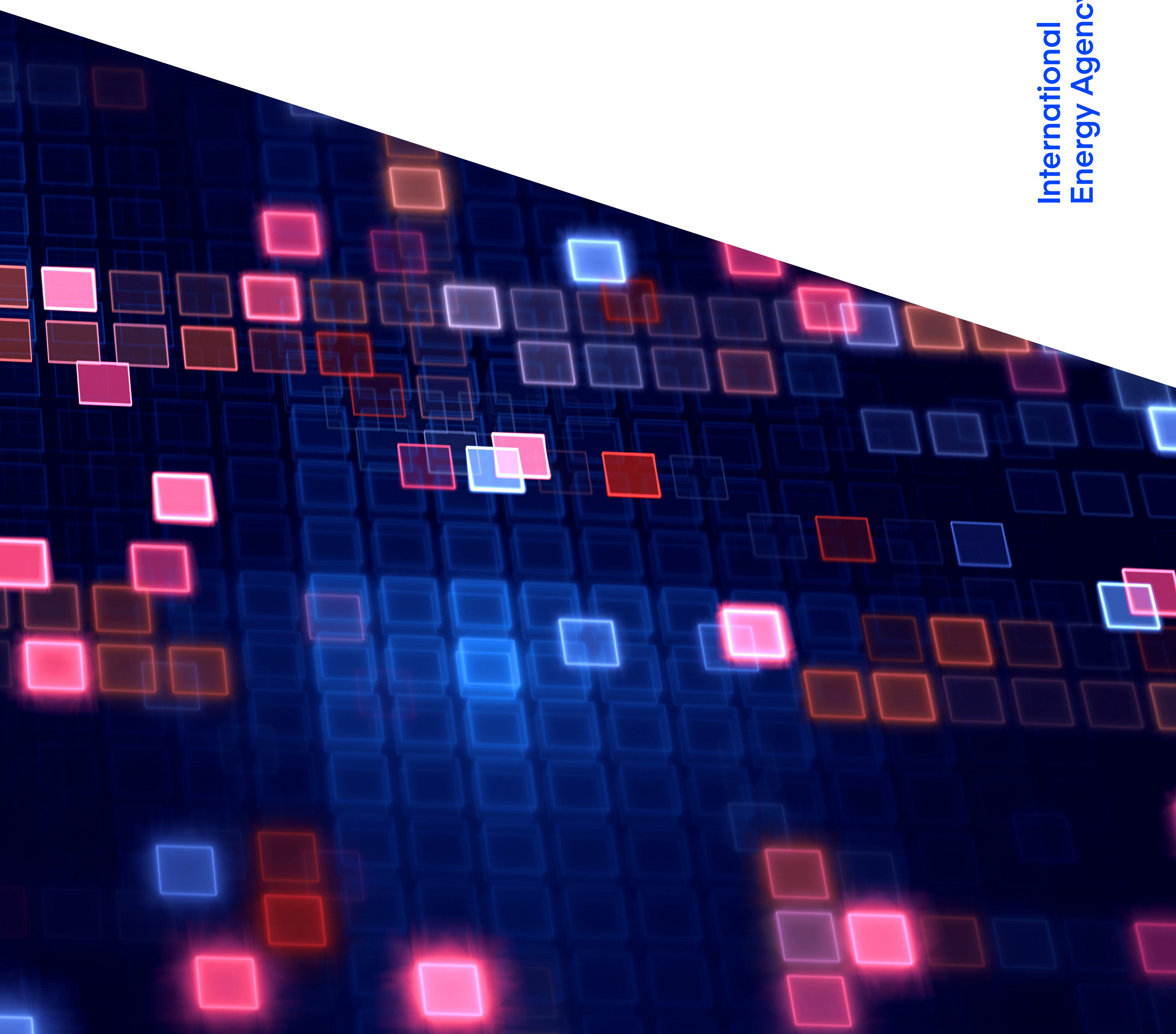
Energy and AI in East Asia



International Energy Agency

INTERNATIONAL ENERGY AGENCY

The IEA examines the full spectrum of energy issues including oil, gas and coal supply and demand, renewable energy technologies, electricity markets, energy efficiency, access to energy, demand side management and much more. Through its work, the IEA advocates policies that will enhance the reliability, affordability and sustainability of energy in its 32 Member countries, 13 Association countries and beyond.

IEA Member countries:

Australia
Austria
Belgium
Canada
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Japan
Korea
Latvia
Lithuania
Luxembourg
Mexico
Netherlands
New Zealand
Norway
Poland
Portugal
Slovak Republic
Spain
Sweden
Switzerland
Republic of Türkiye
United Kingdom
United States

The European Commission also participates in the work of the IEA

IEA Association countries:

Argentina
China
Egypt
India
Indonesia
Kenya
Morocco
Senegal
Singapore
South Africa
Thailand
Ukraine
Viet Nam

Revised version: April 2026
Information notice found at: www.iea.org/corrections

Source: IEA.
International Energy Agency
Website: www.iea.org

Abstract

This report was commissioned by the Korea Energy Economics Institute and was carried out jointly by the International Energy Agency (IEA) and the Korea Energy Economics Institute. The study has three objectives in the context of East Asia. First is to explore the possibilities presented by AI for the energy sector. Second is to examine the expected increase in electricity demand by data centres, and the impact on grid planning and operation. Third is to provide policy recommendations for embracing the opportunities presented by the application of AI to energy, as well as policies for proactively managing the challenges presented by the consumption of energy by AI. The analysis extends and updates the IEA's existing analysis, *Energy and AI*, with a focus on East Asia.

Acknowledgements, contributors and credits

The International Energy Agency (IEA) and the Korea Energy Economics Institute (KEEI) prepared this report. Leading the project was Eren Çam, Energy Analyst at the Electricity Systems and Markets (ESM) Division. Dr Sanglim Lee of the KEEI provided valuable guidance throughout the process.

The lead author was Ja Hyun Kim, with co-ordination support from Eren Çam. Other IEA colleagues provided important contributions, including Rosa Lawrence, Matthew Davis and John Moloney.

Keisuke Sadamori, Director of Energy Markets and Security at the IEA, and Professor Youngduk Pak provided expert comments and senior guidance. Valuable comments and guidance were provided by other senior management within the IEA, including Maria Sicilia, Head of ESM, and Dennis Hesseling, Head of Gas and Coal Markets Division.

IEA colleagues across the agency provided valuable input, comments and feedback, in particular, Jeseung Choi, Thomas Spencer, Siddharth Singh, Vincent Jacamon, Carlos Fernández Alvarez, Camille Paillard, Grace Henry, Augustin Lorne, Floris van Dedem, Teo Lombardo and Gabriel Saive. The report was edited by Justin French-Brooks.

Table of contents

Introduction

Artificial intelligence (AI) holds significant potential to reshape global energy systems at the critical juncture between the energy transition and technological innovation. AI can play a crucial role in digitalising energy systems as the share of variable renewable energy (VRE)[1] increases in the electricity generation mix. At the same time, driven by AI use, the expansion of data centres is expected to result in higher electricity consumption. The timely deployment of low-carbon generation sources and their effective integration into energy systems can allow the growing demand for data centres to be met while staying on track to reach decarbonisation targets.

In 2024, the International Energy Agency (IEA) launched a major set of activities to address important gaps in understanding the implications of AI for the energy sector. The first major event was the Global Conference on Energy, convened by the IEA to catalyse cross-sector discussions. It led to the IEA's contribution to the AI Action Summit in February 2025, followed by the publication of the landmark report Energy and AI in April 2025. This report highlighted the potentially synergistic relationship between AI and the energy sector. Data centres' current share of global electricity consumption is 1.5%. This is projected to more than double by 2030, reaching around 945 TWh or approximately 3% of total global electricity demand. While their share of total electricity demand is limited on a global scale, data centres have historically been highly concentrated spatially in certain regions, which poses significant challenges to local grids given their substantial power draw. Enhancing collaboration between the energy sector, the tech sector and governments can ensure that data centres meet their energy needs affordably and with minimal disruption, and can also unlock AI's full potential to effectively contribute to the energy transition.

Building on the previous report, this report offers a deeper regional perspective on the topic of energy and AI in East Asia. The region plays an integral part in both the global energy transition and the AI industry.

The rise of AI puts East Asian countries in the spotlight. Electricity consumption by the region's data centres is expected to more than double, with the expansion of AI data centres significantly contributing to this growth. Accordingly, their contribution to overall electricity demand growth in the region is expected to rise significantly. At the same time, East Asia is home to prominent semiconductor industries, producing chips that are essential for powering AI systems. Almost six

[1] In this report, VRE refers to solar PV and wind power unless otherwise specified.

out of every ten chips produced, whether graphics processing units (GPUs) or central processing units (CPUs), are manufactured in the region, underscoring East Asia's central role in the global semiconductor supply chain.

While AI is expected to be a significant driver of electricity demand, it also has major use cases in the energy sector and can contribute to more efficient and optimal processes. A major area of use is in the electricity system, specifically more optimal integration of VRE sources such as solar PV and wind. Utilising AI effectively in this field is particularly relevant for East Asia since countries such as Chinese Taipei, Japan, Korea and People's Republic of China : (hereafter, "China") have all set net zero targets, even though the targets and timelines for expanding renewable energy sources vary between countries. Korea aims to produce 29.2% of its electricity from renewables by 2030. Japan has set a similar target for 2030, although its VRE share in 2024 was nearly twice that of Korea. Meanwhile, China set a goal of reaching 1 200 GW of renewable capacity by 2030, achieving this target six years ahead of the date, by 2024.

Against this backdrop, this report aims to provide a comprehensive analysis of the potential for a collaborative relationship between AI and energy in East Asia and policy recommendations on how to maximise it.

Net zero and renewable energy source targets by country

	2025 VRE share	Announced net zero target	2030 renewables target
China	22%	2060	1 200 GW
Chinese Taipei	9%	2050	Offshore wind: 40-55 GW Solar PV: 40-80 GW
Japan	12%	2050	36-38% of total electricity generation
Korea	7%	2050	18.8% of total electricity generation

Notes: VRE: variable renewable energy. 2025 VRE shares are preliminary. China surpassed the 2030 renewables target in 2024; its latest 2035 target is 3 600 GW of wind and solar PV.
Sources: IEA analysis based on IEA (2024), Renewables 2024; Ministry of Environment (2022), Phased Goals and Actions Toward Net-Zero Transition; IEA (2025) Sixth Strategic Energy Plan – 2050 Carbon neutral; The Government of the Republic of Korea (2020), 2050 Carbon Neutral Strategy; Ministry of Trade, Industry and Energy (2025), 11th Basic Electricity Supply and Demand Plan; IEA (2025) Renewable Energy Progress Tracker; Taipower (2025), Thermal Power Status and Performance.

This report consists of three main chapters:

The first chapter – **AI for energy** – explores how AI can accelerate the energy transition and support the integration of VRE into energy systems, with a perspective on its potential in East Asia. AI applications span the entire electricity value chain, from streamlining early-stage project development to enhancing the accuracy of generation forecasting. In power grids, AI enables predictive maintenance and real-time operation, contributing to more efficient and reliable grid operation. On the end-user side, AI helps to optimise electricity consumption across both the residential and industrial sectors. It also plays a key role in advancing innovative energy technologies, such as batteries and other emerging solutions.

The second chapter – **Energy for AI** – focuses on the growing electricity demand from data centres and how to provide low-emissions electricity for them. Both global and regional trends are presented. The chapter also discusses how data centres can be connected to the grid and effectively integrated into the power system.

Drawing on the analysis, the final chapter outlines key policy actions.

What is AI?

Definition of AI

AI is a rapidly evolving concept that can cover widespread applications. Despite having no universally accepted definition, AI can be broadly defined as the "science of making machines that are capable of learning to perform tasks that are traditionally considered to require human intelligence". Official definitions of AI can vary between countries.

Official definitions of AI by key institutions and jurisdictions

Institution / jurisdiction	Definition of AI
OECD	Machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments
Korea	The materialisation of intellectual abilities of human, such as training, inference, judgement and language comprehension, by electronic means
Japan	(Technology that is) necessary for realising functions that substitute, by artificial means, the intellectual abilities related to human cognition, reasoning and judgement, as well as technologies concerning information processing systems that process inputted information using such technologies and realise functions for outputting the results
European Union	A system that is either software-based or embedded in hardware devices, and that displays intelligent behaviour by, inter alia, collecting, processing, analysing and interpreting its environment, and by taking action, with some degree of autonomy, to achieve specific goals
United States	A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments
China	No official legal definition established

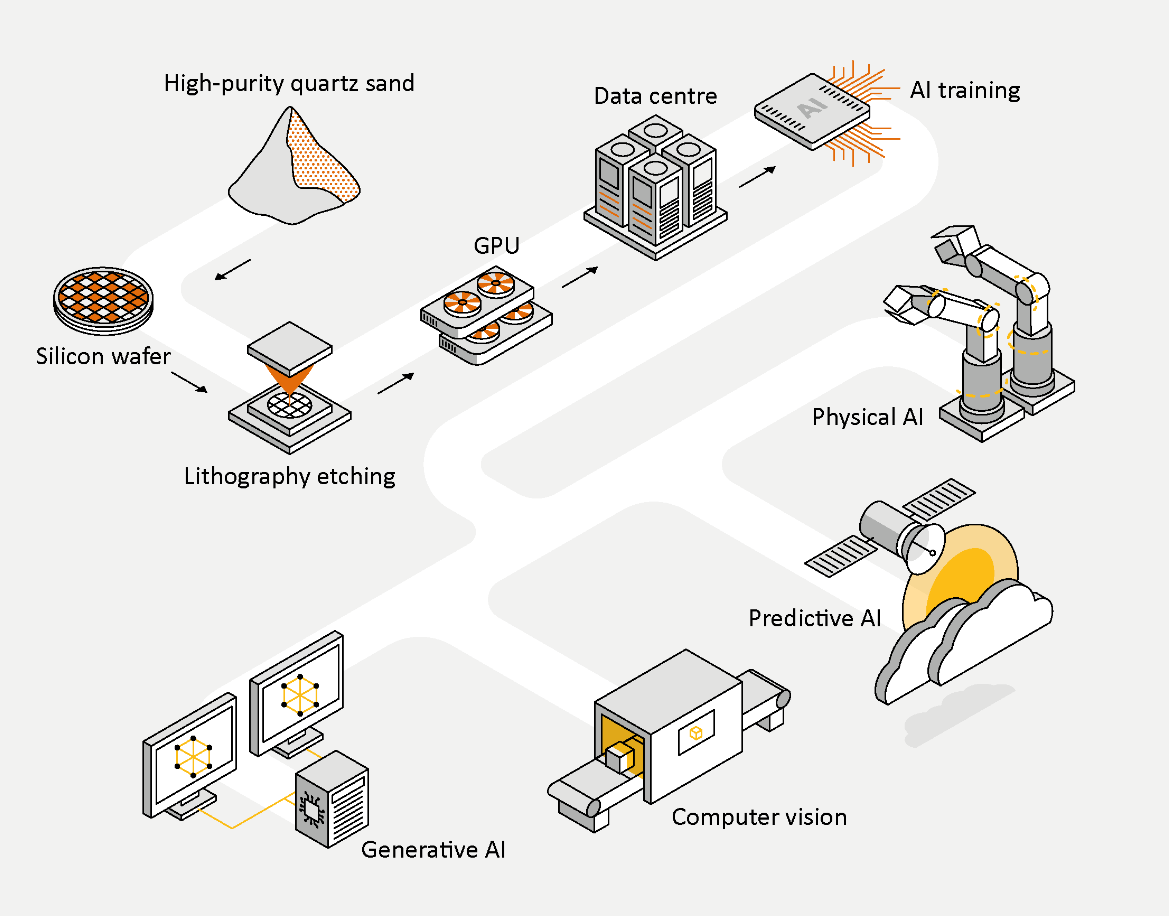
Sources: OECD (2024), Recommendation of the Council of Artificial Intelligence; Government of Korea (2024), Basic Act on the Development of Artificial Intelligence and Establishment of Trust; Government of Japan (2025), Act on Promotion of Research, Development, and Utilisation of Artificial Intelligence-Related Technologies; Legal Information Institute, 15 US Code § 9401.

AI systems can be categorised into three archetypes: (1) rules-based or symbolic AI, which is a relatively early form of AI; (2) machine learning and reinforcement learning, which are the systems that identify patterns and make decisions based on training data; and (3) neural networks and deep learning, with a structure inspired by human brains, where data are passed through layers of interconnected "neurons" to process perceived information. Contemporary AI systems and applications, including AI chatbots, often integrate multiple techniques, making the boundaries between these approaches somewhat ambiguous.

Infrastructure of AI

Physical infrastructure is fundamental to the development of AI. AI infrastructure is composed of several key components, including data centres.

Overview of AI Infrastructure



Source: IEA (2025), Energy and AI.

Data centres lie at the heart of AI infrastructure. They are the "brain" of AI, where the data are processed and stored, and typically where the outputs are produced.

The structure of a data centre may vary depending on its purpose, but a few key components are found in most AI data centres: information technology (IT) equipment (servers, storage systems and networking equipment), cooling systems, and grid-related facilities (uninterruptible power supply [UPS], etc.).

The most integral part of a data centre's IT equipment is the **server**, where all the "intelligence" is created by processing the data in two major operational stages of AI development: **training** and **inference**.

The **training** stage is the process by which the model learns, using large datasets to recognise patterns. Based on the patterns it identifies during training, the model can generate predictions and infer possible outcomes. The **inference** stage is where AI models that have finished training are deployed in service. Based on the data from the training stage, the model develops an answer (inference) to the question (query) that the user asks. These training and inference stages of AI development determine what combinations of processing units are required.

Two operational stages of AI

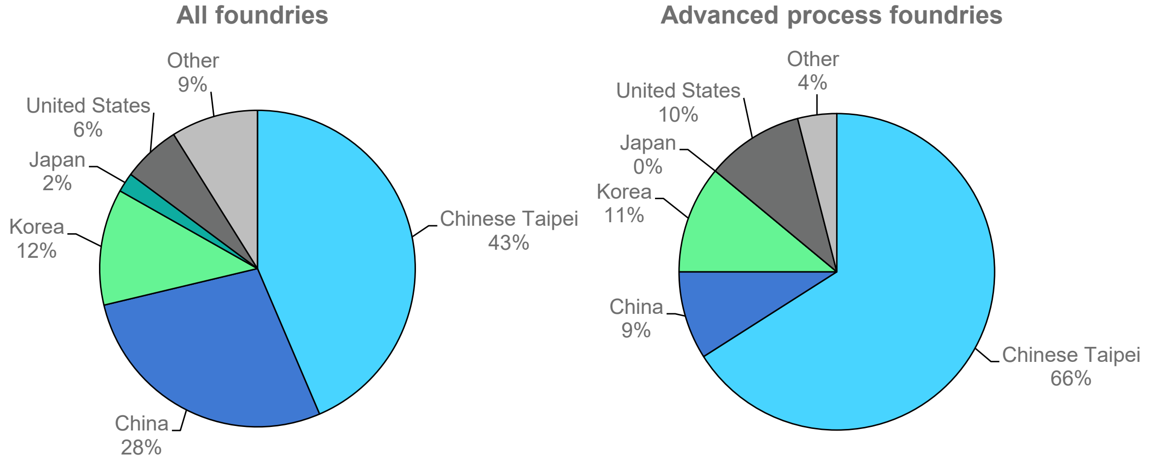
	Training	Inference
Application	Learn from datasets to identify patterns	Develop an answer to the query by calculating from training data
Computing architecture	GPU-intensive	Heterogeneous processor mix (GPU/CPU/TPU/ASIC)
Power consumption	Energy-intensive	Relatively less energy-intensive*
Latency sensitivity	Less latency-sensitive	More latency-sensitive
Power consumption profile	High sustained power consumption, with sub-second swings between training cycle stages	Rapid fluctuations in demand, depending on user behaviour

* Energy intensity refers to power (MW) at any given time. A given model may serve inference queries for months or years, and across more data centres, thereby adding up to more energy (MWh) than training.
Notes: GPU = graphics processing unit; CPU = central processing unit; TPU = tensor processing unit; ASIC = application-specific integrated circuit.
Source: IEA (2025), Energy and AI.

If an AI cluster is mainly for training, such as for deep learning or a large language model (LLM), its infrastructure is typically heavily based on GPUs. Compared to a CPU, which processes data serially, a GPU has a smaller and specialised core that focuses on intensive processing of massive datasets in parallel. AI infrastructure for inference, on the other hand, can have a mixed combination of different types of processing units (e.g. a mix of CPUs, GPUs, tensor processing unit [TPUs] and application-specific integrated circuits [ASICs]).

GPUs have become indispensable to AI development, and since their manufacturing remains highly concentrated in East Asia, this region is strategically important for the continued growth of AI. As of the first quarter of 2025, approximately 90% of the world's ten largest semiconductor manufacturing companies by market share were headquartered in East Asia, with Chinese Taipei accounting for 73%, China 10% and Korea 8%. While their manufacturing facilities (foundries) are globally distributed to some extent, nearly 60% of them are located in East Asia. This means that almost six out of every ten chips produced, whether GPUs or CPUs, are manufactured in the region, underscoring East Asia's central role in the global semiconductor supply chain.

Geographic distribution of foundry capacity, 2024



Source: IEA analysis based on data from TrendForce (2024), Insights.

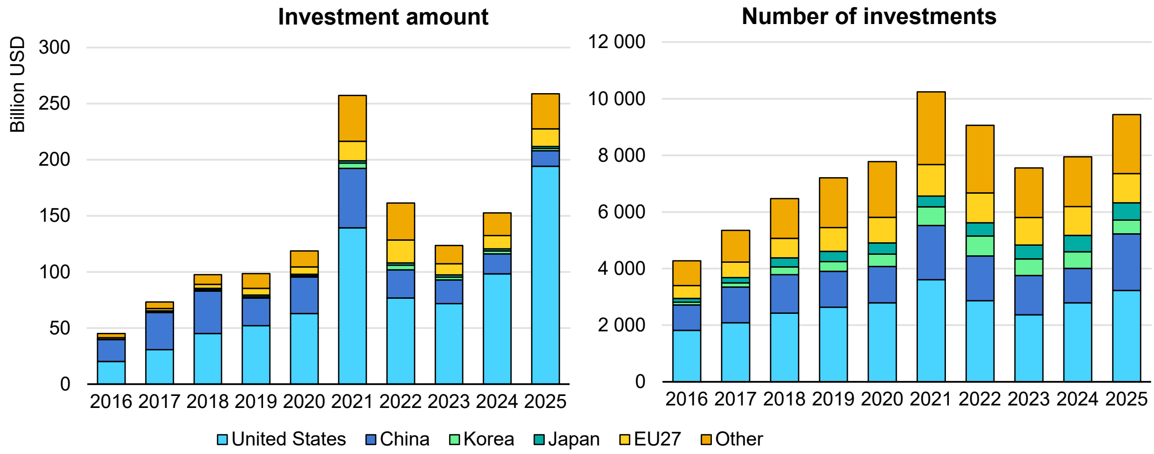
AI in East Asia

The AI industry has been expanding rapidly as its adoption widens dramatically

AI has become a strategic focus for governments and industry leaders across East Asia. Public awareness of its potential expanded after the widely publicised 2016 match between the Korean Go player, Se-dol Lee, and the AI computer Go programme, AlphaGo. In the years that followed, China (2017), Korea (2020) and Japan (2025) announced national AI strategies, signalling a shift toward more structured policy support. The emergence of generative AI applications from 2022 further accelerated AI adoption by expanding use cases to a wider range of end users.

The number of venture capital investments in AI has been on the rise. While the United States has been leading the trend, China, the European Union and Korea have also emerged as major players in the landscape. The amount of money invested in AI by venture capital globally reached a record high of almost USD 260 billion in 2025.

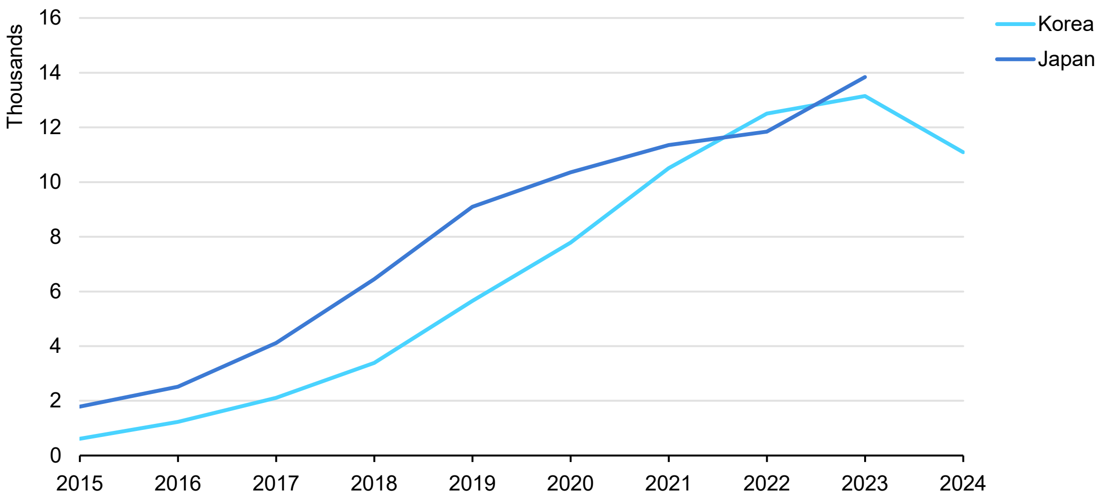
Venture capital investment in AI by country and region, 2016-2025



Source: IEA analysis based on data from OECD (2026), OECD.AI Policy Navigator.

The number of AI-related patents is another powerful proxy to track AI trends. In both Korea and Japan, steady growth is being seen in the number of patent applications for AI-related technologies such as big data processing and image searches. In 2015, patent application numbers were below 2 000 in both countries, but they then continued to increase and reached over 10 000 applications in 2022. Between 2015 and 2022, the number of AI-related patent applications grew at a compound annual growth rate (CAGR) of 46% in Korea and 29% in Japan. Meanwhile, as of 2023 China was the country owning the most AI-related patents granted worldwide, accounting for 69.7% of the global total.

Number of new AI-related patent applications in Korea and Japan, 2015-2024



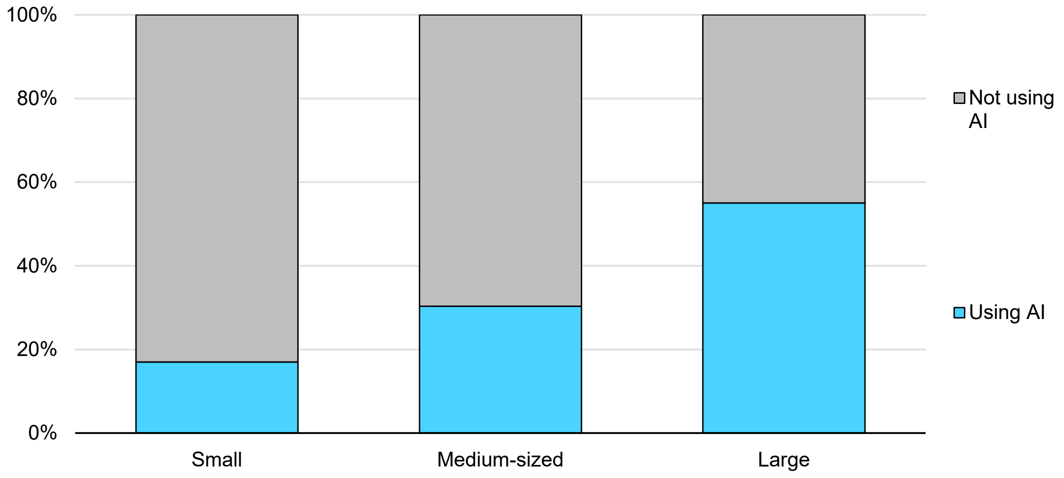
Note: At the time of writing this report, the most recent full year data for Japan and Korea were for 2023 and 2024 respectively.
Sources: Korea Intellectual Property Office (2025), 4th Industrial Revolution-related Patent Statistics; Japan Patent Office (2025), Annual Report on Patent Administration.

AI has seen very rapid adoption among the general public, less so among businesses

AI researchers made breakthroughs in image generation in 2014 with the invention of generative adversarial networks (GANs), and in text generation in 2017 with the introduction of the Transformer.[2] Key products such as ChatGPT were introduced as recently as 2022. Considering how recent these research breakthroughs and products are, the adoption of generative AI in Korea has occurred at a very high pace, nearly twice as fast as the internet during its initial roll-out in the mid-1990s, according to the Bank of Korea. Far quicker than other social media platforms, ChatGPT reached 100 million users globally in just two months, as noted by the Ministry of Internal Affairs and Communications in Japan. As of 2025, 64% of Koreans reported using generative AI for professional or personal purposes. However, corporate adoption is more variable. According to the ministry's recent white paper on AI, only 19% of Japanese corporates are actively utilising AI in their business operations, as compared with 37.4% of US corporations. And a survey recently conducted by the US Census Bureau states that just 9% of the 1.2 million US firms surveyed were using AI technology as of the second quarter of 2025.

Business size is a significant factor that affects the pace of AI adoption. According to the Korean Federation of SMEs, its 2023 survey showed that only 5% of SMEs in Korea had adopted AI or were in a progress of adopting it, whereas 87% had no plan to adopt any AI technology in the foreseeable future. The European Union has also been shedding light on the widening discrepancies in AI adoption between large corporates and SMEs, as the disparity reportedly continues to grow over time.

Proportion of enterprises using AI technologies by size in the European Union, 2025



Notes: Enterprise size breakdown: Small = 10-49 employees and self-employed; Medium-sized = 50-249 employees; Large = 250+ employees.
Source: Eurostat (2025), Use of artificial intelligence in enterprises.

[2] In this context, Transformer refers to the software technique based on attention mechanisms, which underpins large language models (LLMs). This is unrelated to electrical transformers, which change the voltage of AC electricity.

How are energy and AI connected?

AI for energy: Harnessing AI for the energy transition

AI is increasingly being applied in the broader energy field. This can range from the use of AI in the oil and gas sector to its application in thermal power plants and renewable energy sources in the power sector. Transmission and distribution, storage and end-use sectors (e.g. buildings and transport) are relevant sectors that can greatly benefit from AI applications. Reflecting the renewable energy targets in East Asia and the challenges associated with VRE integration, this report focuses on AI applications related to renewable energy sources, grids, end-use sectors and energy innovation related to clean energy sources.

One of the most pertinent issues for integrating higher shares of VRE is addressing its variability. VRE is influenced by temporal weather conditions such as wind speed and solar radiance, which fluctuate daily and between seasons. Therefore, the accurate planning, forecasting and operation of VRE projects is essential to support their effective integration, all of which can be improved by AI applications. AI can contribute to streamlining the VRE project development process, such as siting and permitting. It can facilitate more accurate and granular generation forecasting and optimisation, as well as more efficient operation and maintenance (O&M).

Optimising power grids – transmission and distribution networks – is another key area where AI can play a significant role in enhancing efficiency while maintaining security of supply. Amidst the energy transition, a rapidly changing energy system environment with larger loads (data centres etc.) and weather-dependent supply is requiring grid operators in the region to effectively plan, operate and manage increasingly complex power grids.

Energy for AI: Sustainably powering AI

In the AI lifecycle, the operational stages – training and inference – account for the majority of the energy demand, far exceeding the energy consumption of the associated hardware manufacturing and building construction. Large language models (LLMs), such as ChatGPT, Claude and Copilot, and other generative AI generally consume more energy per model training run, per inference and aggregated across all users than traditional machine learning, such as image recognition, speech to text and weather forecasting. As AI models become more prevalent and data centres continue to expand across many regions, the question of how to sustainably power the data centres is becoming essential, whereby a diverse range of supply sources would be needed across different regions. At the same time, data centres tend to be highly concentrated in certain locations and the expansion can put significant pressure on existing power grids. These aspects are covered in Chapter 2 of this report.

Training large AI models requires substantial GPU clusters to process vast datasets in parallel, driven by growing user demand. Training typically takes hours, days or even weeks. Training is a long-running batch process, which does not involve interaction with end users. Therefore, latency to end users is not a significant concern for training, which can be performed remotely from demand centres. Conversely, the results of many inference tasks are returned to end users or edge devices, so proximity to users may be desirable to reduce latency. Chapter 2 further elaborates on how these stages differ in configuration and energy use and highlights the three key determinants of AI data centres' energy use: power usage effectiveness (PUE), coolants and the current conversion.

PUE measures the operational efficiency of a data centre by comparing IT equipment energy use with total electricity consumption. Cooling systems also play a major role, as a higher GPU density generates more heat. Cooling can range from 7% to 30% of total electricity use depending on data centre efficiency. In various applications, waste heat from data centres can be used effectively for local heating needs, which may need to be considered during the choice of location in the early design stage.

Data centres and other energy-intensive industries have a growing appetite for procuring low-carbon power to satisfy their climate ambitions and hedge price risk. Chapter 2 discusses the unique aspects of Korea's cost-based electricity pool and electricity procurement options, with a focus on the country's nascent power purchase agreement (PPA) landscape.

The increasing energy demand of data centres, combined with their regional concentration, is leading to long grid connection queues in many regions of the world. Once connected to the grid, the operational characteristics of data centres' load profiles can pose system integration challenges. Chapter 2 finishes with a discussion of these challenges and potential solutions.

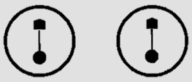
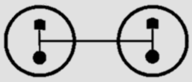
Regional context and challenges

The nexus between AI and energy in East Asia must be understood within the region's unique context, shaped by the geographical and structural characteristics of its energy sector. A lack of cross-border interconnections and the geographical imbalance of electricity supply and demand are potential challenges in this context. Tailoring policies and regulatory frameworks to the local conditions are crucial to fully harness the potential of AI in advancing energy security and sustainability across the region.

The absence of cross-border interconnections presents a challenge to power system operation

The absence of cross-border interconnections in various East Asian countries can pose challenges to power system operation, such as limited options to manage variability and maintain grid stability while integrating VRE. Japan and Chinese Taipei are both island countries. Although Korea is situated on a peninsula, it functions almost like an island country with regard to its power system due to the lack of interconnections. By contrast, China has interconnections with neighbouring Greater Mekong Subregion countries, such as Lao PDR, Myanmar and Viet Nam. It also has various interconnections with the northern part of the continent, including Russia, Mongolia, Kazakhstan and Kyrgyzstan. However, the capacity of these interconnectors is limited and the networks are not as integrated as those in Europe.

Grid systems in East Asia

Island or quasi-island	Large interconnected system
	
Chinese Taipei, Japan, Korea	**China**
Limited resources for short-term supply and demand balance, less diversity and flexibility volumes	Diverse choices of grid management resources

Source: IEA (2024), Integrating Solar and Wind.

The lack of cross-border interconnections in East Asia highlights the potential for AI to address electricity system management challenges within each country. Chapter 1 contains a section dedicated to exploring how AI could provide innovative solutions to allow the grid to be operated effectively and safely, while integrating increasing shares of VRE.

Because of their geographical and geopolitical conditions, there are currently no cross-border interconnections linking the grids of Korea, Japan, China and Chinese Taipei. This disconnection, with most countries relying on isolated grids (China being the exception), limits cross-border electricity trade. Plans are currently underway to develop two projects to connect Korea and Japan. The EEL and FUGU projects are in the planning phase and, once completed, will be approximately one-third and one-half of the length of the world's longest subsea power cable (764 km between Denmark and the United Kingdom), respectively. As the VRE share increases, the potential benefits of cross-border interconnection become more relevant, such as enhancing energy security, facilitating greater VRE integration and increasing economic efficiency.

Japan's power grid is notably fragmented; it operates at two grid frequencies – 50 Hz in the east and 60 Hz in the west − creating challenges for regional integration. Currently, only three frequency converters with a total of 2.1 GW connect these grids. The Organisation for Cross-regional Coordination of Transmission Operators (OCCTO) plans to expand this capacity to 3 GW by 2027.

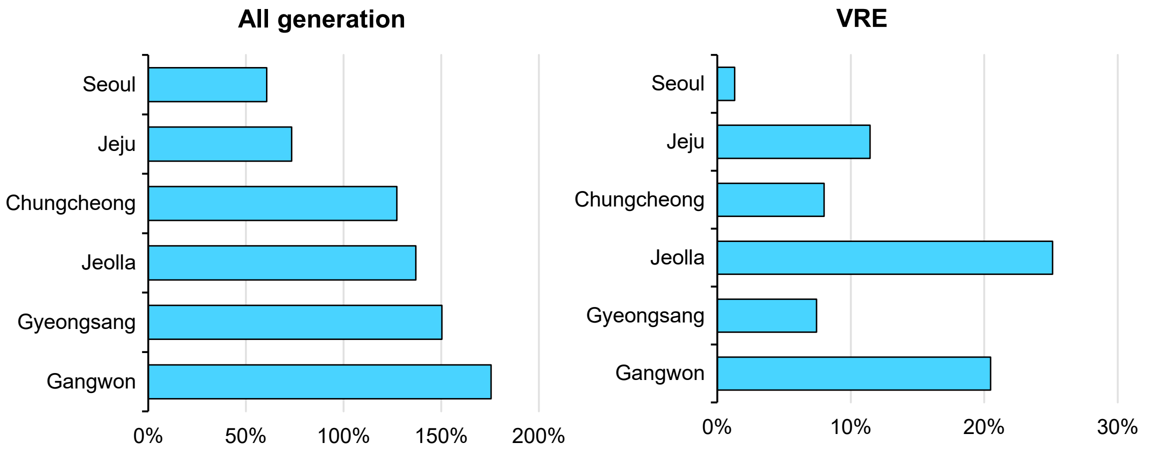
The geographical imbalance between energy supply and demand is widening

As in many other regions, East Asia is experiencing a widening geographical imbalance between energy supply and demand. In most countries, electricity is typically generated in remote areas, near a fuel source such as a coal mine or in a windy area, whereas demand is generally focused in urban and industrialised areas. VRE capacity is expanding rapidly, and so too is demand, including for data centre capacity. However, the construction of transmission lines between them is lagging behind, partly due to the longer construction timeframe, which can eventually lead to significant bottlenecks. This is especially apparent in large metropolitan areas.

Such a disparity can be observed in Korea. The southwest region – Jeolla – is known for its high VRE share from both solar and wind due to its rich solar and wind resources. The Jeolla region accounts for 40% of installed solar PV capacity and 24% of wind turbine capacity in Korea. Except for one nuclear power plant on the west coast (Hanbit nuclear power plant), most nuclear power plants and many industrial consumers are located along the southeast coast. However, around 40% of the generated electricity is consumed in the Seoul Metropolitan Area, where around one-fifth of the Korean population reside.

As of 2024, the electricity self-sufficiency rate of the Seoul Metropolitan Area was 67%, which refers to the proportion of the electricity used by consumers in the area that is generated within the area. All other sub-regions, except for Jeju Island, recorded electricity self-sufficiency of over 100%, which implies that there is more generation capacity than regional demand and the surplus generated electricity is exported to areas of undercapacity, such as the Seoul Metropolitan Area. Such a regional distribution of electricity supply and demand may increase the need for more transmission capacity to transport the electricity to regionally concentrated demand centres.

Electricity self-sufficiency rate by sub-region of Korea, 2024

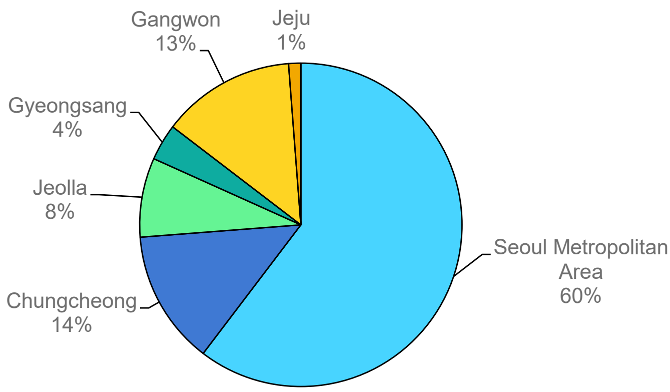


Notes: Electricity self-sufficiency rate = (electricity generated within the region/electricity consumed within the region). VRE self-sufficiency rate = (electricity generated by VRE within the region/electricity consumed within the region). Sub-regions = Seoul Metropolitan Area (Seoul-si, Gyeonggi-do, Incheon-si); Jeju (Jeju-do); Chungcheong (Chungcheongnam-do, Chungcheongbuk-do, Sejong-si); Jeolla (Jeollanam-do, Jeollabuk-do, Gwangju-si); Gyeongsang (Gyeongsangbuk-do, Gyeongsangnam-do, Busan, Daegu, and Ulsan); Gangwon (Gangwon-do).

Source: IEA Analysis based data from KEPCO (2025).

Although it has the lowest electricity self-sufficiency rate, the Seoul Metropolitan Area currently hosts most of the country's data centres. Around 60% of data centres are currently located in either Seoul, Incheon or Gyeonggi, all within the Seoul Metropolitan Area. Chungcheong is home to a large coal-fired generation fleet and has 14% of Korea's data centres. Jeolla has the highest share of renewables but hosts only 8% of existing data centres.

Regional distribution of data centres in Korea, 2024



Notes: Sub-regions = Seoul Metropolitan Area (Seoul-si, Gyeonggi-do, Incheon-si); Jeju (Jeju-do); Chungcheong (Chungcheongnam-do, Chungcheongbuk-do, Sejong-si); Jeolla (Jeollanam-do, Jeollabuk-do, Gwangju-si); Gyeongsang (Gyeongsangbuk-do, Gyeongsangnam-do, Busan, Daegu, and Ulsan); Gangwon (Gangwon-do).

Source: IEA Analysis based data from The Korea Data Center Council (2025).

The Korean government is considering locationally differentiated grid connection tariffs, so that new data centres outside the Seoul Metropolitan Area may receive discounts on grid connection charges, depending on eligibility and tariff design. This locational price signal is intended to ease any localised grid strain from the concentration of data centre loads. The Korean government introduced the Non-metropolitan data centre consulting support centre to provide stakeholders with information on incentives, planning and system capacity. This advice channel is intended to help maximise the effectiveness of relevant policies, which were designed to encourage the location of new data centres in areas with low congestion and high levels of VRE generation.

East Asia has a high reliance on energy imports

Whilst East Asia is largely characterised by a lack of electricity imports, the region is a major importer of other forms of energy, such as coal and gas, compared with other comparable economies. Net energy imports, calculated as imported energy as a share of total energy supply, was over 80% in Korea and Japan, and even up to 95% in Chinese Taipei. China is the largest energy consumer in the world, but it has relatively lower net energy imports – 24% – than other East Asian countries due to its domestic production.

In East Asia, where electricity tariffs are typically tightly regulated and energy systems are heavily dependent on imported fuels, powering energy-intensive AI infrastructure presents distinct challenges. Despite domestic retail price control, electricity costs can be susceptible to global energy market volatility. By better enabling the integration and optimisation of VRE, which is generally less exposed to global fuel price fluctuations, AI can contribute to reducing vulnerability to external energy shocks.

Having high reliance on energy imports is closely linked to the continuing dominance of fossil energy sources in the region, namely coal and gas. Preliminary data show that in 2025 Korea, Japan and China generated 58%, 60% and 58% of their electricity from coal and gas, respectively. Coal was China's largest source of electricity, while Japan had the highest share of gas-fired generation. Many large technology companies active in the AI sector have sustainability targets for their projects. Therefore, the development of low-emissions energy sources, such as solar, wind and nuclear, can act as a signal to entice investors into a country. In Korea, nuclear generation constituted more than 30% of the electricity generation mix in 2025, compared with 10% in Japan and 5% in China.

Chapter 1. AI for energy

Artificial intelligence (AI) applications can span the entire electricity value chain, from supporting the project planning stage and improving the planning of maintenance, to providing more accurate generation forecasts and predicting consumption. On the grid side, AI can help with predictive maintenance, allow for more optimal usage of grids and enhance grid stability. On the end-user side, optimising energy use in buildings and industrial processes is one of the aspects where AI tools can help. On the broader front, AI can support energy innovation, helping with the development of new technologies and solutions in the energy sector. As energy systems become increasingly digitalised and decentralised, AI offers new opportunities to accelerate decarbonisation, reduce costs and improve system reliability.

East Asia is characterised by high energy demand, rapid digital infrastructure development and clear policy commitments to clean energy transitions. These factors make it a region that can become a significant adopter of AI technologies, particularly in the electricity sector. Insights from the region can inform global efforts to harness AI in support of secure, sustainable and inclusive energy transitions.

AI for renewable energy

The electricity sector is characterised by projects with long lead times that typically involve multiple stakeholders, ranging from investors and project developers to utilities, transmission system operators and regulators, among others. This is particularly the case for weather- and location-dependent renewable energy.

This begins from the decision on where to site the generation assets and extends to how to operate and maintain them. In many parts of the world, renewable energy projects face extended permitting processes as well as long connection queues for the grid. Once they are deployed, these weather-dependent generation technologies require high-quality forecasts to manage their variability with efficient real-time balancing. In this context, AI applications are emerging as powerful resources for accelerating the deployment of renewable energy sources by supporting the planning and permitting stage, while also providing effective tools for improved operations.

Project development: AI as a tool to accelerate the deployment of variable renewable energy

AI can be mobilised from the earliest stages of renewable energy project development. Globally, the siting and permitting processes are recognised as major bottlenecks that hinder timely and large-scale deployment of variable renewable energy (VRE). For example, the development of offshore wind projects can last over a decade before reaching operational status, regardless of jurisdiction.

These delays stem from compounding issues, including regulatory and administrative barriers, lack of information and technical capacity, limited grid connection and complex societal engagement. Such obstacles not only extend project timelines but also drive up the costs, ultimately deterring future project investments. While AI can help, pursuing more traditional solutions at the same time is important, including queue prioritisation, tenders for access rights, fees for connection applications and increasing grid investment.

Siting and planning

In renewable energy project development, siting and planning are increasingly recognised as critical stages where AI can be leveraged to accelerate timelines while enhancing accuracy and efficiency. When selecting locations and designing the project for wind turbines or solar PV installations, developers need to evaluate key factors such as resource availability (e.g. wind speed and solar irradiance), grid connectivity and environmental impact. Inaccurate assessments can lead to poor site selection and underperforming projects.

The IEA has highlighted the importance of improving resource assessment, particularly for wind energy. AI can support this process by optimising key variables and building predictive and forecasting models using large and diverse datasets, such as satellite imagery, geospatial information and meteorological data. This potential is especially relevant for the development of a vibrant offshore wind industry in East Asia – one of the world's leading regions for offshore deployment. In 2024, East Asia accounted for 44% of the total awarded offshore wind auction capacity globally (56.3 GW), with major contributions from China, Korea, Chinese Taipei and Japan.

Offshore wind project development in East Asia faces unique challenges, such as seasonal wind patterns (monsoon)[3] and limited grid infrastructure. These factors constrain site availability and add complexity not only to siting, but also planning

[3] A monsoon climate is characterised by a dramatic seasonal change in direction of the prevailing winds, bringing a marked change in rainfall. Unlike Europe and most of the United States, East Asia experiences rapid changes in wind direction as the seasons change. In the winter, dry north-westerly winds originate from the Siberian region, whereas in the summer, warm moisture-laden south-easterly winds originate from the North Pacific Ocean and South China Sea.

and permitting. AI tools can play an important role in navigating these constraints by integrating, organising and streamlining relevant datasets and regulatory frameworks. This goes beyond wind resource maps to include nuanced data, including ecological and economic zones, grid connection points, historical permitting records and local community engagement data, enabling more informed and efficient siting decisions.

For example, seasonal wind patterns require dynamic siting and planning approaches that consider multiple scenarios, involving turbine orientation, tower design, location and grid availability. During the winter, strong north-westerly winds prevail, whereas in the summer the wind direction shifts almost completely, often blowing from the southeast. This seasonal reversal in the region complicates wind farm design and, if not properly accounted for, can significantly affect project profitability. However, these challenges do not diminish the region's vast wind potential. Instead, it underscores the importance of advanced planning tools. Predictive AI algorithms, such as deep learning, can play a critical role in optimising wind farm siting and design to maximise output under varying seasonal conditions.

Region-specific wind data can be integrated into advanced AI models to analyse the impact of upstream wind turbines on downstream turbines in the same farm. One example is the Wind Plant Graph Neural Network (WPGNN), a form of deep learning developed by the US Department of Energy's National Renewable Energy Laboratory.[4] This AI-based surrogate model estimates power production based on different wind farm layouts, using wind wake[5] models, thereby helping to optimise wind farm design. Trained on 250 000 randomised wind farm layouts under varying atmospheric conditions, WPGNN enables developers to simulate multiple configurations and identify optimal designs.

By using WPGNN, developers can optimise land area requirements, reduce the levelised cost of electricity (LCOE) and maximise their revenue capture by plotting turbine wakes. In the case of onshore wind farms in the United States, WPGNN-driven siting and planning simulation shows an average 18% reduction in land requirements per site. It also reduced LCOE by between 1.1% and 2.7% on average, depending on site capacity. Total annual revenue also increased by USD 3.7 million for large-scale farms.

Just as AI models like WPGNN support informed decision-making in wind farm development, similar approaches can be applied to solar PV development, where spatial, environmental and technical factors also require complex, data-driven analysis. Satellite-based AI models now provide solar data in a high resolution, reducing the need for costly ground stations. For example, Google Maps Platform

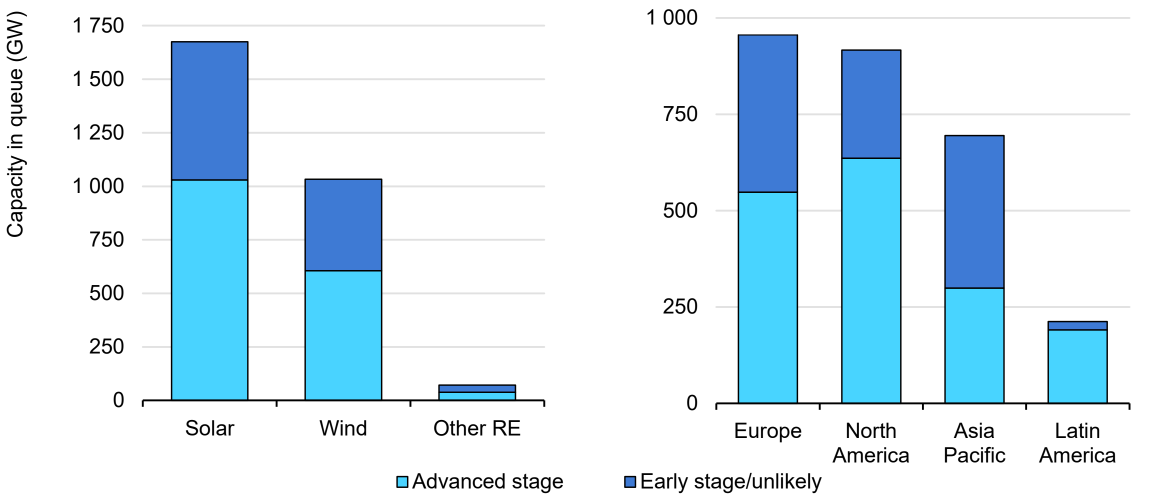
[4] The National Renewable Energy Agency (NREL) has since been renamed as the National Laboratory of the Rockies (NLR).
[5] The wind wake effect refers to a physical phenomenon that occurs behind wind turbines when they generate energy. The trail left by each turbine may decrease the wind's momentum or velocity, while increasing the amount of turbulence. Both impacts lead to power losses in downstream turbines.

Solar API combines satellite imagery, atmospheric data and ground measurements to generate high-resolution solar irradiance maps. It enables businesses of all sizes to make data-driven decisions for solar project design. In Singapore, GetSolar developed automated rooftop solar evaluations using machine learning to detect slope, obstacles, materials and the surface area suitable for solar panels from high-resolution satellite images during angle and direction planning to capture as much sunlight as possible. This kind of decision is well suited to a machine learning solution because it is a “sharp” question, with well-structured numeric inputs and a clear objective. The use of AI to analyse timeseries data is already an established use case.

Permitting and grid connection queues

Grid connection remains a major bottleneck in many regions when expanding renewable energy capacity. Among the countries surveyed by the IEA, around 1 700 GW of renewable energy projects that were at an advanced stage of development in mid-2025 were awaiting grid connection. In East Asia, despite ambitious renewable energy targets, permitting similarly remains a major hurdle. Developers must navigate complex environmental assessments and compliance with national and local regulations, all while negotiating with local communities. Without effective tools, developers often face fragmented procedures, inconsistent requirements and inter-agency co-ordination challenges, which slow deployment and divert resources from actual project development.

Global capacity of renewable energy projects in connection queues by project stage (left) and the regional breakdown of advanced-stage projects in queues (right), 2025



Note: Other RE = Other renewables.
Source: IEA (2025), Renewables 2025.

In Korea, for example, the permitting process for offshore wind farms can take up to around 70-80 months, excluding siting and construction. As of 2023, the majority of offshore wind projects with a generation licence have been stuck in the pipeline awaiting construction permits. These permits must comply with 29 different laws across 10 separate governmental ministries, which may take over a decade.

Recognising the need for reform, Korea, Japan and Chinese Taipei reformed their legal frameworks to address offshore wind permitting bottlenecks. Japan enacted the Renewable Energy Sea Area Utilisation Act in 2018, and Korea followed with the Special Act on Offshore Wind Power in 2025. Chinese Taipei recently announced a significant regulatory overhaul of its permitting review process, reducing the maximum 73 days to just 26 days. These legal frameworks aim to centralise and streamline the permitting process under the leadership of the central government, reducing administrative fragmentation and accelerating project timelines.

While legislative reforms are critical to streamlining permitting, digital innovation – specifically AI – can be a powerful complement to policy efforts. By automating data analysis and administrative tasks, and supporting informed decision-making, AI may help reduce procedural burdens and accelerate the permitting process. Generative AI has the ability to extract and synthesise information from vast regulatory datasets, and flag missing or inconsistent information in draft applications before submission. It has reportedly reduced the time and labour involved in preparing documentation for permitting. Microsoft has developed AI models that produce first drafts of regulatory submissions by analysing historical permitting documents and current legal requirements, reducing the drafting time to just minutes. The US Department of Energy's voltAIc Initiative and PolicyAI project use generative AI to support environmental reviews for permitting processes, particularly aimed at supporting understaffed permitting departments.

Whilst AI has the potential to accelerate the writing of permit applications, it is essential that it is adopted in a mindful way. Permitting processes can have inefficiencies that should be addressed. However, some degree of permitting work is essential for the electricity sector to meet its stringent reliability and safety targets. In a context where errors in small details can cascade into large consequences, the potential for hallucination by large language models needs to be taken into account. As with other new technologies, the risk and benefits of AI solutions should be evaluated on a case-by-case basis.

AI's power can be harnessed to improve system forecasting and optimisation

Among the many benefits of AI for energy systems, enhancing forecasting is one of the most commonly used purposes of AI. While new areas continue to be explored, key technologies such as machine learning and deep learning are identified as the most widely adopted AI subsets in the electricity sector.

The energy transition is driving a shift from centralised, conventional firm generation to distributed VRE sources such as solar PV and wind power. As a result, the current electricity system is adapting to meet evolving flexibility needs that require improved forecasting and optimisation techniques. Greater penetration of VRE increases the need for flexibility on both the supply and demand sides to ensure energy security and system efficiency. On the supply side, VRE requires the system to have flexibility due to its dependence on weather conditions. Meanwhile, demand patterns are also becoming more dynamic, driven by electrification trends such as growing electric vehicle (EV) fleets.

Without accurate forecasting and optimisation, this increasing variability can lead to higher system costs due to greater balancing resource and operating reserve requirements. Increased accuracy in supply and demand forecasting can help optimise reserve power and flexibility resource planning, while managing VRE curtailment, which is rising at a fast pace in many regions. In response, improving forecast accuracy – both in supply and demand – has emerged as a very effective measure to integrate VRE in all phases. Increasing forecasting accuracy is also relevant for the operation of power grids, regardless of VRE deployment levels. AI-driven technologies such as big data analytics, machine learning and predictive algorithms can play an important role in enhancing forecast precision.

Supply and demand forecasting

AI-assisted forecasting recognises patterns from big datasets, such as real-time sensor measurements, satellite images and historical weather and generation data, to calculate generation forecasts. On the demand side, it can forecast based on load patterns, consumption data collected by smart meters and even behavioural data of residential and industrial users. When AI-powered supply forecasts couple with demand forecasts, it can create powerful synergies, offering system operators a clearer view to plan for grid balancing actions, enabling more dynamic and responsive grid management.

Machine learning-based solar forecasting is a well-established use case. In 2014, IBM's Watson programme in the United States forecasted solar generation with 30% greater accuracy than traditional methods. As the VRE share in the grid increases, accurate short-term forecasting can help reduce the costs associated with system balancing.

Deep neural networks and image recognition AI can produce more granular forecasts that adapt to fluctuations. Google's DeepMind applies machine learning algorithms to its wind farms in the central United States, which produce 36-hour-ahead output forecasts based on weather forecasts and turbine data. With the output forecast, the model recommends optimal hourly delivery schedules that can maximise revenues. In 2019, this achieved a 20% increase in economic value compared to a typical wind farm. Since then, it has developed Graphcast, a ten-day machine learning-based weather prediction, offering advantages over traditional numerical-based models in areas such as sub-seasonal heatwave prediction. This model uses almost 37 million parameters to produce weather variables at a 0.25° latitude and longitude resolution.

Based on the high-resolution integrating weather variables, the AI monitoring system can dynamically adjust turbine settings such as blade angle and rotational speed using real-time wind conditions, to maximise energy capture. For instance, AI can optimise the blade angle to enhance aerodynamic efficiency during low wind conditions.

The weather map outputs from this could be used for both solar and wind forecasts. For example, UK-based Open Climate Fix, a non-profit product lab, integrated GraphCast into its transformer-based AI model for predicting solar generation, which proved to be three times more accurate than traditional methods. Furthermore, it has developed AI-based cloudcasting technology, which predicts cloud cover and movements, to improve solar forecasts, with a potential accuracy improvement of 55%.

Machine-learning models conducting support vector regression (a supervised machine learning algorithm used to predict continuous numeric values) are capable of capturing complex non-linear weather–power relationships for wind and solar forecasting in grid-connected microgrids. The model forecasts renewable output by analysing historic production data, weather patterns and real-time grid conditions. In 2019, machine learning-assisted solar forecasting improved the accuracy of forecasts by 33% in Great Britain. Traditional solar forecasts were calculated from solar capacity and solar irradiance, while newer machine learning models consider 80 input variables to achieve a higher degree of granularity.

Demand forecasting is particularly important in East Asia. As previously discussed, East Asian countries are experiencing a growing geographical imbalance between energy supply and demand. As generation assets are often located far from demand centres, and grid infrastructure remains limited or constrained, the need for highly accurate, localised forecasting becomes increasingly critical.

Machine learning and deep learning models are increasingly used in demand-side electricity forecasting. These AI models can enhance forecasting accuracy by integrating diverse factors, including historic consumption patterns, weather conditions, price signals, and even socioeconomic indicators such as population density and commuting behaviour. AI technologies excel in situations where they must process large, structured datasets to answer precise, numerical questions. This makes them well-suited for forecasting electricity demand at granular time intervals, including in real time.

This capability is especially critical as the share of VRE grows, since more accurate demand forecasts enable system operators to better balance supply and demand. For example, Hydro-Quebec, a Canadian utility with over 60 hydroelectric generation stations, implemented new deep neural network-based AI models for short-term load forecasting in 2023. During outliers such as unusually hot days and statutory holidays, the new models performed better than legacy models, correctly predicting a higher evening peak than morning peak. The transition to the new model was deliberately gradual, over a two-year period, to minimise risk. Due to the high stakes and novelty of the solution, the initial deployment was of a "constrained" AI, which was loosened over time as the solution was proven and tested.

Dispatch and scheduling optimisation

Enhanced accuracy in supply and demand forecasting significantly improves the quality of dispatch optimisation modelling. Forecasts provide foundational inputs that help determine how best to meet electricity demand by deciding which generation sources to dispatch, and when. As the energy transition accelerates, dispatch optimisation is becoming increasingly complex, requiring operators to balance multiple priorities.

Dispatch planning prioritises cost-effectiveness while keeping the grid's state within the constraints required for system stability. This is typically achieved through a software tools called unit commitment models, including mixed-integer linear programming, backward dynamic programming and other optimisers. These may or may not be considered AI depending on the definition. The rise in distributed energy sources introduces some challenges, without fundamentally changing this software approach. For example, system operators need to form a view of the real-time power output of rooftop solar systems and of the level of inertia in the system. AI can provide improved methods to estimate these crucial metrics.

Another example of using AI for dispatch and scheduling optimisation is in supporting a virtual power plant (VPP). Essentially, a VPP is an innovative power system management technology that aggregates distributed energy resources, not just VRE and batteries, but also EVs, heat pumps, home appliances and industrial equipment. It operates alongside traditional power plants, offering existing services such as energy and ancillary services, while playing a transformative role in managing power systems more effectively, offering demand flexibility by controlling connected loads.

AI models such as large-scale models, reinforcement learning (RL), deep reinforcement learning (DRL) and multi-agent systems (MASs) can be applied to optimally schedule VPPs. These AI models can effectively co-ordinate and optimise resource allocation throughout various time scales to support secure and economic power delivery. For example, an RL model can optimise the allocation of distributed energy resources effectively after learning through a trial-and-error mechanism, to make the best decision at a given level of uncertainty. Instead of learning from a reference data set, RL models are developed through a trial and error approach in simulations, where a numerical outcome is graded, such as the amount of spot revenue that was earned by the simulated VPP. After tuning the model parameters to maximise the goal, the model can be deployed in the real world to make decisions autonomously.

Many innovative companies are actively using machine learning to optimise their renewable energy resource dispatch. For example, LG CNS has launched Enerdict, an AI-powered VPP software-as-a-service solution. It integrates AI technologies, including machine learning and deep learning, with mathematical optimisation to support VPP operators. Designed specifically for managing small-scale distributed resources, the software enables automated planning and dispatch of aggregated VPP assets. It analyses domestic and international weather forecasts, along with historical weather data, to enhance the accuracy of generation forecasts for distributed VRE assets. Furthermore, it optimises dispatch planning for asset owners, enabling them to maximise their economic incentives in participating in the VPP – by identifying the most profitable time to generate or store energy.

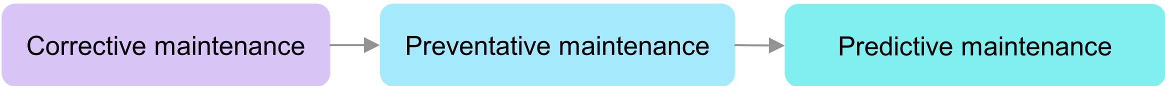
Another example is Kraken Technologies' AI model for its VPP service, which showed 90% accuracy in classifying customers based on their consumption profiles. Powered by machine learning, the platform enables VPP operators to accurately forecast both the demand patterns of participating customers and the output from aggregated utility-scale energy sources. This predictive capability allows operators to optimise operational plans, for example by identifying the best charging schedule for EVs or determining when to dispatch aggregated sources to meet the demand.

AI has the capability to transform operation and maintenance regimes

AI can support the operation and maintenance (O&M) of projects by analysing real-time data from sensors and comparing them with large historic datasets to identify patterns, trends and abnormalities that require adjustments. By leveraging AI, operators can enhance performance, reduce costs and improve the resilience of renewable energy systems.

Traditionally, the O&M of solar farms has relied heavily on ex post corrective maintenance, responding to issues after they occur. As the solar PV sector expanded exponentially, this approach evolved into more proactive preventive maintenance, aiming to address potential problems before they cause disruption. Now, with the integration of AI, predictive maintenance has become more applicable, allowing operators to resolve the issues before they escalate into a full system failure.

Development of solar PV O&M



Source: IEA analysis based on Korea Institute of Science and Technology Information (2022), ASTI Market Insight 2022.

Predictive maintenance

Renewable energy assets operate in diverse and often harsh environments, making maintenance both critical and costly. Reactive approaches can result in unexpected failures and prolonged downtime. AI transforms maintenance strategies by detecting subtle patterns in sensor data that signal early signs of wear or malfunction. This predictive capability allows operators to intervene before issues escalate, improving reliability, extending asset lifespans and reducing operational costs.

AI-assisted predictive maintenance can extend the lifetime of assets by detecting irregularities before they become a critical fault. The Korea Institute of Energy Research has developed AI-powered predictive maintenance software for offshore wind projects with over 90% accuracy. The diagnostic model uses machine learning and curve-fitting algorithms to analyse turbine rotational speed, windspeed and power output from sensors within a turbine to predict bending and stress loads on turbine components.

Machine learning can also analyse sensor data, such as detecting abnormal vibrations, which triggers maintenance alerts, to build a predictive maintenance model for wind turbines. Considering the typically remote location of wind farms, this approach reduces downtime and extends the lifespan of turbines, while also being cost-effective.

Korea Southern Power (KOSPO), one of the six major power generation companies in Korea, developed an in-house generative AI system and implemented it across various work streams including the O&M of generation assets. By leveraging its own generative AI tool – Kospo Evolving Mind Innovation (KEMI) – the company was reportedly able to reduce generation facility failure by an impressive 81% in 2025 compared with 2024. Ørsted uses inspection drones with sensors linked to AI analytics for wind turbine inspections, halving the number of technicians required.

AI-driven predictive maintenance can also be widely used in solar power. Machine-learning algorithms are being used to analyse data from sensors and drones to anticipate equipment failure or maintenance needs, reducing downtime and maintenance costs. Automated quality control is another significant application of AI in solar farms, especially image recognition algorithms. By analysing images captured by drones, AI can detect issues such as dust buildup, panel misalignment or surface damage on solar panels. This automated inspection process ensures that any problems are identified and addressed promptly, maintaining optimal performance.

Researchers at the Korean Institute of Energy Research have developed an AI solar panel defect detection system. The AI model uses historical I-V curve data – which show how much current a solar panel produces at a given voltage – alongside real-time irradiation and temperature inputs to identify performance abnormalities such as degradation and faults. This technology can effectively detect degradation and contamination with 95% accuracy.

Dust and debris can reduce solar panel efficiency with an annual loss of 3-5% of production. AI can be used to predict soiling rates and optimise cleaning schedules based on weather, dust forecasts and energy loss analysis. Technology provider Ecoppia developed robotic solar panel cleaning that uses advanced sensors and machine learning to collect data points and can be used for predictive maintenance before faults appear.

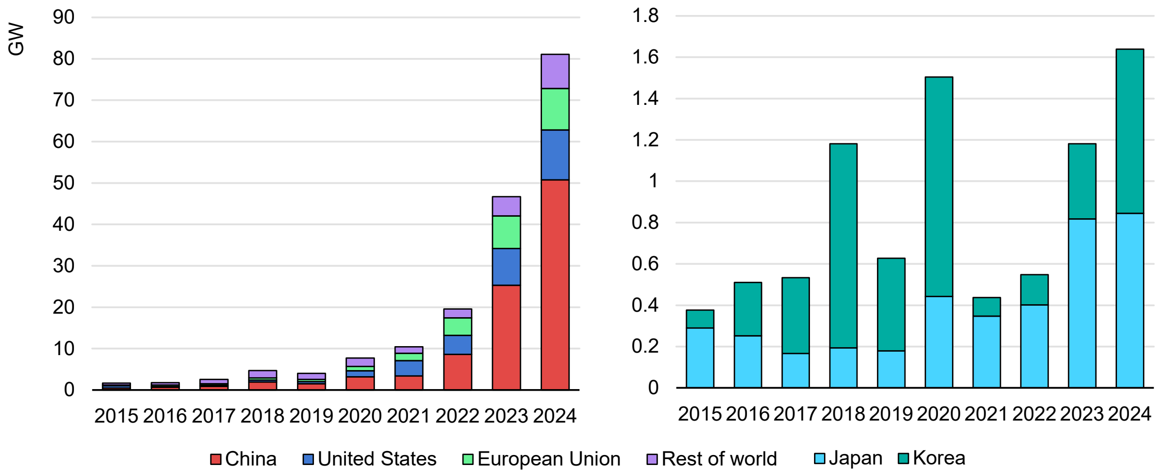
Storage and demand-side flexibility are increasingly essential, and AI can enhance their contribution

In the context of VRE integration, unlocking flexibility from the demand side and storage is essential. AI tools can provide support for the effective application of storage and demand-side operations to provide this flexibility.

Battery energy storage systems (BESS) have an important role to play in balancing variable supply and demand effectively. Like traditional storage resources, such as pumped-storage hydro, batteries can absorb surplus VRE generation during periods of low demand and dispatch it when needed during periods of high demand. Unlike traditional storage resources, batteries can respond in sub-second timescales and can be co-located with VRE generators. In the past, utility-scale batteries have predominantly been used to provide frequency regulation. In recent years, energy shifting and capacity provision have become the main application globally, such that these activities accounted for about 85% of utility-scale battery storage capacity additions in 2023.

Co-optimising revenue for energy, multiple frequency products and other schemes is a complex problem. AI can help integrate these market signals more effectively. In addition to utility-scale projects, AI can also help in more efficient usage of residential battery installations, as well as EV charging and discharging patterns with regard to system conditions, considering both the availability of distributed energy resources and customer consumption patterns.

Battery storage capacity additions worldwide (left) and in Japan and Korea (right), 2015-2024



Source: IEA Analysis based data from IEA (2025), World Energy Outlook 2025.

Japan is one of the top five countries for residential battery installations. This is mainly driven by the prevalence of household solar panels and concerns about maintaining power supply during seismic activity or extreme weather. In Japan, trading company ITOCHU partnered with battery company Lunar to use their Gridshare platform to manage distributed energy resources, including residential batteries, with machine learning and other AI methods. The algorithm creates a customised charging plan based on solar generation forecasts, customer consumption patterns and extreme weather alerts, saving on customer bills and ensuring backup power is available. The concept was expanded in a trial with Japanese electricity companies in which the insights into customer consumption behaviours and accurate generation forecasts were incorporated into a residential VPP. The trial found that on average 91% of the maximum flex power could be dispatched when needed.

Vehicle-to-grid (V2G) systems offer a promising alternative to dedicated grid-scale battery storage, particularly when enhanced by AI-assisted scheduling and optimisation. By enabling EVs to discharge electricity back into the grid during periods of high demand, V2G can help balance intermittent renewable generation and improve grid flexibility.

Driven by ambitious EV deployment targets, East Asia is emerging as a key market for V2G innovation. In Tokyo, TEPCO is developing AI-assisted bidirectional charging infrastructure in collaboration with Diamond & Zebra Electric Mfg. Co., Ltd. The system uses AI algorithms to optimise charging and discharging based on local supply and demand conditions, weather forecasts and time-of-use electricity pricing. Customers can reduce their energy bills by charging during off-peak hours and supplying electricity back to the grid when prices are high. The system also provides backup power during outages, enhancing energy resilience. Historically, V2G adoption has been limited by low EV penetration and unpredictable user charging behaviour. If an EV battery is discharged for V2G purposes, and then a user wishes to drive far unexpectedly, the lack of range may cause challenges for the user. AI could help predict when EV users will next need to unplug and drive, and how much charge they will need. This increase in precision about how much charge is available for V2G use could improve adoption of V2G technologies. Utilities can further accelerate adoption by offering incentives similar to those used for rooftop solar panels.

V2G projects have been commercialised widely across the world. In Europe, successful projects such as Utrecht Energized in the Netherlands and Vattenfall's pilot in Sweden demonstrate the potential for smart V2G systems to support renewable integration. Although these initiatives do not explicitly use AI, they highlight the value of intelligent control systems, many of which can be enhanced by AI, to support renewable integration and grid flexibility.

The application of AI can increase electricity market participation and facilitate trading

AI can play a transformative role in optimising market participation and trading strategies for distributed energy resources. Beyond forecasting supply and demand, AI can enhance the accuracy of market price predictions, enabling market participants to strategise, automate and optimise their bids. This extends to broader trading strategies, helping non-traditional resources such as VRE to navigate complicated market dynamics while contributing to power system stability.

Traditionally, the electricity market relied on relatively predictable outcomes from conventional firm power sources – coal, gas and nuclear. However, the inherent variability of solar PV and wind generation introduces corresponding variabilities in asset performance and electricity spot market prices, which can necessitate more complex bidding strategies and more accurate forecasts.

AI, particularly through machine learning, can analyse vast datasets to enhance the visibility and accuracy of the market environment across different layers of products, including but not limited to short-term, day-ahead, intraday and real-time markets. Different models can extend to integrating weather events and even geopolitical events. This capability allows for more precise forecasting of prices and bidding volumes, and supports the development of optimised bidding strategies that can maximise financial returns for market participants.

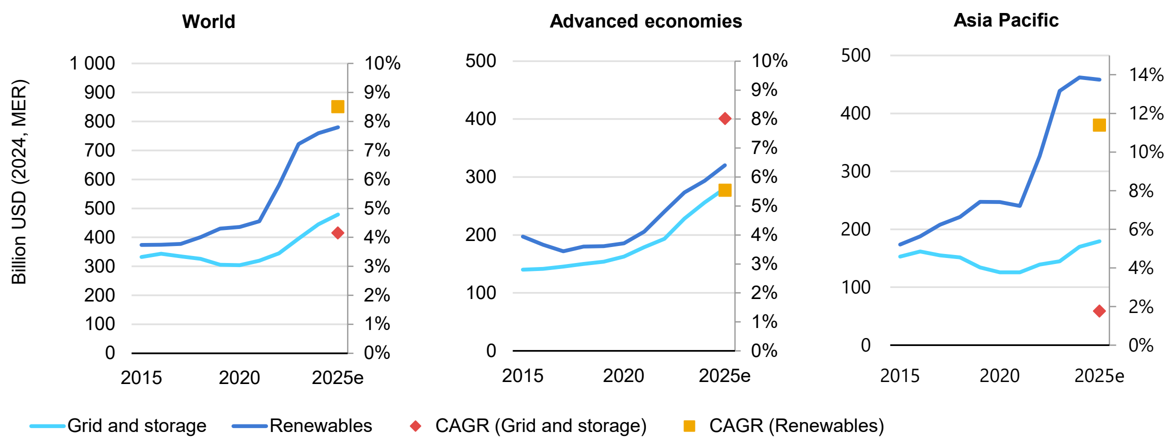
Flexibility is an essential pillar supporting the energy transition. Aggregators and VPPs are an effective way of unlocking the potential of demand response, distributed generation and storage. A recent order by the US Federal Energy Regulatory Commission enables distributed energy resources to better participate in wholesale electricity markets, supported by VPP technologies. Aggregators can use AI to gather, process and act upon data on their portfolio to optimise their bidding.

Having access to clean, comprehensive datasets is important for developing and applying effective AI models. Regulators and market operators can work towards making more of their data accessible and machine-readable, which can contribute to efficiency gains via improved application of AI in markets and trading

AI for the power grid

As the energy transition accelerates electrification and renewable generation projects, it is increasingly important for grid operators to manage their evolving transmission and distribution networks effectively, to avoid grid strain and maintain a secure supply of electricity. In many regions, grid congestion and connection queues for new renewable projects are growing due to bottlenecks in infrastructure siting, planning and permitting. Globally, almost 1 700 GW of renewable energy projects and more than 600 GW of storage projects were in grid connection queues in 2025.

Investment in grid and storage and renewable energy by region, 2015-2025e

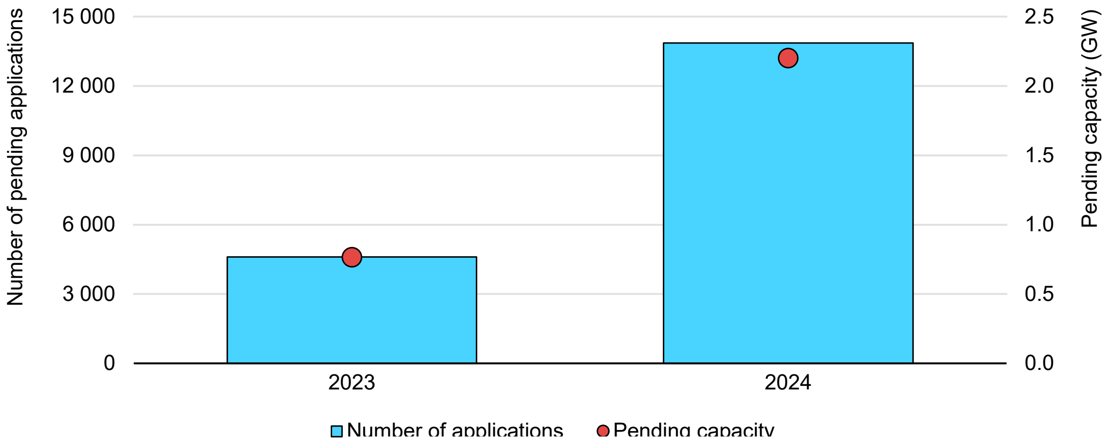


Notes: Values for 2025 are IEA estimates. Grid = transmission and distribution lines, grid equipment (transformers, substations, etc.). Storage = battery storage for buildings and utility-scale. Renewables = solar PV, concentrating solar power, wind, tidal, geothermal, hydro and pumped-storage hydro, bioenergy and renewable waste. MER = Market Exchange Rate.
Source: IEA (2025), World Energy Investment 2025.

Transmission and distribution systems require new infrastructure to connect remotely located renewable generation projects and to keep up with capacity requests for connections to new projects. However, the East Asia region currently has no international interconnections, except between China and Southeast Asia. Moreover, the Japanese grid operates at two frequencies, 50 Hz and 60 Hz depending on the region, limiting east–west power transfers and complicating balancing efforts.

Planning the siting of grid construction projects adds another layer of complexity to grid development, mainly due to resistance from local residents. Korea shows notable cases of local opposition to transmission construction. The longest-delayed project, for example, recently started commercial operation after an extensive 22 years of construction, primarily due to local opposition. The 345 kV Bukdangjin-Shintangjeong transmission line connects major coal power plant fleets on the west coast with large-scale semiconductor factories in the southern Seoul Metropolitan Area.

Grid connection applications in Korea, 2023-2024



Source: IEA analysis based on data provided by the Korea Energy Economics Institute.

AI offers various solutions across three key areas of grid operations: **grid development**, **asset operation and maintenance (O&M)**, and **real-time operation**. The IEA surveyed a group of utilities and system operators from 13 countries, including Europe, the United States, Australia and Asia, for the Energy and AI report about their use of AI in these power system optimisation processes. The key findings of the survey were that only 23% use AI in real-time operation, compared with 54% for grid development planning and almost 70% for asset maintenance and operational planning. This reflects a general reluctance to adopt AI in real-time contexts due to reliability concerns. In contrast, operators are increasingly leveraging AI for long-term planning and asset management, where longer decision windows reduce operational risk.

The capability of AI to assist power systems with an increasing VRE share is being recognised by various institutions. For example, the Korean government announced plans to build an intelligent renewable grid and a taskforce to develop an AI-powered grid to support the integration of renewable sources and handle rising electricity demand. The US government also underlined the role of AI in tackling key challenges in the energy transition, from building resilient and secure power grids to equitable and accessible energy storage deployment. Recognising the significance of AI-related technologies in real-time grid operation, ENTSO-E has set AI-supported power system operation as one of its key milestones of its R&D strategy.

Grid development planning can leverage AI’s capacity to process large datasets in complex systems

The IEA report Electricity Grids and Secure Energy Transitions sheds light on the need for alignment between grid planning and national long-term energy transition plans to implement resilient grids that handle higher shares of VRE. Given the time discrepancy between grid construction and the deployment of generation, storage and loads, it is imperative to plan the grid and for it to be coupled with accurate forecasts for increasing shares of distributed resources, such as VRE, EVs and batteries. Based on the analysis, the IEA recommended significantly improving grid planning and its co-ordination across sectors. Furthermore, public resistance to grid construction is a further challenge requiring attention.

Within this context, long-term scenario planning for the grid can leverage AI’s capacity to process large datasets and identify patterns within complex energy system interactions across different time frames. For example, KEPCO, the state-owned grid company in Korea, is developing an AI model for optimising grid planning and site selection. The model is expected to train using geographic and environmental data, along with related regulations, to optimise grid design, which can minimise resistance among local residents.

System operators are increasingly adopting AI to improve grid operation and maintenance regimes

AI technologies are increasingly being adopted by grid operators to enhance grid O&M, as their versatility can improve the O&M of transmission and distribution networks. Most prominent among the areas to utilise AI are predictive maintenance and maintenance schedule optimisation.

In the previous IEA survey, **grid asset maintenance** was the single area that three East Asian countries – China, Japan and Korea – were actively applying AI technologies.

For **predictive maintenance** in particular, AI can aggregate large datapoints collected from sensors – be they imagery or thermal data – attached to different elements of grid equipment, such as transformers, circuit breakers and switch gears. When analysing these data, machine learning can be a transformative tool to identify patterns and predict future failures.

AI is also being used to support **system expansion** studies by identifying generation and load hypotheses, and to guide budget allocation based on the likelihood of grid component failures. Operators not yet using AI have shown interest in developing tools for demand forecasting, contingency analysis and planning simulations.

In addition to transmission-level maintenance, AI can significantly enhance the efficiency of **distribution network maintenance**. Distribution companies handle millions of maintenance requests annually, often facing inefficiencies due to the lack of accurate pre-diagnosis. To address this, Enedis, a major French electricity distribution operator, is developing natural language processing models aimed at improving intervention planning. By analysing approximately 2.5 million intervention requests the company receives each year, the model identifies and classifies cases to reduce unnecessary site visits, thereby minimising resource wastage and carbon emissions.

Fault management

In recent years, system operators have been actively utilising AI technologies in grid fault management. In particular, deep learning and machine learning models are being widely explored for enhancing the accuracy and speed of fault management – fault detection, classification, localisation, isolation and recovery. Increasing VRE penetration, coupled with the increasing number of extreme weather events, necessitates more accurate and real-time fault management by grid operators. The greatest contributions that AI can make in grid fault management lie in its ability to deliver **predictive analytics**, enhance detection **accuracy** and accelerate **response times**.

With AI, grid operators can detect grid faults proactively and accurately, intervening within minutes, if not in advance, to minimise the repercussions such as unexpected outages. Major grid operators in East Asia are adapting AI technologies for grid asset fault detection. In 2021, KEPCO launched an AI-powered grid fault management system, Substation Equipment Diagnostic and Analysis System (SEDA), to conduct predictive maintenance of its grid assets nationwide. SEDA analyses real-time data collected from sensors installed at substations and assesses grid conditions using AI diagnostic algorithms. In the first two years SEDA detected 59 cases. Three major cases saved a total of approximately USD 22.5 million by intervening in the early stage of the fault.

Chubu Electric Power Grid also recently co-developed an AI-based system – POWER GRID Check – in collaboration with Sensyn Robotics. The system uses drone-captured imagery data to automatically detect abnormalities on overhead transmission lines, analysing the data in real time.

Across East Asia, image recognition is widely used for asset maintenance, including vegetation management and signal processing. In Japan, TEPCO Power Grid is developing automated drones that collect 3D data for equipment diagnostics, using multiple AI systems to analyse image and historic records to support asset maintenance. These innovations enable the creation of digital substation models for simulation and recovery planning. State Grid China also uses AI for image and speech recognition in asset management, and generative AI for fault detection. In asset operation planning, AI helps define and optimise maintenance policies, scheduling and operational modes, contributing to more efficient and resilient grid management by, for example, minimising down time and tackling grid failures in real time.

AI is implemented in grid fault detection not just in East Asia, but also beyond the region. Southern California Edison, one of the largest utilities in the United States, recently developed its Advanced Waveform Anomaly Recognition (AWARE) system. Its machine learning-based system analyses real-time data collected by the digital fault recorders at substations, which can identify faults with more than 80% accuracy and their locations before escalating into major failure.

By rapidly processing vast amounts of data, AI supports real-time grid operation

With the increasing predominance of weather-dependent and inverter-based generation sources, the importance of making optimised decisions in grid operations is becoming more important, particularly in real time. This requires the integration of vast streams of real-time data into grid operation, including more accurate and granular forecasts of electricity supply and demand, along with weather forecasts and contingency scenarios such as transmission line failures. This should enable operators to minimise VRE curtailment, optimise reserve deployment, and eventually help reduce new transmission line buildouts, thereby enhancing the overall efficiency of the energy system. This not only improves security of supply, but also can help reduce reliance on costly reserve capacity – both for adequacy and balancing – often provided by conventional sources such as gas and coal, leading to potential cost savings and emission reductions.

In this context, AI offers a significant advantage by rapidly processing vast amounts of data to support real-time decision-making process for more effective grid operation. AI can provide support to real-time processes such as monitoring and estimating parameters related to grid stability and dynamic line rating.

Grid stability

Monitoring grid frequency, inertia and voltage levels enables the maintenance of stable grid conditions. AI tools can provide support in this context, boosting real-time monitoring and control of various grid parameters. One AI-supported method is to apply a convolutional neural network, a subset of deep learning, to continuously estimate the inertia of power systems containing inverter-based renewables. Japan-based Chubu Electric Power Group integrates a neural network or a random forest model into its voltage control system for grids with high solar output. As a subset of machine learning, a random forest model is widely used for classification or regression. It builds multiple decision trees by randomly selecting subsets of training data. Trained on 30-second grid data, Chubu's model can identify when to intervene and control voltage stabilising devices. It can optimise voltage control by improving efficiency compared with rules-based controls, leading to 20-70% fewer voltage control actions, varying across different operating regions.

Dynamic line rating

Dynamic line rating (DLR) is a technique that determines the real-time or forecasted current-carrying capacity of overhead transmission lines by factoring in meteorological variables and patterns, such as wind speed and ambient temperature.

DLR is a technology that adjusts the "speed limit" of transmission lines in real time, comparable to the analogy of motorways that have variable speed limits depending on conditions, allowing more or less current to flow safely. In this context, AI can effectively help system operators to assess the status of the grid in real time and determine how the line ratings should be optimised accordingly.

Unlike traditional static ratings, which rely on conservative assumptions based on worst-case weather scenarios to prevent overheating, DLR provides a more precise evaluation of a line's operational limits. In practice, this method can boost transmission line capacity by approximately 15-30% for over 90% of operating hours in a year.

DLR systems have long-proven records of their effectiveness in operation. In Belgium and France, for example, DLR systems have been installed on all high-voltage alternating-current (HVAC) interconnection lines and have been operating for over 15 years. As a result, the intraday rated capacity increased by up to 130%.

The IEA estimates that implementing DLR on the 15% most congested overhead transmission lines globally could unlock up to 230 GW of extra capacity, deferring expensive infrastructure upgrades. The main application of AI to DLR is in mitigating uncertainties in weather forecasts and extreme event inputs to better predict transmission line temperature. AI-assisted DLR considers real-time meteorological data and sensor measurements to determine the real-time current-carrying capacity of lines.

Despite DLR's demonstrated effectiveness and the potential benefits of allocating unlocked capacity with AI, many transmission system operators (TSOs) and utilities in East Asia have yet to implement DLR. In contrast, several TSOs – mostly in Europe – have already adopted DLR. Although a few TSOs and utilities outside East Asia are currently using DLR, only one of the Asian respondents answered in the IEA's survey that they are using AI for DLR implementation.

Grid operators using AI applications by category, 2024

Source: IEA (2025), Energy and AI.

AI for end users

Growing electrification of the global energy system is reshaping traditional linear relationships between supply and demand into highly diversified, distributed and multilateral interactions. While the supply side is shifting towards smaller, distributed and weather-dependent energy sources, the demand side is also going through a rapid change towards a more digitalised and flexible load. AI can play a pivotal role in supporting the interaction between electricity suppliers and end users to balance the increasing flexibility in energy systems.

In East Asia, China has been showing rapid progress in electrification, according to IEA analysis. More than half of global electricity demand growth in 2024 occurred in China. This reflects not just higher industrial electricity demand, but also increasing consumption from EV charging, along with expanding data centre load and telecommunication network development, contributing to significant end-user demand growth. Korea and Japan are also expected to experience electricity demand growth after over a decade of plateau.

Given East Asia's high level of digital infrastructure deployment, such as 5G networks, many transformative AI-related technologies are gaining momentum on the demand side.

AI is integral to innovation in residential energy management, contributing flexibility to the system

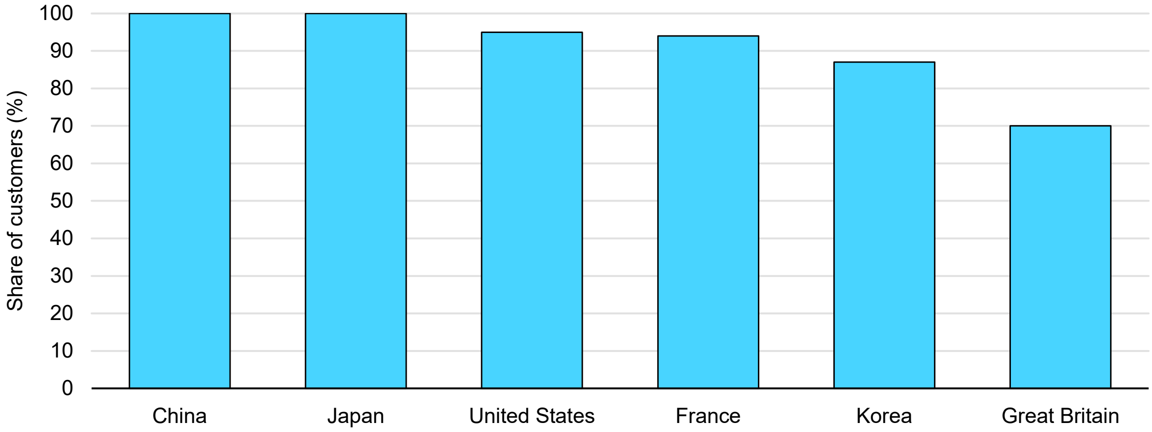
Residential energy management is a broad concept that includes not just electricity, but also heating, cooling, lighting, cooking and more. Based on evolving technologies such as Internet of Things (IoT) technology, residential energy management has the potential to become a major component of demand-side management. AI is an integral part of innovation in residential energy management as it can collect and analyse data, recognise patterns from them and automate the controlling of residential load – EV charging, air conditioning, heating etc. – while aggregating these loads and using them to provide demand-side flexibility to the system.

Smart grids

Smart grids refer to electricity systems that are powered by digital technologies and information and communication technology (ICT). Smart grids can be implemented across all streams of the electricity value chain, from generation and transmission down to the distribution level. With an increasing share of small-scale, distributed and flexible energy sources, as well as more fluctuating load patterns, smart grid solutions hold great value in managing distribution-level electricity, primarily residential usage.

At the core of smart grids lies the smart meter, which enables operators to measure electricity usage remotely in real time. This leads to enhanced grid visibility on both supply and demand sides and, essentially, contributes to more efficient grid management. China reached 100% smart meter penetration at the residential level as of 2021 and the United States is expected to hit 93% by 2027. Korea and Japan are also showing higher levels of smart meter deployment compared with other advanced economies. Traditionally, electricity metering systems could not send out real-time data, but simply recorded electricity usage to be manually checked by the system operator. Smart meters are digitalised systems that enable the monitoring of electricity usage in real time by the relevant stakeholders.

Share of customers with a smart meter by country and region, 2025



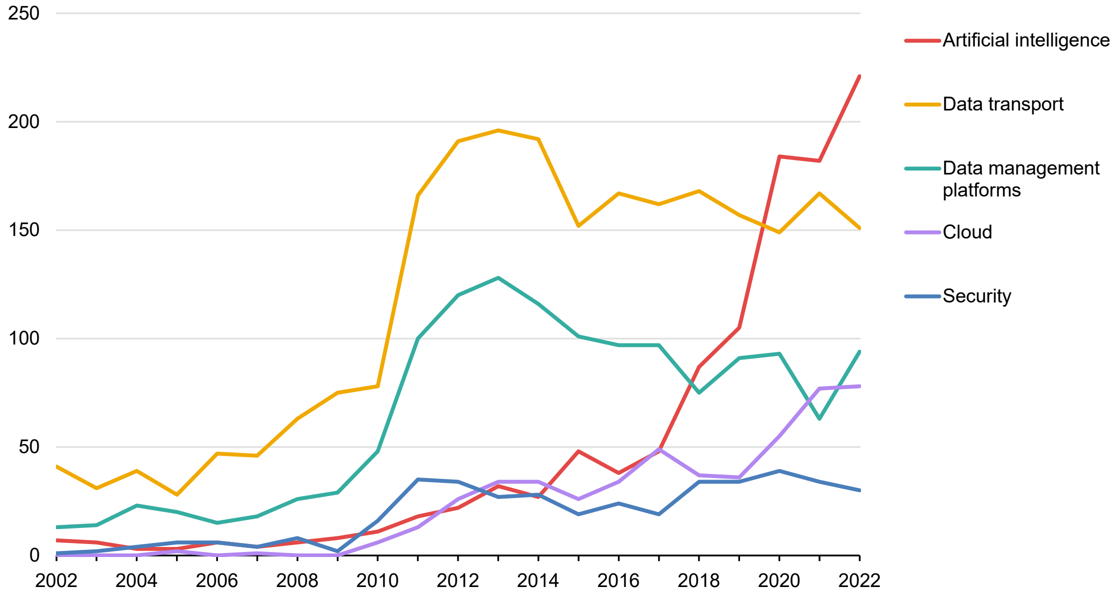
Source: IEA (2026), Electricity 2026.

Smart grids enable demand response and management, allowing utilities to balance supply and demand more effectively. This helps prevent blackouts and ensures a stable power supply even during peak usage times.

Since 2022, AI has become the most patented of the enabling digital technologies connected with smart grid patents, recording around 34% growth in patenting annually since 2016. The greatest strengths of smart grids are in real-time performance and how the data collected from smart grids can contribute to managing the variability and flexibility of evolving low-carbon power grids. AI is a suitable technology to fully unlock such potential as it can collect large amounts of data from various sources, and then aggregate and process them into meaningful datasets for operators in real time.

Forecasting and decision-making are other notable functions of AI in smart grids. In particular by using machine learning, deep learning and artificial neural networks, system operators can make more informed decisions on smart grids based on identified patterns and predictions.

Patenting trends of technologies for smart grids, 2002-2022

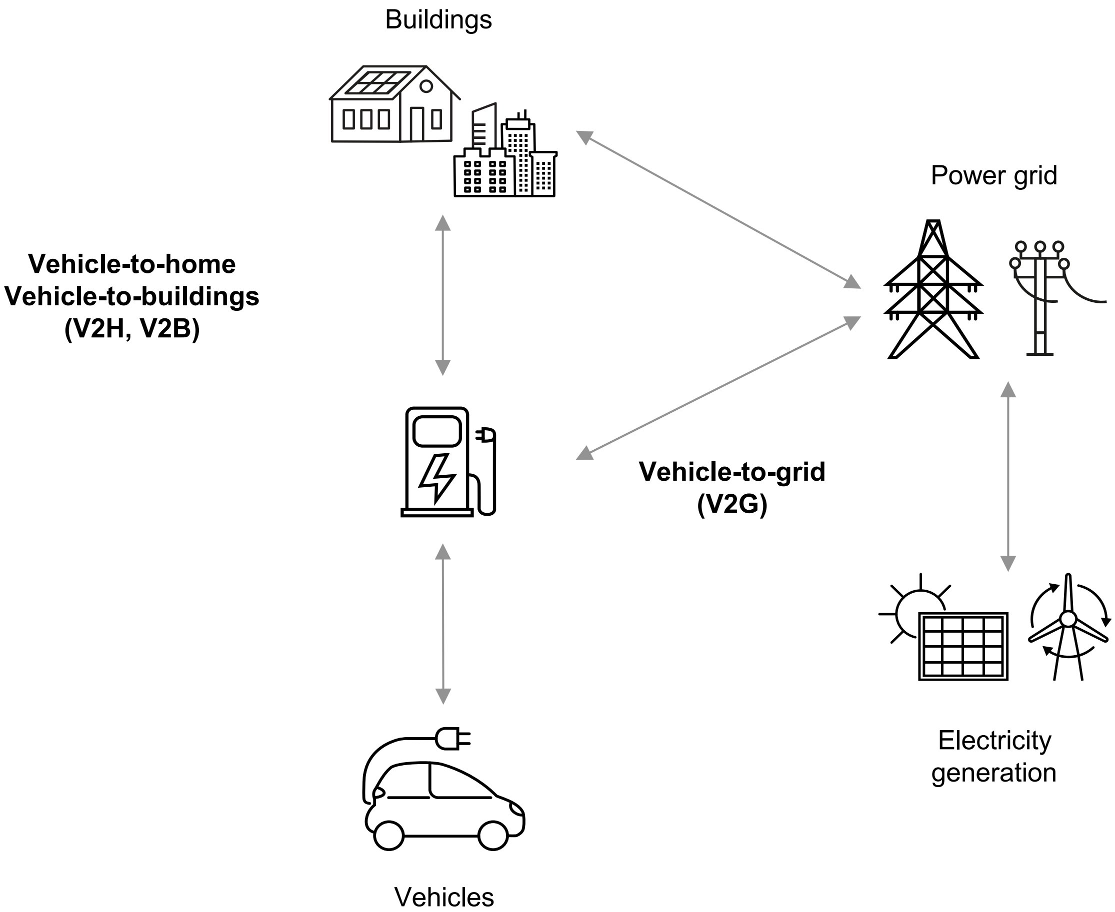


Sources: IEA and European Patent Office (2024), Patents for Enhanced Electricity Grids.

Sector coupling

Sector coupling facilitates the interaction of different energy sectors – electricity, gas, transport, buildings and industry. The main objective of sector coupling is to share the clean energy in one sector with another to accelerate decarbonisation. For example, price-responsive heat pumps (heating sector) can schedule operation for periods when VRE generation is plentiful (electricity sector), to provide heating in households. Similarly, EVs (transport sector) connected to chargers can discharge electricity and contribute to flexible supply and provide ancillary services (electricity sector). These vehicle-to-everything (V2X) services include vehicle-to-building (V2B) and vehicle-to-home (V2H) when the energy is consumed on the premises, or vehicle-to-grid (V2G) when the energy is consumed elsewhere.

Vehicle-to-everything (V2X) overview

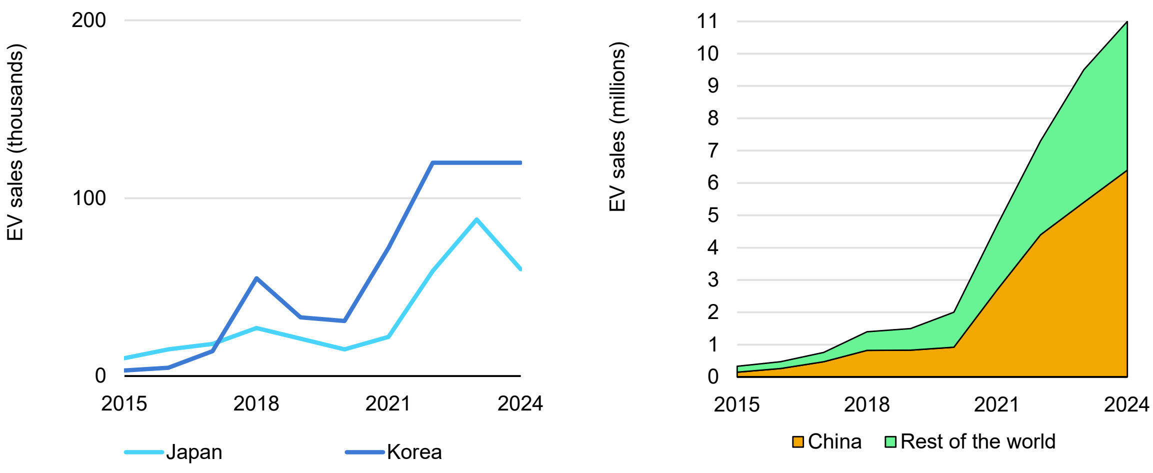


Note: Vehicle-to-everything (V2X) includes V2H, V2B and V2G.
Source: IEA Analysis based data from US Department of Energy, Bidirectional Charging and Electric Vehicles for Mobile Storage.

V2G can be a good example of AI utilisation in sector coupling. AI-powered smart charging technology can effectively facilitate V2G. Its predictive algorithms can optimise charging and discharging schedules, based on supply and demand pattern analysis, to maximise the benefits to both the system operator and to customers. Based on AI-guided schedules, EV owners can charge their vehicles during off-peak hours with lower electricity price, and vice versa. In Malaysia, for instance, AI-based smart charging systems showed a cost reduction of about 20% and energy savings of 30%.

V2G requires a set of physical elements: EVs with bidirectional charging software and hardware, communication technologies that connect the EVs and system operators, and smart charging facilities that can physically connect the EVs to the grid. Along with physical infrastructure, it also requires tailored market structures, such as flexible tariffs, to provide financial incentives for V2G businesses.

EV sales by country and region, 2015-2024



Source: IEA (2025), Global EV Outlook 2025.

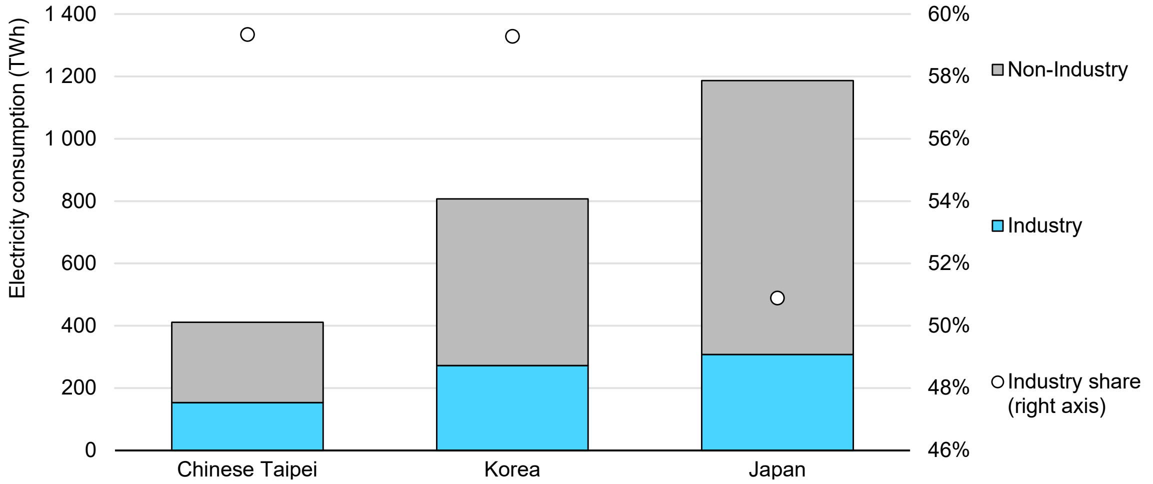
Among these elements, East Asia is prominent in two aspects: EV deployment and communication infrastructure. China is leading in global EV sales, accounting for nearly half of global EV sales in 2024. The share of EVs in total car sales in China is expected to reach 60% in 2025. While China showed annual growth of 46% in 2015-2024, Korea had a similar annual growth rate of 44%. Japan had a CAGR of around 20% during the same period of time. Although the pace of EV sales growth may differ, EV production in Korea and Japan is rising and reaching 1 million vehicles collectively. Although not all models sold in the region support bidirectional charging, the most popular bidirectional models currently on the market are predominantly produced by either China or Korea.

Demand response can optimise system costs by harnessing industrial energy users' flexibility

AI is transforming energy management not just in the residential sector, but also in the industrial sector. This sector accounts for the highest share of electricity consumption in most countries in East Asia. China (60%), Chinese Taipei (60%) and Korea (50%) have a higher industrial sector share of electricity consumption than Japan (35%).

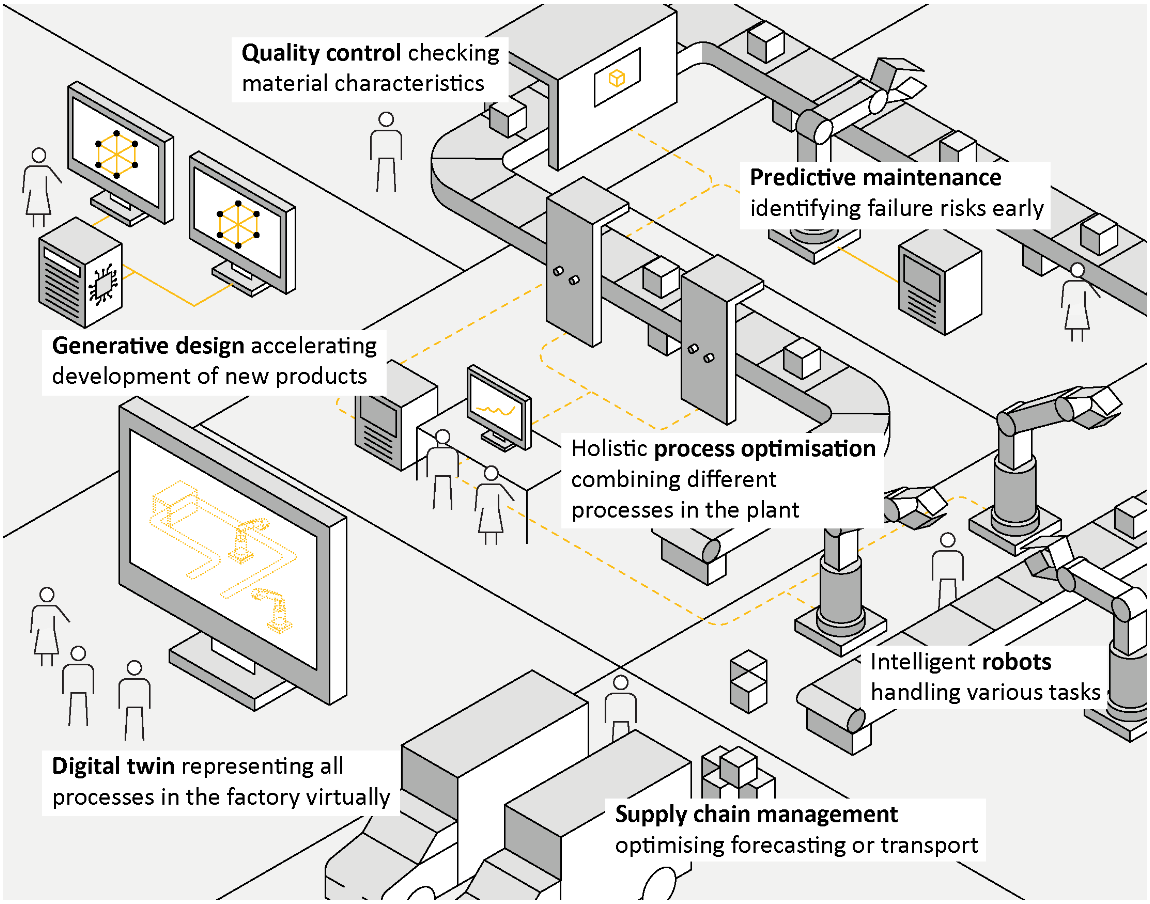
This is driven by energy-intensive industries, such as steelmaking, shipbuilding and petrochemical industries. Large end-user loads also include data centres for hosting AI and cloud computing, and semiconductor manufacturing facilities to provide GPUs for AI data centres. Within this context, optimising the electricity consumption of the industrial sector, especially the energy-intensive industries, is critical for power sector decarbonisation in East Asia.

Industrial sector share of electricity consumption by country, 2023



Source: IEA (2025), World Energy Balances.

AI applications in industry



Source: IEA (2025), Energy and AI.

Demand response

Demand response refers to a set of solutions that balances the demand for electricity by changing end-user behaviour. This includes shifting the time of peak demand to avoid any possible supply–demand imbalance or to reduce system costs. Consumers providing demand response can be rewarded via time-differentiated retail prices or dedicated monetary incentives. Demand response has grown into an important provider of grid flexibility. As a key strategy for balancing supply and demand in modern grids, it plays a crucial role in providing flexibility to the system to integrate increasing shares of VRE.

Demand response first started by targeting large industrial users. In Europe and the United States, however, it is already spreading to the residential sector, led by major players such as Next Kraftwerke (Germany) and Octopus Energy (United Kingdom), along with Tesla (United States and Australia), backed by a high share of EVs. In East Asia, the industrial sector remains the major player in the demand response market.

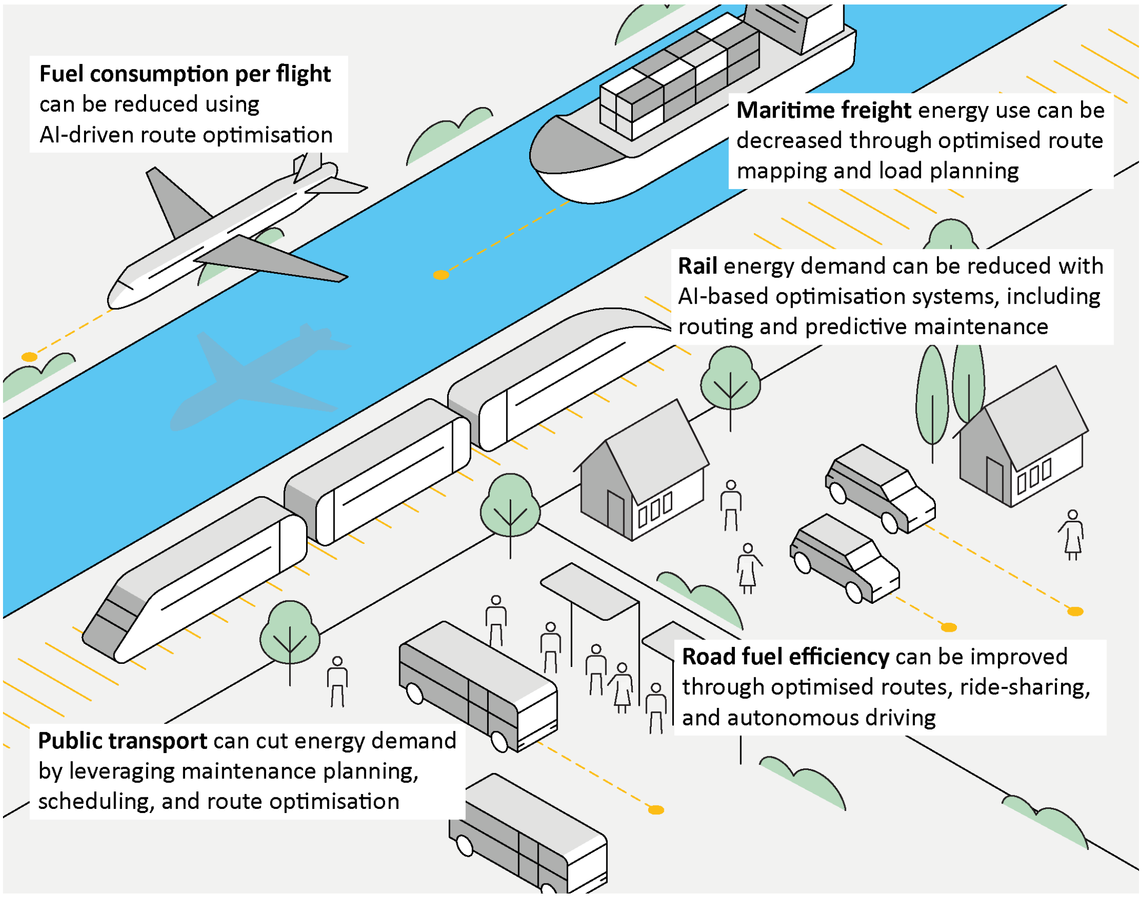
Aggregators must select the appropriate portfolio of end users. This can be a challenging task given the heterogeneity of loads and uncertainties involved. In this context, AI can support the optimisation and operation of demand response schemes. AI models such as artificial neural networks, reinforcement learning and multi-agent systems under game-theoretic environments can contribute to optimisation, informed decision-making and forecasting.

Similar to smart grids, Korea has developed various sets of demand response-related products and market schemes, benefiting from its advanced digital infrastructure. Korea opened a market dedicated to demand response resources in 2014, which made it the first Asian country to do so. The capacity that participates in the market has now reached around 4 GW, which is roughly equivalent to the nameplate capacity of four nuclear power units in Korea. By 2019, the Korea Electrotechnology Research Institute had already developed a predictive machine learning algorithm to enhance the predictability of demand response resources, especially in the end-user sector.

AI can catalyse the decarbonisation of the transport and buildings sectors

AI can synergise with the ongoing electrification of the transport and buildings sectors. The IEA's Electricity 2026 report highlights that these two sectors – alongside the rapid expansion of data centres – are contributing to strong electricity demand growth globally, including in Korea, Japan and China. Based on its strong data processing power and predictive ability, AI can catalyse the decarbonisation of the transport and buildings sectors.

AI applications in transport



IEA. CC BY 4.0.

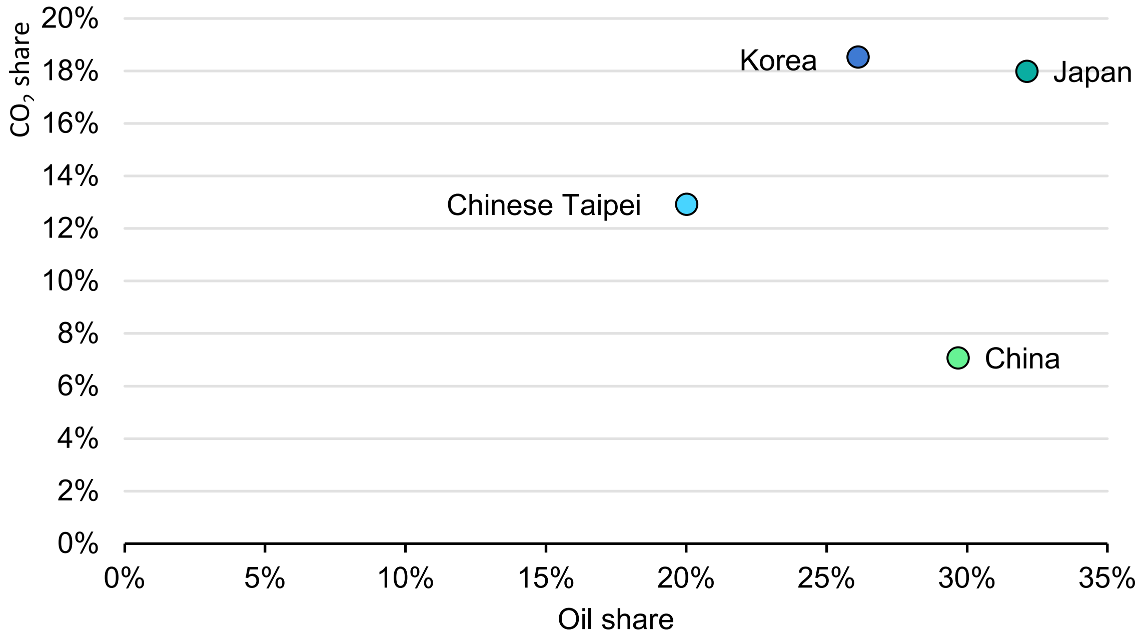
Source: IEA (2025), Energy and AI.

Transport

The transport transition is becoming a centrepiece of the East Asian economy. In road transport, the share of sales comprising EVs in China is expected to exceed 80% by 2030, while in Europe and the United States it is expected to reach almost 60% and 20% respectively. In 2024, 60% of global EV sales took place China, followed by Korea and Japan.

East Asia hosts large-scale shipping and aviation industries due to its substantial share of international trade and its importance as a transport hub. In Korea, for example, the shipping industry accounted for 1.9% of GDP, worth almost USD 25 billion in 2020. Given the large size of the transport sector, there is the potential to dramatically reduce global carbon emissions, primarily by reducing the sector's fossil fuel consumption. Globally, 55% of oil demand and 20% of carbon emissions originate from the transport sector. Within East Asia, Japan shows the highest share of oil demand from the transport sector at 32%, followed by China, Korea and Chinese Taipei. However, the sector's share of annual carbon emissions was highest in Korea, at 18%.

Transport sector share of carbon emissions and oil demand by country, 2024



Notes: Oil products for the transport sector comprise diesel, LPG, ethane and motor gasoline. The emissions are CO_2 only and do not include LULUCF (Land Use, Land-Use Change and Forestry).
Sources: IEA Analysis based data from European Commission (2025), GHG emissions of all world countries; and IEA (2025), Monthly Oil Data Service (MODS) Global Demand by Product.

Coupled with rapid electrification in the transport sector, AI can help reduce costs and increase efficiency. Greenplan, a DHL Express-funded start-up in Germany, uses AI to optimise routes, reducing costs by 20%. This reduction in kilometres driven also results in lower emissions. Another use case is in road freight planning, where AI can increase utilisation. The reduction in empty space has the potential to cut around 5% of road freight emissions. On the human side, AI can offer feedback to optimise driving behaviour. Such nudges to adjust acceleration and braking patterns can reduce fuel use by 2-10%. AI can also greatly reduce energy use and costs by facilitating predictive maintenance.

AI models are trained on various datasets including real-time traffic congestion, maintenance scheduling and cabin conditions. Companies such as Uber and Hyundai are actively implementing AI machine learning technologies to their logistics businesses. Uber Freight, for example, uses AI algorithms for route optimisation, which helps minimise the miles driven by empty trucks by 10-15%. This not only improves operational performance, but also enhances fuel efficiency, which ultimately leads to lower carbon emissions.

AI-powered route optimisation technology has expanded into the shipping industry. HD Hyundai Marine Solution implemented an AI system – OceanSmart – to optimise the routes of major shipping companies such as SK Shipping and

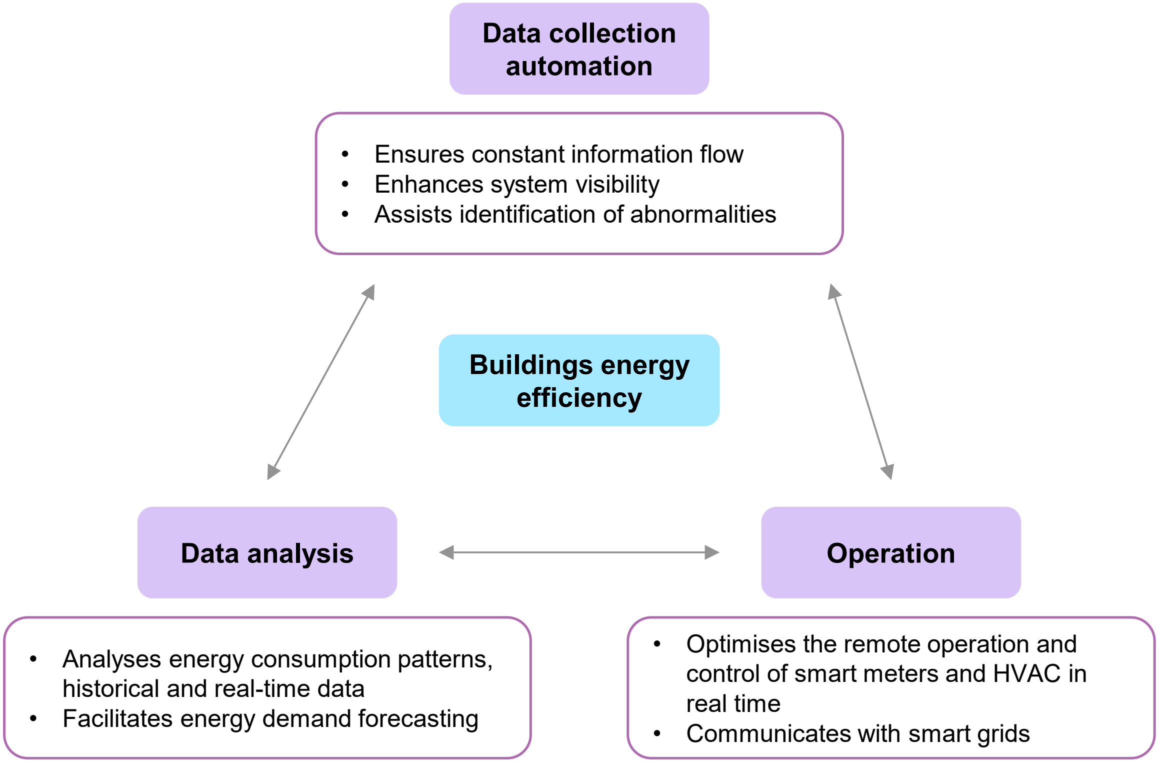
Hyundai Glovis. The optimisation aims to reduce the fuel consumption and greenhouse gas emissions of ships, such as crude oil carriers. Its AI-driven system has reportedly reduced average fuel consumption by 5.3%, worth over USD 240 000. It can also enable shipping companies to better track their carbon footprints transparently.

In the same vein, the aviation sector can also effectively tackle emissions with AI. Geodis has recently developed AirSmart, an AI system that designs optimal cargo routes with fuel-efficient aircraft.

Buildings

When integrated into existing energy management systems, such as a building energy management system (BEMS) or a home energy management system (HEMS), AI can unlock the building sector's energy efficiency and emissions reduction potential. A BEMS or HEMS is a comprehensive system which manages the energy consumption of a building based on ICT coupled IoT. AI can be implemented across various aspects of a BEMS, from automating data collection to operating smart devices such as smart meters.

Key opportunities for AI integration in building energy management systems



Note: HVAC = heating, ventilation and air conditioning.
Source: IEA Analysis based data from IEA (2023), Efficient Grid-Interactive Buildings.

In the context of the buildings sector, the increasing frequency of extreme heat and cold waves in many regions contributes to electricity demand spikes in both commercial and residential buildings due to the need for space cooling and heating. AI can effectively limit these demand spikes by boosting demand flexibility.

Instead of static demand, machine learning-based software, such as Flex2X, allows a building's load profile to be more flexible by controlling energy use. It also makes it possible for buildings to participate in electricity markets by providing more flexibility to the grid. Flexible building loads include heating and cooling, which can be shifted in time to pre-cool or pre-heat a building when electricity prices are low or electricity emissions intensity is low.

Korea has recently tested the Intelligent Building Energy & Environment System (iBEEMS), which is its first-ever AI-based automated BEMS. Based on an AI training model, the system forecasts occupants' behaviour and sets an optimal level of HVAC accordingly. During its initial pilot, the system reduced energy use by an impressive 43%.

Singaporean state-owned utility, SP Group, also applied AI with IoT to enhance energy efficiency in buildings. Given Singapore's tropical climate, its buildings sector accounts for around one-third of the country's total electricity demand. SP Group's micro-climate control solution learns various datasets such as building occupancy status and weather conditions to optimise air-conditioning systems. Based on the collected data, the solution divides the buildings' open spaces into smaller zones to balance each zone's air flow, temperature and CO_2 level. This enhances the building's energy efficiency without compromising occupant experience.

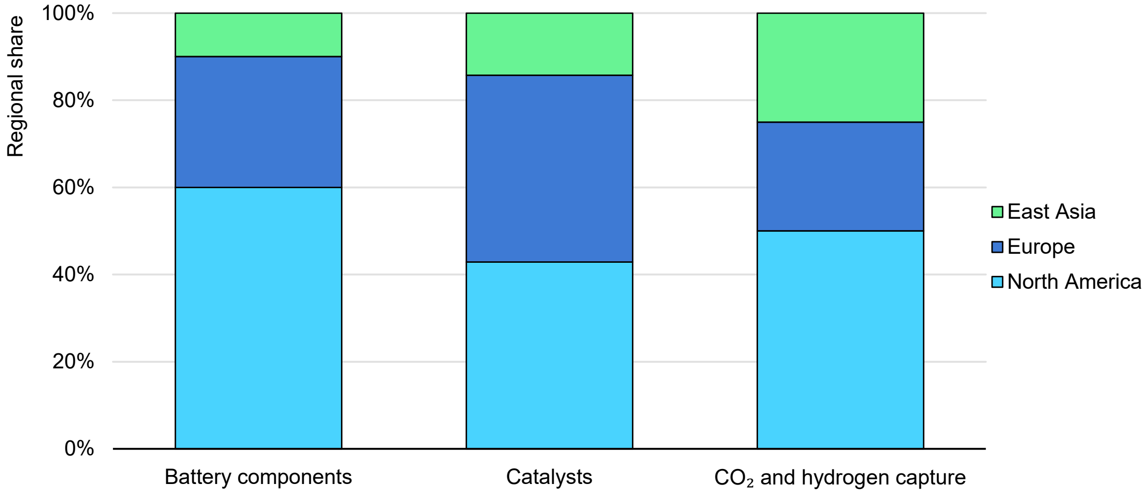
AI for energy innovation

The energy sector is on the frontline of technological innovation and adoption. In East Asia, technologies such as batteries, electrolysis, carbon capture, utilisation and storage (CCUS) and synthetic fuels are contributing to the decarbonisation of hard-to-abate sectors such as the cement and steelmaking industries.

AI can contribute to four key overarching areas across technologies. Firstly, it can help with the discovery of new materials and chemical combinations, such as catalysts, for innovative technologies. AI can also optimise the manufacturing processes to reduce costs and save time. Moreover, AI can increase the precision of quality assessment, while conducting predictive maintenance to ensure proactive interventions. Lastly, AI can optimise the system integration and actual operation of the innovative technologies.

Recent IEA analysis identified a list of companies and start-ups that have commercialised innovative energy technology products, assisted by AI. Amongst that list of innovations using AI, those relating to battery cathode and electrolyte technologies have a higher average technology readiness level (TRL) than hydrogen storage, electrolysis and fuel synthesis.

Regional distribution of companies using AI in selected innovative energy technologies, 2025



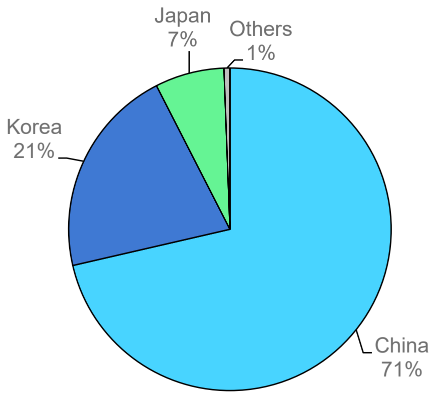
Source: IEA Analysis based data from IEA (2025), The State of Energy Innovation.

Batteries

Batteries have already become an integral part of the energy transition. The technology is rapidly evolving beyond the existing lithium-ion batteries, expanding to new storage types such as redox flow and sodium-ion batteries. These new storage types can be useful for various applications, including longer-duration storage. With rapid technological advancement, AI can effectively support key aspects of battery innovation such as battery material discovery, cell design and operation optimisation. By facilitating automation, computational modelling and forecasting abilities, AI can tackle the long and complex process of developing innovative battery configurations and operation plans.

As one of the biggest players in the industry, East Asia is a vibrant testbed for implementing AI in battery technology innovations, primarily by leading manufacturers in the region, such as CATL (China), LG Energy Solutions (Korea) and Panasonic (Japan). Specifically regarding EV batteries, the ten largest manufacturers are all owned by either China, Korea or Japan. As of 2024, 71% of the top ten EV battery manufacturers were headquartered in China, followed by Korea with a 21% share.

Market share of electric car battery producers by company headquarter country, 2024



Note: Battery refers to lithium-ion battery cells. It reflects the batteries installed in vehicles sold in 2024. Electric car and battery stockpiling are excluded from the analysis. Manufacturers are mapped to countries based on their headquarters location, not factory location.
Source: IEA (2025), Global EV Outlook.

Despite China's market dominance, the regional distribution of manufacturing facilities varies. LG Energy Solutions, for instance, is the third-largest manufacturer globally, following Chinese companies, CATL and BYD. However, it accounts for over half of the manufacturing capacity in Europe.

The company developed its own generative AI model to simulate and find optimal battery design that reflects the clients' Request for X (RfX), which are technical specifications from LG's clients. This model, trained with around 100 000 cell designs, significantly reduced the design period from a maximum of two weeks down to just one day. Another major actor, Siemens, is also implementing its generative AI, Xcelerator, to cut the battery design and production time by 50%.

Electrolysers

Hydrogen is a versatile fuel that can be produced through a wide range of technologies. Currently, global hydrogen demand is met almost exclusively from fossil fuels. Electrolysis is an alternative approach, where electricity can be used to split water into hydrogen and oxygen. If the electricity is generated from low-carbon energy sources, the resulting hydrogen is low-carbon hydrogen.

Electrolysers can be used to target two pertinent issues in the energy transition – the decarbonisation of hard-to-abate sectors, and VRE integration. Low-carbon electrolysers produce hydrogen by splitting water into hydrogen and oxygen, using electricity generated from low-carbon energy sources such as renewables or nuclear. Surplus low-carbon electricity can be used, rather than being curtailed, to produce low-carbon hydrogen.

As of 2020, electrolysis accounted for 0.03% of worldwide hydrogen supply. AI can play an important role in the decarbonisation of this hard-to-abate sector. It can accelerate the commercialisation of electrolysers by streamlining catalyst discovery and modelling, rapidly testing numerous chemical combinations to identify optimal solutions. Its predictive capabilities also help optimise production schedules by forecasting load conditions based on VRE inputs, such as solar and wind power.

Tokyo Gas uses machine learning to search for more cost-efficient and durable catalysts for electrolysers. Proton exchange membrane (PEM) water electrolysis requires iridium as a catalyst on the anode side. However, iridium's price is extremely high due to its scarcity. This motivated Tokyo Gas to explore the possibility of reducing the iridium loading and finding new alternatives. The company's subsidiary, H2U, collected approximately 20 000 catalyst samples to be trained with machine learning algorithms to evaluate their performance and durability.

Synthetic fuels

Synthetic fuels are expected to play a pivotal role in several hard-to-abate sectors, especially long-distance transport sectors such as aviation and shipping. These fuels, such as synthetic hydrocarbon fuels, are produced by combining hydrogen and CO_2. Existing transport infrastructure can utilise these fuels as they are in liquid and gaseous form (drop-in fuels).

AI can be used to accelerate the design of catalysts for synthetic fuel production, such as Fischer-Tropsch (FT) synthesis.[6] FT synthesis is highly energy-intensive, which requires better catalyst designs that can make the process more effective. In this regard, AI can support in identifying catalysts with better performance by training on a large set of performance data. One of the methods of modelling catalyst performance is the density functional theory (DFT), a traditional quantum physics-based model used in computational chemistry. It helps to discover new materials and predict their performance. Despite being computationally intense, the model can provide valuable training data for predictive AI, which then uses machine learning or neural networks to estimate catalyst performance in a fraction of the time.

Following developments in the application of AI in biochemistry innovation, generative AI models are being developed to propose new candidate materials that meet specific catalysis criteria. Tools like AutoMat and MAGECS use

[6] A catalytic progress that converts hydrogen and carbon into synthetic hydrocarbons, a petroleum-like product.

predictive and generative models to identify high-performance catalysts. Open datasets such as the Open Catalyst Project support this innovation.

Carbon capture, utilisation and storage

Carbon capture, utilisation and storage (CCUS) technology is another technology that can benefit from AI. One of the main challenges in implementing CCUS lies in the fact that the existing carbon capture materials are not yet sufficiently cost- or energy-efficient, making the overall process economically less competitive. Within this context, AI can catalyse the discovery of more cost- and energy-efficient carbon capture materials that can efficiently select and extract CO_2 from a gas mixture or the atmosphere.

AI's strong computational power can support predictive and generative models to process various datasets on the performance of capture materials, ranging from amines to metal organic frameworks (MOFs). For example, a generative AI framework – GHP-MOFassemble – can create new combinations by assembling pre-selected nodes into MOFs, while testing each combination's performance, such as its structural validity. It validated 102 satisfactory options within 11 hours. Although this technology is still at an early stage of development – currently at TRL 2 – it serves as a compelling example of how AI can be leveraged to accelerate the discovery of new MOFs through its computational and predictive capabilities.

AI applications in the energy sector

	Material and design optimisation	Manufacturing and process optimisation	Quality assessment and predictive maintenance	System integration and operation optimisation
Batteries	Anode, cathode, precursor and electrolyte material selection	Digital twin for manufacturing optimisation	Battery health monitoring and improvement	Storage optimisation for various durations
	Design simulation	Coating process optimisation	Cell degradation modelling	
	Cell testing	Scrap rate reduction	Performance forecasting	
		Defect detection		
Electrolysers	Electrocatalyst selection and modelling	Operation and manufacturing parameter optimisation	Degradation modelling	Load conditions prediction
			Scheduling	Hydrogen production optimisation

	Material and design optimisation	Manufacturing and process optimisation	Quality assessment and predictive maintenance	System integration and operation optimisation
Synthetic fuels	Fuel and catalyst design	Fischer-Tropsch synthesis optimisation	Catalyst design	
	Fuel blending	Density functional theory		
CCUS	Solvent, membrane and catalyst selection and design	Capture process optimisation	Life cycle sustainability assessment	CO_2 monitoring
		Conversion efficiency improvement	Leak detection	

Note: CCUS = Carbon capture, utilisation and storage.

Sources: IEA Analysis based data from IDTechEx (2024), AI Is Well Set to Disrupt the Battery Supply Chain and Life Cycle; Seo et al. (2025), High quality large-scale nickel-rich layered oxides precursor co-precipitation via domain adaptation-based machine learning; BMW (2025), Efficient, sustainable, digital: The new BMW Group Plant Debrecen; Siemens (2025), Precision Meets Digitalization: How Siemens Is Redefining Battery Production with the Coating App; NREL (2025), Artificial Intelligence Models Improve Efficiency of Battery Diagnostics; Moon et al. (2024), Active learning guides discovery of a champion four-metal perovskite oxide for oxygen evolution electrocatalysis; Enapter (2025), Enapter presents the world's first AI-powered electrolyzer; Ma et al. (2025), Perspective on artificial intelligence for carbon capture utilization and storage (CCUS) in Petrochemical Industry; Tawalbeh et al. (2025), Artificial intelligence and material design in carbon capture and utilization: A review of emerging synergies.

Chapter 2. Energy for AI

The world is entering the Age of Electricity, as highlighted in the IEA's World Energy Outlook 2025. Data centres, especially for artificial intelligence (AI), are one of the drivers of rising global electricity demand. Compared with other load types, data centres show strong geographical concentration. Hence, this growth can put additional pressures on power systems.

Aligning AI growth with secure and clean energy goals in East Asia requires understanding data centres' electricity demand trends, the development of supply to meet this rising demand and the integration of data centres into the power system. Against this backdrop, this chapter examines the key characteristics of data centre electricity use and discusses their future electricity requirements. It reviews options for meeting this rising demand, including power purchase agreements, behind-the-meter generation and innovative low-emissions supply sources. The chapter also considers how large and concentrated loads from AI data centres can be connected to power systems in ways that maintain reliability and contribute to grid flexibility.

Energy use in data centres

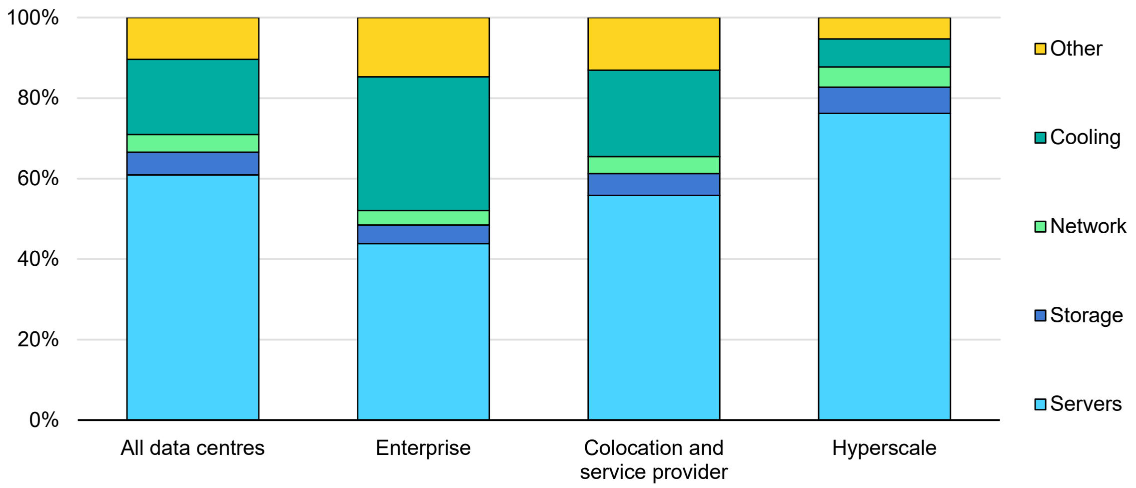
Data centres are complex infrastructure; they consume energy not only for computing hardware, but also for the systems that support it, such as cooling, data storage and network facilities. Compared with conventional data centres, AI data centres use similar components but require more power-dense hardware, longer and more variable computing cycles and cooling systems capable of managing high thermal loads.

Estimating data centres' energy use can be challenging as it depends on multiple factors. On the hardware side, the density of graphics processing units (GPUs) or accelerators within each rack and the efficiency of the cooling systems can affect overall energy usage. Whether the hardware is optimised for training or inference affects the energy usage characteristics, with future trends on the inference side being more uncertain. In respect of software, the model size, model type, training parameters, training duration and training scheduling also affect energy usage. The atmospheric environment, including the external temperature and humidity, drive substantial differences in energy usage, through changes in cooling requirements.

Digital AI depends on physical silicon, concrete and steel

Data centres consist of multiple physical systems that consume electricity in different ways. The key hardware systems that account for most electricity use in data centres include servers, cooling systems, storage and networking equipment, and backup power infrastructure.

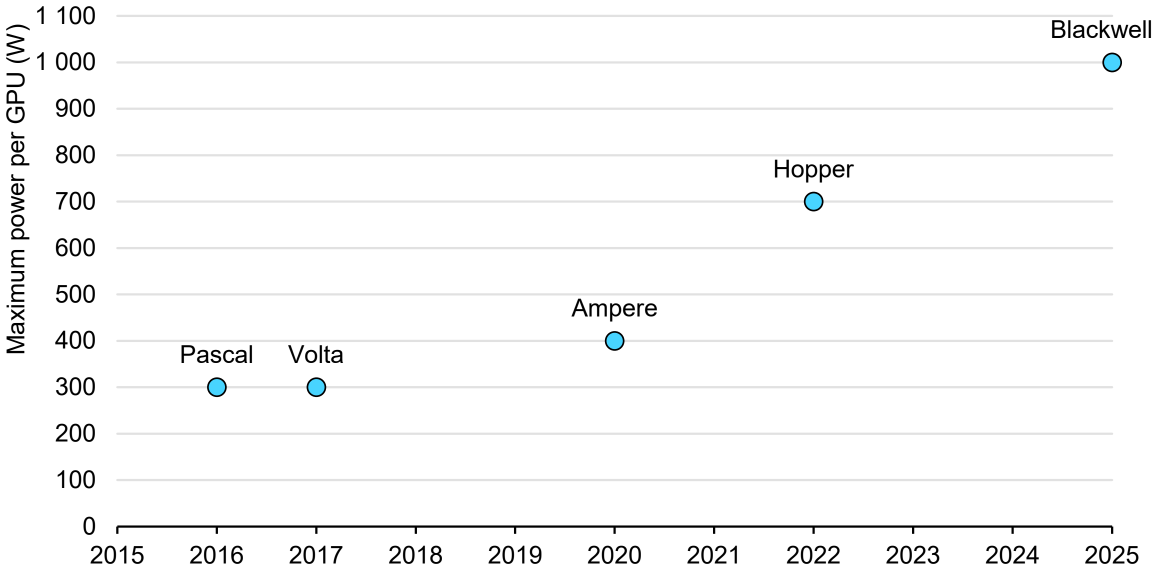
Share of electricity consumption by data centre and equipment type, 2024



Source: IEA (2025), Energy and AI.

Servers account for the largest share of electricity use in traditional data centres at around 60% of total demand on average, although this share may change in future AI data centres as designs are rapidly evolving. Servers are the core information technology (IT) infrastructure containing computer chips and memory, providing the processing power needed to support the development and operation of AI models – training and inference. Traditionally, servers relied on central processing units (CPUs), which are designed for a wide range of computation tasks. However, as AI models have grown in size and complexity, data centres have shifted towards using accelerated processors, mostly GPUs. These processors are designed to handle large volumes of parallel computations, which enables the model to process massive datasets in a shorter amount of time, making them more suitable for training advanced AI models. While each new generation of GPUs tends to be more efficient than the last, the total power draw per GPU still tends to be larger, as does the total number of GPUs per rack.

The power consumption of flagship NVIDIA graphics processing units



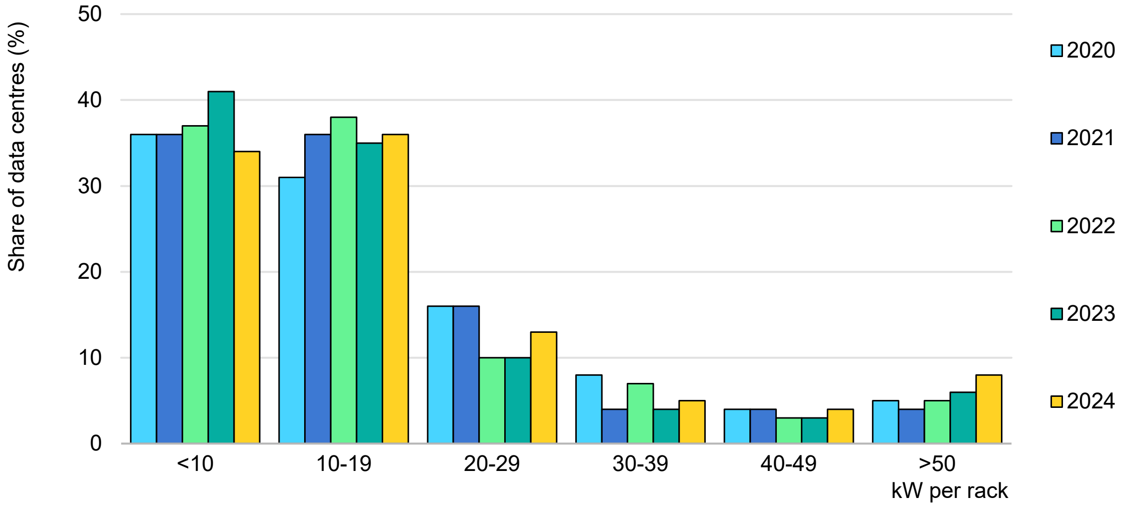
Sources: IEA Analysis based data from NVIDIA datasheets (2026), HGX B200, H100 SXM, A100, V100 SXM2, P100.

To maximise computing capacity, AI data centres stack multiple GPUs into one rack and deploy several high-density racks across the facility. This configuration significantly increases computing outcome, but raises energy intensity as well. Higher density racks consume more energy per rack because they require additional cooling and power delivery systems. However, they also deliver more compute output per kilowatt (kW) by running more parallel jobs simultaneously and reducing loss from latency or fragmentation across racks. Increasing AI workloads are one of the primary drivers of increasing rack density. More computing power requires more GPUs to be clustered, hence higher rack density.

This intensity at scale is how hyperscale data centres, mostly operated by major technology firms, can facilitate some of the most demanding AI workloads. In these data centres, servers alone can account for around 75% of the total electricity consumption.

While there is no globally agreed definition of a hyperscale data centre, the industry generally refers to them as facilities that are equipped with at least 5 000 servers, requiring a vast amount of land – at least over 1 km^2. Recently deployed hyperscale data centres' rack density can climb to over 140 kW per rack, which can significantly affect many aspects of data centre operation such as electricity consumption and cooling systems. The rack density of traditional enterprise data centres is in the range of 5-10 kW per rack.

Distribution of maximum rack density in data centres, 2020-2024



Sources: IEA Analysis based data from Uptime Institute (2022), Global Data Centre Survey 2022; Uptime Institute (2024), Global Data Centre Survey 2024.

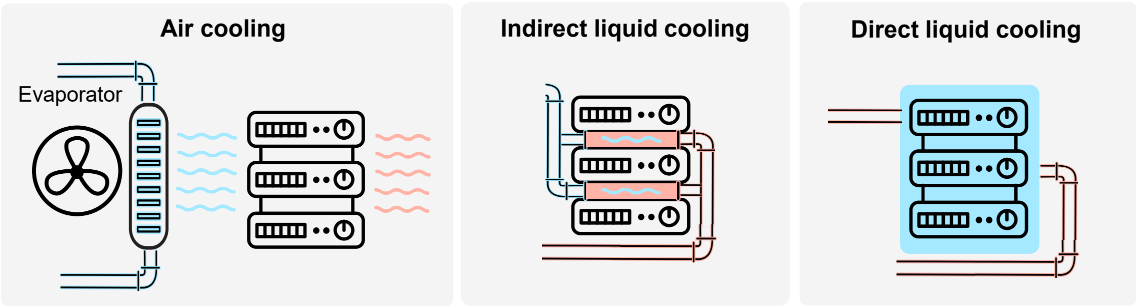
Responses to the Uptime Institute's annual global survey (of all types of data centres) show that the vast majority of data centre operators have increased their rack density over the past few years. Racks less than 10 kW are the most common, while the facilities with a higher rack density of between 10 kW and 19 kW have been continuously increasing for the past six years. Many AI workloads may run in lower-density environments, especially inference workloads and smaller models that do not require physically close GPU setups. Air-cooling setups have lower rack density than liquid or immersion cooling setups. At the higher end, installations with densities over 50 kW per rack have also been continuously growing for the past six years, but remain in the minority. In 2024, around 30% of data centres surveyed globally ran on rack densities of 20 kW per rack or higher.

Following servers, **cooling systems** account for the second-largest share of total electricity demand across all types of data centres, greatly ranging from 7% up to 30%, depending on the type of data centre and its efficiency. Due to higher rack densities, high-performance data centres generate higher temperature waste heat than traditional data centres.

Intense AI workloads may push rack densities beyond the physical limits of existing air cooling systems. Air cooling remains widely used in the industry, but it can become impractical and less cost- and energy-efficient when the rack density surpasses a certain threshold, especially in high-performance hyperscale or large-scale enterprise data centres.

For this reason, AI data centres are seeking innovative cooling technologies, such as liquid cooling systems that include direct-to-chip and immersion cooling. Liquid cooling is gaining attention as an effective cooling system for intensive computing infrastructures as it has better cooling performance than air-based systems. Liquid cooling can be combined with other cooling systems such as rear door heat exchangers (RDHXs) to accommodate higher rack densities. Cold plate and immersion coolant systems are both forms of liquid cooling, which have lower energy consumption than air coolant systems. However liquid cooling is more challenging to deploy, especially in existing data centres.

Overview of air cooling and liquid cooling systems in data centres



Notes: Indirect liquid cooling uses cold plate cooling as the example; direct liquid cooling uses immersion cooling as the example.
Source: Reproduced with permission from Kong, R. et al. (2024), Enhancing data center cooling efficiency and ability: A comprehensive review of direct liquid cooling technologies - ScienceDirect.

Atmospheric variability, including ambient temperature and humidity, is an important factor influencing cooling efficiency in data centres, making site climate profiles a key consideration at the design stage. Cooling requirements fluctuate with daily and seasonal variations and the size of the load. While industry-wide cooling demand is rising, there are notable differences between Asia and Europe.

Practices such as "free cooling", which leverage cooler outdoor temperatures, especially overnight, are becoming less effective during increasingly hot summers, while winter conditions in East Asia remain favourable due to colder temperatures. In the case of larger seasonal variations and more heat- and cold waves, having adaptive, flexible cooling strategies becomes important to maintain efficiency and reliability.

Some data centres recycle their waste heat by using it to warm local communities. This revenue source should be considered during siting decisions. Helen, an energy company in Finland, is capturing waste heat from Equinix's data centre to use in local heating networks. In Denmark, Meta repurposes 100 GWh of waste heat per year, while Exergi, a district heating and cooling provider, has a similar project in Stockholm. In Ireland, heat from an Amazon data centre is recycled in the Tallaght district. Hemiko, a UK-based developer, investor and operator of district heating networks, warms 9 000 homes in London with waste heat. Nearby, Gatwick Airport recycles heat from its data centre. During the 2024 Paris Olympics, Equinix heated Olympic swimming pools with data centre heat. In Lhasa, Tibet, data centre heat is reused for agricultural purposes.

AI training can take hours or days, which dwarfs the transmission delay of sending data around the world. For use cases without stringent data sovereignty rules, organisations may decide to train models in other countries with cooler weather and grids with a higher proportion of VRE. BMW shifted some of its high performance computing applications to Iceland, reducing its costs by 82%. Thanks to Iceland's 100% renewable grid, this also allowed BMW to reduce its emissions by 3 570 tonnes annually.

Storage and network equipment can collectively make up more than 10% of the typical data centre's total electricity demand, with potentially higher amounts for systems optimised for AI training. A server equipped with the most advanced GPUs cannot perform to its full potential without being coupled with high-performance storage and network capacity.

Fast network infrastructure is another integral ingredient for high computing performance. High network speed enables faster data exchange between servers and storage equipment, faster training and more co-ordinated communication between GPUs, ultimately leading to less latency to deliver the final inferred output to the end users. Data centres and racks designed for AI inference and training can have substantially different network topologies than traditional cloud computing, with differing numbers of switches per rack, and differing technology choices such as Ethernet and InfiniBand, which affects the power draw of the switches in aggregate. These design decisions are rapidly changing, and future network designs in AI data centres are expected to evolve accordingly. Electricity demand for storage and network equipment, however, is not only determined by their capacity, but also by other compounding aspects such as applicable technologies, workloads and utilisation patterns.

Backup power infrastructure generally refers to the uninterruptible power supply (UPS) and backup generators that provide emergency electricity to the data centre if its main power source fails. Most modern data centres aim to reach a very high reliability – 99.999% – which makes backup power facilities more important. Backup generators are used to power a data centre during rare periods when the grid supply is unavailable. They are engine-based, often powered by diesel or gas, and may take several minutes to start following a grid outage. The purpose of a UPS is to instantly provide battery power to the servers during the short period before the generator starts.

Lead-acid batteries have traditionally been used for UPS units. In recent years lithium-ion batteries have gained in popularity and are rapidly being deployed as UPS, mostly due to their higher energy density.

Batteries alone are rarely used to provide backup power to large data centres. It is typically costly to scale them for the power and durations required. Diesel generators are a cost-effective and well-established choice, with several downsides. When running, diesel generators produce noise which can disturb neighbours, as well as emitting relatively large amounts of CO_2, NO_X and CO per unit of energy. To minimise the start-up delay of backup generators, they may be kept warm while not running, using block heaters powered by the grid. This contributes to the power draw of the site.

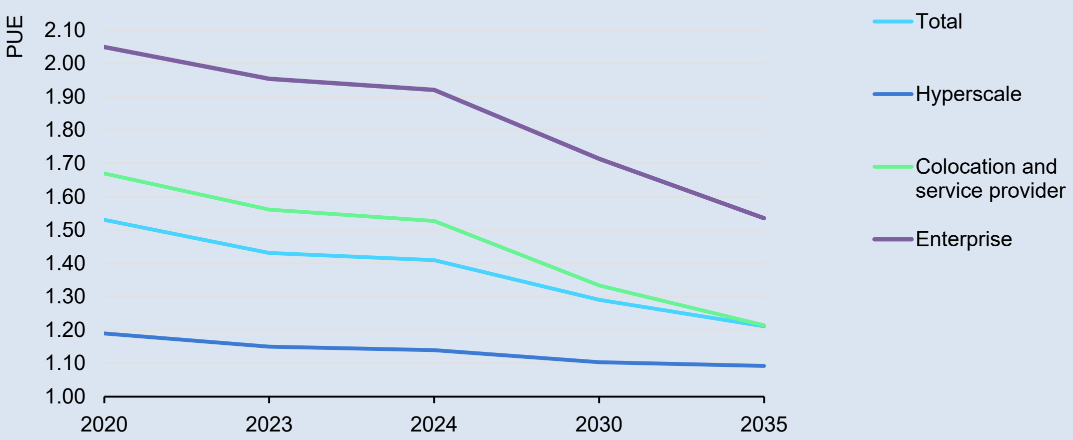
Power usage effectiveness: An indicator of data centre energy efficiency

Power usage effectiveness (PUE) is the ratio of a facility's total energy consumption to the energy consumed by its IT equipment, such as the server. A perfect PUE value would be 1, meaning all energy consumed by the facility is used by IT equipment, with no overhead for other infrastructure. In practice, this is not achievable because some energy is always required for cooling and other support systems. A higher PUE value means a larger proportion of energy is consumed by other infrastructure relative to IT equipment.

Over the past decade, the annual average PUE has significantly improved from 2.5 in 2007 down to 1.41 in 2024. However, at the industry level the PUE has plateaued since 2020. Despite more efficient data centres coming online, a significant share of existing legacy facilities remains with higher PUEs. According to recent IEA analysis, the global weighted average PUE is expected to decline continuously, reaching 1.29 in 2030 in the Base Case. This could save around 90 TWh of data centre electricity demand for the IT load growth forecast in the Base Case, equivalent to reducing the cooling requirement per unit of IT electricity by 30%.

PUE varies depending on various factors such as the facility's type and computing performance, as well as the climate of its geographical location. Different types of data centre show notable differences in PUE. Hyperscale data centres are more energy efficient than others, at 1.14 in 2024 and are expected to drop down to 1.09 by 2035. Google reportedly reached a trailing twelve-month (TTM) PUE of 1.09 on average across its global fleet of data centres in 2024.

Data centre PUE by type, 2020-2035



Note: Data for 2025-2035 are forecast values from the Base Case in the source.

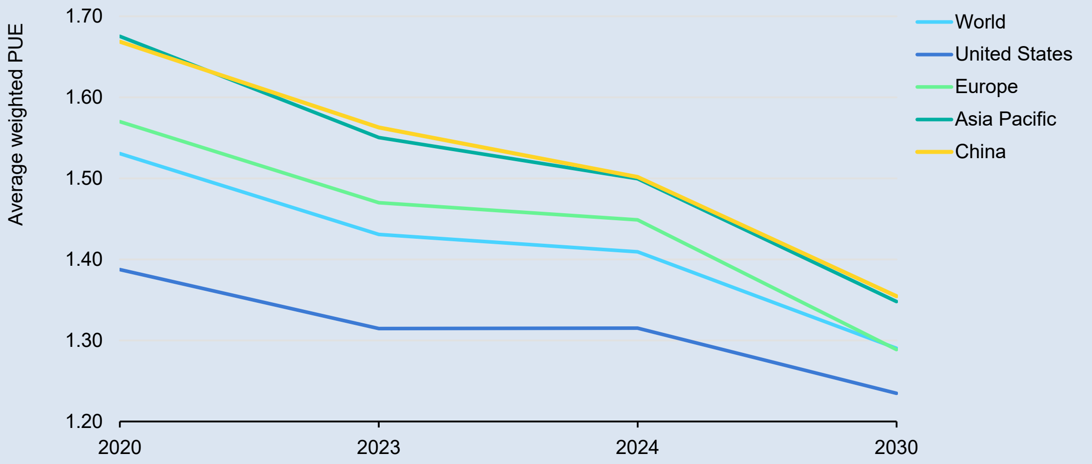
Source: IEA (2025), Energy and AI.

Although global PUE values are projected to decline, the pace of improvement varies by region. Under the IEA's Base Case, PUE decreases across all regions through to 2030, with existing disparities between regions narrowing. By 2030, the global average PUE is projected to be lower than 1.3.

This gap in future PUE improvements is largely driven by climate-related factors that significantly affect energy consumption for cooling (and in some cases heating) data centres. Asia Pacific's hot and humid conditions substantially increase cooling requirements, especially as facilities scale up. As discussed earlier, "free cooling", which relies on outside air, is less feasible in this region, compared with Europe and the United States, where temperate climates allow operators to leverage such techniques more effectively.

However, it is important to note that many regions in the United States and Europe are increasingly experiencing extreme weather events, particularly heatwaves during the summer. The growing uncertainty around severe climate conditions could influence future PUE projections across all regions.

Data centre PUE by region, 2020-2030



Source: IEA (2025) Energy and AI.

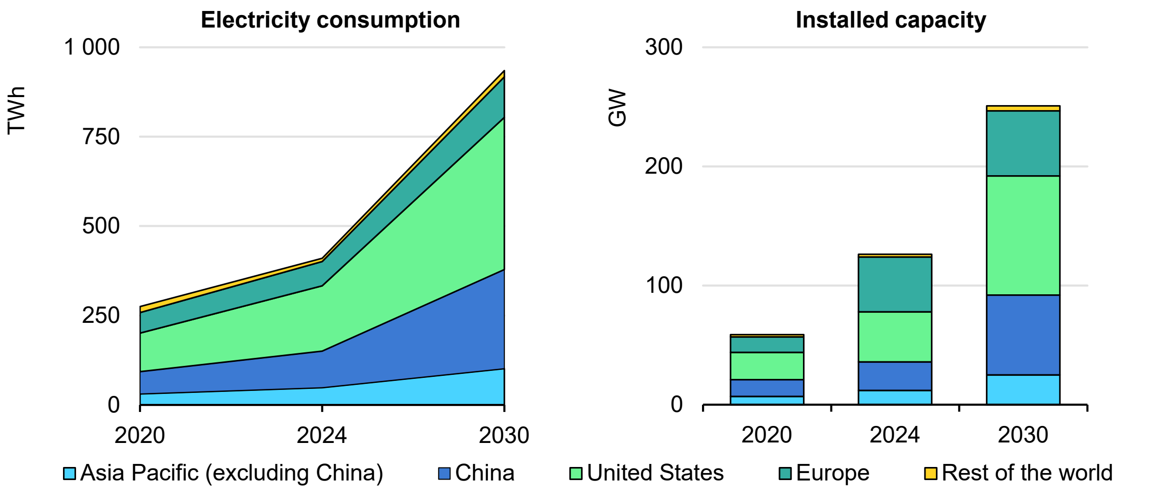
Outlook for electricity consumption for AI in East Asia and beyond

Understanding AI's future electricity consumption is essential to prepare power systems around the world, as well as to align policies and regulations accordingly. Recognising this, the IEA recently presented a set of scenarios for electricity consumption by data centres, based on a bottom-up methodology. This uses the industry's IT equipment projections and is based on future demand as well as supply constraints. The main drivers of the scenarios include the future total server stock, including accelerated servers, the overall installed capacity of data centres, and hardware efficiency improvements, such as advancements in cooling systems.

In the Base Case, the global electricity consumption of data centres more than doubles from around 415 TWh in 2024 to 945 TWh by 2030. Regionally, the Asia Pacific region – excluding China – is projected to consume around 100 TWh for data centres in 2030, which accounts for 11% of future global electricity consumption by data centres. This is comparable to the projected future demand from data centres in Europe, which is projected to grow to around 115 TWh, accounting for approximately 12% of global demand.

China is expected to maintain its position as the leading source of demand in East Asia by 2030 with regard to electricity consumption by data centres. China currently accounts for the second-largest share of data centre electricity consumption (after the Unites States) at 25% of the global 415 TWh of consumption by data centres as of 2024. In 2030, at around 280 TWh, China is projected to account for almost 30% of global consumption.

Global data centre electricity consumption and installed capacity by region, 2020-2030

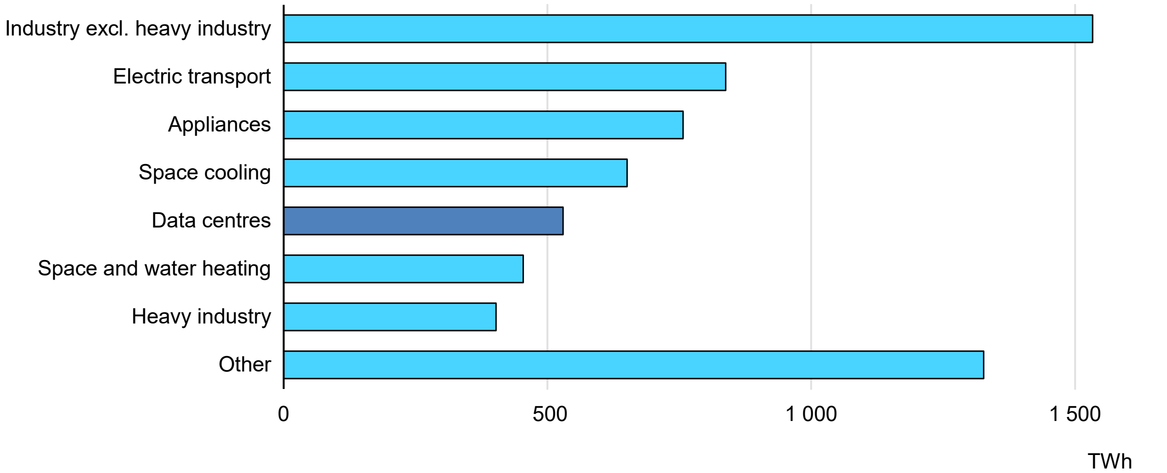


Note: The data are based on the Base Case. Data for 2025-2030 are forecast values.
Source: IEA (2025), Energy and AI.

The surge in data centre electricity demand is tied to the increase in total installed data centre capacity – globally, it is expected to more than double from 97 GW in 2024 to 225 GW in 2030. China shows the highest growth in capacity at around 180%, reaching almost 70 GW in total by 2030. China and the United States are projected to account for more than 70% of global data centre capacity in 2030. The Asia Pacific region, excluding China, is also projected to see strong growth, doubling the total capacity of its data centres from 12 GW in 2024 to 25 GW in 2030.

Despite this strong increase, electricity demand from data centres accounts for less than 3% of total global electricity consumption in 2030. The majority of electricity demand growth is expected to be driven by the industrial sector as well as rapid electrification, including the deployment of electric vehicle (EV) fleets and the wider diffusion of air conditioning.

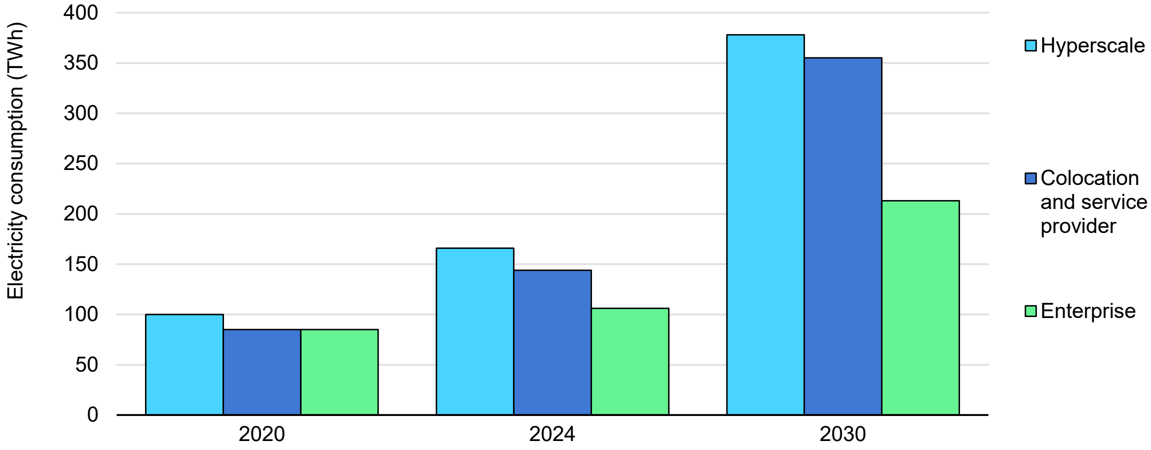
Increase in global electricity demand by sector, 2024-2030



Note: The data are based on the Base Case.
Source: IEA (2025), Energy and AI.

Due to variations in energy efficiency and technical configurations, different types of data centres show varying projections for total electricity consumption globally. In 2020, all types of data centres – hyperscale, colocation and enterprise – consumed a similar amount of electricity, with hyperscalers using only about 18% more than the other two types. By 2030, however, electricity consumption is expected to rise sharply for both hyperscale and colocation data centres. Notably, globally, hyperscale data centres are projected to consume approximately 77% more electricity than the other two types, highlighting their growing dominance in overall demand.

Data centre electricity consumption scenario by data centre type, 2020-2030



Notes: The data are based on the Base Case. Data for 2030 are forecast values.
Source: IEA (2025), Energy and AI.

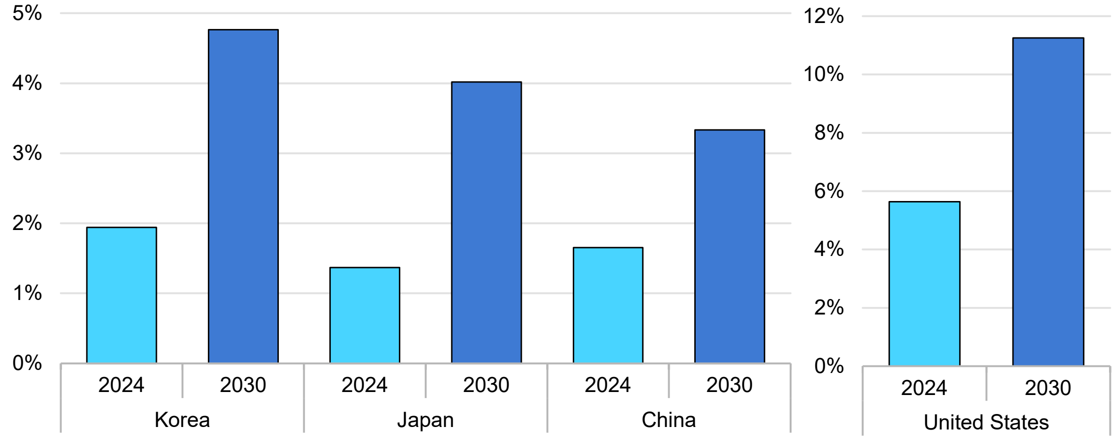
Data centres' share of peak electricity demand is set to increase

To measure the scale of data centres' growing electricity demand, it is useful to compare it with system peak demand. As data centres continue to expand, their share of peak demand is projected to increase significantly. This trend underscores their system-wide impact.

The projection for the United States shows a significant rise in the influence of data centres on power systems. In 2025, their peak capacity was equivalent to 5% of the nation's peak demand. In 2030, their share will grow to more than 10% of projected peak demand. This is mainly due to hosting larger and more concentrated data centre clusters.

In Korea, Japan and China, data centres' peak capacity was only comparable to a little over 1% of each countries' peak demand in 2025. In 2030, however, this is expected to show a major increase to around 4% of peak demand in 2030.

Data centre capacity as a share of peak demand by selected country, 2024-2030



Notes: The data are based on the Base Case, except for Korea and Japan. Data for 2030 are forecast values. Data centre capacity corresponds to total installed nameplate capacity. Coincident peak contribution is likely to be materially lower because utilisation factors vary, loads are geographically distributed and system peaks may not align temporally with maximum data centre operation. Data centres may also participate in demand response or operate on-site generation, which can further reduce grid-supplied load during peak conditions. These ratios should be interpreted as an upper-bound indicator of potential peak contribution rather than realised peak load. Data for Chinese Taipei are not available.

Sources: IEA Analysis based data from Ministry of Trade, Industry and Energy (2025), Ministry of Trade, Industry and Energy, 11th Plan for Electricity Supply and Demand, Power supply and demand this summer and the outlook and operations for this winter and beyond; OCCTO (2025), Aggregation of Electricity Supply Plans for Fiscal Year 2025; Wood Mackenzie(2025), Japan's data centre boom to drive 60% of power demand growth; US EIA (2025), US electricity peak demand set new records twice in July; IEA (2025), Energy and AI, Electricity 2025; IEA (2026), Electricity 2026.

Regulatory frameworks for data centre energy efficiency

As data centres expand to meet the demands of AI development, their increasing consumption of resources, especially electricity, is raising concerns as policy makers recognise their growing role as key infrastructures.

Despite the growing need to manage the impact of data centres’ surging electricity demand, regulatory frameworks have not kept pace with the sector’s rapid development. Most sustainability disclosures are currently via company self-reporting, which varies widely in scope and detail. Global data centre operators typically publish regular sustainability reports that include related data such as the average PUE, greenhouse gas emissions, water consumption and the share of electricity sourced from renewables, but not the detailed data on actual electricity consumption.

A few international standards organisations, such as the International Electrotechnical Commission (IEC), are continuously updating their guidelines to include data centre performance metrics, such as PUE. However, participation remains voluntary and is largely driven by industry and market actors. Improvements in the global regulatory landscape could help businesses and governments more accurately assess data centres’ energy-related performance.

Examples of global sustainability standards for data centres

Standard	Description
ISO/IEC 30134	Globalised standard developed by the International Electrotechnical Commission (IEC), including key indicators of data centre sustainability: - ISO/IEC 30134-2 PUE - ISO/IEC 30134-3 REF (renewable energy factor) - ISO/IEC 30134-6 ERF (energy reuse factor) - ISO/IEC 30134-7 CER (cooling efficiency ratio) - ISO/IEC 30134-8 CUE (carbon usage effectiveness) - ISO/IEC 30134-9 WUE (water usage effectiveness)
Leadership in Energy and Environmental Design (LEED)	Global green building rating system, operated by a US-based non-profit organisation, the US Green Building Council Updated LEED v5 to reflect data centres' characteristics

Sources: Future-tech (2024), Introduction: The ISO/IEC 30134 Series of Standardised KPIS; US Green Building Council (2026), LEED certification for new buildings and major renovations.

Regionally, some countries and regions are initiating regulatory frameworks to enhance the visibility of data centre energy efficiency performance.

China is the first jurisdiction in the world to adapt a nationwide standard for data centre energy efficiency. In 2021, the nationwide standard GB 40879-2021 defined performance standards for data centres. This has developed into the government's action plan on developing green data centres, announced in 2024. The plan has set a PUE of 1.5 as the target threshold by 2025.

The European Union has agreed the first example of legally binding obligations for data centre energy efficiency. Its recently revised Energy Efficiency Directive (EU/2023/1791) provides the first-ever legal standing for EU countries to impose an obligation to track and disclose energy performance of data centres.

In relation to the EU directive, Germany also established a legal obligation via the federal law Energieeffizienzgesetz (EnEfG) on energy efficiency, which includes data centres. The law stipulates a legal obligation on data centres that exceed a certain capacity standard to validate or certify their energy management system, to measure their energy demand and improve overall efficiency. The European Union and Germany have not set any specific PUE value for data centres to comply with.

The Netherlands and Singapore also have data centre energy efficiency frameworks in place, but via policy initiatives not legal instruments, setting a target PUE value of 1.16 and 1.3, respectively. Japan's Agency for Natural Resources and Energy, affiliated with the Ministry of Economy, Trade and Industry, has also set a target PUE of 1.4 for data centres above a certain capacity.

While some countries stipulate the energy efficiency target for data centres by law or policy, data centres in other countries such as Korea and the United States are complying with global voluntary standards. Industry associations for data centres are also providing private voluntary sustainability schemes, such as the Korea Data Centre Council's Green Data Centre Certification.

Examples of data centre regulations and standards in selected countries

Country	Type	Title	Target PUE
China	Policy initiative	Green Data Centre Action Plan National Standard (GB 40879-2021)	1.5
European Union	Directive	Energy Efficiency Directive (EU/2023/1791)	To be set by member states
Germany	Law	Energy Efficiency Act (BGBl. 2023 I Nr. 309)	N/A
Netherlands	Policy initiative	Guidelines for Noord-Holland	1.16
Singapore	Policy initiative	Green Data Centre Roadmap	1.3
Japan	Policy initiative	Agency for Natural Resources and Energy	1.4
Korea	Voluntary certification	Green Data Centre Certification	N/A
United States	Voluntary certification	ENERGY STAR	N/A

Sources: IEA Analysis based data from Energy in Building and Communities Programme (2022), International review of energy efficiency in Data Centres; UN Environment Programme (2025), Sustainable Procurement Guidelines for Data Centres and Servers.

Providing low-emissions electricity for data centres

Globally, governments and companies are doubling down on their efforts to diversify electricity procurement strategies to meet the surging demand from AI data centres. While many data centres are scaling up to support increasingly complex AI workloads, their electricity requirements are growing at an unprecedented pace. In response, major AI companies have pledged to power operations with low-emissions electricity, such as renewables and nuclear power. Corporate-level commitments to use 100% renewable electricity are being extended to data centre operations, and have already been achieved by major players such as Google, Meta and Amazon, which host AI models (Gemini, Llama) and cloud services (AWS). RE100 is an example of an initiative to standardise the accounting of corporate climate promises by large energy users, to incentivise improved reporting, and to advocate the removal of obstacles to achieving high corporate decarbonisation targets.

Examples of major data centre companies with renewable electricity pledges

Technology company	Type	Renewable energy targets
Amazon Web Services	Cloud provider	100% renewables matching by 2030 (achieved in 2023)
Microsoft Azure	Cloud provider	100% zero-carbon energy matching by 2030
Google	Cloud provider	24/7 carbon-free energy on every grid that it operates by 2030
Equinix	Colocation provider	100% clean and renewable energy by 2030
Meta	Hyperscaler (internal usage)	100% renewables matching since 2020

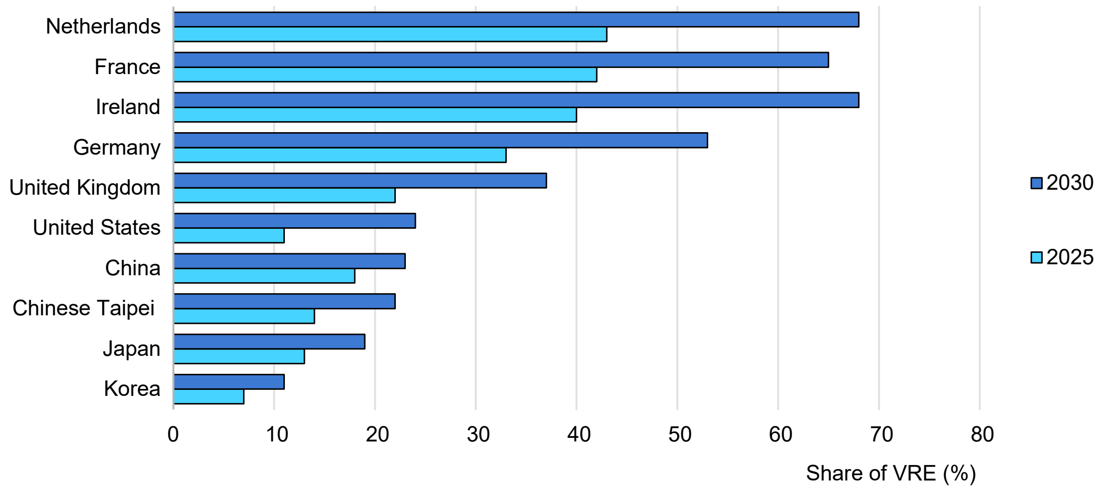
East Asia has a variety of low-emissions electricity sources to power its data centres

While East Asian countries – China, Korea, Japan and Chinese Taipei – are rapidly expanding their data centre capacity, their renewable energy share remains relatively low compared with Europe and the United States. This poses challenges for companies within the region to procure sufficient renewable electricity to meet their clean energy targets.

Compared with markets in Europe, East Asia shows a relatively lower share of solar and wind in its energy mix. In 2025, China had the region's highest share of variable renewable energy (VRE) (solar PV and wind) in its electricity generation mix at around 18% (a similar level to the United Kingdom), a level that is projected to grow to 23% by 2030, driven primarily by solar PV. Japan and Chinese Taipei currently have roughly similar VRE shares, but by 2030 Chinese Taipei is expected to surpass the 20% threshold, while Japan reaches around 19%. The share of VRE in electricity generation was 7% in Korea in 2025, and is expected rise to just over 10% by 2030.

Korea has the largest share of nuclear power in the generation mix in the region at above 30%, compared with 10% in Japan and 5% in China. Nuclear power generation in Korea is expected to grow further, contributing 37% in 2030. The role of nuclear power in Japan is also expected to increase as reactor restarts continue, growing its share from 10% in 2025 to 17% in 2030.

Share of VRE in energy mix in selected countries, 2025-2030



Source: IEA (2026), Electricity 2026.

The Korean government has recently updated its national renewables target, which aims to reach 100 GW of installed renewable capacity by 2030, a significant increase from its previous goal of around 70 GW.

Japan is also striving to increase renewables. The Mitsubishi Research Institute proposed that Japanese data centres run on local renewable energy, training AI models with surplus renewable electricity. AI training jobs can take weeks to run. Tooling to co-ordinate large training runs already uses checkpoints to save intermediate results for fault-tolerance. By leveraging this existing checkpoint mechanism, AI training jobs could be paused when the emissions intensity of the grid is high, to reduce emissions and costs. Additionally, parameters such as batch size and GPU power limits can be dynamically adjusted to reduce emissions by up to 75% with only moderate increases in overall training duration.

Inference can be performed on an end user's device to minimise latency or to protect privacy. Common examples include speech-to-text and optical character recognition. For a given AI model, regardless of whether the inference is completed in a data centre or on the end user's device, the same mathematical calculations are performed. Laptops and mobile phones use less energy than a typical server, although they are also less computationally powerful. In terms of efficiency, as a ratio of AI output to energy input, laptops and mobile phones typically use more power per calculation than servers. Data centres are highly optimised thanks to economies of scale. Virtualisation and multi-tenancy result in very high utilisation, which increases operational efficiency and amortises embedded emissions from hardware over more users. Mobile devices running off battery power have an energy overhead due to the round trip inefficiency of the battery. Therefore, it is unlikely that shifting inference to an end device will reduce total energy consumption for a given model. However, such a shift would decentralise the power consumption. Additionally, the size constraints of mobile devices may force AI developers to choose smaller models, which would reduce energy usage.

For inference without stringent latency and data sovereignty requirements, the workload could be shifted between multiple data centres around the world, based on where electricity is the cheapest or has the lowest emissions intensity. Google has built a carbon-intelligent computing platform to shift some of their non-urgent compute tasks to when the carbon intensity of the grid is lowest. It plans to expand this to shifting loads between data centres, effectively chasing the sun around the world, which could reduce emissions by 40%.

Globally, renewables currently provide 27% of data centres' electricity consumption, mostly from wind, solar PV and hydropower generation. The electricity generated to power data centres is expected to more than double, from 460 TWh in 2024 to over 1 000 TWh in 2030. The share of renewables is also projected to grow significantly by 2030, reaching nearly half of the additional power supplied to data centres, followed by gas and coal. Nuclear power also takes off as one of the major power sources for data centres in 2030. Some data centre demand growth is met with coal- and gas-fired generation. Robust growth in gas demand to power data centres in various regions represents an opportunity for East Asian manufacturers of gas turbines.

Data centre operators are increasingly diversifying their electricity procurement strategies

Electricity procurement strategies for data centres vary widely, shaped by factors such as the local energy mix, regulatory environment and technical configurations. For AI-driven facilities, where maintaining high uptime and continuous load are important, securing a reliable and cost-effective power supply is particularly important.

To achieve this, operators are increasingly diversifying their procurement strategies by combining different options, rather than relying on a single solution. In evaluating various strategies, decisions are typically based on three strategic dimensions: the type of contracts, the temporal accounting methods, and the level of physical reliability. This multidimensional approach enables data centres to balance sustainability, cost and operational resilience, while managing their rapidly growing electricity demand.

Temporal accounting: Annual vs hourly matching

Defining how to account for data centre consumption of low-emissions electricity is another key factor in building sustainable procurement strategies. Renewable electricity generation is inherently variable, which poses challenges for data centres that require an almost constant amount of power with no interruption.

While many major technology firms pledge 100% renewable energy, it is not possible to control or measure which particular generator powers a particular load. Instead, consumers can procure a quantity of renewable energy that matches what they consume. Commonly, companies that sign power purchase agreements (PPAs) can claim year-round renewable usage when their total annual electricity consumption matches the contracted renewable volume. However, the real-time power output of VRE, especially any individual generator, varies throughout each day. The spatial and temporal separation of renewable generation from data centre electricity consumption can necessitate reliance on alternative generation sources, including natural gas or coal. Consequently, the physical electricity mix may diverge from the procured, or "financial", electricity mix.

By contrast, hourly matching entails purchasing renewable power with a profile that matches or exceeds a consumer's consumption for every hour of the year. In the case of hourly matching, data centres may sign PPAs with an individual VRE generator whose output may vary substantially throughout each day due to local weather, maintenance outages and network constraints. Diversifying across many sites and technologies can help smooth out this variation.

The requirement to match consumption with renewable generation hourly requires consideration of design choices that can affect the daily profile of a generator. For example, single-axis solar involves mounting panels on motorised axles that can change the orientation of the panels to follow the sun throughout the day. This is more expensive per megawatt of capacity than fixed-tilt solar panels, due to the cost of the motor and moving parts. However, by moving the panels to follow the sun, single-axis solar produces more in the early morning and late afternoon than standard fixed-tilt, thus flattening the power generation curve to more closely match a data centre's load profile. Vertical solar panels can complement conventional fixed-tilt solar panels, flattening the curve further. When choosing a wind farm with which to sign a PPA, a data centre could select a wind farm in a location with less wind overall, but in a location where the wind is uncorrelated to existing wind farms in its portfolio. The requirement to achieve hourly matching can help drive these decisions to build the generation infrastructure of tomorrow in a way that addresses the most difficult gaps in renewable integration.

Hourly matching may require special attention from data centres because of their short-term demand inelasticity. Data centres have stringent uptime targets, meaning that most workloads cannot be cancelled or delayed to curtail consumption to match a VRE portfolio. Google's carbon-intelligent computing platform is an example of shifting non-urgent server load throughout the day to follow VRE output.

Many data centre companies are interested in nuclear power to achieve their hourly matching. Compared to VRE, nuclear power generators have an output that more closely matches the load profile of a data centre. The use of nuclear power in corporate emissions reduction and cost hedging goals is contingent on a welcoming regulatory environment and social acceptance. This may be plausible in countries such as Korea, but less so in other markets with prohibitions on nuclear power or differing community sentiment.

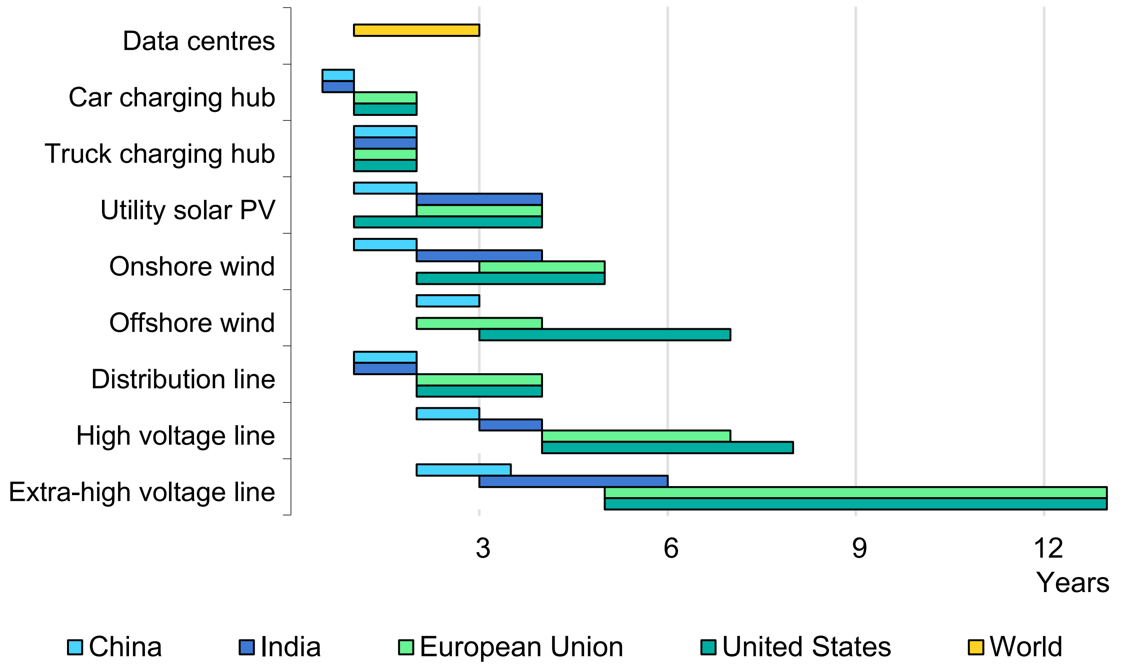
One important aspect of hourly matching is the lack of real-time emissions and load data. Accurate and time-specific data are essential to facilitate hourly matching, which aims to ensure that low-emissions electricity supply fully matches data centres' round-the-clock load profile. Some independent system operators (ISOs), such as the United Kingdom's NESO, offer near-real-time carbon intensity data and forecasts; however, most do not. Granular carbon intensity calculations can be challenging due to complexities related to imports and exports, as well as storage. Improving transparency and data availability in this space can yield substantial benefits. Once data centres have obtained granular grid emissions intensity estimates, they would need to develop data pipelines to ingest and process their own electricity consumption data in real-time. This can be complicated by the regulatory and IT boundaries between meter providers, retailers, distributors and consumers. The Green Software Foundation is an example of a collaboration to improve the transparency, accuracy and usability of grid and cloud data.

Timespan challenges in co-ordinated planning between data centres and grids

In the AI industry, companies compete by developing and deploying new models as quickly as possible. However, securing sufficient electricity and a grid connection on time is becoming more challenging for data centres, primarily due to the lag between their surging demand and transmission network buildout.

Planning, permitting and completing new grid infrastructure can often take between 5 and 13 years, whereas project deployment on the supply side and demand side can be much faster, at 1-5 years for renewables projects, 1-3 years for data centres, and 1-2 years for EV charging infrastructure. This mismatch in delivery timelines is leading to growing connection queues globally for generation and supply. Additionally, network and data centre infrastructure have different depreciation rates, and forward-looking demand forecasts, influenced by near-term spikes in connection requests, may not materialise with the desired permanence. Network operators may adopt a prudent approach towards capital-intensive upgrades, recognising the structural volatility in the technology sector, and cognisant of the potential for asset stranding.

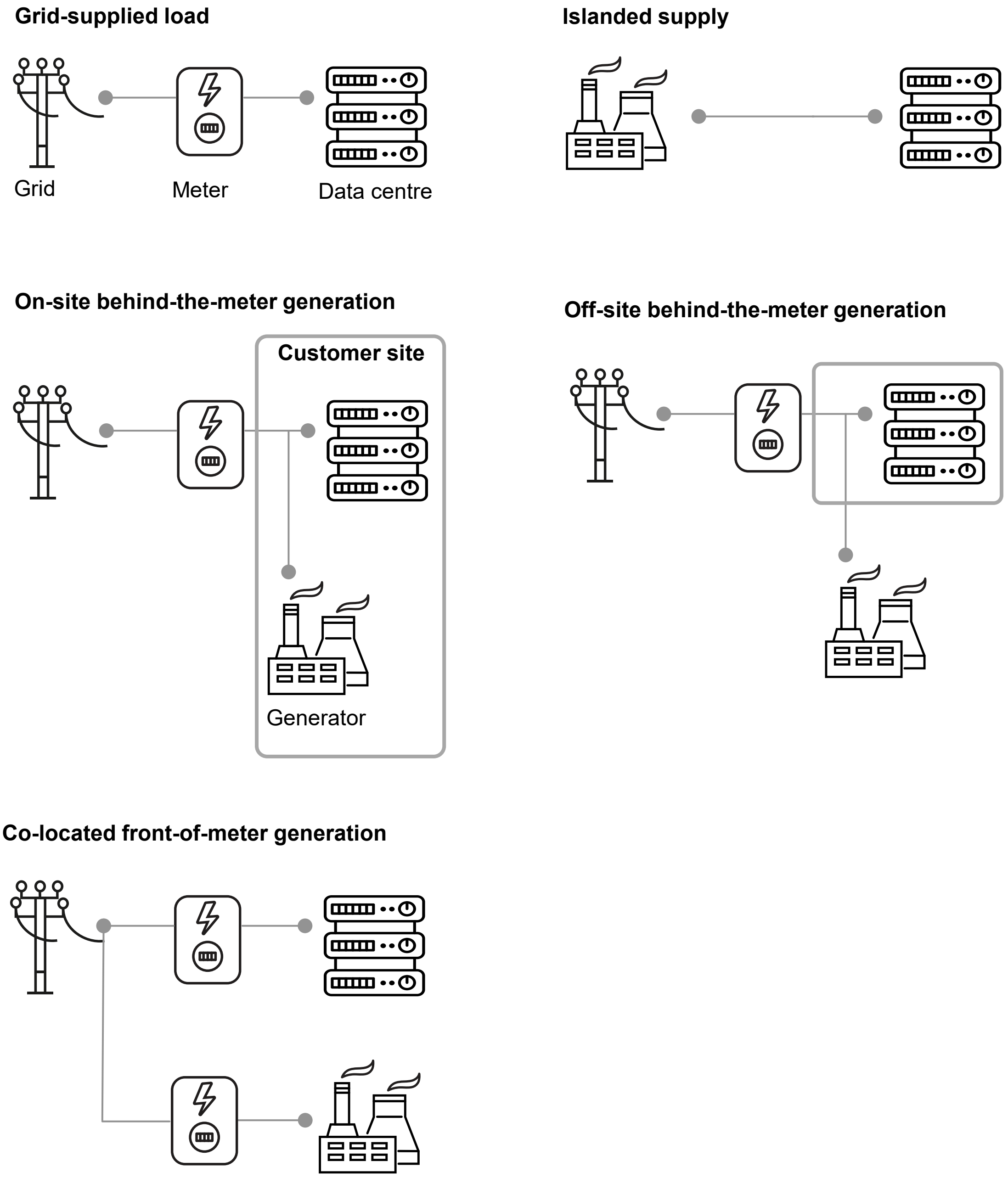
Typical project deployment time of electricity grids, solar PV, wind and EV charging stations



Notes: Ranges reflect typical projects commissioned during 2021-2023. Distribution line = 1-36 kV overhead line. Transmission is split between high-voltage line = 36-220 kV overhead line, and extra-high-voltage line = 220-765 kV overhead line. To date, India has not developed offshore wind projects.
Source: IEA (2023), Electricity Grids and Secure Energy Transitions.

As a result, in several regions of the world there is increasing interest among data centres in opting for **behind-the-meter** (BTM) solutions to ensure timely, reliable and adequate power supply. BTM assets are connected to the consumer-side meter, serving load directly rather than via the transmission or distribution network. Facilities using BTM assets are rarely fully off-grid, typically retaining a grid connection for redundancy. By leveraging BTM solutions in regions with long connection queues, data centres can effectively mitigate the uncertainty stemming from long grid connection queues, powering their load sooner by requesting less grid access. Data centre developments have faced increased community pushback in recent years due to a range of residents' concerns, including a perceived risk of increasing residential electricity bills or decreasing electricity reliability. Some projects are designed to be fully off-grid, to mitigate these concerns.

Power supply configurations for data centres relative to the customer meter



While many BTM assets are located on-site at data centres, the terms BTM and on-site generation are often conflated. On-site (co-location) generation can be still connected to the grid through its own meter, whereas some BTM assets may be off-site but directly linked to the customer's meter through private infrastructure. In some of the jurisdictions that have introduced "bring your own" (BYO) generator requirements for data centres, the generator is not required to be on-site. This section focuses on BTM generation assets, regardless of site location, if they operate behind the customer's meter.

For low-emissions BTM generation, data centres have commonly deployed on-site (or rooftop) solar PVs with batteries. For example, in China, Tencent's High Tech Cloud Data Centre is powered partially by solar PV. However, even with battery storage, on-site solar PV generally cannot fully meet the continuous, large-scale power demand for data centres. For rooftop PV specifically, data centre roof space is often occupied with heating, ventilation and air-conditioning (HVAC) equipment. Even for ground-based PV, the potential solar capacity of nearby land is generally significantly smaller than the demand from the data centre itself, due to the high spatial intensity of demand, and the fact that data centres tend to be installed in metropolitan areas where available land is less abundant than at typical off-site grid-scale solar projects. To increase the portion of demand met by BTM generation, a recent trend for data centres is the increasing amount of BTM projects that provide firm capacity at scale. The Stargate project, a series of major data centre development projects for OpenAI, is known to power its sites in Texas and New Mexico in the United States with gas turbines, connected via microgrids, coupled with solar PV and batteries. xAI's Colossus 2 data centres are also powering their data centres with off-site gas power plants, connected via private transmission infrastructure.

While BTM generation can accelerate the speed of data centre deployment, it may bring with it various challenges. The data centre operator may want to remain connected to the grid for redundancy, to manage planned turbine maintenance outages and unplanned failures. In such cases, the challenges of obtaining grid connection cannot be completely avoided. Additionally, local residents may object to increased air pollution from on-site combustion. From the grid operator's perspective, where the data centre has BTM generation and a grid connection, the reduced visibility of the generation asset can complicate certain processes such as balancing, forecasting and contingency planning. It can also complicate resource planning as well as regulatory aspects such as the allocation of network charges. Some jurisdictions mandate that generators over certain size thresholds must connect to the grid and participate in central dispatch or bidding. Furthermore, regulators may require retaining interconnectedness and integrated operation for reasons such as enhanced overall reliability, cost efficiency and emissions efficiency through shared contingency reserves and averaging out variations in supply and demand profiles.

New fuel sources to supply data centres' appetite for energy

When sourcing power from the grid, in addition to VRE, data centre operators are actively exploring several low-emissions generation sources that can provide firm dispatchable capacity, including nuclear power (particularly small modular reactors), hydropower and advanced geothermal technologies.

Nuclear power, specifically **small modular reactor** (SMR) technology, is being actively considered as a low-emissions firm power source for AI data centres. Its steady load profile aligns well with data centres' flat load profile, placing nuclear at the nexus between reliability and decarbonisation. Since 2024, several AI-related SMR projects have been announced, mostly in the United States and Europe, though most remain in the early stages of development. Given that SMRs are only expected to be commercialised by 2030, the reopening of retired nuclear power plants is a further option seeing interest in various countries such as the United States.

In 2025, several SMR projects for data centres, mostly in the United States and Europe, selected sites, scaled up planned capacity and extended PPAs to longer time horizons. In the United States, restarts of retired nuclear plants are expected within the next few years, potentially as early as 2027. The reopening of the Three Mile Island Unit 1 in Pennsylvania has been expedited to 2027, and the Duane Arnold nuclear power plant in Iowa has recently received its grid connection permit, with plans to secure PPA deals with regional AI data centres. However, most new SMR projects are not expected to come online until the 2030s, resulting in PPA contracts lasting until the 2050s.

Korea and Japan have also recently decided to extend the lifetime of existing nuclear power plants to meet growing electricity demand, to which data centres are contributing. In Korea, the regulator recently made a final decision to extend the lifespan of the Kori 2 nuclear reactor by 2033, which began operating in 1983. The Kori 2 reactor had stopped operating in 2023 as its operation permit expired after 30 years of operation. The reactor has a capacity of 685 MW_e. Japan has recently made a similar decision on reviving nuclear generation. The government has passed a law that allows ageing nuclear reactors to continue operating even after the existing 60-year limit. Similarly, Tokyo Electric Power Company (TEPCO) is working towards the restart of two units at Kashiwazaki-Kariwa, the country's largest nuclear power plant with seven units with a total capacity of 8 GW.

Geothermal is gaining traction as a low-emissions technology to provide renewable, dispatchable power for data centres. Major technology firms like Google, Microsoft and Meta are actively pursuing geothermal for 24/7 low-emissions power. Along with existing traditional technology, advanced geothermal is also becoming one of the viable options for data centres, tapping into the drilling techniques that were previously used in oil and gas projects. In 2023, Google became the first company to launch an advanced geothermal power plant, located in Nevada. Announced geothermal capacity dedicated to data centres exceeded 5.5 GW globally in 2025, mainly in the United States, with projects also in Chinese Taipei, Indonesia and Kenya. In East Asia, Google recently signed a 10 MW geothermal PPA in Chinese Taipei. Growth is expected in Japan and Kenya by 2030, supported by next-generation geothermal technologies.

Hydropower is also another firm dispatchable low-emissions source suitable for supplying data centres. Google secured a record 3 GW of hydropower under a

20-year PPA, while smaller deals, such as Chelan PUD's 18 MW contract, are anticipated. Norway is receiving attention for hydropower projects for data centres due to its abundance of hydropower, attracting AI developers for both power and cooling benefits. Recently, three AI companies – Microsoft, Nscale and Aker – have created a consortium to develop a hydropowered data centre complex in Narvik, Norway. Open AI has also announced its first European data centre initiative, Stargate Norway, in the same city of Narvik, for a 500 MW computing campus, operating from early 2026.

Power purchase agreements are increasingly popular for corporate renewable energy procurement in Korea

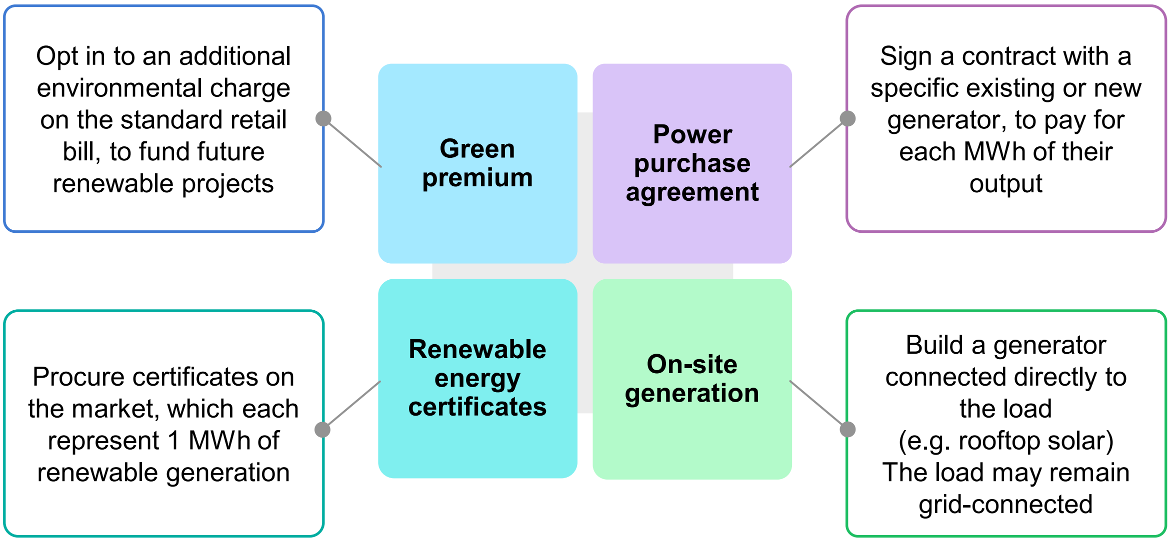
As the energy demand from the data centre sector grows, reducing emissions is becoming increasingly important. By supporting energy-intensive companies achieve their corporate climate targets through renewable energy procurement options, policy makers can help unlock capital to support the energy transition. Additionally, renewable energy procurement approaches can help energy-intensive sectors, such as data centres and semiconductor manufacturing, hedge against long term electricity price risk, and maintain industrial competitiveness in the context of international carbon border adjustment mechanisms. Power Purchase Agreements (PPAs) are one approach for helping companies achieve these objectives.

In Korea, corporate demand for renewable energy is growing. According to a Solution For Our Climate survey of 585 companies in the Korea RE100 Consortium,[7] more than half of respondents identified sustainability goals as their main reason for procuring renewable energy. Additionally, approximately one-quarter identified long-term cost management as a motivating factor, and one-quarter[8] are responding to international carbon border adjustment mechanisms to maintain international competitiveness.

Unlike most other goods, a characteristic of large, interconnected power systems is that the physical flow of electrons cannot be easily traced from a particular generator to a particular consumer. With the exception of on-site generation, renewable procurement methods are financial and accounting processes designed to incentivise the construction of carbon-abating renewables, and to establish a verifiable linkage between renewable generators and consumers.

[7] RE100 is an initiative to standardise the accounting of corporate climate promises by large energy users, to incentivise improved reporting, and to advocate the removal of obstacles to achieving high corporate decarbonisation targets.
[8] Respondents were allowed to select multiple answers, so the results add up to more than 100%.

The four main approaches to corporate renewable energy procurement in Korea



The type of electricity procurement contract is a key determinant in managing financial risk. In Korea, there are four principal channels through which large consumers can procure renewable electricity.

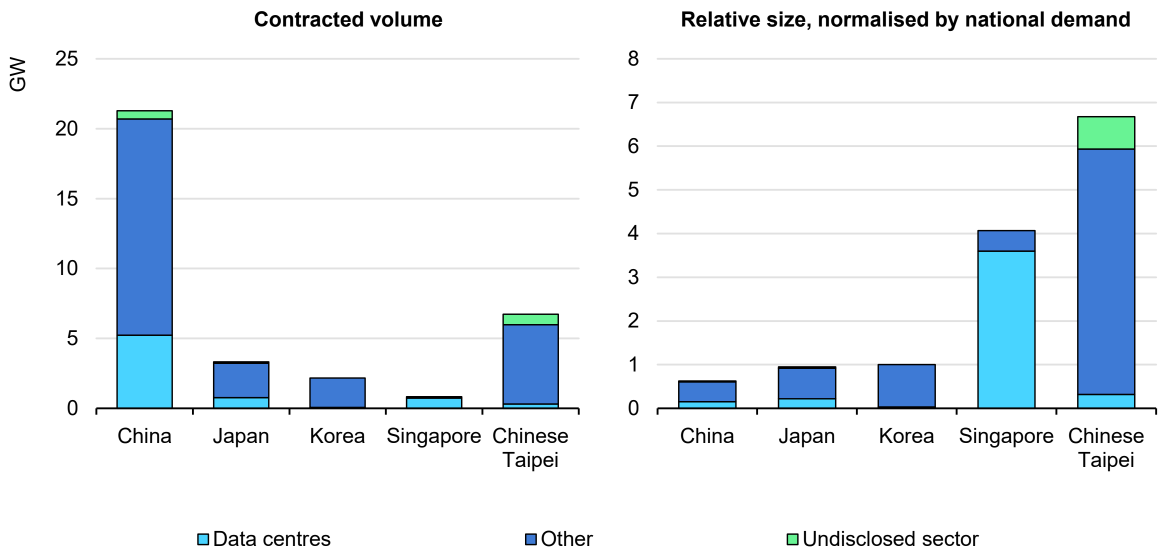
The **Green Premium** is the simplest and most popular approach to renewable energy procurement in Korea. Consumers continue to procure their energy from Korea Electric Power Corporation (KEPCO), and they opt in to a small additional green charge to allow them to claim credit for renewable energy consumption. This charge is in addition to the climate and environmental charge that all customers pay. The funds are passed to the Korean Energy Agency (KEA) to invest in new renewable capacity. In this way, the charge is used to fund future renewable energy and does not necessary correspond to a volume that matches the consumed volume. This is the most widely used procurement approach, reported to account for 98% of renewable purchases in 2024, due to its simplicity and low price, reported at around KRW 10-15 /kWh in recent auction rounds. The price is set through a competitive tender.

Power purchase agreements (PPAs) are an increasingly popular procurement approach where consumers sign contracts to procure power from a particular generator. Globally, PPAs are a common contract type used by data centres. Data centres and other large loads can procure low-emissions electricity reliably for a longer time frame, often more than a decade. PPAs can also effectively provide price stability with long-term cost certainty.

In Korea, there are two main types of PPAs, **third-party trilateral PPAs**, introduced in 2021, and **direct bilateral PPAs**, introduced later in 2021, with implementing details effective in 2022. Due to Korea's unique regulatory environment, both of these PPA structures differ from PPAs in other regions.

The third-party trilateral PPA was legalised first, as a structure that allows electricity procurement while complying with market regulations that restrict end users from directly purchasing from generators. Amid rising corporate demand for large-capacity PPAs, including from data centres, Korea's electricity market regulator (KOREC) made amendments to the Electricity Utility Act and related implementing rules, easing restrictions on direct bilateral PPAs. PPAs are less common in Korea than in other countries, possibly because they were only introduced in 2021. Amongst RE100 companies in Korea in 2024, less than 2% of corporate renewable energy procurement was with PPAs, a relatively small share compared with many other regions globally.

Renewable corporate PPAs by region and offtaker sector, 2018-2026



Note: Contracts signed between 1st January 2018 and 28 February 2026 are included. Data includes only deals with publicly available information, so may underrepresent the market. Normalised values adjust for each region's size, by dividing the capacity of PPA contracts (GW) by the region's annual consumption (TWh), then scaling to make Korea's value equal to 1. The data centres category includes data related services.
Source: IEA Analysis based data from S&P Global Commodity Insights (2026)

Renewable Energy Certificates (RECs) are issued to renewable energy generators at a rate of one certificate for 1 MWh. The Renewable Energy Portfolio Standard (RPS) imposes a mandate on generation companies to provide a minimum proportion of their portfolio from renewable generation. They may do so by constructing their own renewable generators or by procuring RECs from another generator's portfolio. Large consumers may choose to purchase RECs to certify that all or some of their consumption is matched by an equivalent amount of renewable energy.

The remaining renewable procurement option for large consumers is **on-site generation**. Loads can install generation, such as rooftop solar panels, **on site**, to be self-consumed. The site typically remains grid connected, to allow any shortfall to be imported. Excess can be exported at the system marginal price (SMP), or, for sites smaller than 1 MW, can be sold at the retail price.

How power purchase agreements work in Korea

PPAs are increasingly being considered and used in Korea as energy-intensive consumers seek to achieve sustainability goals, retain international competitiveness in the face of carbon border adjustment mechanisms and hedge long-term input costs. In Korea, generators are generally not permitted to sign electricity contracts directly with consumers outside permitted PPA frameworks. Large customers wishing to procure renewable energy with PPAs generally do so via an intermediary. Consumers have a choice between two types of PPAs:

Third-party trilateral PPAs: Consumers sign a contract with the sole electricity retailer, KEPCO, which in turn signs a similar contract with the generator.

Direct bilateral PPAs : Consumers sign a contract with an intermediary other than KEPCO, which in turn signs a similar contract with a generator. Despite what the English translation of the contract name suggests, the load and generator typically do not sign a contract directly with one another, as a third-party intermediary is still involved. A key difference between the two types of PPA is that in the direct bilateral PPA the intermediary is not KEPCO. The intermediary and the generator cannot be the same legal entity, but may be legally related, such as through the creation of a special purpose company. While KEPCO is not involved in the trading of energy in such a PPA, it is still involved as the transmission and distribution network operator, receiving payments for network usage, typically from the generator, who passes the cost on to the consumer.

Both of these PPA types are classified as **physical** PPAs, operating alongside Korea's Cost-Based Pool (CBP). In regions with a mandatory gross pool, such as Australia's National Electricity Market, PPAs are typically **virtual**. Despite what the names may suggest, in both physical and virtual PPAs the flow of electrons is not routed from a particular generator to a particular load. The physical characteristics of large grids preclude such visibility. Instead, PPAs are a contractual overlay, to certify that the volume of consumed energy matches the volume of generated energy, on an annual, daily or perhaps hourly timescale. The difference between virtual and physical PPAs is about how the settlements are split between transactions within and outside the spot market.

In both the third-party trilateral PPA and the direct bilateral case, the physical flow of energy is identical, and the consumer pays an intermediary, who pays the generator. The key difference is whether that intermediary is KEPCO or a renewable energy provider. Additionally, there are differences relating to grid usage fees, credit risk allocation, minimum generation guarantees, and the settlement and transfer of RECs associated with over-generation.

In the third-party trilateral case, the consumer and generator sign an additional contract directly with each other to make the generator liable for damages if it fails to produce the agreed amount of energy. In the direct bilateral contract, the renewable energy provider may provide some credit support. With the third-party trilateral PPA, consumers are responsible for finding a suitable generator, whereas in the direct bilateral approach, the renewable energy provider offers such project discovery as part of its value proposition. There are also differences in network charge settlement.

In Korea, PPAs are generally "pay-as-produced", as opposed to fixed volume (baseload) contracts, "pay-as-consumed" and other shapes, which allocate supply volume risk differently according to different baselines for shortfall and overproduction accounting.

For a pay-as-produced contract, when the load consumes more power than the generator supplies, the shortfall must be paid for. In the direct bilateral PPA, renewable energy providers are not permitted to procure this shortfall energy from the spot market. Consumers must procure it themselves at the typical time-of-use (TOU) retail rate. (They are allowed to procure it from the spot market directly, but to date no consumer has done so.) If the generator's output exceeds the consumer's demand, under both a direct bilateral PPA and third-party trilateral PPA, the excess is typically sold on KPX at the SMP, and that revenue may be passed back to the consumer, although the details vary between contracts.

PPA contract prices (the strike price) tend to be derived from the sum of the expected average of the wholesale price (SMP), fees and a risk premium. In the Korean electricity market, retail tariffs have historically been set at a price point that is materially below the average of the SMP. This difference can reduce the economic appeal of PPAs for consumers from a cost perspective. Recent increases in retail tariffs have closed or perhaps reversed this gap.

According to regulations in Korea, PPAs must adhere to a range of restrictions, particularly for direct bilateral agreements. For example, while it is now possible for consumers to sign N:1 contracts[9] to procure renewable energy from multiple generators to increase volume and mitigate shape risk, a supplier under a 1:N contract is currently not permitted to share any remaining capacity with other customers. Direct bilateral PPA consumers are allowed to procure any shortfall from the spot market, but only if the PPA is for more than 10% of their total contracted volume. PPA contracts must be registered, with notice periods of at least one month. Generators and consumers are subject to minimum size

[9] N:1 refers to N number of generators and 1 consumer.

requirements for PPA contracts. This minimum size differs between the consumer (300 kW) and generator (1 MW). In contrast, there are no minimum size restrictions or registration requirements for PPAs in Japan.

Physical PPAs entail complexities in terms of managing the discrepancy between the generation profile and load profile. In virtual PPAs (also known as financial PPAs), generators and consumers initially sell and buy electricity as if they had no PPA, and then hedge spot price risk using a contract for difference with a pay-as-produced quantity. This style of contract is allowed in Korea, but is rare. Examples include Brite Energy Partners with LG Innotek in 2024, SK E&S with Amazon Web Services in 2023 and Bukchon Seomo Wind Partner with Amorepacific in 2022. Virtual PPAs come in many different volume profiles, such as pay-as-produced, baseload (flat/fixed) and shaped (pay-as-consumed). In some markets, such as Europe, there are several spot markets (e.g. day-ahead and intraday), and the differences between them should be considered by consumers choosing a style of PPA.

Recent policy developments in power purchase agreements

In February 2025, several rule changes were proposed in the Notice on Direct Power Transactions of Renewable Energy Electricity Providers. In October 2025, the responsibility for electricity policy and oversight of the electricity market regulator (KOREC) transitioned from the Ministry of Trade, Industry and Energy to the newly formed Ministry of Climate, Energy and Environment. The proposed rule changes include a PPA brokerage platform, to be operated by KEA, to provide a matching process between consumers and generators. At the time of writing this report, this platform is still being formulated. Proposals to reduce the minimum PPA size for generators from 1 MW to 500 kW have not yet been implemented. The minimum size for off-grid energy storage was reduced to 500 kW.

Energy-intensive industrial consumers seeking to procure renewable energy through PPAs must manage the discrepancy between the intraday profile of generation and load. Recent changes allow PPA consumers to average out generation over a monthly timespan before calculating discrepancies. For companies with renewable procurement targets based on monthly matching, as opposed to hourly, this may reduce the financial risk of managing intraday profile misalignment. Such averaging for settlement purposes may reduce the effectiveness of temporal price signals from the SMP by reducing the temporal precision of the signal.

Comparison of Korean power purchase agreements with other markets

Korea's overall electricity market structure and physical PPA arrangements have various unique characteristics. The vertically integrated role of KEPCO, as the sole retailer, distribution network, transmission network and majority generator, is unusual among liberalised electricity markets, where these roles are typically unbundled. Korea is planning to undertake a large number of market reforms over the coming few years, such as the introduction of a real-time two-way bidding market, which would make the electricity sector more similar to other liberalised markets.

Korea's electricity market operates as a Cost-Based Pool (CBP), using technical, centralised estimates of fixed and variable costs. This differs from bid-based pools such as in Europe, the United States, Australia and India, where generators place bids based on their individual, decentralised cost estimates and commercial bidding strategies. Korea operates under a single-buyer model, where KEPCO purchases all energy from generators and acts as the sole retailer to customers, except for direct bilateral PPAs, where KEPCO charges a fee to transport the electricity through its network without buying and reselling the energy. The Korean approach differs from optional pool models, such as in Europe, where a substantial fraction of served energy is not traded through the centralised markets, and generators perform self-dispatch. In Korea, dispatch is centrally managed. In this respect Korea is similar to Australia's National Electricity Market. However, in another aspect, Korean PPAs differ from PPAs in Australia's mandatory pool, because energy for the direct bilateral PPAs is traded through approved intermediaries other than KEPCO and KPX, whereas in Australia all energy contracted through PPAs must also be traded through the centralised spot market. In India there are also two kinds of intermediated PPAs, with the average contract price differing between the two based on the relative credit risk of the intermediary, although this is due to country-specific circumstances.

In countries with mandatory pools, such as Australia, New Zealand and most of North America, all energy that physically flows through the grid must be traded in the wholesale market at a centralised exchange. This provides the ISO with comprehensive visibility and centralises merit-based dispatch. Generators and loads are still able to sign virtual PPA contracts under such a framework. Virtual PPAs are still possible in other market structures such as in Korea and Japan, although they may be more contractually complex.

In Korea, PPA signatories must notify regulatory authorities of the signing of the contract, and of certain changes. This is similar to Poland, where the largest energy generation companies must report their physical PPAs to the President of the Energy Regulatory Office. In the Philippines, energy is purchased under power supply agreements, and networks must seek regulatory approval for each agreement, which is more stringent than Korea's notification requirements. In contrast, in Australia and the United Kingdom there is no centralised collection of PPA contract information, which is less stringent than in Korea. Korean consumers sign substantially fewer PPAs than many other countries, largely because the Green Premium has historically been at a more commercially appealing price point, and because PPAs were introduced more recently than in many other countries.

Examples of low-carbon PPAs signed in 2025 and 2026, for data centres, globally

Seller	Buyer	Fuel type	Size	Timing	Location
AES	Meta	Solar	650 MW	Not disclosed	Texas and Kansas, United States
Clearway Energy Group	Google	Carbon-free energy (more specific category not disclosed)	1.17 GW	Expected to come online in 2027, contracted for up to 20 years	Missouri, Texas and West Virginia, United States
Jera	Google	Solar	15 MW	20 years	Chiba Prefecture, Japan
NextEra	Meta	Solar and battery	2.5 GW	Expected to come online between 2026 and 2028	ERCOT, SPP and MISO, United States
RWE	Meta	Solar	200 MW	Construction started in late 2025	Texas, United States
RWE	Telehouse International Corporation of Europe	Offshore wind	630 MW	10 years	London, United Kingdom
Shizen	Microsoft	Solar	100 MW	Not disclosed	Kyushu and Chugoku, Japan
Talen Energy	Amazon	Nuclear	1.92 GW	To 2042, delivery by 2032	Pennsylvania, United States
TotalEnergies	Data4	Wind and solar	610 GWh	10 year duration, starting January 2026	Spain
TotalEnergies	Google	Solar	20 MW, 1 TWh	21 years	Kedah province, Malaysia

Notes: Some examples cover multiple deals. Differences in data coverage and specificity are due to differences in public disclosure by the involved parties. This table only covers nuclear and renewable generation. Recent data centre energy procurement trends include thermal generation through PPAs and other procurement methods. These are out of scope of this table.

Connecting data centres to the grid

Power grids are becoming a major bottleneck in the energy transition. Increasing numbers of clean energy projects are entering the queues, waiting for months, and even years, to be connected to the grid. As briefly mentioned in the first chapter, IEA analysis identified that at least 1 700 GW of renewable projects (solar PV, wind, hydropower and bioenergy) at advanced stages are currently queueing for grid connections globally.

This challenge is also extending to data centres, as surging grid connection requests from data centres add further strain to power systems. A sizable proportion of the capacity in these queues relates to projects that are unlikely to materialise or are even duplicates as developers tend to apply for the same project in different jurisdictions. This complicates the process of estimating the increase in demand from data centres in the coming years. IEA analysis indicates that at least 150 GW of data centre projects are in the advanced stages of connection queues globally.

As a first step, it is important to distinguish between data centre projects that are more likely to advance versus those that act more as "placeholders" or "phantom projects" blocking connection queues. The high regional concentration of data centres is another aspect that requires attention. While accelerating grid expansion is essential, being able to locate new data centre connections in locations with less saturated grid conditions can provide a timely solution. In any case, effective system integration of data centres is crucial for a reliable operation of the whole system. This requires the consideration and monitoring of certain technical aspects before the data centres are connected as well as during their operation.

Community engagement is essential for timely development of infrastructure

While building additional grids is vital to connect both more renewable energy capacity and data centres, building new grids takes a significant amount of time. Local opposition can cause additional delays, as experienced in certain countries. For example, in the case of Korea, one of the major challenges in grid expansion is strong community opposition, especially to new overhead transmission lines, which often causes significant permitting delays for rural siting. Over the past decade, the average delay for major transmission projects has been around 70 months, nearly six years, primarily due to local resistance.

Examples of major cases of grid construction delay in Korea

Capacity	Project	Added delay (months)
500 kV	Eastcoast-Shingaphyeong HVDC	88
345 kV	Bukdangjin-Shintangjeong	137
345 kV	Dangjin Thermal Power Plant-Shinsongsan Transmission Line	78
345 kV	Shin-Dangjin - Buk-Dangjin Underground Transmission Line	54
345 kV	Godeok-Seosanseong Transmission Line	13
345 kV	Shin-Siheung - Shin-songdong Underground Transmission Line	59
345 kV	Shin-Jangseong Substation	62

Note: HVDC = high-voltage direct-current.
Source: IEEFA (2025), Bottlenecks to renewable energy integration in South Korea.

In 2025, the Electricity Network Expansion Bill was introduced, as part of the Energy Trifecta Bill, to help streamline community approval processes. This bill reduced the local consultation period to 60 days, to help provide more certainty upfront about project timelines. Additionally, the bill enables KEPCO to offer more compensation to residents when buying their land for transmission projects.

Addressing "phantom" data centre projects can help unlock grid connection queues

As grid connection queues grow in many regions, especially for data centres, policy makers are exploring options to accelerate new connections for generators, storage and loads. Efforts to debottleneck the head of the queue are becoming common. Additionally, connection application throughput can be improved by addressing the so-called "phantom projects" that are applying for new grid capacity without realistic prospect of being connected. These projects typically emerge because of speculative behaviour seeking to secure strategic locations for future real estate development of data centres. This is also closely linked to current contractual arrangements: once grid capacity is approved it is reserved for the applicant, even if unused, and can prevent or delay reallocation to other customers.

The connection of data centres to the power grid is a complex process which involves an application process, contractual agreement with the transmission system operator (TSO) and integration of new infrastructure. In the application process, data centres may overestimate or inflate their capacity predictions. This may be due to factoring in a contingency, unrealistic expectations and uncertainty over future demand. Such uncertainty could lead to the increase in these "phantom projects".

A key barrier to tackling this issue is the lack of transparency in grid connection data. In some jurisdictions in the United States and a few countries with advanced databases in system operation, such as Australia and the United Kingdom, preliminary data are presented on the speculative viability of data centre projects within the queue. However, the classification standards for determining development prospects vary widely and remain unclear in many cases. As a result, the ratio of initial grid applications to realistically feasible projects is around 20% across the United States and Australia. According to KEPCO, the share of realistically feasible data centre projects in Korea is around 70% of grid connection applications.

Regulatory toolkits can help with clearing the queues

Various policy measures and regulatory frameworks can be helpful in clearing grid connection queues and maximising available capacity. A range of countries are moving from incentive-based policies towards obligation-driven policies for data centre grid connections. Historically, many policy mixes relied on incentives, such as tax relief and expedited permitting. However, the growing strain on transmission systems and connection queues, driven by data centres, are prompting regulators to adopt more prescriptive measures.

These mechanisms vary by jurisdiction, but generally fall into three categories: **financial disadvantage** for speculative queue applications (e.g. higher queue application deposits and grid upgrade payments), **readiness-based assessment** that prioritises projects with proven feasibility, and **generation-load coupling** that ensures the large loads “bring their own generation” by aligning with available capacity. For example, in 2023 KEPCO introduced stricter screening requirements for large-load connection applications, requiring evidence of end use, project timelines and construction readiness to limit speculative capacity reservations. Such policies can tie expedited access to grid-supporting behaviours, such as accepting flexible operation, on-site storage and locational signals that steer large loads toward nodes with available grid capacity.

Such a global trend implies that governments are increasingly recognising data centres, especially AI-driven facilities, as significant stakeholders in future power systems and expect them to take responsibility for grid reliability, given their growing influence in grid operation. The increasing demand from data centres can lead to uncertainty for system operators in system planning and real-time operation.

Examples of grid-related policies and regulatory frameworks for data centres in selected regions and countries

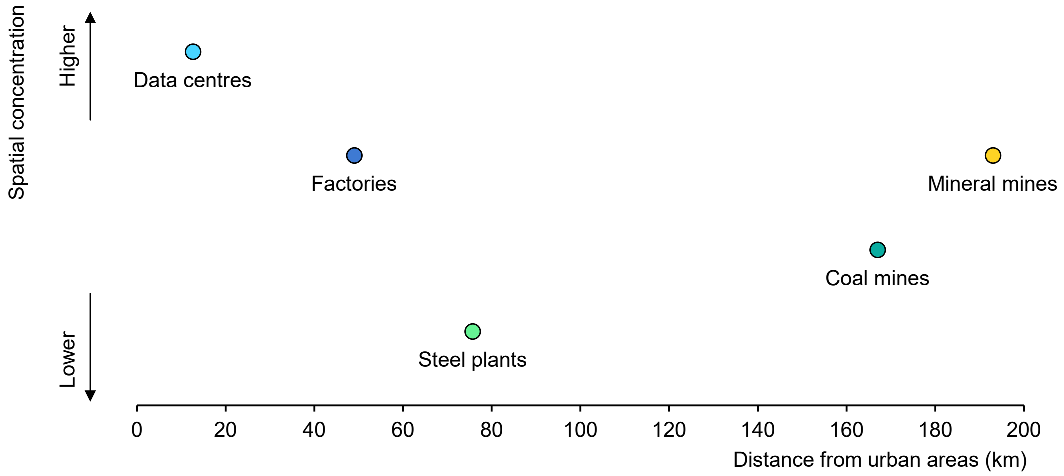
Market/ regulator (country)	Policy and regulation	Description
Department of Energy (United States)	Advanced Notice of Proposed Rulemaking (ANOPR) for the interconnection of large electricity loads	Loads that are 20+ MW and wish to be connected to transmission networks should be subject to: 1) Joint connection studies with large loads and generation assets 2) Standardised connection procedure 3) Expedited studies for the large loads that agree to be curtailed (up to 60 days) 4) Compliance and financial contribution to requested network upgrade Utilities serving large loads should provide ancillary service.
MISO (United States)	Definitive Planning Phase (DPP)	Large load customers should provide key evidence of development milestones to receive the final approval for grid connections Capacity cap on grid connection queue studies
NESO (United Kingdom)	Gate 2 to Whole Queue (G2TWQ)	Only projects meeting specific readiness and Strategic Alignment Criteria progress through the reformed queue: 1) Gate 1: projects that do not meet the Gate 2 criteria – no confirmed connection date 2) Gate 2: projects that meet the new requirements – can be confirmed with grid connection date, connection point and queue position
Ireland	Large Energy User connection policy	New data centres must provide dispatchable generation or storage matching import capacity, and source at least 80% of annual electricity demand from renewable generation
Korea	Special Act on Activation of Distributed Energy	Loads that are 10+ MW should pass the grid impact assessment evaluation to receive final approval for grid connection

Regional concentration of data centres brings its own challenges for grid management

Data centre concentration: Status and outlook

Data centres are commonly built in clusters, creating concentrations that can put significant strain on local grids due to their constant high loads. In some states of the United States, data centre clusters are taking up over 10% of electricity demand; in Ireland it is around 20%. Compared with other major large-load facilities, such as industrial plants, data centres are more spatially concentrated and located closer to urban areas, where electricity demand is already high due to dense populations. While hyperscale data centres for AI training are often located in more remote regions to meet large land requirements, many inference-driven or edge computing facilities prefer proximity to end users.

Spatial concentration and urban proximity of major large-load projects

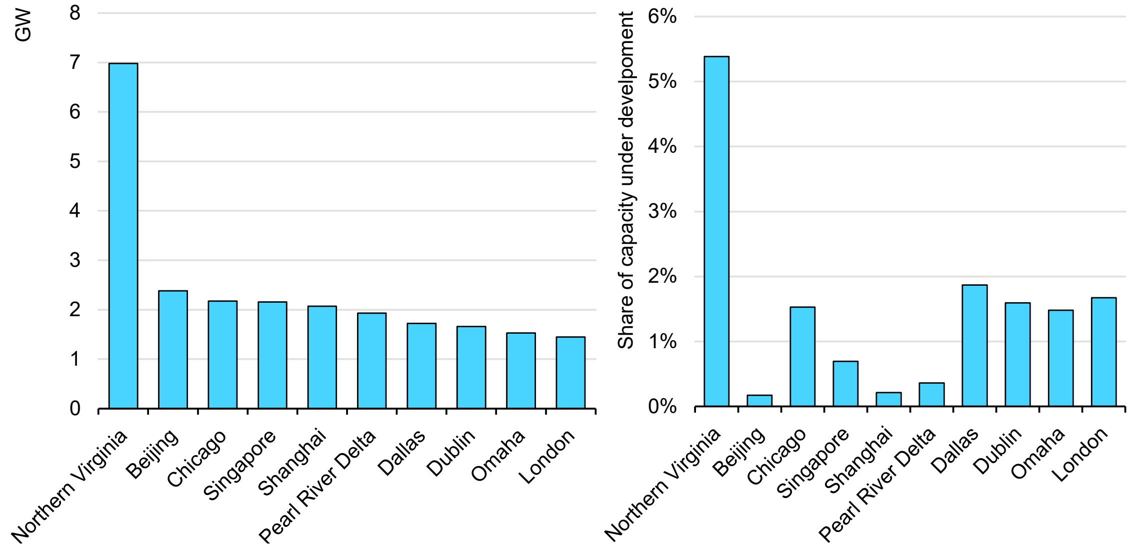


Source: IEA (2025), Energy and AI.

Locating in more remote, less saturated areas may allow data centres to secure more grid availability, although this may not always be the case and depends on grid availability and data centre load.

Globally, the United States, China and Europe are home to the most data centre-dense cities in the world. Over 15% of data centre projects in the pipeline are concentrated in the ten cities with the largest existing capacities. Northern Virginia in the United States leads globally, with the highest installed and developing capacity making it the most data centre-dense region worldwide. In East Asia, Beijing, Singapore and Shanghai rank among the top ten, each with around 2 GW of installed capacity.

Total installed data centre capacity and the share of global capacity under development, 2024

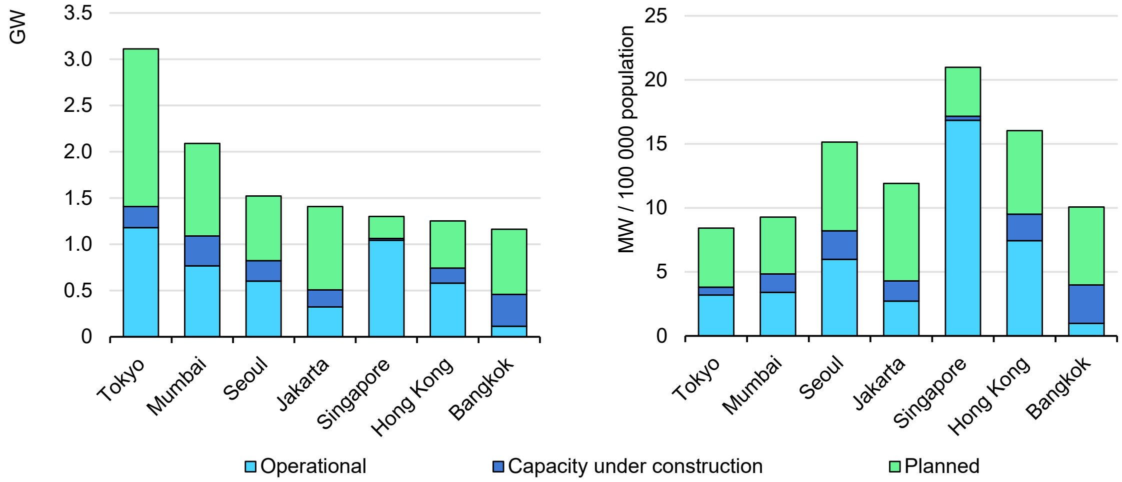


IEA. CC BY 4.0.

Note: Shares are the amount of data centre capacity under development in each region divided by the amount of data centre capacity under development globally. The Pearl River Delta encompasses the combined capacity of Guangzhou, Shenzhen and Hong Kong, China. The geographies considered represent the ten largest clusters in the world. Capacity under development is based on announced projects.
Source: IEA Analysis based on data from OMDIA (2025).

Comparing data centre capacity relative to city population provides another perspective on regional concentration. For every 100 000 people, Seoul hosts around 6 MW of operational data centre capacity, more than Tokyo, even though Tokyo has a larger data centre portfolio in absolute terms. Among the FLAP-D markets (Frankfurt, London, Amsterdam, Paris and Dublin), Frankfurt leads with 85 MW, while Paris ranks the lowest at 6 MW, the same as Seoul. In comparison, Virginia, home to the "data centre valley", has 125 MW per 100 000 people, while having over 30 GW in the pipeline.

Data centre capacity per 100 000 population in selected cities in Asia, 2025

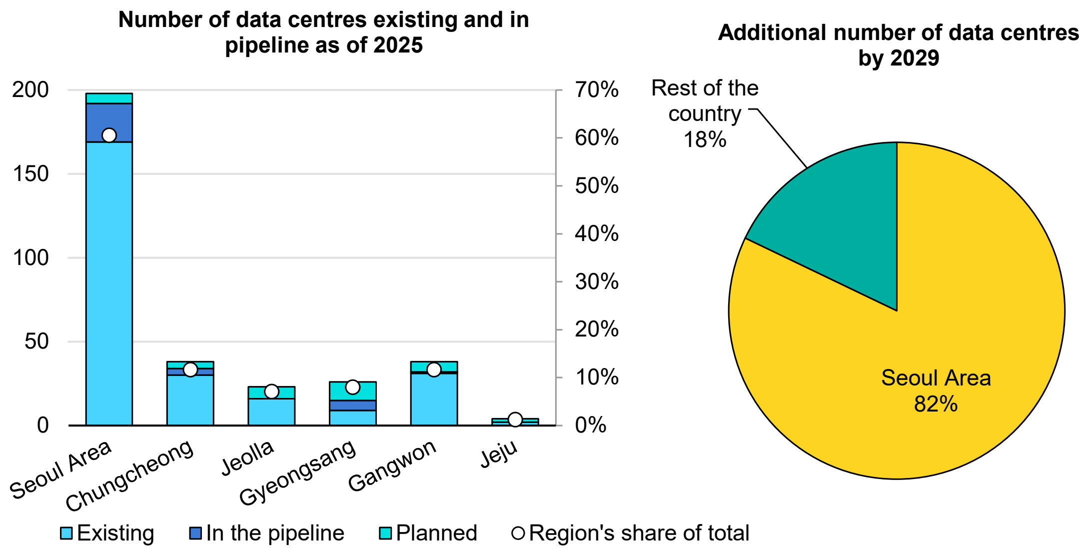


IEA. CC BY 4.0.

Notes: Only cities with publicly available data are shown.
Sources: IEA Analysis based on data from Cushman & Wakefield (2025), APAC Data Centre Update for data centre capacities breakdown; World Population Review (2025) for population statistics.

In Korea, the gap in data centre capacity between the Seoul Metropolitan Area and other regions is expected to widen. As of 2025, Seoul accounted for 60% of the country's data centres, with nearly 200 facilities. According to official Korean sources, an additional 732 data centres may be added nationwide by 2029, requiring almost 49 GW of electricity. Up to 82% of these new facilities and about 81% of the associated increase in electricity demand are projected to be concentrated in the Seoul Metropolitan Area. While some overlap may exist between the 2025 pipeline and the 2029 projections, the trend toward Seoul-concentrated data centre expansion is expected to persist if no appropriate policy measures are taken.

Regional distribution data centres in Korea, 2025-2029



Notes: Seoul Area refers to the Seoul Metropolitan Area. Sub-regions = Seoul Area (Seoul-si, Gyeonggi-do, Incheon-si); Jeju (Jeju-do); Chungcheong (Chungcheongnam-do, Chungcheongbuk-do, Sejong-si); Jeolla (Jeollanam-do, Jeollabuk-do, Gwangju-si); Gyeongsang (Gyeongsangbuk-do, Gyeongsangnam-do, Busan, Daegu, and Ulsan); Gangwon (Gangwon-do). Data for 2029 are forecasted values. Overlap between 2025 pipeline and 2029 projects may exist.

Sources: IEA Analysis based on data from Ministry of Trade, Industry and Energy (2023), Measures to alleviate the concentration of data centres in the Seoul Metropolitan Area; Korea Data Center Council (2025), 2024 Korea Datacenter Market Report.

Reasons for the regional concentration of data centres

Data centres often cluster near major hubs to ensure reliable access to power infrastructure, including generation assets and transmission networks, while minimising latency and maintaining strong connectivity. Proximity to urban areas also helps operators to tap into skilled talent pools. However, being close to urban areas and regional concentration are not always prerequisites for data centre deployment. In fact, locating facilities closer to renewable generation sites can offer significant benefits, such as avoiding grid connection delays and easier procurement of renewable electricity.

For example, in Korea renewable integration is constrained by limited transmission capacity and curtailment-prone nodes. Strategically placing flexible or power-intensive data centres near such regions with better local grid capacity can benefit both data centres and the local grid. For data centres, they can procure abundant renewable resources, especially in the southwestern Jeolla area with its growing solar fleet, and they can absorb surplus output, which enables local grids to better mitigate grid congestion and reduce curtailment.

As flagship data centres continue to grow in scale beyond 1 GW, technology companies could start to distribute model training across multiple data centres, similar to how other cloud services are horizontally scaled. Such distributed training introduces challenges and costs. This may require substantial investment in new, dedicated fibre connections between data centres. However, the cost of such cable laying is likely to be dwarfed by the cost of each data centre.

Some data centre workloads, such as e-commerce and video gaming, have strict latency requirements. However, generative AI workloads may not have the same constraints. Large language model text responses take several seconds, or even tens of seconds to generate, particularly when chained with tools such as web search. Image and video generation may take minutes. In comparison, shifting a data centre from Seoul to the relatively renewables-heavy Jeolla region in the south of Korea would increase latency to users in Seoul by less than 10 milliseconds (ms).[10] Therefore, a unique characteristic of AI loads is that they have the potential to be shifted further from end users without a noticeable impact on user experience.

The system integration of data centres requires attention

Once connected to the grid, large data centres may have increasing operational impacts on electricity system operation. The integration of modern AI data centres into power systems poses unique challenges compared to traditional compute workloads, such as relatively fast and large steps in aggregate load.

In a survey by the Electric Power Research Institute, several network utilities reported issues related to data centres involving thermal violations (22%), voltage violations (17%), harmonic concerns (9%), fault ride-through issues (9%) and ramp rate issues (26%).

As data centres grow, the unexpected loss of their load from the grid can become noteworthy for power grid operators and planners. Data centre operators have stringent reliability targets. Disturbances on the power grid not caused by data centres, such as voltage dips, may result in data centres switching to backup power with their UPS. This removes a large load from the grid. In turn, this may

[10] Assuming a distance of 400 km, propagation speed of 5 μs/km, and 5 additional hops of 1 ms each.

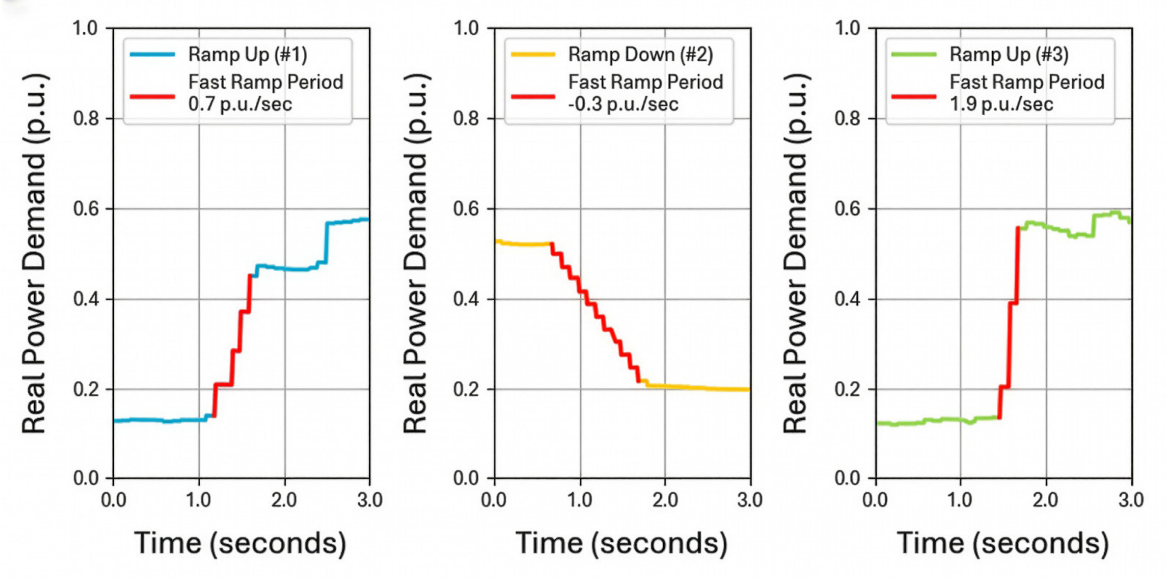
cause subsequent changes in the grid voltage or frequency, particularly if many data centres react simultaneously. These changes may induce additional data centres to disconnect, in a feedback loop. The North American Electric Reliability Corporation (NERC) has reported an incident where a transmission line failed for reasons unrelated to data centres; then data centres responded to the disturbance in a way that removed 1.5 GW of load from the grid. This incident did not pose a reliability risk; however, grid operators should be aware of this possibility to ensure that such load losses do not reach unmanageable levels. Many power systems, such as Ireland's, are investigating and implementing improvements to standards and monitoring for ride-through and related parameters in data centre connections.

AI workloads differ from traditional compute in both scale and behaviour. High-performance compute infrastructure is capital-intensive, so hyperscalers maximise utilisation of general-purpose data centres by diversifying their workload across many customers. In contrast, large AI data centres may have only one end user, which may result in more correlated demand across servers and more pronounced variation of aggregate load.

Large AI training jobs present a demand profile that can be challenging to smooth out across time. During data transfers between batches and during checkpointing, GPU load can dip as networking becomes the bottleneck. This leads to rapid fluctuations in power draw. Meta reports that during the training of its Llama 3 model, power consumption in its data centre varied rapidly by tens of megawatts in less than one second. Google reports that its load can ramp by 15 MW in less than one second. NERC has observed a reduction of 410 MW in 36 seconds. On the supply side, it is difficult for generators to ramp at a comparable speed. As data centres and the size of flagship model training jobs grow, grid planners and data centre operators should give consideration to a potential future relationship between operational step changes in AI data centre load and deviations of grid frequency outside acceptable bands.

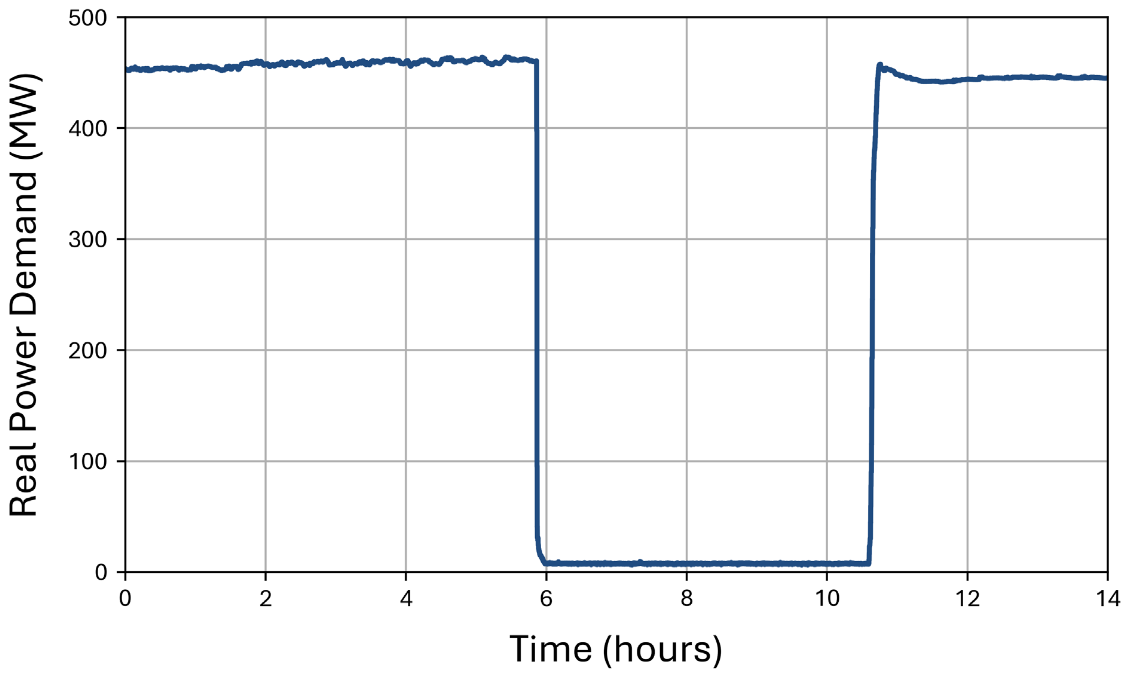
It is important that policy makers and grid planners put in place frameworks to incentivise data centres to implement solutions to smooth out such variations. On the software side, AI trainers can use dummy workloads to fill in lulls in the power consumption of individual GPUs. On the hardware side, data centres could invest in a hierarchy of electrolytic capacitors, supercapacitors, batteries and synchronous condensers as responsive buffers. Policy levers can help drive this investment.

Load behaviour in AI model training



Source: Reproduced from North American Electric Reliability Corporation (July 2025), Characteristics and Risks of Emerging Large Loads: Large Loads Task Force White Paper.

Example of load fluctuation curve in a modern data centre



Note: In this example, the data centre ramped down from 450 MW to 40 MW within 36 seconds, then later ramped up within several minutes.

Source: Reproduced from North American Electric Reliability Corporation (July 2025), Characteristics and Risks of Emerging Large Loads: Large Loads Task Force White Paper.

Chapter 3. High-level policy recommendations

This chapter provides high-level policy recommendations on how to maximise the potential of the collaborative relationship between AI and energy. The recommendations are presented in two parts: the first, on the question of AI for energy, puts forward policies on how AI uptake in the energy sector can be achieved in an effective way; the second, on the question of energy for AI, addresses improving the energy efficiency, energy supply and grid connection of data centres.

Policy recommendations on AI for energy

Recognise that every country may have its own characteristics and priorities regarding AI for energy

Countries differ in their energy mix and policy targets. Applying AI in the energy sector would need to take these into account accordingly. For example, in countries that already have high shares of variable solar PV and wind in their electricity generation mix (e.g. Denmark and Germany in Europe, or China in East Asia), AI applications to increase system flexibility may have a higher priority to enable greater integration of more VRE on the system. These would include applications to improve forecasting, grid flexibility, balancing and demand response. By contrast, in countries with more limited land availability for siting renewables (e.g. the Netherlands in Europe, or Japan and Korea in East Asia), where the VRE share is rising, AI applications to assist optimal siting, accelerate permitting and unlock grid capacity may have a higher impact in the short term, followed by the previously mentioned aspects in the medium term.

East Asia is characterised by fragmented national power grids. Japan and Korea in particular function as power system islands. This characteristic, compared with well-interconnected power systems in Europe and the United States, can necessitate more tailor-made policies for the application of AI on the power system. The role of AI in supporting increasing power system flexibility deserves particular attention given the lack of cross-border electricity exchange.

Countries may also differ with regard to their existing digital infrastructure, which can influence the pace of AI uptake. Countries with robust digital infrastructure and data ecosystems can benefit more from real-time AI analytics. Conversely systems with limited connectivity may need to expand their digital infrastructure first to benefit from such real-time AI applications effectively. Additionally, countries may vary in their approach to data sovereignty, which could influence siting and procurement decisions on data centres and AI.

Accelerate digitalisation for AI adoption across the energy system

In a 2025 survey in Korea, 64% of people reported that they use generative AI for work or personal purposes, with an adoption rate nearly twice as fast as the internet during its initial rollout. Many corporations are exploring and utilising AI in business operations, particularly larger enterprises. The number of venture capital investments in the IT infrastructure and hosting industry, which includes data centres, has risen in East Asia in recent years.

In this context, policy makers should recognise that the effective deployment of AI applications depends on the availability of enabling digital infrastructure. Irrespective of the differences between country characteristics, accelerating digitalisation is essential for a fast-paced and effective uptake of AI across the energy system.

Investment frameworks should therefore prioritise the rollout of sensors, data platforms, analytics and control systems that allow the large-scale data collection and automation required for AI. Investing in digitalisation of grids, generation and end-use sectors is important to allow the fostering of AI applications in these sectors. Developing national data-sharing frameworks and standards can also help establish a data ecosystem with enhanced data security and privacy.

Governments can also play an important role in fostering AI innovation for energy optimisation by funding R&D and demonstration projects. Such projects can encompass grid management, predictive maintenance, demand response and energy efficiency, among other relevant fields of application. Enabling co-operation between energy companies, technology firms and research institutions will be essential in this regard.

Update regulatory and policy frameworks to enable AI deployment

Reviewing and adapting regulations to enable AI deployment can facilitate the effective uptake of AI across the energy sector. Establishing clear guidelines for data privacy, cybersecurity and ethical AI use in energy are essential for a robust framework for AI deployment. When doing so, it is important to avoid creating disproportionate compliance burdens that discourage innovation. Many actors in the energy sector already operate under comprehensive safety regulations. In this regard, integrating AI regulations into these existing frameworks rather than creating parallel structures may be more effective and should be considered.

Designing regulation that contains incentives to promote efficiency gains via AI is important to encourage investment in AI deployment. This is because the deployment of AI to solve energy challenges will also depend in part on the alignment of incentives. Uptake is likely to be strongest where the use of AI is in line with goals such as reducing operational costs, lowering emissions, increasing resilience and improving safety. Like all new technologies, regulation promoting AI opportunities should be designed with commensurate guardrails to manage the risks of premature deployment and unproven solutions, as well as consideration of liability frameworks, fit for high-stakes critical infrastructure.

Bolster dialogue and collaboration between policy makers, the technology sector and the energy industry

The application of AI in the energy sector touches upon multiple stakeholders, including policy makers, technology companies and actors from the energy industry such as utilities and grid operators. Therefore, it is essential to foster collaboration between these stakeholders. This is especially important as energy regulators and ministries may have limited overview of the fast-evolving technology landscape, whereas technology companies can lack information on the regulatory and safety practices in the energy landscape. Before policymaking, it may also be beneficial to have open public consultations to get feedback from a diverse set of stakeholders. Regulators can also provide certain voluntary guidance as well, such as those on ethical AI use in the energy sector.

Capitalise on the promising opportunity for synergies between AI and local manufacturing strengths

AI for energy efficiency can be positioned in East Asia as a pillar of industrial competitiveness, beyond its direct potential for reducing energy use. The region is a major producer of batteries and an essential supplier within the global semiconductor value chain. Efficiency improvements in these electricity-intensive industries can therefore scale rapidly and strengthen cost competitiveness. In power systems that are dense and have limited cross-border balancing, lower electricity intensity reduces pressure on constrained grids. Policy design can align industrial digitalisation support with energy objectives by linking incentives and demonstration initiatives to verified reductions in electricity intensity.

AI-enabled optimisation in semiconductor manufacturing offers a route to lower electricity intensity because energy demand in fabrication plants is strongly influenced by facility systems and the need for continuous, high-reliability operation. The IEEE International Roadmap for Devices and Systems notes that rising process complexity increases tool counts and cleanroom area, which pushes total energy demand upwards unless efficiency gains keep pace. AI applications such as predictive maintenance and advanced supervisory control for HVAC and cooling, including digital twin-based control, can reduce facility electricity use and avoid energy waste. Linking industrial support measures such as R&D funding, tax credits and accelerated depreciation to documented reductions in electricity use per unit of output can provide a transparent way to connect AI deployment to lower system-level electricity demand.

AI in battery manufacturing and battery operation can reduce production energy intensity while strengthening system flexibility. Recent manufacturing assessments show that drying processes and dry room conditioning account for a large share of energy use in cell production, which makes process optimisation an important lever for reducing electricity intensity. AI can help reduce energy per cell manufactured through tighter process control, improved quality assurance and optimised cell formation processes. Operational algorithms that enhance dispatch decisions, state of charge management and degradation monitoring can also improve the performance of storage fleets. These combined effects can reduce industrial electricity intensity and increase power system flexibility, reinforcing AI-enabled efficiency as a strategic advantage for East Asia's energy and manufacturing systems.

Policy recommendations on energy for AI

Incentivise energy efficiency in data centres

Data centres are projected to contribute 50% of increased electricity demand in the United States in the next five years. Within East Asia, data centre growth in China will be large in magnitude, and Japan and Korea are also expected to see significant power demand growth from data centres. Considering this rapid growth in demand, incentivising more energy-efficient data centres can lower their electricity consumption, helping to alleviate the strain on power systems as well as improving data centres' carbon footprint. For this purpose, setting minimum energy performance standards can help, where certain thresholds for power usage effectiveness (PUE) can be proposed. This could apply solely to new data centres or existing data centres too, depending on policymaking priorities. Incentivising best-practice cooling technologies and waste heat recovery is also essential. Using waste heat from data centres in district heating networks or industrial processes can improve overall system efficiency, particularly in urban and industrial zones where there is a strong match between heat demand and data centre location. Promoting energy audits and public reporting on energy and water use at data centres is another measure that can increase transparency in the sector and facilitate best practices.

Promote and enable clean energy procurement

Countries can establish policy frameworks that actively promote clean, reliable and efficient energy procurement for data centres, while aligning these facilities with broader decarbonisation and grid-resilience goals. In this regard, supporting the use of flexible, long-term power purchase agreements (PPAs) for renewable electricity to supply the power needs of data centres is essential. In markets where the power sector is dominated by a single actor and PPA markets are not yet very liquid, additional measures and regulatory adjustments may be necessary to facilitate a healthy PPA market with sufficient uptake. In some markets, navigating complex regulatory environments is difficult. Policy makers in such markets should prioritise simplifying and streamlining the rules and regulations relating to PPAs, as well as strengthening governance and clarifying decisions that may apply to some types of PPAs but not others. Rules governing virtual PPAs could provide more support, to give consumers more options to manage hourly profile differences. Such changes would enable data centres and other large loads to flexibly procure renewable energy from generators, thereby supporting industry and generation alike.

East Asia's energy-intensive industries, including data centres, have a growing appetite for renewable energy. While most East Asian countries already generate substantial low-carbon capacity in the form of nuclear power, some corporate climate objectives and procurement frameworks may exclude this source. For example, companies who wish to claim compliance with certain corporate initiatives, including RE100, may not be able to do so using electricity from nuclear generators. In the case that industrial demand for renewable power grows over the coming few years at a rate that exceeds the growth of renewable power supply in some East Asian countries, the risk of some companies not being able to achieve their renewable procurement goals may emerge. This may be detrimental to investment in industry, thereby affecting international competitiveness. Policy changes to increase the supply of renewable energy and regulatory enhancements for flexible PPAs are essential to alleviate these concerns, and to mitigate the risk that exporters are unable to avoid paying under carbon border adjustment mechanisms in countries importing their goods.

Korea's comparably more limited uptake of solar power is partly driven by its small land area and high population density. When comparing the ratio of installed solar capacity to land area, Korea is among the countries with the highest density of solar power in the world. Despite this, regulatory changes are possible to use scarce land more efficiently. Land use permitting rules in Korea commonly impose a minimum distance between solar farms and infrastructure such as roads and residential areas of between 300 m and 1 km. In aggregate, this means that solar panels are not allowed to be installed on a large share of Korea's land. Reducing these minimum distances, while taking into account potential concerns and complexities, could unlock substantial solar capacity.

Depending on grid conditions and policymaking priorities, complementary measures can be taken to support renewable generation and storage at or near data centre locations, with the aim of enabling a resilient and diversified mix of energy sources capable of providing uninterrupted power to critical AI infrastructure. In addition to renewables, low-emissions sources such as nuclear, geothermal and bioenergy can be also considered as relevant contributors to supply. Emerging technologies such as small modular nuclear reactors and advanced geothermal, where socially and environmentally acceptable, are also potential options in the longer term.

In parallel, accelerating investment in electricity grids and related infrastructure, with a focus on flexibility, congestion management and digitalisation, is crucial to accommodate the rising demand from data centres in a secure manner while also integrating higher shares of renewables. Dedicated large-scale energy infrastructure projects, such as Korea's "energy highway" plans for large-scale grid expansion, can facilitate the integration of new large loads such as data centres and new supply sources.

Incentivise siting data centres in a grid-friendly manner

Data centres have historically been concentrated in certain locations due to factors such as latency, infrastructure and workforce availability. While data centres that primarily cater to the finance industry and e-commerce largely benefit from being close to the customer and having low latency, some AI-optimised data centres could be located further from customers due to the relatively long time taken to generate responses compared to the transmission delay. At the same time, many regions are seeing increased grid bottlenecks amid rapidly rising generation sources such as solar PV and wind farms as well as large load connections, such as those from data centres. Therefore, incentivising the siting of new data centres in areas with grid capacity, and if desired, abundant low-emissions power, becomes essential for reducing grid costs and maintaining system security while also satisfying demand for rapid data centre expansion.

In East Asian countries such as Japan and Korea that are effectively islanded power systems due to lack of interconnections, encouraging a more spatially balanced distribution of power demand is particularly important. As power imbalances cannot be managed via imports and exports of electricity in such regions, incentivising more dispersed deployment of large loads such as data centres in available grid connection points can allow for more effective usage of system flexibility, with reduced risk of congestion. Enabling the necessary price signals to encourage this regional decentralisation therefore deserves greater attention. For countries seeking to introduce regionally differentiated tariffs or prices, establishing an appropriate legal framework and ensuring timely implementation are essential. Nevertheless, the risks and benefits of such reforms warrant careful assessment, with consideration of countries' unique regulatory environment and market design.

In general, increasing the geographical resolution of price signals can unlock greater value from electricity systems by reflecting local differences and encouraging participants to make efficient adjustments to their behaviour. This can reduce the need for redispatch, reserves and other costly operational measures. Granular price signals can also guide investment in generation and industrial loads, such as data centres, towards areas where capacity or flexibility is needed the most. Whilst grid investment is still needed, aligning price signals to physical realities can help reduce system costs.

For example, in 2011, Norway and Sweden split their wholesale price regions from a single zone each into five and four zones, respectively. In these narrow and long countries, prices in the hydro- and wind-abundant north are typically lower than in the south, reflecting the physical reality of southern demand centres connected to northern generators through transmission bottlenecks. Lower prices in the north reflect the lower marginal cost of supplying demand in the north, and higher prices in the south reflect the higher marginal benefit of installing generation in the south. Higher prices in the south may incentivise streamlining wind farm permitting to be as easy as in the north, which may help address growing challenges with power balance in the south. Svenska kraftnät, Sweden's TSO, expects new large industrial facilities to be drawn to the north to take advantage of cost savings from plentiful generation capacity. Nordic price regions recently underwent a review by the European Network of Transmission System Operators, ENTSO-E, which found that the other considered alternatives were less economically efficient.

Price region disaggregation can be taken further than intra-country zones to the creation of more granular price nodes, such as is common in the United States. For example, ERCOT switched to nodal pricing in 2010, after several years of implementation delays, at a cost of USD 500 million, delivering USD 300 million of benefits in the first year alone. When carefully designed and implemented, price disaggregation can deliver benefits, and challenges can be addressed with careful consideration and planning.

Similarly, retail price disaggregation offers similar benefits. Energy-intensive industries, such as data centres, would give strong consideration to regional differences in retail electricity tariffs when choosing where to locate within a country. If new loads and generation are constructed in better-matched geographical locations, transmission upgrade costs could be reduced. Additionally, since low-carbon technologies such as solar, wind and nuclear have lower marginal costs than gas, increasing the geographical resolution of retail prices also provides a market signal that encourages the siting of loads in areas with more low-emissions renewables.

While locational disaggregation of electricity prices can provide benefits by better aligning market signals with network realities, it can also introduce risks and complexities that need to be weighed by policy makers. Increased local volatility may create the need for hedging instruments, such as financial transmission rights, and reduced liquidity may necessitate more robust market power monitoring.

In addition to locational price signals at the wholesale and retail tariff level, distribution connection and usage charges can also be geographically differentiated to provide investment signals that more closely reflect the scarcity and cost of additional load-supporting infrastructure. In Korea, new data centres built outside the Seoul Metropolitan Area may receive discounts on grid connection charges, depending on eligibility and tariff design. The effect of such policies can be enhanced with initiatives such as the Non-Metropolitan Data Center Consulting Support Center, which provides stakeholders with information on incentives, planning and system capacity, to maximise the effectiveness of these policy signals.

Manage the rapid increase in data centre connection applications

In recent years, network operators around the world have been facing a rapidly increasing number of connection applications for new generation, storage and load, such as data centres, which are rising more rapidly than the pace of grid expansion. A lack of grid capacity is emerging as a bottleneck for data centre and generation deployment in many regions of the world, resulting in higher levels of grid congestion and connection queues for both generation and loads, particularly data centres. Policy makers should focus on efforts to alleviate these growing queues, including streamlining processes to speed up the head of the queue, queue prioritisation, tenders for access rights, fees for connection applications and increasing grid investment.

Additionally, many countries are facing over-reservation of capacity in their grid connection queues due to duplicate or so-called “phantom” projects. This is also increasingly becoming an issue in Korea. While many countries, including Korea, have introduced institutional measures to limit such non-viable grid connection applications, the impact of these measures may be constrained amid limited enforcement. Hence, more effective application of existing rules may be necessary. Enforcing the withdrawal of connection rights when project milestones are not met, clearer prioritisation of projects that are ready to proceed, and appropriately calibrated connection deposits can be essential in reducing speculative behaviour.

Grid planners and operators need to consider data centres in their planning and operations to ensure that the transmission system is adequately planned with respect to such large loads. This also includes ensuring that the system has sufficient reserves and adequacy under different conditions. In the case of a supply or connection bottleneck, policy makers may require that data centres bring supply with them under the principle of additionality, depending on local grid conditions. They may also impose certain local criteria and matching restrictions. While these policies and measures can contribute to additional supply (and especially renewables, if the policy is tailored accordingly) and improved security of supply, they can also be potentially restrictive for certain data centre developers.

Data centres are traditionally concentrated in certain hubs not only because of abundant power supply, but also because of robust telecommunications infrastructure, availability of dedicated workforce and favourable business practices. In this sense, incentivising new data centres to locate towards locations with better grid and supply availability may require addressing such aspects as well.

Streamline approvals for grid construction

As highlighted by the IEA, meeting rising demand in the Age of Electricity requires investment in grids to increase. As with all infrastructure projects, generator and grid buildout needs to be planned with consideration of local community needs. The balance between community engagement and national infrastructure objectives is a delicate one, because the benefits of transmission and generation infrastructure are realised across a broader geographical scope than localised downsides, such as the obstruction of views and construction noise.

Opposition from local communities is one of the main causes of transmission project delays in Korea, with many projects being delayed for many years due to resident objections. The 345 kV Bukdangjin-Shintangjeong transmission line was planned for completion by December 2012, but was in fact completed over 12 years later due to residents' concerns. Systemic multi-year delays to strategic transmission projects hinder climate goals and may threaten the industrial competitiveness of local industries.

Policy makers may need to consider the appropriate balance between national infrastructure priorities and local residents' concerns. When planning large-scale generation and transmission infrastructure projects that can yield substantial benefits for society, concerns of the local community may need to be weighed against national benefits, supported by transparent processes and appropriate compensation. Delays and project cancellations due to community concerns may hinder broader public goals, such as energy security and climate objectives. Policy makers designing community engagement frameworks may give consideration to the differing societal cost of delays to or the cancellation of different types of projects.

To help shift this balance in Korea, the Energy Trifecta Bill, passed in February 2025, includes the Electricity Network Expansion Bill. This reduces the local consultation period to 60 days and increases the amount of compensation that KEPCO may pay to residents when buying their land for transmission projects.

Improve transparency and data collection

Estimates of data centre electricity consumption have varied significantly between sources, largely due to the absence of consistent and reliable data in many countries. This leads to considerable uncertainty about actual consumption levels. Global trends in data centre deployment, the shift towards more power-intensive servers and recent plans for new construction highlight the need for more accurate tracking of data centre power usage.

Enhancing the collection of energy consumption data is crucial for understanding past developments and for gaining clearer insights into future trends, which in turn support more informed public discussion. For this purpose, governments can mandate the development of national or regional registries of data centre energy use. This can apply to data centres above a certain size. Standardising the reporting frameworks and promoting voluntary benchmarking and best practice sharing can also be beneficial in this regard. Enabling and incentivising relevant stakeholders to provide accurate near-real-time grid emissions intensity data, along with short-term forecasts, can help companies undertaking AI training and batch inference to shift their consumption to periods with a higher share of low-carbon generation.

Annexes

Abbreviations and acronyms

AC	alternating current
AI	artificial intelligence
ASIC	application-specific integrated circuit
AWS	Amazon Web Services
BEMS	building energy management system
BESS	battery energy storage systems
BTM	behind-the-meter
CAGR	compound annual growth rate
CBP	Cost-Based Pool
CCUS	carbon capture, utilisation and storage
CPU	central processing unit
DC	direct current
DLR	dynamic line rating
DRL	deep reinforcement learning
EV	electric vehicle
FT	Fischer-Tropsch
GDP	gross domestic product
GPU	graphics processing unit
HEMS	home energy management system
HVAC	heating, ventilation and air conditioning
ICT	information and communication technology
IEA	International Energy Agency
IEC	International Electrotechnical Commission
IoT	Internet of Things
ISO	International Organization for Standardization
IT	information technology
KEA	Korea Energy Agency
KEEI	Korea Energy Economics Institute
KEPCO	Korea Electric Power Corporation
KOREC	Korea Electricity Regulatory Commission
KPX	Korea Power Exchange
LCOE	levelised cost of electricity
LLM	large language model
MAS	multi-agent system
MOF	metal organic framework
NERC	North American Electric Reliability Corporation
O&M	operation and maintenance
PPA	power purchase agreement
PUE	power usage effectiveness

PV	photovoltaic
REC	renewable energy certificate
RL	reinforcement learning
SEDA	Substation Equipment Diagnostic and Analysis System
SME	small and medium-sized enterprise
SMP	system marginal price
SMR	small modular reactor
TPU	tensor processing unit
TRL	technology readiness level
TSO	transmission system operator
UPS	uninterruptible power supply
VPP	virtual power plant
VRE	variable renewable energy
V2G	vehicle-to-grid
V2X	vehicle-to-everything
WPGNN	Wind Plant Graph Neural Network

Units of measure

GW	gigawatt
GWh	gigawatt hour
Hz	hertz
km	kilometre
km^2	square kilometre
kV	kilovolt
kW	kilowatt
μs	microsecond
ms	millisecond
MW	megawatt
MW_e	megawatt electrical
MWh	megawatt hour
TWh	terawatt hour

See the IEA glossary for a further explanation of many of the terms used in this report.

International Energy Agency (IEA)

IEA Publications
International Energy Agency
Website: www.iea.org
Contact information: www.iea.org/contact
Typeset in France by IEA - March 2026
Cover design: IEA
Photo credits: © ShutterStock

Revised version: April 2026
Information notice found at:
www.iea.org/corrections