CHIPPING POINT
TRACKING ELECTRICITY CONSUMPTION AND EMISSIONS FROM AI CHIP MANUFACTURING

GREENPEACE

Authors

Katrin W.
Summer J.
Rachel Y.
Alex de Vries

Acknowledgement and special thanks to

Erin Newport, Yeonho Yang, Lena Chang, Insung Lee, Dayon Kang, Coco Wu, Shaolei Ren, Tsai-Chi Kuo

Editor

Kathryn Miller

Disclaimer

This report was originally written in English, and was subsequently translated into briefings in Chinese, Korean. In the event of any discrepancies, the English version prevails. This report is ONLY for the purposes of information sharing, environmental protection and the wider public interest. It should not be used as a reference for any investments, or any other such decision making process. If so used, Greenpeace is exempt from any liabilities arising from such use.

The content of this report is based on officially published information obtained independently from publicly available sources disclosed by the company. Greenpeace East Asia does not guarantee the promptness, accuracy and integrity of the information contained in this report.

Contents

04 **Executive Summary**

06 **Key Findings**

09 **Introduction**

10 The Growing AI Industry

13 **The Energy Demand of AI**

14 Energy Demand from Data Centers

16 Energy Demand from AI Chipmaking

25 **Recommendations**

Detailed information regarding the research methodology employed to investigate AI chip manufacturing energy demand is provided in the supplementary document: **Appendix A. Methodology: A Bottom-Up Analysis of Energy Consumption and Carbon Footprint in AI Chip Production**

Executive Summary

Executive Summary

The artificial intelligence (AI) race is fueling billions of dollars in investment in AI infrastructure, led by companies such as Nvidia, Microsoft, Meta, and Google. A significant portion of this spending is allocated to the manufacture of AI chips, specialized hardware that is essential to train and deploy advanced AI models.

As the use of AI grows, there has been increasing focus on the electricity and climate impact of AI data centers, including growth in fossil fuel power generation. However, much less attention has been paid to the climate impact of AI chipmaking, a process that is concentrated in East Asia, especially in South Korea and Taiwan.

East Asian firms dominate the manufacture of advanced graphics processing unit (GPU) chips, which are essential for the operation of AI models. The Taiwan Semiconductor Manufacturing Company Limited (TSMC) leads the manufacturing of logic chips for AI-specific GPUs, which are produced in its Taiwan fabs. SK hynix, Samsung, and Micron manufacture the memory chips used in AI hardware, primarily in South Korea and Japan.

East Asia's electricity grids rely heavily on fossil fuels, including coal, oil and gas, creating a challenge for AI chipmakers and an opportunity for renewable energy development. In 2023, 58.5% of the electricity supply in South Korea was generated from fossil fuels, 68.6% in Japan, and 83.1% in Taiwan.[1,2,3]

To meet the new demand from electronics manufacturers, dependence on fossil fuel electricity generation is intensifying. In 2024, South Korea's government approved the construction of a 1 gigawatt (GW) liquid natural gas (LNG) combined heat and power plant for SK hynix in the Yongin General Industrial complex. The government also plans to build 3 GW of LNG capacity for Samsung in the National Semiconductor Cluster. When operational, both projects will lead to an increase in greenhouse gas emissions and air pollution, environmental impacts that could be avoided if new demand is met with renewable energy.

To assess the global climate impact of AI chipmaking, this research estimates the electricity consumption of the dominant AI chip models (Nvidia A100, H100, H200, B100/200 and AMD MI300X) in 2023 and 2024 and emissions from their electricity consumption using a bottom-up approach.

The analysis uses publicly available data and market analyses on production and wafer demand for AI models as well as published electricity requirements for the production process. By identifying suppliers and manufacturing locations, the study determines the electricity mix used in production and the associated carbon emissions. Based on forecasted demand and supply for AI logic and memory wafers, the estimations are then extended to 2030.

Given the projected growth of AI and the crucial role of chip manufacturing, increased scrutiny of this sector's carbon footprint is urgently needed. While some major tech brands have strengthened their commitments to renewable energy in recent years, the AI chip sector lags far behind on sustainability.

The research also includes recommendations for how electronics manufacturers can reduce emissions from AI chipmaking. East Asia-based chipmakers can source renewable energy locally through high-impact sourcing methods, such as building and investing in renewable energy generation facilities and signing long-term power purchase agreements (PPAs). Likewise, tech companies such as Nvidia, Microsoft, Meta, and Google must support their suppliers to increase renewable energy procurement and should target 100% renewable energy across their supply chains by 2030.

1. EPSIS, "Generation Output by Energy Source," 2024, accessed February 10, 2025, https://epsis.kpx.or.kr/epsisnew/selectEkgeGepGesGrid.do.
2. Ministry of Economic Affairs, Energy Administration (MOEAEA), "Electricity Statistics," 2024, accessed February 10, 2025, https://www.esist.org.tw/database/search/electric-generation.
3. Ministry of Economy, Trade and Industry (METI), "General Energy Statistics," 2024, accessed February 10, 2025, https://www.enecho.meti.go.jp/statistics/total_energy/results.html.

Key Findings

Key Findings

By 2030, global electricity demand for artificial intelligence (AI) chipmaking is estimated to reach as much as 37,238 gigawatt hours (GWh), a 170-fold increase compared to 2023 and more than the total current electricity consumption of Ireland.[4]

- East Asia has been the primary manufacturing location for key hardware components of commercially available AI models. For this reason, growing electricity demand places a disproportionate burden on electricity grids in East Asia, especially in South Korea and Taiwan.

Global electricity consumption from AI chipmaking increased by more than 350% year on year, from 218 GWh in 2023 to nearly 984 GWh in 2024.

- In 2024, electricity consumption from AI chip manufacturing in Taiwan reached 375.8 GWh, equivalent to that of roughly 93,000 Taiwan households in 2023, a 350.6% increase year on year. Electricity usage in Taiwan is projected to increase by 12%-13% by 2030, in large part to meet surging demand from the semiconductor industry.[5]
- In South Korea, electricity consumption from AI chip manufacturing more than doubled between 2023 and 2024, from 134.6 GWh to 315.2 GWh. By 2050, the Yongin area, which is densely packed with semiconductor fabs, is expected to require over 10 gigawatts of electricity demand, accounting for 25% of the total power demand in the metropolitan area and potentially causing a burden on the national electricity grid.[6]

Global emissions from electricity consumption related to AI chipmaking increased more than 4-fold in 2024, rising from 99,200 metric tons of CO_2 equivalent in 2023 to 453,600 metric tons of CO_2 equivalent in 2024. The emissions were driven in large part by heavy reliance on fossil fuels in East Asia's power grids.

- Emissions from AI chipmaking in Taiwan rose from 41,200 metric tons in 2023 to 185,700 metric tons in 2024, a more than 4-fold increase.
- In South Korea, emissions related to AI chipmaking more than doubled year on year, growing from 58,000 metric tons in 2023 to 135,900 metric tons in 2024.
- In 2024, emissions from AI chip manufacturing in Japan reached 132,100 metric tons.

Across East Asia, growth in electricity consumption from chipmaking is being used to justify false climate solutions such as gas and nuclear.

- South Korea's government approved the construction of a 1.05 gigawatt (GW) liquefied natural gas (LNG) combined heat and power plant for SK hynix in the Yongin General Industrial complex. The government also plans to build 3 GW of LNG capacity for Samsung in the National Semiconductor Cluster.[7,8]
- In Taiwan, the burgeoning AI sector has been used as a justification for an expansion of fossil fuel capacity. Proposed construction of a new LNG receiving terminal at Keelung Port involves converting four 500 megawatt (MW) oil units into two 2.6 GW gas turbines.[9] Industry representatives have argued that stable power is crucial to fuel the rise of AI, saying that Nvidia plans to set up its Taiwan headquarters in the north.[10] Likewise, the chairman of Pegatron Corp, a major AI manufacturer, has publicly advocated for more nuclear power capacity in Taiwan.[11]

To prevent an increase in emissions from East Asia, new electricity demand for AI chipmaking must be met with renewable energy.

- Tech companies such as Nvidia, Microsoft, Meta, and Google must strengthen their supply chain emissions reduction efforts by achieving 100% renewable energy by 2030 across their supply chains.
- Chip manufacturers need to set a 100% renewable energy target by 2030 in response to clients' requirements and source renewable energy locally through high-impact sourcing approaches, such as building and investing in renewable energy generation facilities, exploring and signing long-term power purchase agreements (PPAs).

4. Central Statistics Office of Ireland, "Large Energy Users accounted for 30% of Metered Electricity Consumption in 2023," July 23, 2024, accessed February 28, 2025, https://www.cso.ie/en/releasesandpublications/ep/p-mec/meteredelectricityconsumption2023/keyfindings/.
5. Ministry of Economic Affairs, Energy Administration (MOEAEA), "2023 Power Resources Supply and Demand Report," July 15, 2024, accessed March 5, 2025, https://www.moeaea.gov.tw/ECW/populace/news/News.aspx?kind=1&menu_id=41&news_id=33815.
6. KBS, "Semiconductor cluster to 'use 1/4 of metropolitan electricity," July 7, 2023, accessed March 19, 2025, https://news.kbs.co.kr/news/pc/view/view.do?ncd=7718342
7. Business Korea, "Green Light for Yongin Semiconductor Cluster's 1.05GW LNG CHP Plant," August 9, 2024, accessed February 10, 2025, https://www.businesskorea.co.kr/news/articleView.html?idxno=222719.
8. Newstree Korea, "Samsung and Hynix to Move into Yongin Semiconductor... What to Do About Renewable Energy," August 5, 2024, accessed March 31, 2025, https://www.newstree.kr/newsView/ntr202408020010.
9. Environmental Information Center, "Four cases were received and two were submitted to the Environmental Impact Assessment Conference," January 20, 2025, accessed February 28, 2025, https://e-info.org.tw/node/240623.
10. Recessary, "Keelung's fourth environmental impact assessment is coming soon! The business community shouts "Industry needs electricity" to support the reconstruction and seize the opportunity of Nvidia to set up its headquarters," January 13, 2025, accessed February 19, 2025, https://www.reccessary.com/zh-tw/news/world-market/lng-receiving-terminal-dispute?fbclid=IwY2xjawIb96FIeHRuA2FlbQIxMAABHb6WnrLQqQQyCqbR7QCq1oD0IjAkPH0TjOe91IbM0uGfb46i06Z6RODKFQ_aem_cMgNGJeQjAN7ydAo1Q_WKg.
11. TVBS, "Pegatron chair backs nuclear energy to cut carbon emissions," November 12, 2024, accessed February 19, 2025, https://news.tvbs.com.tw/english/2683129.

Introduction

Introduction

The Growing AI Industry

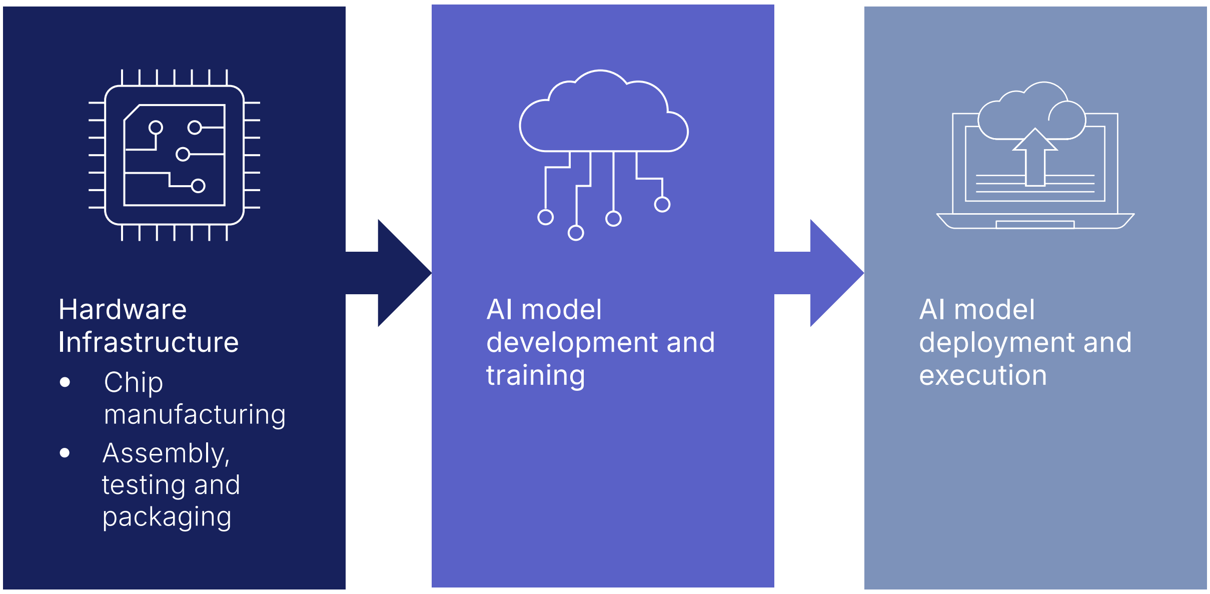
The field of artificial intelligence (AI) has made rapid progress over the past few years. The advanced learning and reasoning capacity of AI is transforming industries, driving innovation, efficiency, and cost savings, and it will continue to play an important role in the years to come. The global AI market is expected to expand at a compound annual growth rate (CAGR)[12] of between 40% and 55% from 2023 to 2027, reaching up to US$990 billion.[13]

Alongside the opportunities presented by AI are significant environmental, social and ethical implications. The substantial energy consumption and environmental footprint of large-scale AI deployments are becoming major concerns.

Current discussions have largely centered on the electricity demand of data centers. These facilities, which support the AI training and inference processes, are placing increasing stress on electricity grids and sustainability due to their surging electricity capacity demand. According to the International Energy Agency (IEA), data centers currently account for around 1% of global electricity consumption.[14]

However, the assessment of AI energy consumption should extend beyond data centers to encompass the broader landscape of the AI industry, notably the energy-intensive process of upstream chip manufacturing.

Figure 1.1 **The energy footprint of AI: From chip to execution**



12. Compound annual growth rate (CAGR) measures the mean annualized growth rate for compounding values over a given time period.
13. Bain & Company, "AI's Trillion-Dollar Opportunity," tech report, 2024, accessed February 10, 2025, https://www.bain.com/insights/ais-trillion-dollar-opportunity-tech-report-2024/.
14. International Energy Agency (IEA), "What the Data Centre and AI Boom Could Mean for the Energy Sector," IEA Commentaries, October 18, 2024, accessed February 10, 2025, https://www.iea.org/commentaries/what-the-data-centre-and-ai-boom-could-mean-for-the-energy-sector.

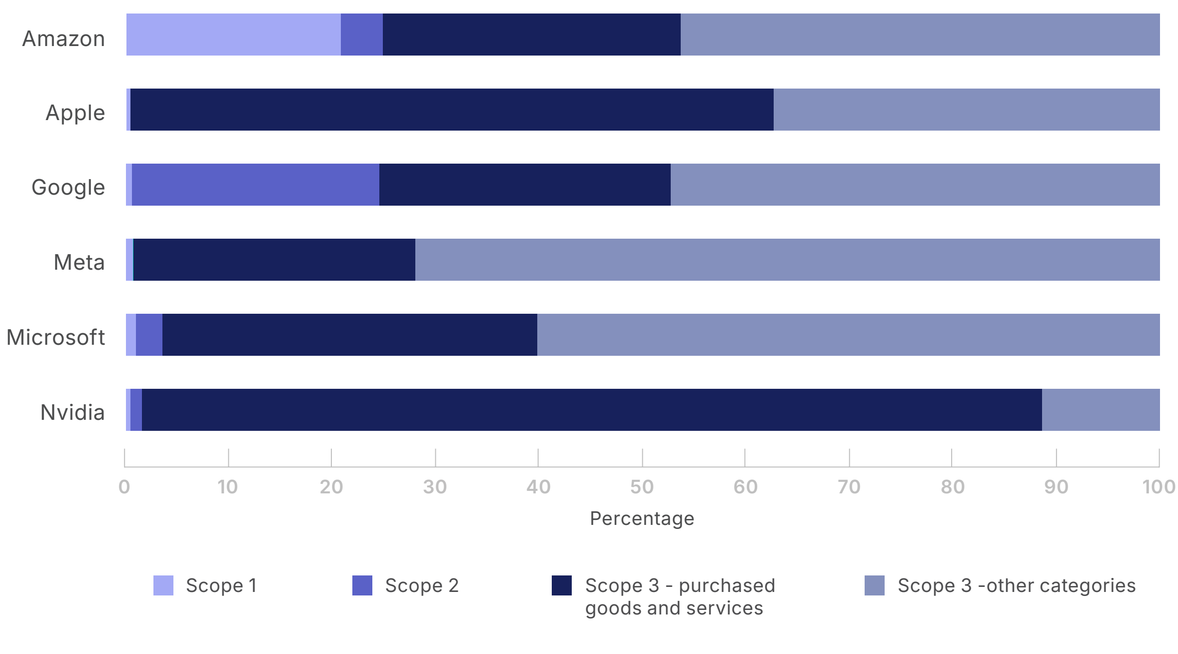
Chip manufacturing is essential to the development of AI. AI relies heavily on specialized advanced chips such as graphics processing units (GPUs) and high bandwidth memory (HBM). The escalating demands of modern AI models are met by the combined power of GPUs and HBM. GPUs provide the necessary parallel processing power and offer immense computational power compared to traditional CPUs, while HBM overcomes memory bandwidth limitations inherent in large models, enabling more complex AI systems.

While some tech giants are pioneering to develop their own in-house AI chips, the prevalent practice among tech companies is to procure standard or customized AI chip models designed by fabless design companies, like Nvidia, which commission the production of their products to chipmakers (foundries). Chip manufacturing is extremely energy intensive. A large foundry can consume up to 100 megawatt-hours (MWh) of electricity per hour.[15]

Tech giants are at the forefront of the rapidly expanding AI industry, investing heavily in AI research and infrastructure. In 2024 alone, Google, Microsoft, Amazon, and Meta collectively spent US$246 billion, fueling demand for energy-intensive AI chips.[16] This high demand for energy contributes to the tech companies' Scope 3 emissions (specifically the 'Purchased Goods and Services' category), representing a substantial portion of their overall greenhouse gas footprint **(Figure 1.2).**[17,18]

The scale of tech companies' Scope 3 emissions typically eclipses that of their operational emissions, underscoring the urgent need for supply chain decarbonization. Despite this, the adoption of strategies to achieve 100% renewable energy across supply chains remains limited within the industry. The shortfall in supply chain management is likely to exacerbate emissions across the AI industry, undermining other sustainability efforts.

Figure 1.2 **Relative weight of leading tech companies' emissions in the 2023 fiscal year**



Source: Company disclosures.

15. Sánchez J. et al., "IDSEM, an invoices database of the Spanish electricity market," *Scientific Data* 9, 786 (December 26, 2022), https://doi.org/10.1038/s41597-022-01885-3.
16. Financial Times, "Big Tech lines up over $300bn in AI spending for 2025," February 7, 2025, accessed February 20, 2025, https://www.ft.com/content/634b7ec5-10c3-44d3-ae49-2a5b9ad566fa.
17. McKinsey & Company, "Bringing Fab Energy Efficiency," white paper, accessed February 10, 2025, https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/operations/pdfs/bringing_fabenergyefficiency.ashx.
18. Tech company GHG emissions: Scope 1 emissions are direct emissions emit from sources that are owned or controlled by the company, such as company vehicles or generators at offices and data centers; Scope 2 emissions are indirect emissions deriving from the purchase of electricity, heating/cooling, hot/chilled water or steam for own consumption; Scope 3 emissions are indirect emissions arise from other sources in value chain as a consequence of the activities of the company, such as from suppliers, the use of consumer devices, and business travel.

Major suppliers of AI-related chips, such as the Taiwan Semiconductor Manufacturing Company Limited (TSMC), SK hynix, Samsung and Micron, base their major manufacturing facilities in East Asia. However, the region's electricity grid is dominated by fossil fuels (coal, oil and natural gas), with 2023 figures showing 83.1% reliance in Taiwan, 68.6% in Japan, and 58.5% in South Korea.[19,20,21] Although these chip manufacturers have been transitioning to renewable energy through renewable energy sourcing, the speed of transition is slow.

Greenpeace East Asia collaborated with Alex de Vries[22] to research the overlooked electricity consumption from AI-related chip manufacturing processes and the environmental impact on the East Asia region. The outcome of the research suggested that tech companies work with AI hardware suppliers towards 100% renewable energy across supply chains by 2030.

19. EPSIS, "Generation Output by Energy Source," 2024, accessed February 10, 2025, https://epsis.kpx.or.kr/epsisnew/selectEkgeGepGesGrid.do.
20. Ministry of Economic Affairs, Energy Administration (MOEAEA), "Electricity Statistics," 2024, accessed February 10, 2025, https://www.esist.org.tw/database/search/electric-generation.
21. Ministry of Economy, Trade and Industry (METI), "General Energy Statistics," 2024, accessed February 10, 2025, https://www.enecho.meti.go.jp/statistics/total_energy/results.html.
22. The founder of Digiconomist and a PhD candidate at VU Amsterdam, where he focuses on investigating the environmental sustainability of emerging technologies.

The Energy Demand of AI

The Energy Demand of AI

Energy Demand from Data Centers

In 2022, the electricity consumption of global data centers was estimated at 240-340 terawatt hours (TWh), representing approximately 1-1.3% of the global electricity demand.[23]

Data Center Trends

- **Growth in Capacity –** The AI boom has driven up the average capacity of a data center, from facilities ranging from 5-10 MW, to larger 30 MW, to current hyperscale projects with energy capacity in excess of 100 MW.[24]
- **Growth in Numbers –** Companies are racing to build more data centers to meet the capacity demands of AI. In the United States, the number of data centers has expanded from around 2,600 in 2021 to more than 5,300 facilities in 2024.[25,26] The capacity of global hyperscale data centers has doubled in just four years.[27]

Despite efficiency improvements, the rapid growth in energy-intensive workload is likely to drive substantial electricity demand growth over the next few years. Goldman Sachs, a multinational investment bank, estimated that the data center electricity consumption driven by AI will increase at the rate of 200 TWh per year between 2023 and 2030, which could translate to a more than doubling of data center CO_2 emissions by 2030 compared to 2022.[28,29]

From a hardware and equipment perspective, data center electricity demand is driven by three main categories: computing equipment; cooling systems; and other components. Computing hardware (servers, storage systems, and network infrastructure) is the largest consumer of power, accounting for up to half of the total electricity consumption. Cooling systems, such as liquid cooling and air conditioning, typically account for around 40% of a data center's energy consumption, and are essential for maintaining an optimal operating temperature and preventing hardware failures. The remaining electricity consumption is attributed to operational and maintenance activities such as power distribution, lighting, and monitoring.[30,31]

From a workload perspective, energy consumption of AI data centers can be categorized into model development, training and inference. While model development accounts for up to 10% of total energy usage, the remaining 90% is consumed by power-intensive training and inference phases.[32] The significant increase in data center demand has been largely attributable to AI model training, where vast volumes of data are processed by the algorithm to learn patterns and relationships. Inference was already dominant in Google's AI-related energy consumption before the AI hype.[33] Nevertheless, AI inferencing is expected to surpass training and eventually become a larger workload in the longer term, and account for a higher share of energy.[34]

---

23. International Energy Agency (IEA), "Data Centres and Data Transmission Networks," IEA Energy System, accessed February 10, 2025, https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks.
24. IBM, "Hyperscale Data Center," IBM Think Topics, accessed February 10, 2025, https://www.ibm.com/think/topics/hyperscale-data-center.
25. United States International Trade Commission (USITC), "Data Centers Around the World," Executive Briefing, May 2021, accessed February 10, 2025, https://www.usitc.gov/publications/332/executive_briefings/ebot_data_centers_around_the_world.pdf.
26. Cloudscene, "United States of America | Data Center Market Overview | Cloudscene," accessed February 10, 2025, https://cloudscene.com/market/data-centers-in-united-states/all.
27. "Hyperscale Data Centers Hit the Thousand Mark; Total Capacity is Doubling Every Four Years," SRG Research, April 17, 2024, accessed February 10, 2025, https://www.srgresearch.com/articles/hyperscale-data-centers-hit-the-thousand-mark-total-capacity-is-doubling-every-four-years.
28. Goldman Sachs, "AI, data centers and the coming US power demand surge," equity research, April 28, 2024, accessed February 20, 2025, https://www.goldmansachs.com/pdfs/insights/pages/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf.
29. Goldman Sachs, "AI/data centers' global power surge and the Sustainability impact," April 28, 2024, accessed February 10, 2025, https://www.goldmansachs.com/images/migrated/insights/pages/gs-research/gs-sustain-generational-growth-ai-data-centers-global-power-surge-and-the-sustainability-impact/sustain-data-center-redaction.pdf.
30. Cai S. and Gou Z., "Towards Energy-Efficient Data Centers: A Comprehensive Review of Passive and Active Cooling Strategies," *Energy and Built Environment* (September 1, 2024), https://doi.org/10.1016/j.enbenv.2024.08.009.
31. DataSpan, "Data Center Cooling Costs", August 23, 2022, accessed February 10, 2025, https://dataspan.com/blog/data-center-cooling-costs/.
32. IEA, "Data Centres and Data Transmission Networks", accessed February 10, 2025, https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks.

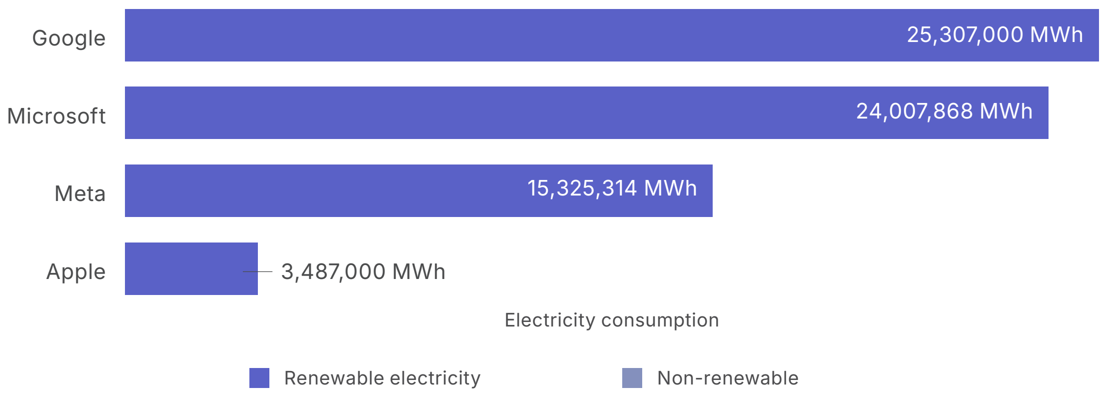
Based on the latest available data, North America held a dominant position in the global data center market, accounting for approximately 39% of revenue (US$85.84 billion), and around 54% of facilities (more than 5,700).[35, 36, 37] The United States alone accounted for more than 5,300 data centers, while Europe housed around 3,300 and East Asia around 855.[38,39]

Top AI players' electricity consumption

Several tech companies have set ambitious climate targets and have been actively transitioning their operational energy supply towards renewable sources. Common strategies include purchasing renewable energy credits (RECs), signing power purchase agreements (PPAs), and investing in on-site renewable power generation. Moving beyond the prevalent net-zero emissions strategy, some companies have set goals to match energy generation and consumption on a 24/7 hourly basis.[40]

In the 2023 fiscal year, Apple, Meta, Google and Microsoft matched 100% of their electricity usage with renewable electricity sources **(Figure 2.1)**, lowering to zero their market-based Scope 2 emissions from electricity.[41] For some players, this was achieved through the purchase of unbundled RECs. Unbundled RECs are sold separately from the associated green electricity, allowing buyers to claim environmental benefits without actually using green electricity, raising concerns about the actual contribution to emissions reduction.

Figure 2.1 **Google, Microsoft, Meta and Apple's electricity consumption and electricity mix in the 2023 fiscal year (in MWh)**

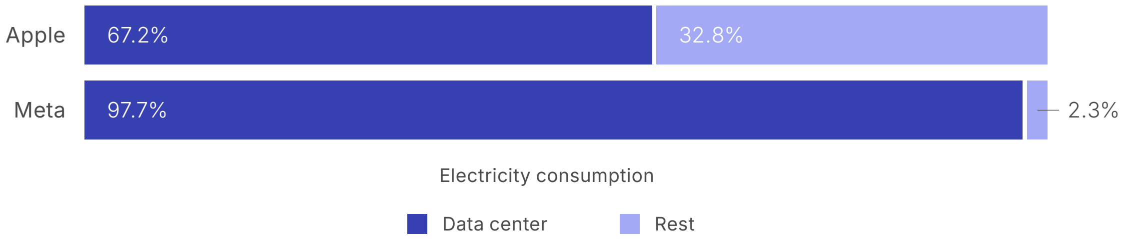


Source: Company disclosures.

33. Google Research, "Good News About the Carbon Footprint of Machine Learning Training," February 15, 2022, accessed February 17, 2025, https://research.google/blog/good-news-about-the-carbon-footprint-of-machine-learning-training/.
34. Desislavov R. et al., "Trends in AI Inference Energy Consumption: Beyond the Performance-vs-Parameter Laws of Deep Learning," *Sustainable Computing: Informatics and Systems* 38 (February 26, 2023), https://doi.org/10.1016/j.suscom.2023.100857.
35. Fortune Business Insights, "Data Center Market Size, Share & Industry Analysis," January 20, 2025, accessed February 10, 2025, https://www.fortunebusinessinsights.com/data-center-market-109851.
36. EPRI, "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption," 2024, accessed February 10, 2025, https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf.
37. Cloudscene, "North America," accessed December 20, 2024, https://cloudscene.com/region/datacenters-in-north-america.
38. Cloudscene, "Europe," accessed December 20, 2024, https://cloudscene.com/region/datacenters-in-europe.
39. Cloudscene, "Aisa Pacific, " accessed December 20, 2024, https://cloudscene.com/region/datacenters-in-asia-pacific.
40. Locational and 24/7 carbon-free electricity procurement:every kilowatt-hour of electricity consumption is met with carbon-free electricity sources, every hour of every day, from the same local grid where they consume power.
41. A location-based reporting method reflects the average carbon intensity of electric grids where energy consumption occurs ; A market-based method reflects emissions from electricity that incorporate company's procurement choices, such as RECs.

The substantial amount of electricity consumed from tech companies' self-operated data centers contributes to these businesses' Scope 2 emissions (indirect emissions from purchased energy) **(Figure 2.2)**. However, in many cases, tech companies only report electricity consumption at the company or location level, and disclosure of data center electricity consumption remains insufficient.

Figure 2.2 **Apple and Meta's percentage of electricity used by data centers in the 2023 fiscal year**



Source: Company disclosures.

Leading tech companies have purchased more than 45,000 megawatts (MW) of renewable energy to match electricity consumption from their on-premises data centers.[42]

Meanwhile, environmental regulations regarding the energy usage of data centers are emerging. Germany, for example, has been mandating that data center operators cover 50% of their electricity from renewable sources from January 1, 2024, and 100% from January 1, 2027.[43]

Energy Demand from AI Chipmaking

The energy demands of AI are attracting growing attention, which is largely directed at the model training and inference phases of the AI hardware lifecycle.

AI chips are fundamental to AI digital infrastructure and drive AI development (the definition of "AI chips" includes specialized computer parts, i.e., AI hardware, such as AI accelerators and high bandwidth memory (HBM), which facilitate the efficient processing of AI workloads for generative AI applications).[44] Before reaching end users, AI hardware undergoes design and manufacturing. Chip manufacturing, encompassing raw material extraction, wafer fabrication, assembly, and testing, is energy-intensive, with wafer fabrication being the largest electricity consumer and thus the main source of carbon emissions.[45] Despite the surge in demand and expansive efforts, chip manufacturing remains surprisingly overlooked and the lack of comprehensive disclosure hinders accurate quantification and assessment of its environmental impact.

During the fabrication process, the major sources of greenhouse gas (GHG) emissions are the use of chemicals and energy usage in chip production. Electricity accounts for the largest share of this energy usage and is therefore the biggest single source of GHG emissions in semiconductor manufacturing. While direct, on-site emissions from the high global warming potential (GWP) gases used during manufacturing are a primary source, they are largely driven by production demand; whereas manufacturers have more direct control over decarbonizing their electricity usage through strategies such as renewable energy procurement, energy efficiency improvements, and advocacy for cleaner energy grids.[46]

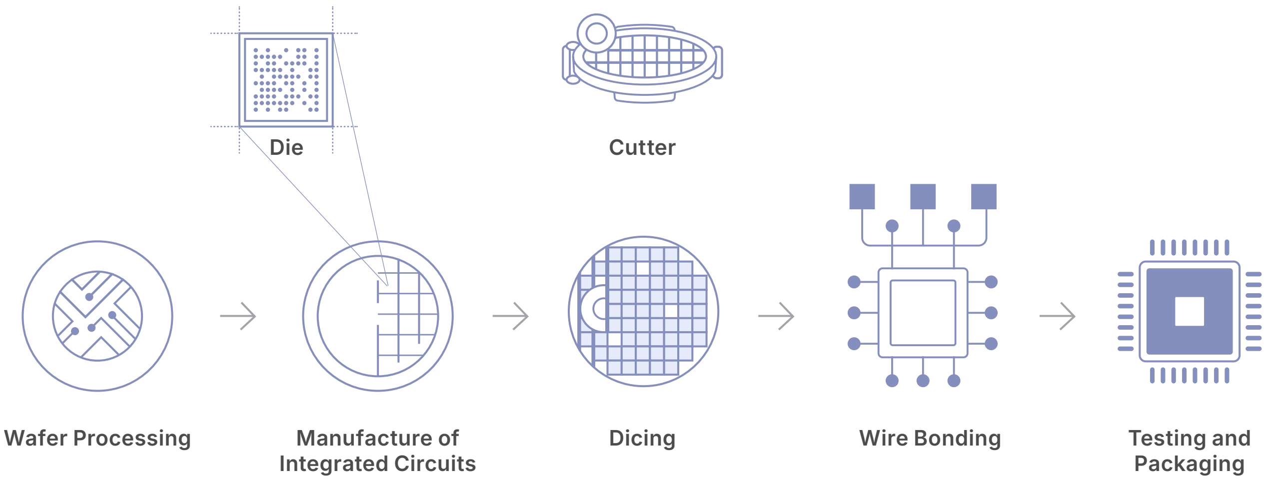
42. S&P Global, "Datacenter companies continue renewable buying spree, surpassing 40 GW in US," S&P Global Market Intelligence, March 28, 2023, accessed February 10, 2025, https://www.spglobal.com/market-intelligence/en/news-insights/research/datacenter-companies-continue-renewable-buying-spree-surpassing-40-gw-in-us.
43. Bundesrepublik Deutschland. Gesetz zur Digitalisierung der Energiewende (Energiewendegesetz). Accessed February 11, 2025. https://www.gesetze-im-internet.de/enefg/BJNR1350B0023.html.
44. The AI chip definition does not include general purpose components such as CPUs, DRAM, and NAND storage, which are typically used with AI accelerators in complete server systems.
45. Marcello R., "The Chip Manufacturing Industry: Environmental Impacts and Eco-Efficiency Analysis," *Science of The Total Environment* 858 (November 2, 2023), https://doi.org/10.1016/j.scitotenv.2022.159873.
46. Inference, "Chip Production's Ecological Footprint: Mapping Climate and Environmental Impact," June 20, 2024, accessed March 5, 2025, https://www.interface-eu.org/publications/chip-productions-ecological-footprint.

As semiconductor technology advances, the manufacturing processes become more complex and require more repetitions, leading to increased electricity consumption.[47] Given that chip production generates substantial emissions, the projected rise in demand and global expansion of semiconductor manufacturing capacity will inevitably contribute to a growing carbon footprint.

Key energy-consuming processes in chip manufacturing

The process begins with a silicon wafer. Wafers are sliced from a bar of silicon, and then cleaned, polished, and prepared to be used as a substrate for the creation of the electronic components. Next, wafers are printed with highly intricate circuit designs through patterning, doping, deposition and lithography. After that, the wafer is diced into individual chips using a diamond saw, so called “dice”. The size of dice varies depending on the purpose of the chip. The final step is packaging, where the chip die is placed onto a substrate and packaged into a final product that can be used in electronic devices **(Figure 2.3)**.

Figure 2.3 Illustration of the wafer fabrication[48]



47. Interuniversity Microelectronics Centre, “Sustainable semiconductor technologies and systems: The green transition of the IC industry,” accessed March 5, 2025, https://www.imec-int.com/en/expertise/cmos-advanced/sustainable-semiconductor-technologies-and-systems-ssts/stss-white-paper.
48. Nazir H. Z. et al., “Robust Adaptive Exponentially Weighted Moving Average Control Charts with Applications of Manufacturing Processes,” *Int. J. .Adv. Manufact. Technol*:105 (August 14, 2019), https://doi.org/10.1007/s00170-019-04206-y.

A bottom-up approach was used to estimate electricity consumption and associated carbon emissions from AI-related chip manufacturing in 2023-2024. These estimates were extended to 2030 based on a forecast of AI-driven wafer supply and demand from McKinsey. The analysis incorporated publicly available data and market analyses on production and wafer demand for flagship AI models, electricity requirements of production process, and the electricity mix of manufacturing facilities' local power grids.

A detailed description of the methodology of this analysis can be found in the accompanying methodology document (Appendix A).

Summary of the research approach

1. **AI hardware production estimates:**
   Analyst insights and supply chain constraints were combined to create a holistic distribution of AI hardware production estimates.

2. **Product specifications:**
   Information on chip dimensions, manufacturer details, and technology processes was collected for each device type under consideration.

3. **Wafer demand estimation:**
   Total AI-driven wafer demand was calculated by estimating the number of chips required for AI hardware production.

4. **Electricity consumption estimation:**
   Academic insights on the electricity intensity of wafer production were combined with the estimated wafer demand to determine the aggregate power demand for wafer production.

5. **Environmental impact assessment:**
   The carbon emissions from electricity consumption in AI-driven wafer production were estimated by analyzing the power grids supplying the manufacturing facilities.

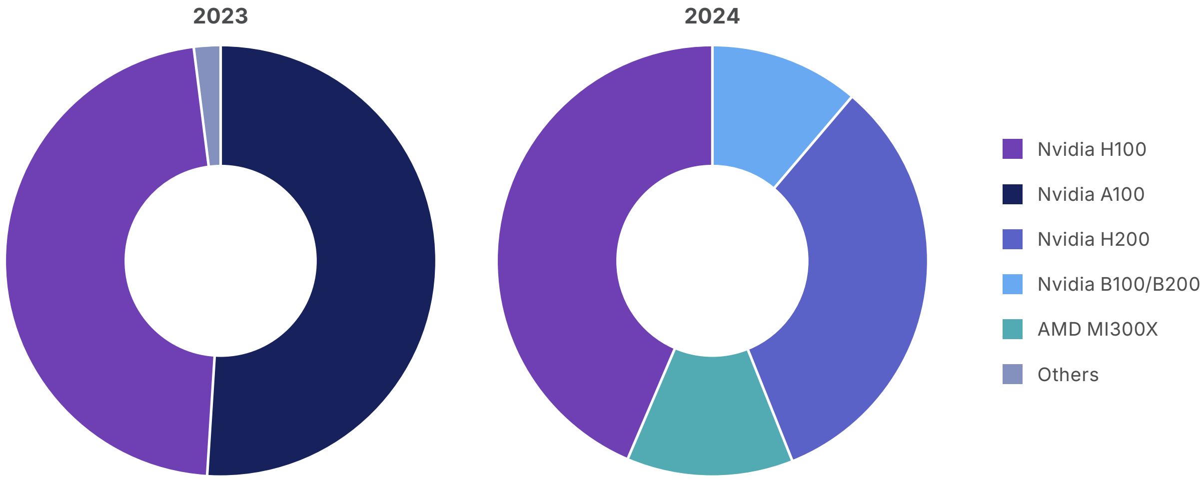
During 2023-2024, a key bottleneck in the AI hardware supply chain was the packaging technology to integrate HBMs and processors within the same package. Research suggests that the limited available packaging capacity has mainly been used by Nvidia and AMD's AI models during this period.[49,50] Our analysis focuses on their AI models: Nvidia's A100, H100/200 and B100/200 and AMD's MI300X **(Figure 2.4)**, as well as their key component suppliers.

The estimated number of AI model units produced was derived from market share and packaging capacity data. Over the past two years, Nvidia's market dominance in the data-center GPU sector and its H100 alone occupied a significant share in both 2023 and 2024.

Some tech companies, such as Google, are developing their own AI chips, but these chips are typically not commercially available. Consequently, obtaining product specifications and production data for these accelerators is not feasible within the scope of this analysis.

49. Nvidia, "Investor Presentation," 2023, accessed February 10, 2025, https://s201.q4cdn.com/141608511/files/doc_presentations/2023/Oct/01/ndr_presentation_oct_2023_final.pdf.
50. WCCFTech, "AMD Instinct AI Accelerator Lineup Gets MI325X Refresh in Q4, 3nm MI350 'CDNA 4' in 2025, CDNA MI400 'CDNA Next' in 2026," June 2, 2024, accessed February 10, 2025, https://wccftech.com/amd-instinct-ai-accelerator-lineup-mi325x-refresh-q4-3nm-mi350-cdna-4-2025-cdna-mi400-cdna-next-2026/.

Figure 2.4 **Estimated market share of AI models in 2023 and 2024**



Analysis of the estimated number of major AI models sold in 2023-2024 and their product specifications suggests the supply chain implications for the various components used in their packaging. Both Nvidia and AMD rely on TSMC for their GPUs, accelerator chiplets (XCDs), and IO dice (IODs), but differ in HBM sourcing: Nvidia's HBMs were supplied by SK hynix and Micron, while AMD reportedly obtains its HBM from Samsung **(Figure 2.5)**.

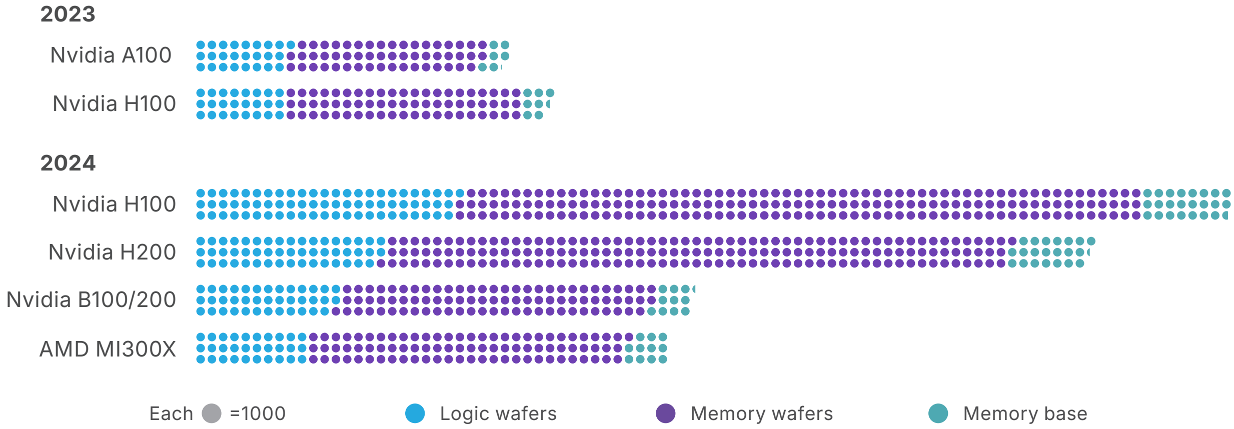
Figure 2.5 **Manufacturers and manufacturing facility locations of the AI models analyzed**[51]

Model	Nvidia A100	Nvidia H100	Nvidia H200	Nvidia B100/B200	AMD MI300X
GPU/XCD/IOD Manufacturer	TSMC	TSMC	TSMC	TSMC	TSMC
GPU/XCD/IOD manufacturing facility location	Taiwan	Taiwan	Taiwan	Taiwan	Taiwan
Memory manufacturer	SK hynix	SK hynix	Micron	Micron	Samsung
Memory manufacturing facility location	South Korea	South Korea	Japan	Japan	South Korea

51. While several variants of a given chip may exist (e.g., Nvidia's H800 and H200 exhibit similarities to the H100), this analysis does not differentiate between such variants. These chips possess comparable production requirements, rendering them effectively interchangeable for the purposes of calculations pertaining to 2023 and 2024.

Using the specifications of these AI models, with the consideration of production yields, we estimated the wafer demand related to device production **(Figure 2.6,** with the details of the estimation methodology in **Appendix A)**. In 2023, the manufacturing of Nvidia A100 and H100 AI models consumed over 47,000 logic wafers, 115,000 memory wafers, and 14,000 memory bases. In 2024, the wafer demand for AI models more than tripled, reaching more than 185,000, 521,000 and 65,000 wafers, respectively.

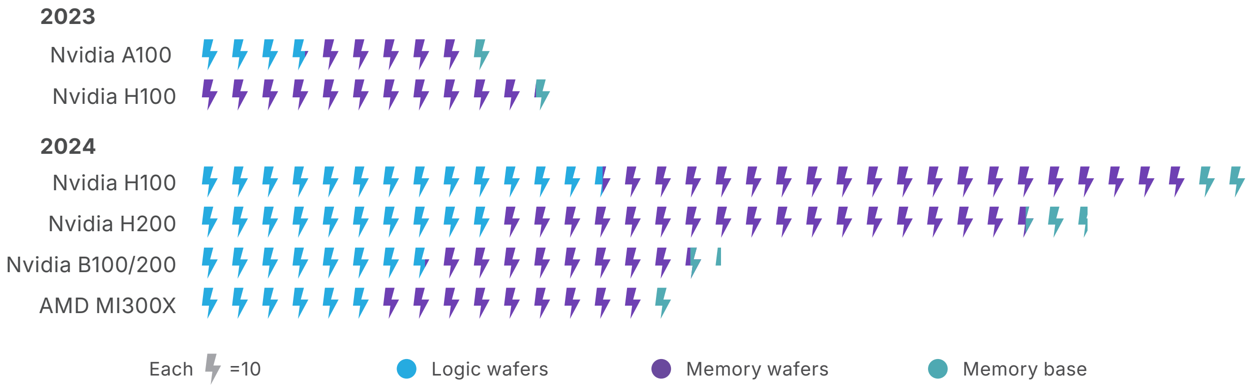
Figure 2.6 **Estimated wafer demand of the analyzed AI models in 2023 and 2024**



Energy demand

Based on our estimated wafer demand of these models, we calculated the electricity demand for the wafer production and found the energy needs increased significantly from 218 GWh in 2023 to 983.9 GWh in 2024, representing a 351% rise **(Figure 2.7)**.

Figure 2.7 **Estimated electricity consumption for manufacturing the AI-related wafers in 2023 and 2024 (in GWh)**

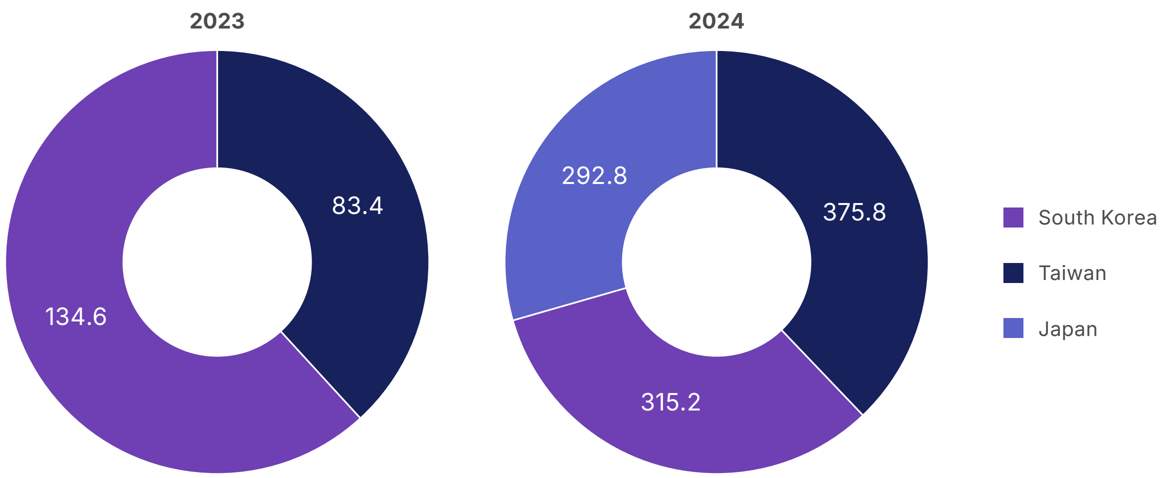


With the continued expansion of production capacity, rapidly rising AI-related manufacturing may soon constitute a significant portion of the chip market.

The estimated electricity consumption is attributed to the location of the manufacturing facilities **(Figure 2.8)**. Although detailed information regarding the proportion of production capacity dedicated to AI-related hardware remains unavailable, the analysis of GPU and memory chip suppliers suggests that the relevant manufacturing facilities are concentrated in the East Asia region **(Figure 2.5)**:

- GPUs of both Nvidia and AMD's aforementioned models were manufactured by TSMC in Taiwan.
- The memory chips of Nvidia's A100 and H100 were manufactured by SK hynix in South Korea; Nvidia's H200 and B100/200 memory chips were manufactured by Micron in Japan; and memory chips of AMD's MI300X by Samsung in South Korea.

Figure 2.8 **Estimated electricity consumption for AI-related chip manufacturing by region in 2023 and 2024 (in GWh)**



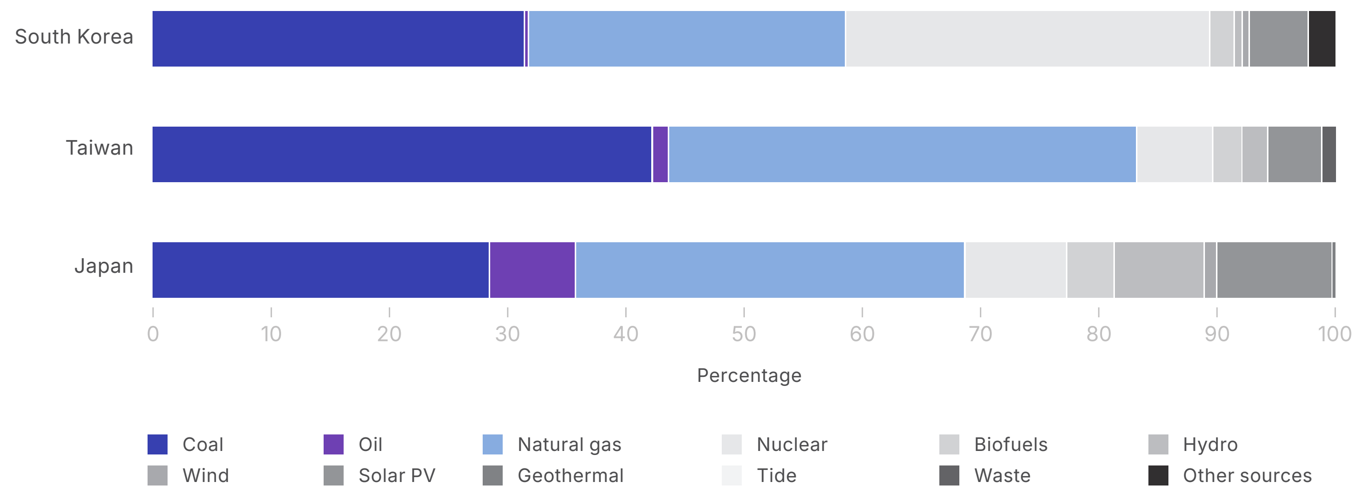
The electricity consumption from AI chip manufacturing in Taiwan showed the biggest increase in the region, jumping from 83.4 GWh to 375.8 GWh, equivalent to that of more than 92,650 Taiwan households in 2023, a 350.6% increase over one year.[52] In South Korea, the electricity consumption of AI chip manufacturing increased more than two times from 2023 to 2024, from 134.6 GWh to 315.2 GWh.

52. Industrial Technology Research Institute, "Residential Sector Energy Statistics 2023," July 2024, accessed February 17, 2025, https://energypark.org.tw/upload/propaganda/202408131715095.pdf.

Carbon emissions

Currently, the electricity grid in East Asia is mainly powered by fossil fuels, which leads to AI chips that are mainly manufactured by coal, oil and gas, with 2023 figures showing 83.1% reliance in Taiwan, 68.6% in Japan, and 58.5% in South Korea **(Figure 2.9)**. In 2023, electricity consumption in AI chip manufacturing generated 99,200 metric tons of CO_2 equivalent in East Asia, and the number increased more than 4.5 times in 2024, accounting for more than 453,600 metric tons of emissions.

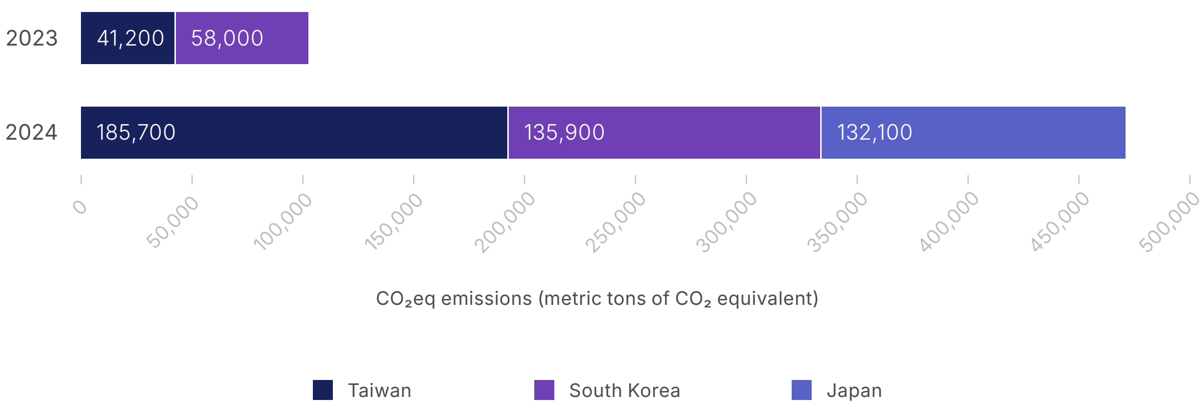
Figure 2.9 **Electricity mix of the power grids of Taiwan, South Korea and Japan in 2023**



Source: EPSIS[53], MOEAEA[54], METI[55]

Among all the key AI chip manufacturing locations, the emissions of AI chip manufacturing in Taiwan had the biggest surge from 41,200 metric tons in 2023 to 185,700 metric tons in 2024, which is more than a four-fold increase. In South Korea, the emissions related to AI chipmaking more than doubled, from 58,000 metric tons to 135,900 metric tons in 2024 **(Figure 2.10)**.

Figure 2.10 **Estimated carbon emissions from AI-related chip manufacturing in 2023 and 2024**

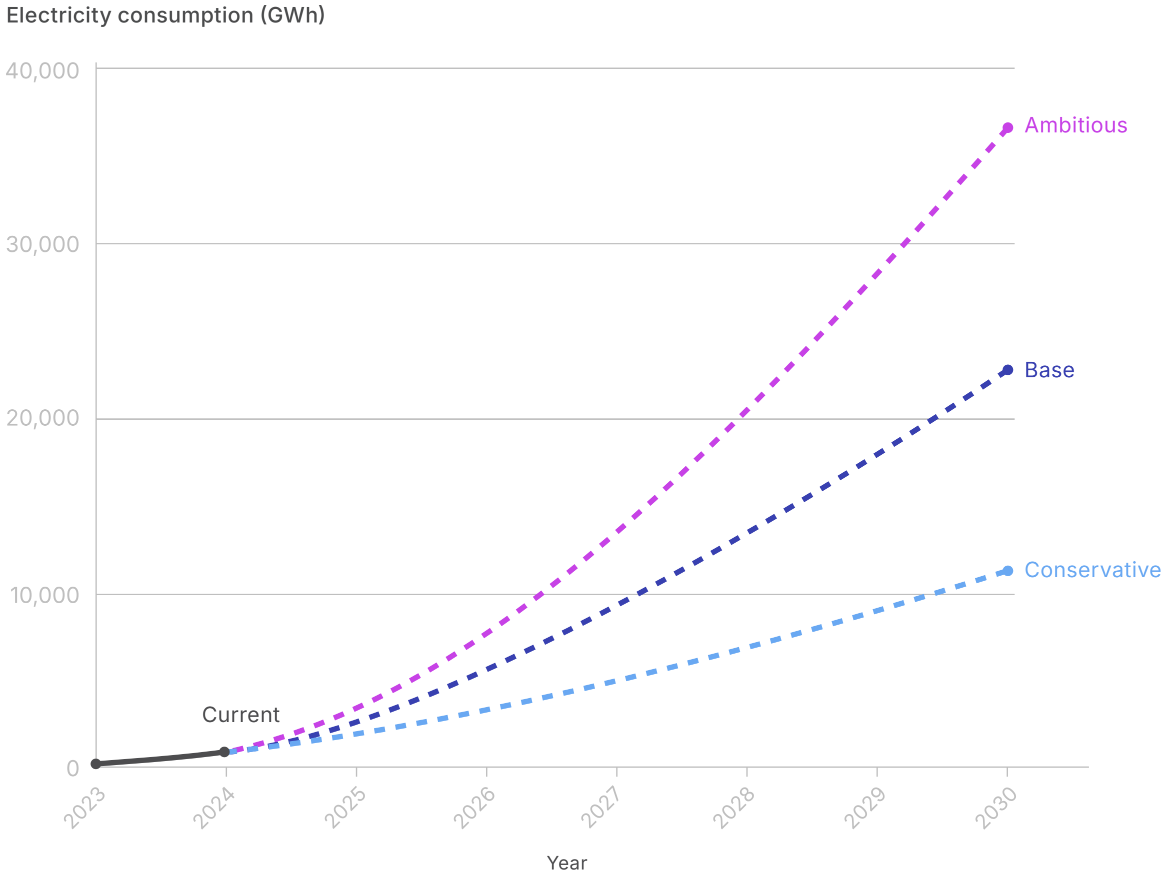


53. EPSIS, "Generation Output by Energy Source," 2024, accessed February 10, 2025, https://epsis.kpx.or.kr/epsisnew/selectEkgeGepGesGrid.do.
54. MOEAEA, "Electricity Statistics," 2024, accessed February 10, 2025, https://www.esist.org.tw/database/search/electric-generation.
55. METI, "General Energy Statistics," 2024, accessed February 10, 2025, https://www.enecho.meti.go.jp/statistics/total_energy/results.html.

The potential impact of AI hardware in 2030

The demand for AI hardware is expected to continue expanding through 2030. According to McKinsey's estimates, the projected growth in demand for AI-related chips will necessitate the production of an additional 1.2 to 3.6 million logic wafers, 4.5 to 21 million dynamic random-access memory (DRAM) wafers, and 1.7 to 7.9 million NAND flash memory wafers (more details can be found in **Appendix A**).[56,57] Combining this projection with current electricity demand estimates suggests that to meet such a demand, the electricity consumption would be expected to increase to between 11,550 GWh and 37,238 GWh **(Figure 2.11)**, up to 170 times more than the demand in 2023, which would exceed Ireland's current consumption by 2030.[58]

Figure 2.11 **Projected electricity consumption for AI-related chip manufacturing in 2030**[59]



56. McKinsey, "Generative AI: The next S-curve for the semiconductor industry?" March 29, 2024, accessed February 10, 2025, https://www.mckinsey.com/industries/semiconductors/our-insights/generative-ai-the-next-s-curve-for-the-semiconductor-industry.
57. McKinsey's forecast focused on demand of three wafer types, does not distinguish between AI models.
58. Central Statistics Office of Ireland, "Last Updated Tables," accessed February 10, 2025, https://data.cso.ie/.
59. The three senarios: Conservative, Base and Ambitious was according to the senarios in McKinsey's exhibit "Global logic and memory wafer demand and supply in 2030", https://www.mckinsey.com/industries/semiconductors/our-insights/generative-ai-the-next-s-curve-for-the-semiconductor-industry.

As AI chip production increases, however, so do emissions. If the ratios of carbon intensity and fossil fuel-based electricity generation in the power grid remain at the current level (see detailed discussions in Appendix A), the projected energy consumption could result in up to 16 million metric tons of CO_2 equivalent emissions (ranging from 5.2 to 16.8 million metric tons across conservative, base-case, and ambitious scenarios).

Powering AI with false solutions

To ride on the wave of AI and meet the energy demand of the industry, chip manufacturers and other AI suppliers are seeking solutions such as nuclear reactors, liquefied natural gas (LNG), or purchasing a large number of renewable energy credits (RECs), rather than increasing renewable energy sourcing through high-quality renewable energy sources. Growth in electricity demand from chipmaking is being used to justify an expansion of fossil fuel and nuclear infrastructure in East Asia, hindering the region's progress towards energy transition goals.

- In the summer of 2024, South Korea's government approved the construction of a 1.05 GW LNG combined heat and power plant in the Yongin General Industrial complex for SK hynix. The government also has planned to build 3GW LNG power plants in the National Semiconductor Cluster for Samsung.[60,61]
- In Taiwan, the rise of AI has been used as a justification for the expansion of fossil fuel capacity. Proposed construction of a new LNG receiving terminal at Keelung Port involves converting four 500 MW oil units into two 2.6 GW gas turbines. Industry representatives have argued that stable power is crucial with the rise of AI, saying that Nvidia plans to set up its Taiwan headquarters in the north.[62] Likewise, the chairman of Pegatron Corp, a major AI manufacturer, has publicly advocated for more nuclear power capacity in Taiwan.[63]

RECs have been widely purchased by the tech and semiconductor sectors to offset fossil-based electricity consumption and reduce their market-based emissions. However, this practice does not inherently result in a corresponding reduction or elimination of companies' underlying emissions nor stimulate additional renewable energy production.[64] Instead, rising electricity demands may continue to elevate emissions under the location-based approach — without reducing the company's overall carbon footprint.

Limitations

The analysis of AI chip manufacturing energy consumption is subject to certain limitations regarding its scope.

- The definition of "AI chips" is intentionally narrow, encompassing specialized AI hardware such as AI accelerators and high bandwidth memory (HBM) that directly enhance the performance of generative AI applications. General-purpose components commonly integrated with AI accelerators in complete server systems, are excluded from this definition.
- In-house solutions, such as Google's Tensor Processing Units (TPUs), and the broader demand for generic components used in conjunction with the aforementioned specialized devices, fall outside the scope of this research.
- Due to a lack of standardization and significant variations in emissions reporting methodologies across manufacturers, a market-based emissions analysis could not be reliably performed and is therefore not included in the research scope.

60. Business Korea, "Green Light for Yongin Semiconductor Cluster's 1.05GW LNG CHP Plant," August 9, 2024, accessed February 10, 2025, https://www.businesskorea.co.kr/news/articleView.html?idxno=222719
61. Newstree Korea, "Samsung and Hynix to Move into Yongin Semiconductor... What to Do About Renewable Energy," August 5, 2024, accessed March 31, 2025, https://www.newstree.kr/newsView/ntr202408020010.
62. Recessary, "Keelung's fourth environmental impact assessment is coming soon! The business community shouts "Industry needs electricity" to support the reconstruction and seize the opportunity of Nvidia to set up its headquarters," January 13, 2025, accessed February 19, 2025, https://www.reccessary.com/zh-tw/news/world-market/lng-receiving-terminal-dispute?fbclid=IwY2xjawIb96FIeHRuA2FlbQIxMAABHb6WnrLQqQQyCqbR7QCq1oD0IjAkPH0TjOe91lbM0uGfb46i06Z6RODKFQ_aem_cMgNGJeQjAN7ydAo1Q_WKg.
63. TVBS, "Pegatron chair backs nuclear energy to cut carbon emissions," November 12, 2024, accessed February 19, 2025, https://news.tvbs.com.tw/english/2683129.
64. Interface, "Chip Production's Ecological Footprint: Mapping Climate and Environmental Impact," June 20, 2024, accessed February 10, 2025, https://www.interface-eu.org/publications/chip-productions-ecological-footprint.

Recommendations

Recommendations

Leading tech companies and AI chip design companies, such as Nvidia and AMD, need to target 100% renewable energy across the supply chain by 2030.

AI players must design a clear pathway for supply chain emissions reduction. The majority of emissions are from electricity consumption, therefore it is essential that chipmakers and other AI suppliers adopt renewable energy. Companies must set up a 100% renewable energy target across the supply chain by 2030.

AI chipmakers and other AI suppliers need to proactively set up their own 100% renewable energy targets by 2030.

When expanding semiconductor factories or creating new clusters in response to the AI boom, renewable energy should be prioritized over fossil fuels or nuclear power. Companies should establish and quickly implement a renewable energy-based power supply and demand roadmap in collaboration with clients, local governments, and power companies.

Leading tech companies and AI chip design companies should actively engage with suppliers on renewable energy procurement and strategies to reduce emissions.

Active engagement with suppliers is imperative to drive the renewable energy transition and achieve Scope 3 emissions reduction goals. Companies should provide financial support and incentives to their suppliers, engage meaningfully through training and reporting, and actively require key suppliers to set their own renewable energy and emissions reduction targets.

Companies need to choose high-impact sourcing methods when it comes to renewable energy procurement.

High-impact renewable energy sourcing options, such as power purchase agreements (PPAs), renewable energy investment, and onsite generation should be the primary options for a company to achieve renewable energy targets because these methods have clear additionality and trackability. Renewable energy certificates can be an additional choice for companies to meet their targets. When companies set the renewable energy targets for the supply chain, high-impact sourcing methods need to be clearly stated.

AI players including tech companies, chip designing companies and AI chipmakers, should use their position to actively engage with policymakers and government institutions to develop renewable energy-friendly policies.

When large corporations voice their demands, policymakers listen. AI players must challenge and engage policymakers to remove barriers to renewable energy procurement and to streamline solutions.

Government institutions should prioritise the supply of renewable energy sources to semiconductor manufacturers.

Due to the expansion of semiconductor manufacturing in the East Asia region, relevant government institutions should prioritize renewable energy supplies, especially to semiconductor facilities. Additionally, semiconductor facilities should be decentralised to prevent power overload in specific regions.

Published April 2025 by Greenpeace East Asia

Greenpeace is an independent campaigning organization that uses peaceful protest and creative communication to expose global environmental problems and to promote solutions that are essential to a green and peaceful future.