanding exponentially, driven by innovation in generative AI, cloud and super rapid computing, NVIDIA's GPU supply is falling short of demand. In order to remain the leader in the market and stay ahead of the competition, it's imperative that NVIDIA speeds up formation of alliances toward cloud computing and implementing AI into companies while investing in next-generation chip development to continue to scale up and become more sustainable.

The AI industry

History:

The artificial intelligence market has changed substantially over the past 30 years, having moved from just research to becoming a driver of global businesses' development. Early machine learning methods in the 1990's were the foundation for AI to be greatly improved upon. However, a bigger breakthrough occurred in 2012 as the capabilities of learning models surpassed traditional image recognition, which started the beginning of a new era of technology. Which was facilitated by global improvement in cloud computing and rapidly improving graphics processing units (GPUs). In the 2020s, generative AI technologies in large language models have been applied through various industries, from finance and health to auto and academia, displaying the global possibility for advancement and the large amount of uses for the new technology. AI has become a strategic industry for countries, further increasing demand for computing power and making the industry even more competitive. NVIDIA GPUs remain at the top of the totem pole, as they're the fundamental drivers of the adoption and demand for new technology through many industries. (Our World in Data, 2023).

Size and Key Markets:

The global artificial intelligence (AI) market is expanding exponentially, it was valued at $233 billion in 2024, and it's projected to surpass $1.77 trillion by 2032, with a compound annual growth rate of 29% (Fortune Business Insights, 2024).

Market Projection

Year	AI Market in USD (Billions)
2025	294
2026	380
2027	491
2028	634
2029	819

2030	1059
2031	1368
2032	1770

Even though North America leads the space with its secure digital infrastructure and high demand, the highest rate of growth is in Pacific Asia, with China's, India's, and Japan's huge investments and developments. Government regulations like the E.U. 's AI Act and the significant impact it has on AI ethical standards are influencing the development and release of new AI (European Parliament, 2024). AI is also being used for low-cost purposes in emerging economies in Southeast Asia, Africa, and Latin America, demonstrating the new technologies' global relevance and reach.

Nature of the Industry and Key Competitors:

The artificial intelligence industry is rich in innovation, and full of high upfront investment with harsh competition at the global level. NVIDIA's key competitors include Advanced Micro Devices (AMD), Intel, and Amazon Web Services (AWS), Microsoft Azure, and Google Cloud. AMD's MI300 accelerators became popular for big workloads that require high performance, showed in 2024 revenues of USD 22.7 billion (AMD, 2024). Intel competes with its semiconductors and Gaudi AI accelerators even after falling behind in discrete GPUs (Intel, 2024). At the same time, cloud vendors are aiming toward vertical integration, integrating patented chips globally to differentiate their products and retain customers.

There is vast competition in foreign markets. In China, Huawei and Alibaba are investing heavily in AI hardware, while in the European Union, startups like Graphcore are receiving aid to diversify supply. The political aspect is worsened by U.S. export restrictions on China’s access to high-end chips, redefining accessibility and the dynamics of competition. As AI supply globalization occurs, success becomes not just about technological innovation but also about navigating regulations, supply chain inefficiencies, and partnerships. For NVIDIA to remain the leader, there needs to be a balance in hardware innovation with global strategizing that fulfills new demand while acknowledging geopolitical situations.

Competitors’ Revenue in 2024

Company:	Revenue in USD (Billions):
NVIDIA	60.9
AMD	25.8
Intel	53.1
TSMC	88.34
Amazon AWS	107.6
Microsoft Cloud	46.7

Technological Changes

NVIDIA has to evolve with the fast rate of transformation for the AI industry. In order to maintain pace with newer alternatives like Google's TPUs to accelerate machine learning, NVIDIA must move beyond conventional GPU design and focus on efficiency, scalability, and AI accelerators because AI will increasingly rely on resources and will keep on consuming more power. Since software environments of AI, compilers, and tools are as crucial as the hardware, new computing development ought to be met with quick, energy-expensive models for smaller devices that operate at high speeds instead of solely focusing on internet dependent models (Smith, 2023). Additionally, as hyperscalers like Microsoft and AWS are developing their own AI infrastructure, NVIDIA must respond by developing its software ecosystem, offering integrated AI platforms, boosting partnerships, and investing in up-and-coming technologies to stay afloat in the years to come (NVIDIA, 2024; Jouppi et al., 2023). As much as quantum computing is still in the process of development, it has the potential to transform the AI computing world in its entirety. In the meantime, NVIDIA will need to be flexible in engineering and accommodating to shifting competition laws of AI worldwide. Technical vision adaptability will also be needed as a way of maintaining long-term leadership in the sector.

Current Technological Challenges:

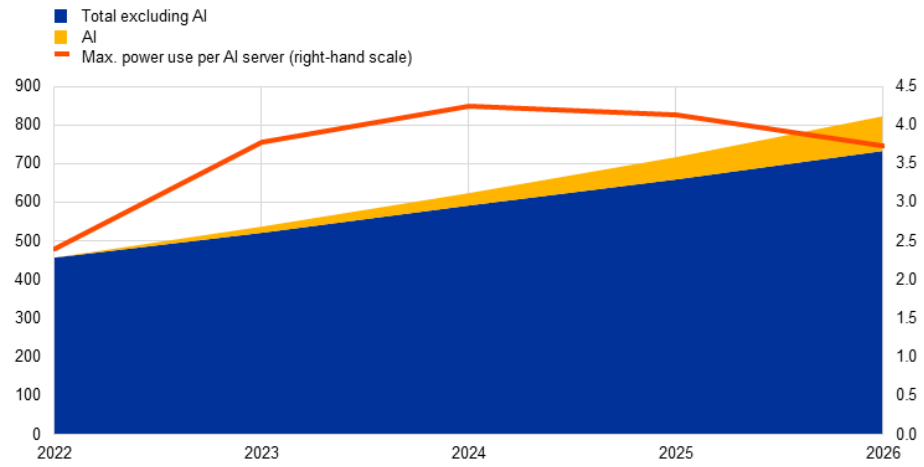
Many technological issues pose threats for NVIDIA. One of the biggest issues is increasing costs for training new large models. New systems such as GPT-4 require thousands of GPUs along with a lot of energy and special data centers (NVIDIA, 2023) which is met with demand for them to be more efficient and able to handle large tasks. These requirements lead to supply bottlenecks because of the limited availability and how hard it’s becoming to innovate in this space. It’s also important keeping in mind NVIDIA's reliance on manufacturing partners like TSMC which is in a very politically uncertain country. Especially because companies such as Amazon Web Services are increasing their home-grown chip supply to reduce reliance on NVIDIA gear because of this. While NVIDIA is one of the leaders still, the industry is furiously changing due to companies and countries looking for sovereignty.

Energy Issue:

Sustainability is also an issue. AI data center’s huge energy requirements are targets for government regulation, especially in the E.U., which aids companies that reduce emissions. With ever increasing AI energy consumption and energy prices, NVIDIA has responded with breakthroughs such as the Hopper and Grace Hopper superchips, which show the technical ability to become more energy efficient. Within developing markets, where power infrastructure might be less powerful, supplying even more cost-effective hardware will be required for NVIDIA AI to reach every corner of the world.

Projected growth in the electricity demand of data centres

(left-hand scale: terawatt-hours; right-hand scale: kilowatts)



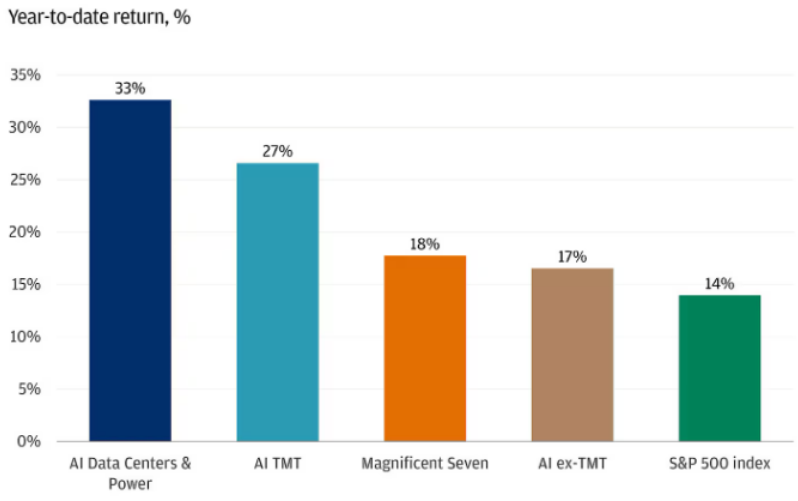
Sources: IEA, Goldman Sachs and ECB staff calculations.

Notes: AI electricity demand between 2022 and 2026 is linearly interpolated with values for 2026 taken from IEA forecasts. The orange line shows the projected maximum power use of an AI server, highlighting expected efficiency gains.

Future Technological Opportunities and Challenges:

In the future, a couple challenges and opportunities will define NVIDIA's journey. Scaling limits of current technology can become a challenge: as models increase in size, modern parallelization might become less efficient. This, however, opens opportunities for leadership in new fields like quantum computing that are being researched globally (García-Martín et al., 2022). There is a second opportunity in edge AI with autonomous vehicles and healthcare. These applications will drive global demand for expert processors outside of mainstream data centers.

Keeping dominance over the industry will depend on being able to globalize technology. NVIDIA will have to continue balancing high-performance innovation with energy efficiency, supply chain strengths and ability to work around regulations. History’s technology change has benefitted NVIDIA, but the future lies in breakthroughs in global business in an increasingly competitive and complex world.



Sources: J.P.Morgan; Bloomberg Finance L.P. Data as of September 25, 2025. Note: U.S. companies.

Cultural, Geopolitical, and Environmental Factors

The global AI industry is influenced by cultural, geopolitical, and environmental situations that all affect NVIDIA's global business plan. Geopolitically, the rivalry between powers like the U.S. and China has already affected NVIDIA's market access. U.S. export regulations to China limit GPU sales, whilst China decreases its reliance on foreign technology (U.S. Department of Commerce, 2023).

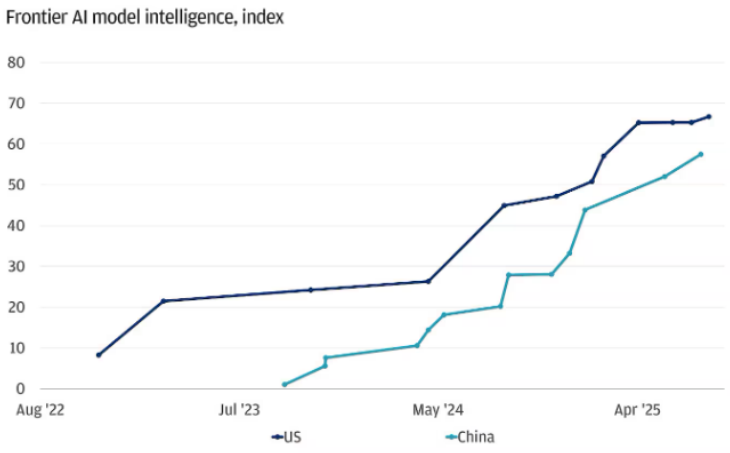
The U.S. China technological rivalry has intensified supply chain instability for NVIDIA as well as the entire technology market. Tariffs on components in semiconductors along with limitations on exports of advanced chips have increased production costs and limited market accessibility for international companies. Change is forcing firms to set up supply chains in countries like South Korea, Malaysia, and Vietnam. The decoupling of technological sectors has also encouraged China to invest even more aggressively in domestic semiconductor manufacturing companies like Huawei and SMIC, reducing reliance on U.S. companies. The U.S. *CHIPS and Science Act* incentivizes domestic chip manufacturing but doesn't offset the present costs of reshoring production. Because of this, NVIDIA faces a challenge: reconfiguring supply routes while staying competitive through market fragmentation and protectionism.

Geopolitical threats also exist in Taiwan, where TSMC manufactures most of NVIDIA's chips, which exposes the company to supply shocks in the event of war tensions increasing. Corruption and poor governance of emerging markets also threaten business with added instability, as well as ongoing conflict in Eastern Europe and the Middle East creating international supply chain volatility.

Trends:

Social and cultural trends change the industry's trajectory. Consumer, business, and governmental expectations for sustainability and responsible use of AI are changing. For example, the European Union's AI Act prioritizes consumer protection and transparency influences the pricing and marketing of AI related products throughout the union (European Commission, 2023). In Asia, demand is highest for low-cost and "mobile-first" AI. Specifically

in India and Southeast Asia's sizable markets, where consumption is very sensitive to price, and technology isn't necessarily available to all. In North America on the other hand, AI aids in productivity which influences marketing strategies to emphasize innovation and performance. Acknowledging these regional consumer sentiments are critical for NVIDIA to strategically position itself in different environments.



Sources: Artificial Analysis. Data as of September 18, 2025. Note: : Artificial Analysis Intelligence Index v3.0 incorporates 10 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME 2025, IFBench, AA-LCR, Terminal-Bench Hard, τ^2-Bench Telecom

Environment:

Consideration for the environment is gaining influence in shaping demand and regulation. Training large AI models requires ridiculous amounts of electricity, and the impact on the environment is under scrutiny. Consumers in Europe and Asia largely desire technologies that are environmentally mindful (PwC, 2022). Data centers and firms also seek power-efficient hardware to maximize profit and minimize emissions. NVIDIA's focus on power-efficiency puts the company in a position to satisfy these dreams, but more innovation will need to take place.

Implications:

For NVIDIA, these cultural, geopolitical, and environmental factors have huge implications. Geopolitical conflict and trade prohibitions in markets like China or Taiwan mean that it's necessary to secure supply chains and politically stable markets. Technology adoption differs between countries too, so AI marketing and choice should differ. For example, data privacy in Europe, affordability in emerging economies like India, and automation industrial economies like Germany and Japan. Similarly, environmental concerns provide an opportunity for NVIDIA to market its energy efficient GPUs as the solution while working under the paradigms of sustainability and government allowance. In the end, growth will require more than technological breakthroughs, like proactive measures like limiting geopolitical risk, adapting products to local demands, and using green innovation to interest governments, businesses, and consumers globally.

Recommendations:

Emphasize diversification, innovation, and alignment with global market trends.

Diversify Supply Chains Outside Taiwan (Investment: $3–5B; Timeline: 3–5 years).

NVIDIA is dependent on TSMC and exposed to geopolitical risk (U.S. Department of Commerce, 2023). Diversification is necessary, especially with competitors like AMD and Intel also reorganizing manufacturing. Agreements with Samsung and GlobalFoundries, with U.S. manufacturing increases with the CHIPS Act, will reduce risk and ensure NVIDIA's competitiveness (White House, 2022).

Enhance Energy-Efficient Development (Investment: $2–3B/year; Ongoing).

Continued development of Hopper and Grace Hopper models and research into liquid cooling will improve efficiency (PwC, 2022), which is possible but must stay in front of competitors like AMD who are pursuing energy efficiency.

Expand Edge AI Product Lines (Investment: $1–2B; 2–3 years).

Specialized chips like Jetson will be useful in autonomous cars and healthcare practices (García-Martín et al., 2022), with huge return potential. NVIDIA will have to balance price and performance in order to gain a share of the market where Intel and Qualcomm are already established.

Strengthen Global Strategic Partnerships (Investment: $500M–$1B; 1–3 years).

Deals with cloud providers, smart car manufacturers, and health institutions will include NVIDIA products in AI solutions (Fortune Business Insights, 2024). NVIDIA must prioritize important partnerships in the face of strong AWS, Microsoft, and Huawei ecosystems.

Advanced Micro Devices. (2024). *Annual report 2024*. https://www.amd.com

Advanced Micro Devices. (2025, January 30). *AMD reports fourth quarter and full year 2024 financial results* [Press release]. AMD Investor Relations. https://ir.amd.com

Amazon.com, Inc. (2025, February 1). *Amazon.com announces fourth quarter results / 2024 annual report* [Press release]. Amazon Investor Relations. https://ir.aboutamazon.com

European Commission. (2023). *Artificial Intelligence Act: Ensuring safe and trustworthy AI*. https://digital-strategy.ec.europa.eu

Fortune Business Insights. (2024). *Artificial intelligence market size, share & COVID-19 impact analysis, by component, by technology, by end-use, and regional forecast, 2024–2032*. https://www.fortunebusinessinsights.com

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2022). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing, 162*, 152–164. https://doi.org/10.1016/j.jpdc.2022.01.006

Intel Corporation. (2024). *Annual report 2023*. https://www.intel.com

Intel Corporation. (2025, January 25). *Intel reports fourth-quarter and full-year 2024 financial results* [Press release]. Intel Investor Relations. https://www.intc.com

Jouppi, N. P., Young, C., Patil, N., et al. (2023). A domain-specific architecture for AI acceleration. *Communications of the ACM, 66*(7), 56–67. https://cacm.acm.org/magazines/2023/7/273246-domain-specific-architecture-for-ai-acceleration/fulltext

J.P. Morgan. (2024, July 19). *4 trends we're watching in the artificial intelligence investing space.* J.P. Morgan Insights. https://www.jpmorgan.com/insights/markets-and-economy/top-market-takeaways/tmt-4-trends-were-watching-in-the-artificial-intelligence-investing-space

Microsoft Corporation. (2024, July 25). *FY24 Intelligent Cloud performance* [Investor earnings release]. Microsoft Investor Relations. https://www.microsoft.com/en-us/Investor

NVIDIA. (2023). *Accelerating AI with GPUs: The path forward*. NVIDIA Corporation. https://www.nvidia.com

NVIDIA. (2024). *NVIDIA AI platform: Accelerating AI innovation*. Retrieved September 30, 2025, from https://www.nvidia.com/en-us/ai-platform

OpenAI. (2018). *AI and compute*. OpenAI Blog. https://openai.com/research/ai-and-compute

Our World in Data. (2023). *Artificial intelligence and computing power*. https://ourworldindata.org

PwC. (2022). *AI and sustainability: Responsible innovation in the green transition*. PricewaterhouseCoopers. https://www.pwc.com

Smith, J. (2023). Integrated AI platforms: The new frontier for enterprise AI. *Artificial Intelligence Review, 18*(3), 112–129. https://link.springer.com/article/10.1007/s10462-023-10123-4

Taiwan Semiconductor Manufacturing Company (TSMC). (2025). *2024 annual report*. TSMC Investor Relations. https://www.tsmc.com/investors

U.S. Department of Commerce. (2023, October 17). *Commerce implements new export controls on advanced computing and semiconductor manufacturing items to the People's Republic of China (PRC)* [Press release]. U.S. Department of Commerce. https://www.commerce.gov/news/press-releases/2023/10/commerce-implements-new-export-controls-advanced-computing-and

White House. (2022, August 9). *CHIPS and Science Act of 2022: A guide to strengthening America's semiconductor supply chains*. The White House. https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/

World Economic Forum. (2023). *Global risks report 2023*. https://www.weforum.org