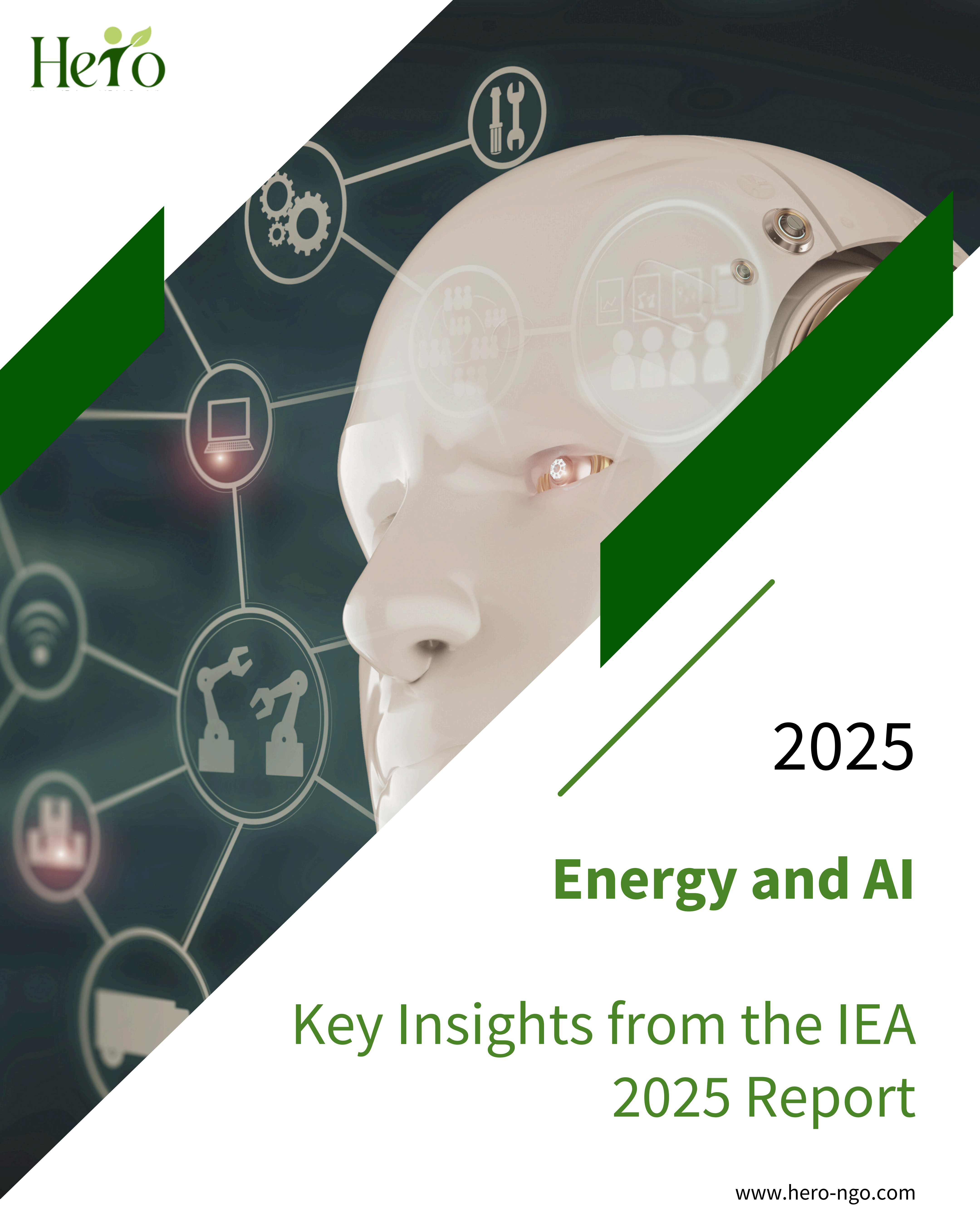
Hero



2025

Energy and AI

Key Insights from the IEA 2025 Report

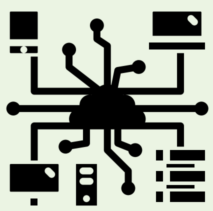
www.hero-ngo.com

AI's Emergence and Impact on the Energy Sector

In recent years, artificial intelligence (AI) has soared to the top of the political and business agenda. Once a mostly academic pursuit, it has evolved into an industry with trillions of dollars at stake. Despite significant uncertainties, it is now very clear: AI is coming. In many sectors, it is already here.

This has major consequences for the global energy sector. There is no AI without energy, specifically electricity. At the same time, AI has the potential to transform the sector’s future. However, policy makers and the market have often lacked the tools to fully understand these wide-ranging impacts.

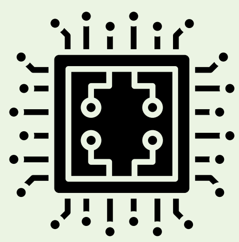
This study performed by the International Energy Agency (IEA) addresses this gap by leveraging data collecting and analysis, as well as their expertise and knowledge to inform and strengthen the global dialogue on these issues.



Since 2014, the computation required to train state-of-the-art AI models has surged by **350,000×**.

A **USD 12 trillion of the USD 16 trillion** increase in S&P 500 market capitalization since 2022 is attributed to AI-related companies.

Corporate adoption continues to rise sharply: large-scale firms using AI grew from 15% in 2020 to nearly **40%** in 2024, while household use has become global with more than **40%** of internet users.

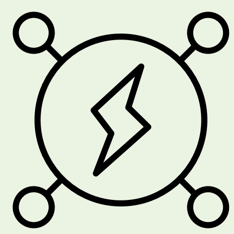


East Asia dominates the advanced chip fabrication (holding **~65%** of global share).

Meanwhile, the United States leads AI software development and deployment, with China rapidly expanding its capabilities.

Since 2022, global investment in data centers has doubled, reaching **USD 0.5 trillion** in 2024.

These centers consumed 1.5% of global electricity (≈ 415 TWh) in 2024, a figure expected to double by 2030, covering almost **10%** of worldwide demand.



While renewables and natural gas currently supply most of this electricity, grid constraints already pose challenges with around **20%** of planned data center projects potentially at risk of delay unless grid expansions and efficiency measures accelerate.

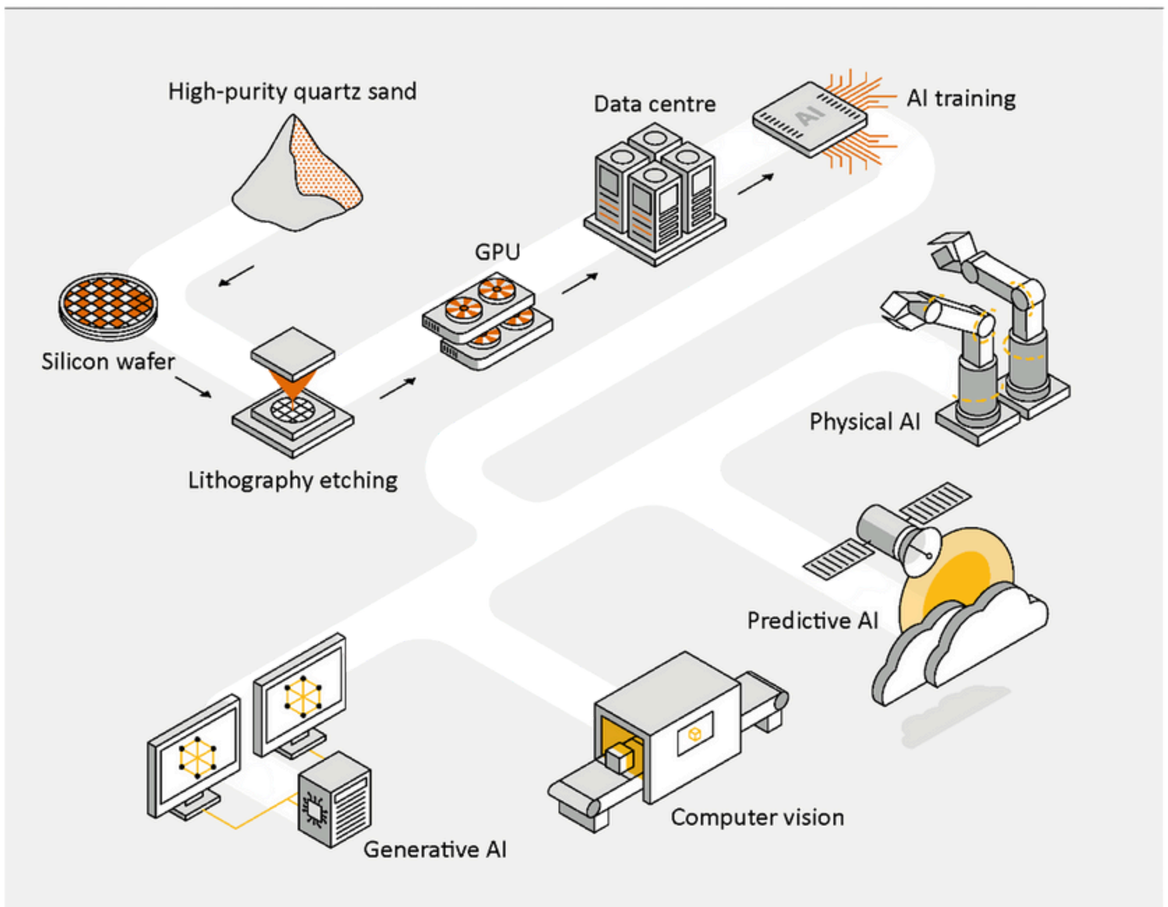
Understanding AI Types and Global Supply Chain

Artificial Intelligence (AI) has no single universal definition, as its meaning has evolved alongside the technology. In essence, AI is the science of enabling machines to learn, reason, and act in ways that mimic human intelligence.

Types of AI and their Applications

1. **Predictive AI:** Anticipates future outcomes using data (e.g., weather forecasting, predictive maintenance, or AlphaFold for protein structure prediction).
2. **Generative AI:** Creates new content such as text, images, or video through models like ChatGPT. Subtypes include:
    - Language Models (text generation)
    - Multimodal Models (text, image, and video)
    - Large Reasoning Models (complex logic, coding, or math - e.g., OpenAI o1, DeepSeek-R1)
    - Computer Vision: Enables machines to "see" and interpret visual data, used in autonomous vehicles, medical imaging, and security systems.
3. **Physical AI:** Embodied systems such as robots, drones, and self-driving cars that interact with the real world, increasingly learning from their environment.
4. **Agentic AI:** Task-oriented autonomous agents that manage workflows, like voice assistants or energy management systems for smart buildings and EV charging.

Figure 1: Select AI Infrastructure and Types of Applications



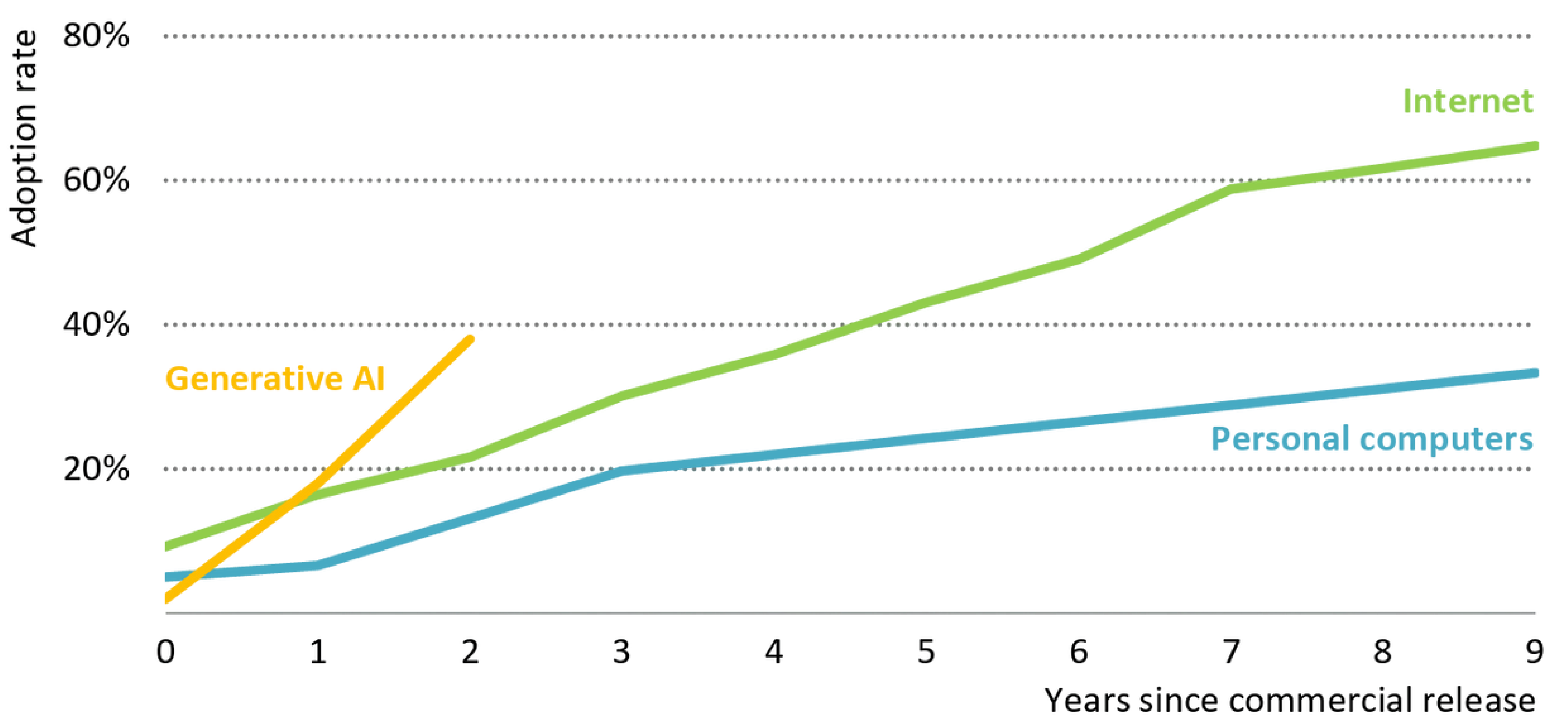
The AI Supply Chain

The AI ecosystem is a complex, globally distributed network encompassing:

- Data Centers: Massive computing hubs consuming power equivalent to small towns; mainly located in the US, Europe, and China.
- Specialized Hardware: AI relies on GPUs and TPUs, capable of fast parallel processing.
- Chip Manufacturing: Design is dominated by US firms (NVIDIA, AMD, Intel, Broadcom), while production is concentrated in East Asia, led by TSMC with a **65%** global foundry market share in 2024.
- Critical Materials & Equipment: High-purity silicon, rare earth elements, and ultra-precise lithography machines form the foundation of chip fabrication.

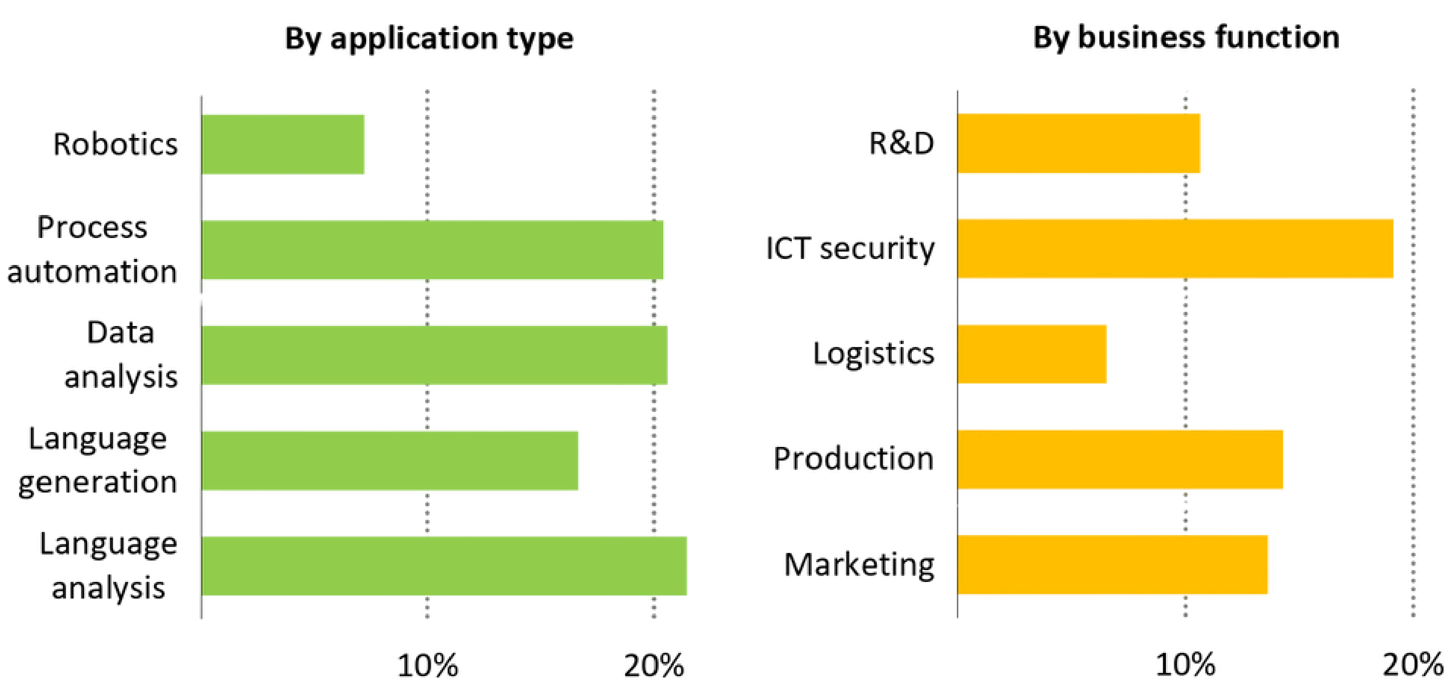
Explosive Adoption of Generative AI in the Workplace

Figure 2: Growth in the use of digital technologies in the workplace since the year of first commercial release, United States.



Generative AI shows a remarkably steep adoption curve, reaching nearly **40%** within just 2 years, far surpassing the early adoption rates of both the Internet and personal computers.

Figure 3: Percentage of large firms reporting using AI by application type and business function, European Union, 2024.



AI is being leveraged across diverse functions and technical domains: **process automation and language generation** leading AI use by application type, **ICT security and production leading** the use by business function.

Global Disparities in Data Center Distribution

North America, parts of Europe and East Asia dominate the map with large clusters, with very few large data centres in Africa, South America, and parts of the Middle East (there is a clear **global digital divide**).

These data centres are heavily concentrated in urbanized or economically strategic areas, which may create strain on local electricity grids.

Figure 4: Global map of large data center clusters, 2024.

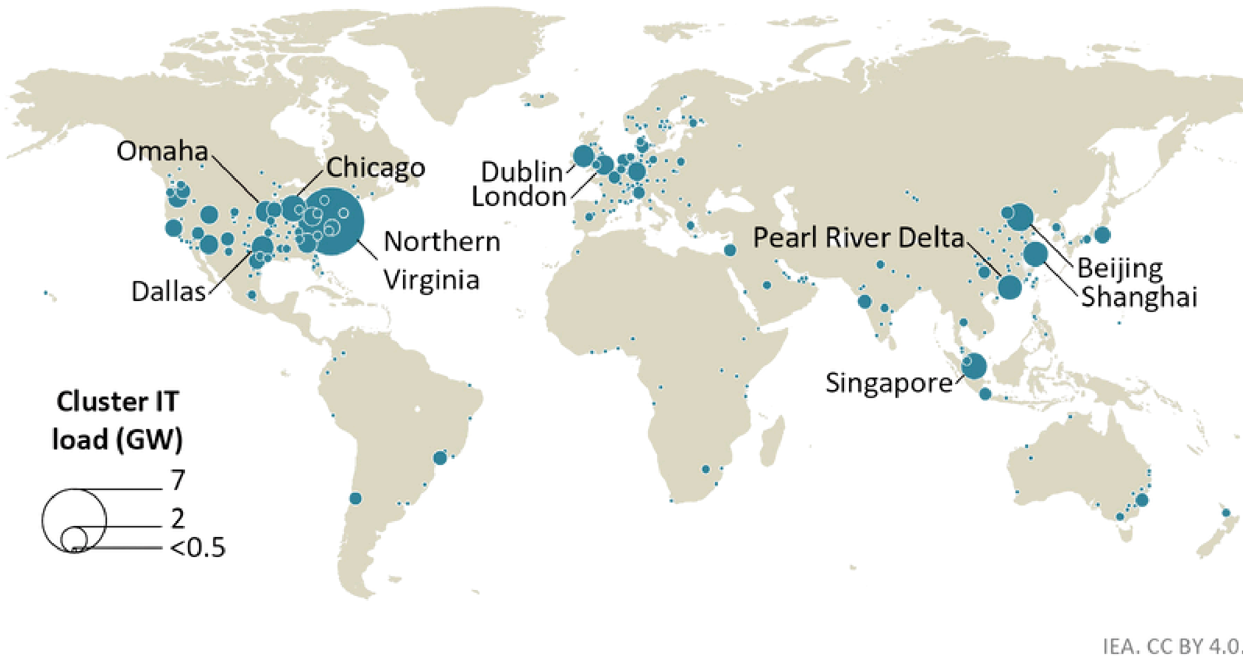
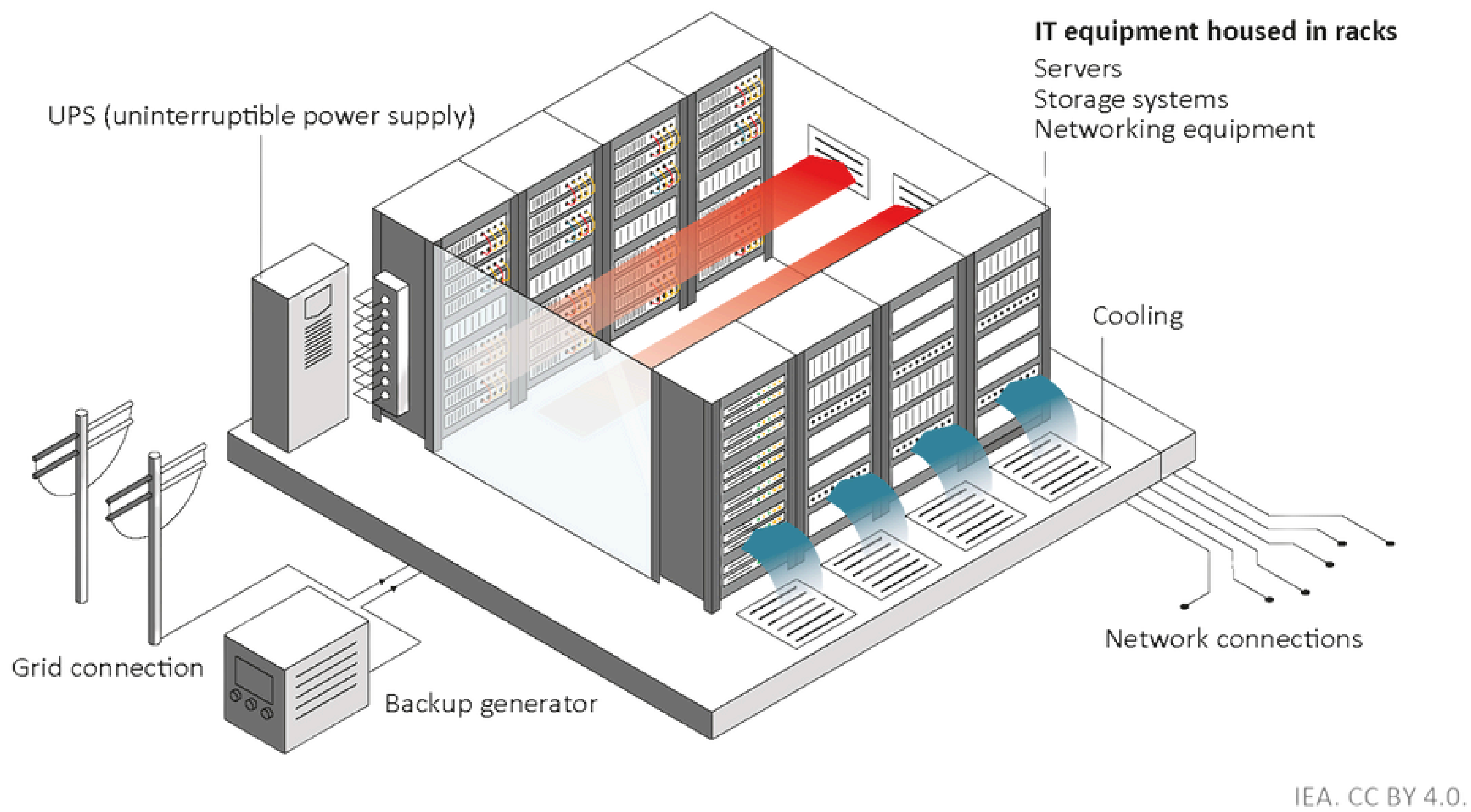


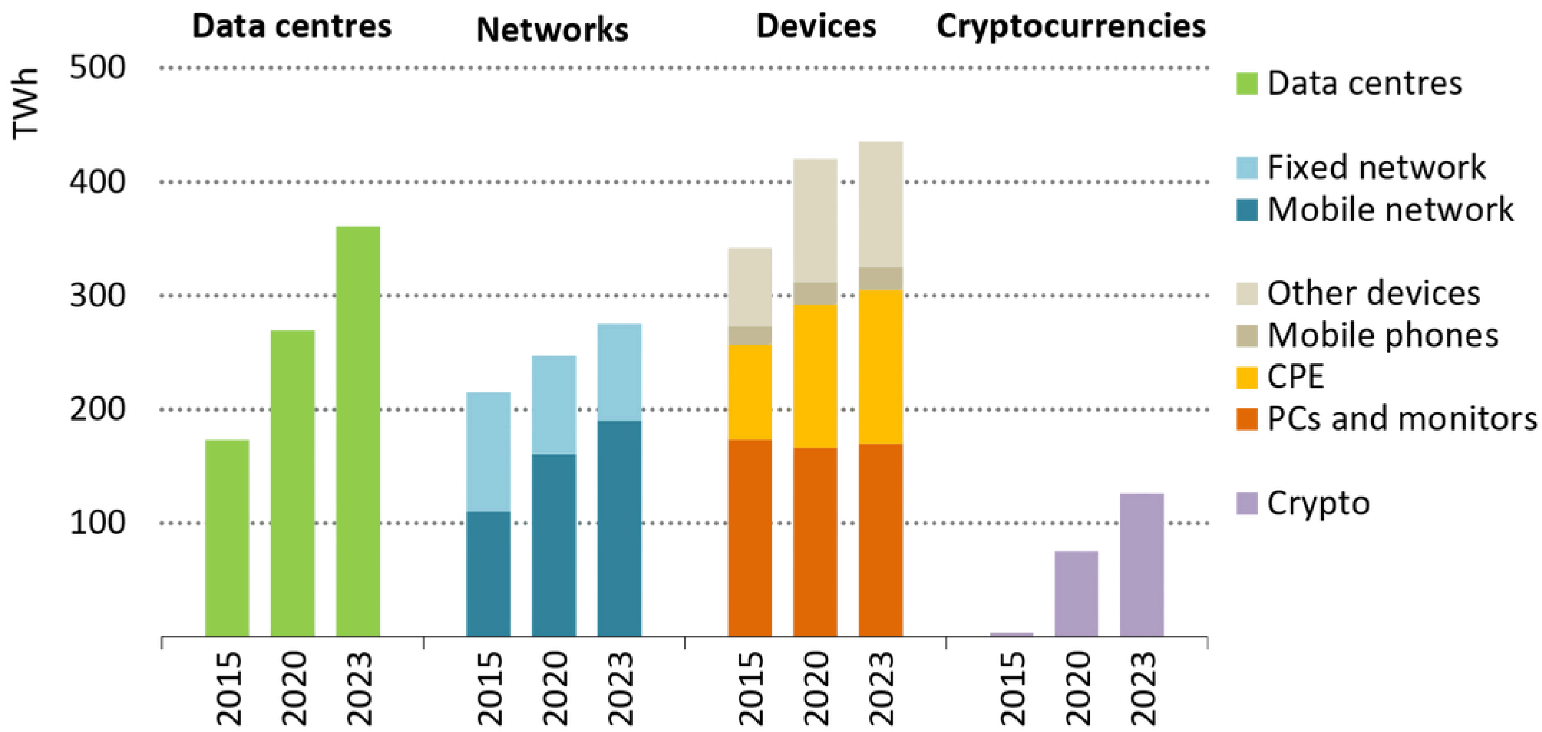
Figure 5: Major data centre components.



A modern data center typically includes IT equipment housed in racks (servers, storage systems, and networking gear), which are the operational core. These are supported by cooling systems to manage heat, uninterruptible power supply (UPS) systems, grid connections, backup generators, and network connections to ensure continuous and efficient operation.

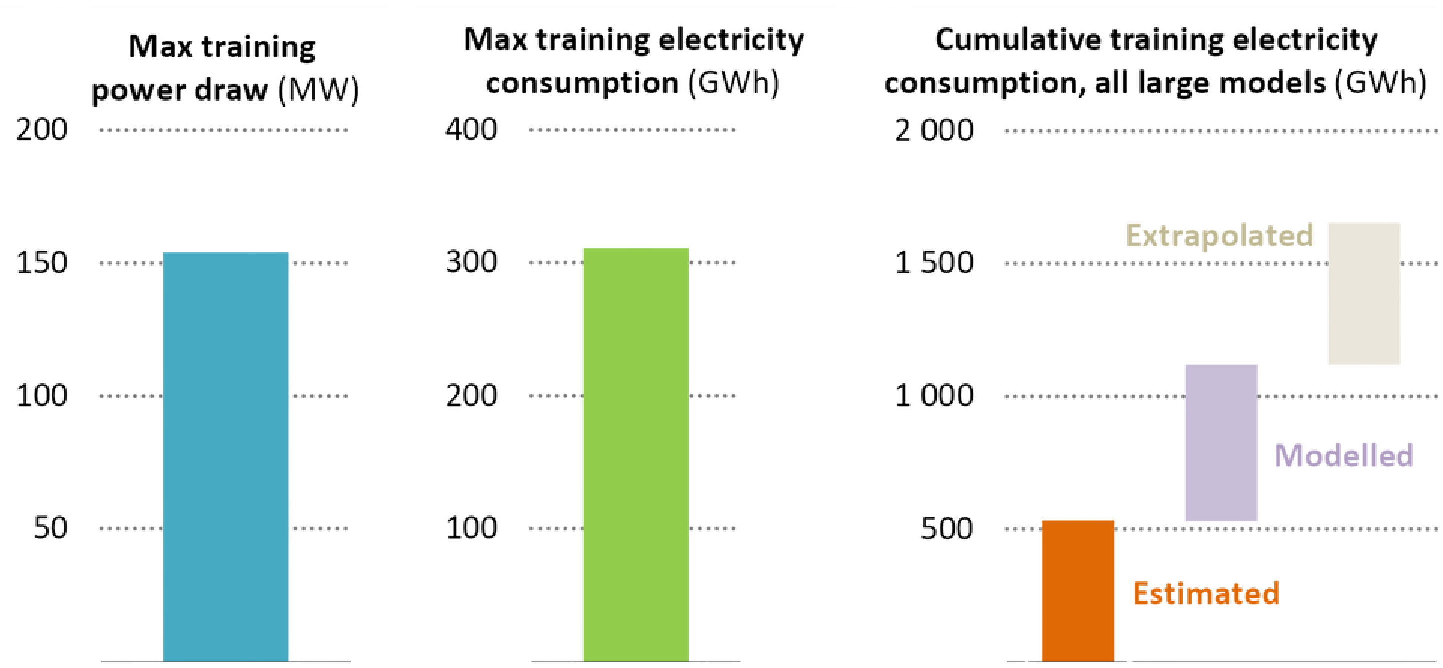
Global Electricity Demand from Digital Infrastructure

Figure 6: Global electricty demand from data centres, data trasmission networks, devices and cryptocurrency mining, 2015-2023.



The increasing load of digital infrastructure led by data centers and cryprocurrency mining poses significant challenges for global energy systems.

Figure 7: Estimated training-related maximum power draw, electricity consumption and cumulative electricity consumption for a set of large AI models since 2020.



The largest AI model today requires a maximum power draw of around **154 MW** and up to **300 GWh** of electricity for a **single training cycle**. When considering the cumulative training consumption of all large models, the electricity use approaches **1,700 GWh**, including modelled, extrapolated, and estimated data.

Table 1: Sources of electricity to match the needs to data centers.

Electricity source	Construction period	Variable or dispatchable	Global average CO_2 intensity (g CO_2/kWh)	Global average LCOE (USD/MWh)
Utility solar PV	1-4 years	Variable	0	60
Wind onshore	2-5 years	Variable	0	50
Wind offshore	3-7 years	Variable	0	110
Hydropower plant	5-15 years	Variable (run-of-river) Dispatchable (reservoir)	0	80
Conventional geothermal	3-8 years	Dispatchable	0	80
Nuclear (new)	5-15 years	Dispatchable	0	90
Nuclear (restart)	2-5 years	Dispatchable	0	60
Coal	3-6 years	Dispatchable	960	80
Gas CCGT	2-4 years	Dispatchable	390	80
Gas GT	1-3 years	Dispatchable	620	220
Grid connection	3-7+ years	Dispatchable	United States: 350 China: 600 Southeast Asia: 610 Europe: 240 World: 460	-

Notes: CO_2 = carbon dioxide; g CO_2/kWh = grammes of carbon dioxide per kilowatt hour; CCGT = combined-cycle gas turbine; GT = gas turbine; LCOE = levelised cost of electricity; MWh = megawatt hour. Construction period refers to typical projects, excluding supply chain equipment delays. Average emissions intensity is assessed on direct emissions from the average mix between 2021 and 2023. Other assumptions come from the *WEO-2024* (IEA, 2024). Nuclear (new) includes small modular reactors.

To decarbonize data centres effectively, power sourcing should prioritize dispatchable, low-carbon options such as geothermal, nuclear restarts, and hydropower (reservoir-based), while minimizing reliance on high-emission fossil fuels.

Renewable Growth vs Future Demand

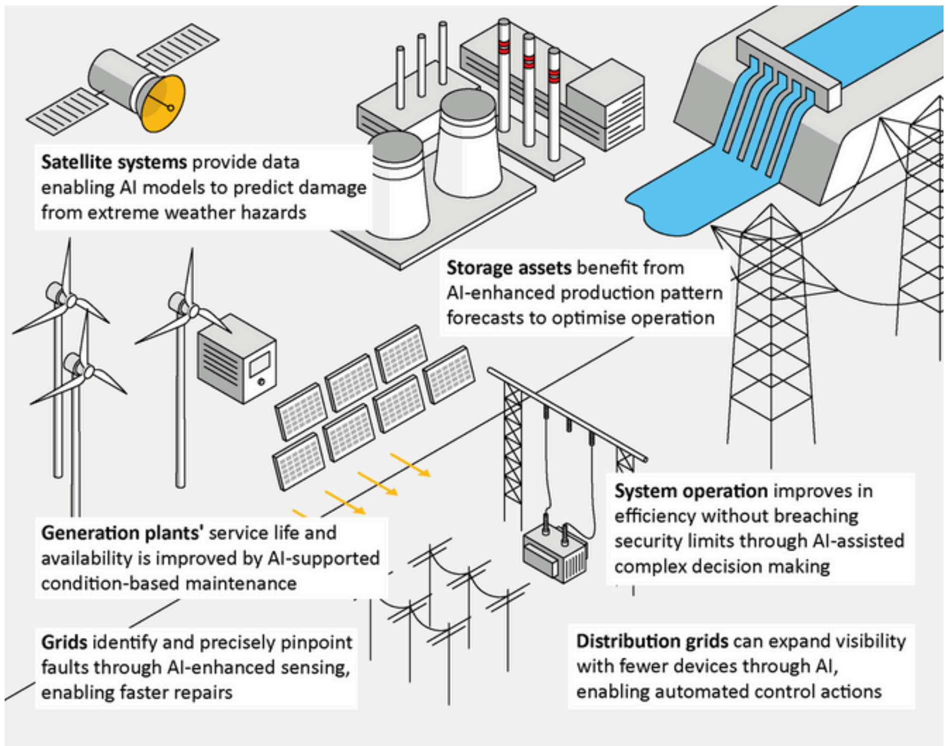
Although the adoption of renewable energy (RE) is accelerating globally, it is important to recognize that overall energy demand, particularly from data centres, AI training, and digital infrastructure, is also increasing rapidly. As a result:

- Simply expanding renewable energy capacity in parallel with fossil fuel generation is not enough. The growth in RE must **exceed the rate of demand growth**.
- This requires massive scaling of dispatchable renewables, grid flexibility, storage solutions, and policy support for fossil phaseout.

Failure to do so risks locking in dual dependency: rising clean capacity plus continued fossil fuel use, especially during peak loads or intermittency gaps.

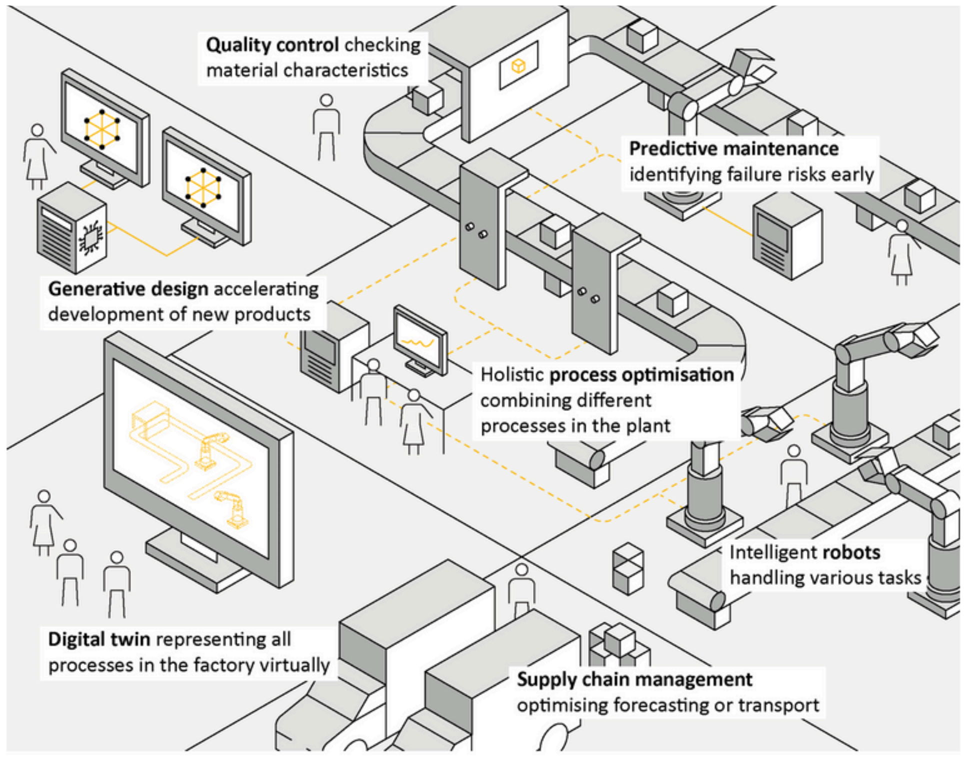
Some of AI Applications

Figure 8: AI applications in electricity generation and transmission.



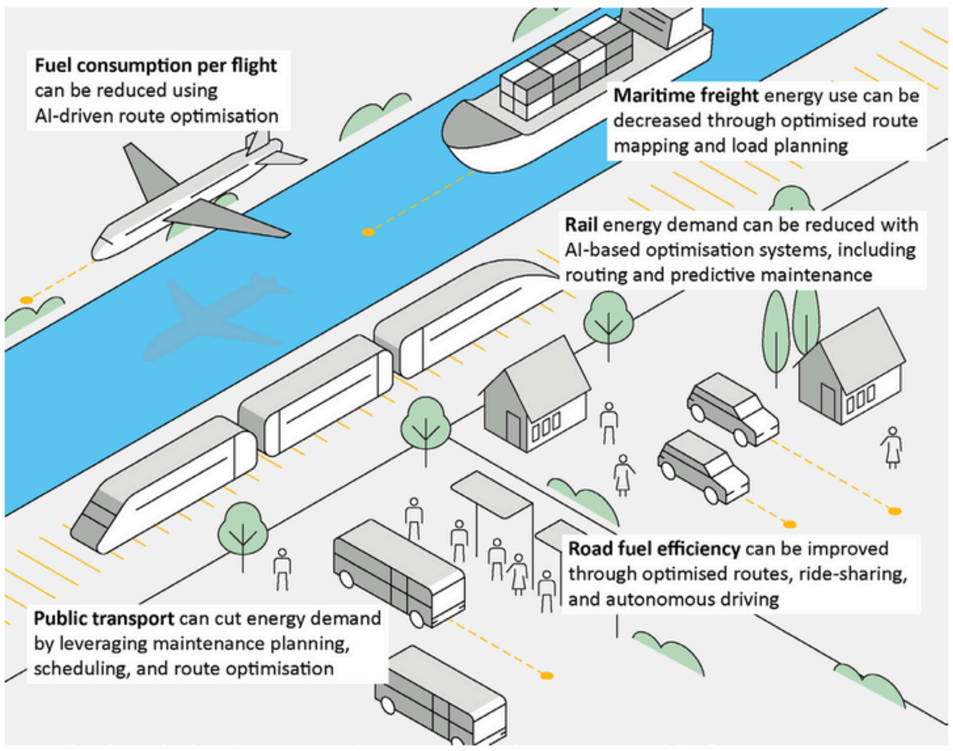
Power systems are increasingly complex, with more distributed sources and a wider set of flexibility sources, requiring more advanced operation methods that would benefit from AI.

Figure 9: AI applications in industry.



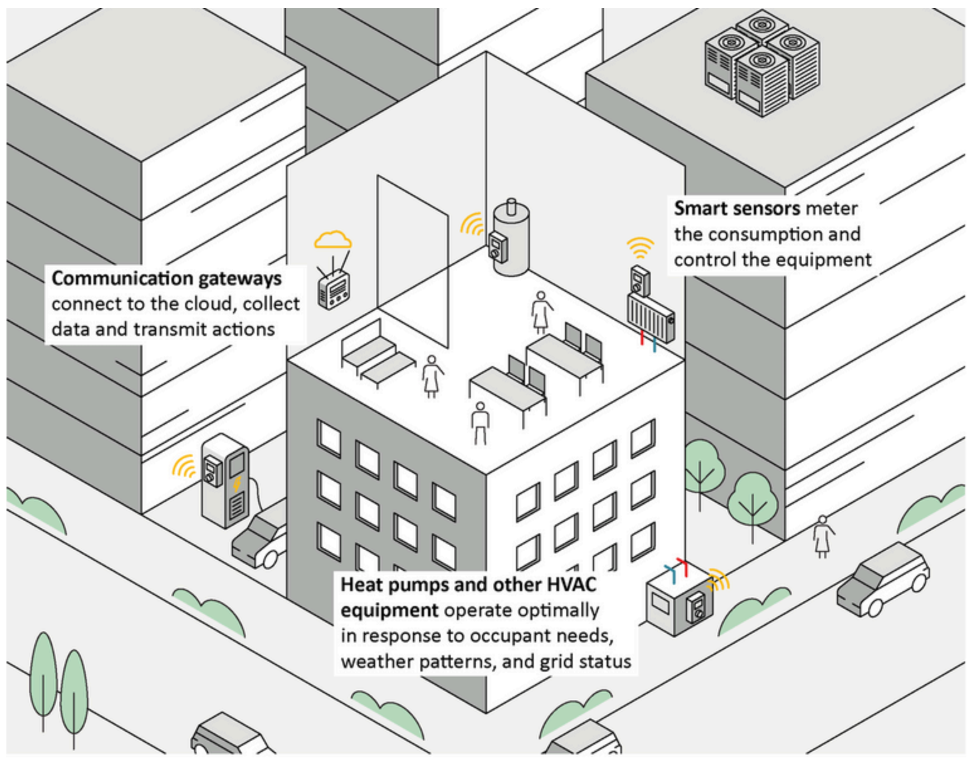
AI can improve many steps in industrial production, but optimising either single processes or the entire plant process has the most direct impact on energy demand.

Figure 10: AI applications in transport.



AI can deliver optimisation and improved operations across multiple end-uses, especially for passenger and freight urban mobility.

Figure 11: AI applications in buildings that lack network-enabled HVAC controls.



AI applications in buildings can enable energy savings even with the deployment of only a limited number of connected devices and sensors.