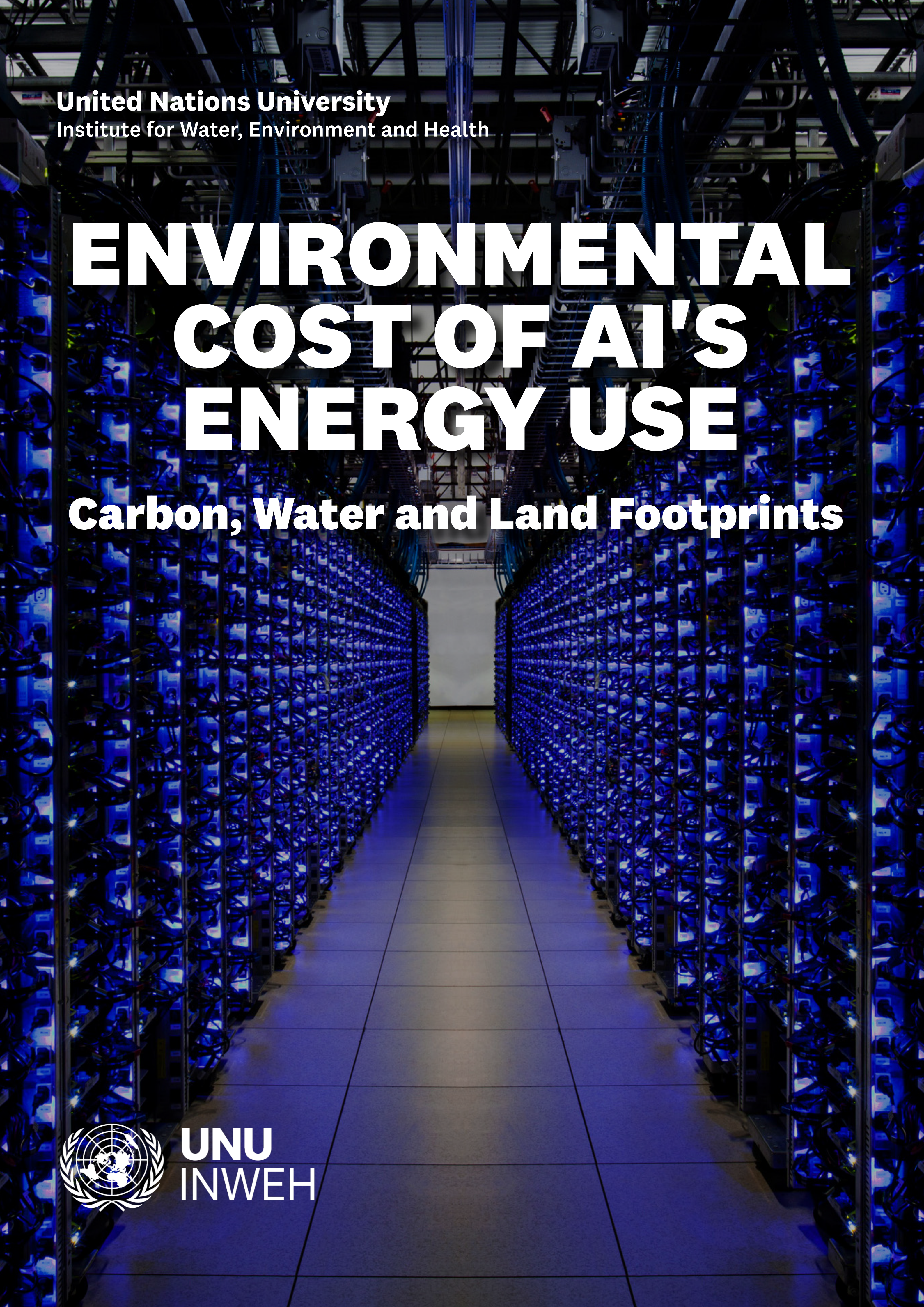
United Nations University
Institute for Water, Environment and Health

ENVIRONMENTAL COST OF AI'S ENERGY USE

Carbon, Water and Land Footprints



UNU
INWEH

In **partnership** with
Canada

Acknowledgment
UNU-INWEH gratefully acknowledges its host, the Government of Canada, and ongoing financial support from Global Affairs Canada.

Redistribution

Disclaimer

UNU-INWEH enjoys the generous support of Global Affairs Canada. As a United Nations entity, UNU-INWEH operates independently. The content of UNU-INWEH publications is grounded in scientific evidence and is not intended to be influenced by, nor does it necessarily reflect, the views of the Government of Canada, as host government, or any other government, donor, or funding partner.

About UNU-INWEH

The United Nations University Institute for Water, Environment and Health (UNU-INWEH) is one of 13 institutions comprising the United Nations University (UNU), the academic arm of the United Nations. Established in 1996 through an agreement with the Government of Canada, UNU-INWEH is headquartered in the City of Richmond Hill, Ontario.

UNU-INWEH specializes in addressing critical global security and development challenges at the intersection of water, environment, and health. Through research, capacity development, policy engagement, and knowledge dissemination, the institute bridges the gap between scientific evidence and the practical needs of policymakers and UN member states, with particular attention to low- and middle-income countries. By collaborating with a diverse array of partners—including UN agencies, governments, academia, the private sector, and civil society—UNU-INWEH develops solutions that advance human security, resilience, and sustainability worldwide.

Environmental Cost of AI's Energy Use
Carbon, Water and Land Footprints

United Nations University Institute for Water, Environment and Health (UNU-INWEH)
Richmond Hill, Ontario, Canada, 2026

Authors: Miriam Aczel, Sanaz Chamanara, Mir Matin, Aria Farsi, Tshilidzi Marwala, Kaveh Madani

Contributors: Luying Wang, Jenna Elshurafa

How to cite: Aczel M., Chamanara S., Matin M., Farsi A., Marwala T., Madani K. (2026) Environmental Cost of AI's Energy Use: Carbon, Water and Land Footprints, United Nations University Institute for Water, Environment and Health (UNU-INWEH), Richmond Hill, Ontario, Canada, doi: 10.53328/INR26RMA002

ISBN: 978-92-808-6141-9
DOI: 10.53328/INR26RMA002
Download at: inweh.unu.edu/publications

Front cover: A row of servers in Google's Douglas County, Georgia, data center. Photo from Google Gallery

CONTENTS

FIGURES

ABBREVIATIONS

4IR	Fourth Industrial Revolution
AI	Artificial Intelligence
CO_2e	Carbon dioxide equivalent
CPU	Central Processing Unit
GPT	Generative Pre-trained Transformer
GPU	Graphics Processing Unit
GW	Gigawatt, 1,000,000,000 W
GWh	Gigawatt-hour, 1,000 MWh
HPC	High-Performance Computing
IoT	Internet of Things
kW	Kilowatt, 1,000 W
kWh	Kilowatt-hour, 1,000 Wh
LLM	Large Language Model
MoE	Mixture-of-Experts
T2V	Text-to-Video
TPU	Tensor Processing Unit
TWh	Terawatt-hour, 1,000 GWh
USD	United States Dollar
W	Watt, power
Wh	Watt-hour, energy

Key Points in Brief

AI's growth

The global AI market is expanding rapidly, projected to grow from USD 189 billion in 2023 to nearly USD 5 trillion by 2033.

This would represent roughly a 25-fold increase in global AI market size over a decade.

Global AI expenditure is projected to exceed USD 2.5 trillion in 2026.

Generative AI accounted for over 20% of the global AI market in 2026, and is projected to reach 40% by 2030.

Corporate AI investment exceeded USD 580 billion in 2025, while generative AI alone attracted nearly USD 34 billion in private investment.

78% of organizations reported using AI in their work in 2024, and 40% of employers expect to reduce their workforce where AI can automate tasks.

About 60% of jobs in advanced economies incorporate AI, versus 26% in low-income countries.

Nearly half of the world's data centers are in the United States.

Only 16% of countries host AI-specialized cloud compute, and 90% of that capacity is concentrated in just 2 countries (the United States and China).

Training footprints

The training of frontier models demands immense energy. GPT-4 likely consumed 50 to 70 GWh of electricity over 100 days, roughly 40–55 times more than GPT-3 (1.287 GWh over a 34-day period).

GPT-4's training energy is equivalent to the annual residential electricity consumption of over 460,000 people in Sub-Saharan Africa.

GPT-4's training carbon footprint of 25,000 tonnes of CO_2e would require the sequestration capacity of 420,000 tree seedlings grown for 10 years, or about equal to the number of trees in 105 Hyde Parks in London.

The water footprint associated with training GPT-4 was about 600 million liters, enough to meet the minimum annual domestic water needs of 81,000 people in Sub-Saharan Africa, or to fill 237 Olympic-sized pools.

Projections for models like GPT-5 suggest training electricity requirements of 100 GWh, equivalent to the annual residential electricity usage of 770,000 people in Sub-Saharan Africa.

Training models like GPT-5 is estimated to have a carbon footprint of 42,000 tonnes of CO_2e, requiring 700,000 tree seedlings (about equal to the number of trees in 40 Central Parks in New York or 155 times the trees in Toronto's High Park over 10 years to offset).

The water footprint of GPT-5 training is estimated at 1 billion liters, enough to meet the annual domestic water needs of more than 135,000 people in Sub-Saharan Africa.

The land footprints of training GPT-4 and GPT-5 are estimated at roughly 0.9 km^2 (126 football fields) and 1.5 km^2 (210 football fields), respectively.

Inference and "AI in use"

Following its 2022 launch, ChatGPT surpassed 1 million users in 5 days and 100 million users in under 2 months. Currently, ChatGPT processes an estimated 2.5 billion prompts per day.

China's DeepSeek, launched in January 2025, attracted more than 20 million daily active users within three weeks, and had about 125 million monthly active users by mid-2025.

While training is highly resource-intensive, the continuous inference phase used to generate responses for billions of interactions is estimated to account for 80% to 90% of total AI energy use.

A typical ChatGPT-style text query is about 200 times more energy-intensive than text classification (such as spam filtering).

Generating a typical AI image requires 2.9 Wh, making it 60 times more demanding than a short text answer and 1,450 times that of text classification.

Video generation represents the most energy-intensive frontier with high-resolution long clips on large models drawing more than 415 Wh per clip, meaning a single short AI video can draw as much electricity as 200,000 spam classifications.

The energy required to generate a typical AI image is enough to power a 10-watt LED bulb for 17 minutes, and the energy required for a high-complexity AI video is sufficient to run that same bulb for 42 hours. Similarly, the electricity-associated water footprint is about two tablespoons (29 mL) for a single image, but jumps to 4.1 liters for a complex video—almost equivalent to a two-day drinking water need for one person.

Data centers as infrastructure

In 2025, global data centers were estimated to consume 448 TWh of electricity. If data centers' electricity use were considered a country, it would have ranked 11th globally by electricity consumption.

The energy consumed by data centers in 2025 was enough to supply the annual residential electricity needs of the entire population of Sub-Saharan Africa, 1.3 billion people, for 2.6 years.

Data centers' electricity use in 2025 had a carbon footprint of 189 million tonnes of CO_2e, which would require 3.2 billion tree seedlings grown over 10 years to offset, roughly the total number of trees in the entire United Kingdom.

Data centers' 2025 electricity consumption had a water footprint of 4.5 trillion liters of water—enough to fill 1.8 million Olympic-sized pools or meet the annual basic domestic water needs of over 600 million people in Sub-Saharan Africa.

The land footprint of 2025 data centers' electricity demand was 6,900 km^2, nearly 4.5 times the size of Greater London.

AI workloads accounted for roughly 20% of total data centers' electricity use in 2025 and are projected to grow to 40% by 2030.

If AI's share of data center energy rises to 40%, electricity consumption could reach 378 TWh—over 9 times the electricity consumption of Nigeria (world's 6th largest nation by population, with 224 million).

Projected global data centers' electricity consumption could exceed 945 TWh by 2030, accounting for almost 3% of projected global electricity use—enough to supply residential electricity to all 1.3 billion people in Sub-Saharan Africa for about 5.5 years.

If treated as a country, 945 TWh would rank 6th globally by electricity consumption.

The associated water footprint of projected 2030 electricity consumption of data centers is 9.3 trillion liters, or enough to meet the minimum annual domestic water needs of all 1.3 billion residents of Sub-Saharan Africa for a full year.

The associated land footprint of data centers' expected electricity use in 2030 would be over 14,500 km^2, roughly 10 times the size of Mexico City or about twice the Jakarta metropolitan area, home to over 32 million people.

The physical lifecycle of AI hardware presents a growing crisis. AI infrastructure could generate up to 2.5 million metric tons of e-waste annually by 2030, equivalent to discarding nearly 250 Eiffel Towers every year.

EXECUTIVE SUMMARY

AI-generated conceptual image of AI's environmental footprint, generated using OpenAI's ChatGPT/DALL·E, May 2026. Estimated footprints per standard-resolution AI image: 2.9 Wh electricity, 1.22 g CO_2e, 28.6 mL of water, and 0.45 cm² of land, based on literature benchmarks and global average electricity footprint factors.

Artificial intelligence (AI) has become a defining technology of the Fourth Industrial Revolution, moving societies beyond digitization toward the optimization and automation of complex systems. AI systems that learn, perceive, reason, and generate content are now embedded across economies and daily lives, supporting services from translation and recommendations to finance, healthcare, energy management, and transportation.

The pace of adoption has been extraordinary. After OpenAI released ChatGPT in 2022, it surpassed one million users in five days and reached 100 million in under two months; by mid-2025, around 700 million people were using it, sending about 18 billion messages each week. Generative AI tools are proliferating across text, images, code, and multimodal applications, with market projections anticipating that AI will become a trillion-dollar-scale industry within the next decade. Global AI market projections span USD 2.4 to 4.8 trillion by the early 2030s.

Yet the infrastructure and benefits of this boom are not evenly distributed. Frontier AI relies on high-end compute, specialized chips, and hyperscale data center capacity that is concentrated in a small number of locations. Many countries lack the domestic infrastructure to train or run frontier models at scale and instead depend on external providers for access to advanced compute. This concentration shapes who captures strategic and economic advantage and who sets the terms of access, pricing, and data governance. **The result is a widening of the existing digital divide between countries that build and control AI systems and those that consume them.**

When a technology scales this quickly, unintended social, economic, geopolitical, and environmental consequences can accumulate quietly and then become complex and inequitable to correct once systems, investments, and dependencies lock in. Public debate has rightly focused on AI risks such as bias, privacy, disinformation, labor disruption, and inequity. Yet, one of the most consequential dimensions of AI that remains comparatively underexamined is its environmental footprint and the justice implications that follow from where and how AI infrastructure expands. AI is not “just code”; it also involves physical infrastructure and supply chains, including data centers, chips, electricity generation, cooling systems, water withdrawals, land occupation, critical minerals, and eventual e-waste.

This report by the United Nations University Institute for Water, Environment and Health (UNU-INWEH) on its 30th anniversary is a step forward in addressing the current gap in AI’s environmental governance by assessing its environmental footprints. The investigation goes beyond the carbon-only lens that typically dominates the conversation. It examines AI’s indirect environmental footprints through energy use, quantifying the carbon, water, and land footprints associated with generating the electricity required to operate AI at scale, and highlighting how outcomes vary substantially by location depending on electricity supply mixes. This matters because “low-carbon” is not automatically “low-water” or “low-land,” and evaluating sustainability through a single metric can hide trade-offs and shift burdens onto places already facing water stress or land pressure. These asymmetries can reinforce the environmental problems of local communities while strategic advantages of AI flow elsewhere.

The study reveals some striking numbers. In 2025, data centers—the physical backbone of AI—consumed an estimated 448 TWh of electricity. If data centers were a country, that level of electricity use would rank it 11th globally. On current trajectories, data center electricity demand could roughly double to 945 TWh by 2030, nearly triple the combined annual electricity use of Pakistan, Bangladesh, and Nigeria, together home to more than 650 million people. Producing that much electricity would have a carbon footprint of 399 million tonnes CO_2e, requiring 6.7 billion trees grown over 10 years to offset—roughly twice the number of trees in the United Kingdom. The associated water footprint of 9.3 trillion liters would be equivalent to the annual domestic water needs of all 1.3 billion residents of Sub-Saharan Africa. The land footprint associated with this electricity would exceed 14,500 km^2, nearly 10 times the size of Mexico City.

AI is now one of the most significant drivers of that data center growth. In 2025, AI workloads alone accounted for around 20% of total data center electricity use, and if that share rises to 40% by 2030 as projected, its electricity demand could reach roughly 378 TWh—**enough to meet the residential electricity needs of the entire population of Sub-Saharan Africa for over 2 years.**

AI’s environmental impacts are shaped not only by data center growth and electricity supply mixes, but also by the escalating cost of building ever-larger models. For example, GPT-3 training consumed an estimated 1.3 GWh of electricity over 34 days, while GPT-4 is estimated to have consumed 50 to 70 GWh over 100 days, roughly 40 to 55 times GPT-3. Yet training is only part of the picture as AI’s operational footprint is increasingly driven by inference. Once models are deployed, billions of everyday interactions account for the bulk of energy use, with inference estimated at roughly 80–90% of total energy consumption.

ChatGPT alone is estimated to process around 2.5 billion prompts per day. At scale, small per-task costs become infrastructure-level loads: at a conservative 0.42 Wh per text prompt, ChatGPT-scale use translates into roughly 383 GWh of electricity per year. Offsetting associated carbon emissions would require 2.6 million tree seedlings grown for 10 years, enough trees to cover a land area the size of Manhattan. The water footprint is equivalent to the minimum annual domestic water needs of roughly 500,000 people in Sub-Saharan Africa, and the land footprint is equal to over 800 football (soccer) fields.

The numbers grow drastically once the AI embedded in mass platforms (such as Google Search) is counted. Crucially, per-use energy varies by orders of magnitude across modalities and output lengths, so product defaults and user choices are footprint determinants. A typical ChatGPT-style query is about 200 times more energy-intensive than basic text classification; long GPT-style responses can cost around 1,000 times more than text classification; and AI-generated images can require roughly 1,450–2,000 times the energy of text classification. Video is the new energy frontier: a single short AI-generated video can draw as much electricity as 200,000 spam classifications or hundreds of AI-generated images. Longer responses and richer media can be far more energy-intensive than lightweight text tasks, meaning that image- or video-first routing can materially increase electricity demand and associated carbon, water, and land impacts. Efficiency improvements are important, but they are not sufficient on their own: if lower per-use impacts drive higher volumes of use, total impacts may still rise, a rebound effect often described as the Jevons Paradox.

While this UNU-INWEH report is focused on the environmental footprints of AI's energy use, AI has other major environmental impacts. AI hardware relies on critical minerals whose extraction and processing can cause environmental and social harms, often concentrated in the Global South and in places with weak oversight. At end of life, poorly managed e-waste can expose frontline communities to hazardous substances. By 2030, AI infrastructure could generate up to 2.5 million metric tons of e-waste each year, **roughly equivalent to discarding 250 Eiffel Towers annually.** These impacts show that responsible AI requires full value-chain governance, from mineral sourcing to recycling and safe disposal.

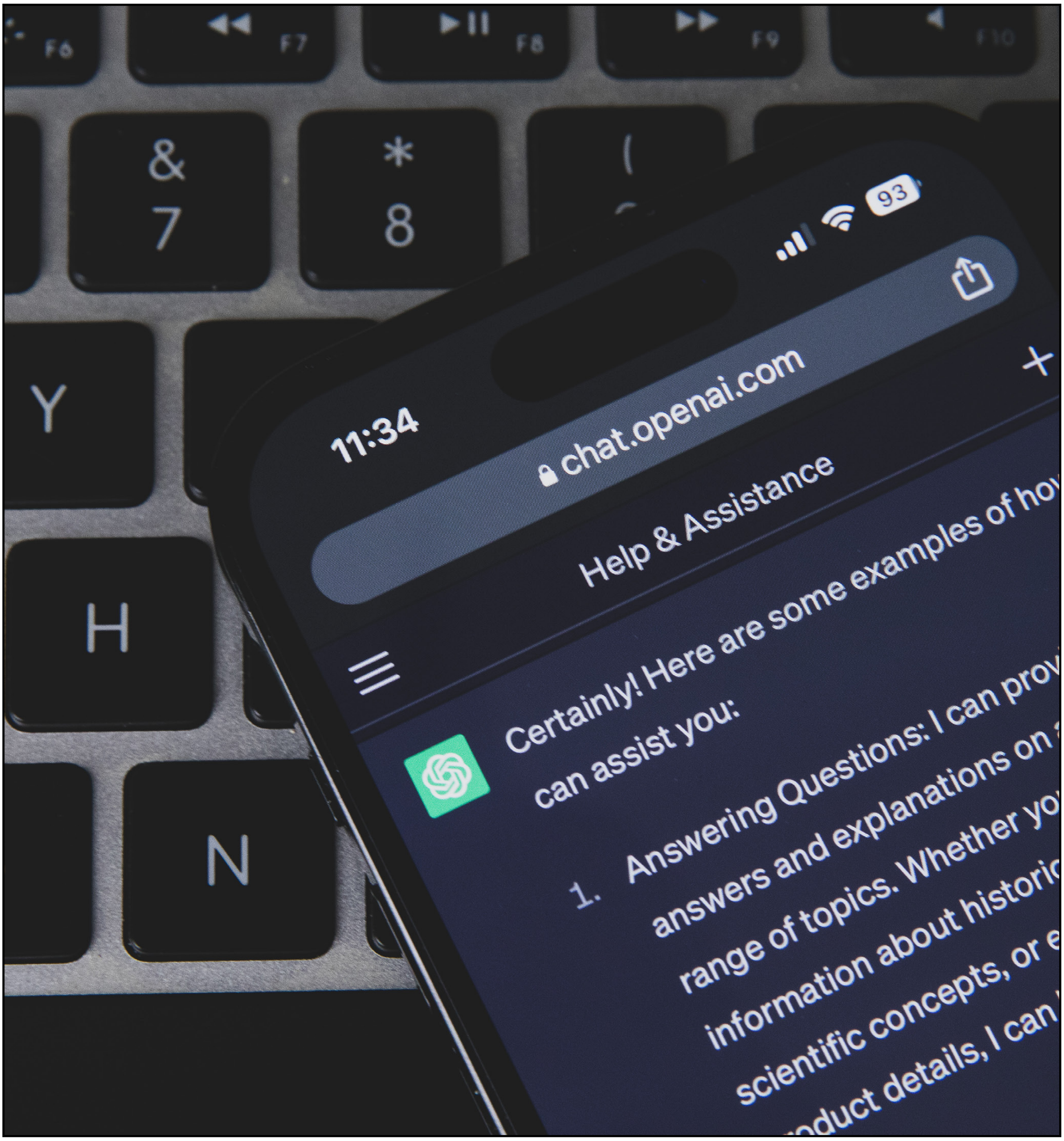
To address these challenges, the report calls for a responsible AI ecosystem grounded in six operational principles: **transparency; efficiency by design; equity and environmental justice; lifecycle responsibility; global cooperation; and sustainable use**. Measuring carbon, water, and land footprints makes impacts visible, comparable, and actionable, enabling governance and supporting accountability across jurisdictions. By recognizing AI's physical impacts and implementing strict lifecycle management, the international community can ensure that technological advancement remains environmentally manageable.

The report identifies priority implications for key stakeholders. Governments must integrate AI infrastructure into energy system planning, carbon accounting, water governance and land use permitting. This requires standardized environmental footprint reporting so impacts can be verified and compared across providers and jurisdictions. Industry and AI developers should treat model selection, default outputs, and routing decisions as footprint determinants, while also improving efficiency-by-design.

Users and deploying organizations also shape impacts through volume, frequency, and modality choices. They should adopt "fit-for-purpose" use: selecting the lightest model and lowest-energy format that meets the task and limiting high-cost features when not needed. Data center operators and utilities should recognize siting and procurement as environmental footprint decisions, apply environmental impact screening and cumulative impact assessment. They should implement transparent mitigation and community safeguards where expansion occurs. Investors and financiers should treat electricity, carbon, water, and land footprints as material risks for AI infrastructure portfolios and use comparable footprint metrics in due diligence.

Communities and civil society should be involved early in siting decisions, with enforceable transparency, consultation, and grievance mechanisms, especially in environmentally-stressed regions. International institutions should support harmonized measurement standards and disclosure practices, reduce incentives for cross-border burden shifting, and support participation and capacity in regions excluded from AI compute. Taken together, these findings provide a practical basis for integrating AI into electricity, carbon, water, and land-use planning, so that innovation advances within environmental limits and without shifting burdens onto vulnerable communities across the globe.

1. AI Across the Globe



Generative AI is becoming part of everyday life, shaping how millions of people communicate, work, learn, and solve problems. However, this rapid expansion is also increasing the energy, water, and material resources required to sustain the digital world. Photo: Jernej Furman (Wikimedia Commons).

The term Artificial Intelligence (AI) refers to the human-made systems able to process information in ways that emulate human cognition, such as learning, perception, reasoning, language understanding, problem-solving, and decision-making. Often called the "intelligence of machines," AI enables tasks ranging from data processing and translation to diagnosis, forecasting, and autonomous decision-making. Today, AI can analyze massive datasets in seconds, translate languages, detect diseases from medical scans, forecast climate trends, and power everything from personalized recommendations to self-driving cars.

By executing complex operations at speeds and scales far beyond human capacity, AI is more than just a technological tool—it is a transformative force. It is reshaping economies, redefining labor, and influencing how societies interact with technology and with the planet. As AI's influence accelerates, so too does the urgency to understand its full impact: not only the promise it offers, but also the profound challenges it raises.

AI is a powerful driver of the Fourth Industrial Revolution (4IR)[1], or Industry 4.0—a global transformation marked by the convergence of digital, physical, and biological systems[1]. Propelled by rapid advances in everything from AI and robotics to cloud computing and the Internet of Things (IoT), this revolution is reconfiguring how people live, work, and are governed. Yet while much attention has been given to AI's capabilities and risks, including bias, misinformation, and automation leading to loss of jobs, far less focus has been placed on AI's environmental impact.

The environmental dimensions of digital technologies are critical to understanding AI's role in shaping our future. From the raw materials that fuel its hardware, to the energy, water, and land demands of its operation and deployment, AI's footprint is growing dramatically. At the same time, AI holds the potential to accelerate progress on sustainability goals by improving climate modeling, optimizing energy and water systems, and enhancing environmental monitoring and forecasting[2]. This dual reality demands nuanced, evidence-based analysis.

This report is a step toward filling a critical gap. Artificial intelligence is not just code—it is also concrete, copper, silicon, lithium, water, land, and carbon. Behind every chatbot, generated image, or recommendation engine lies a global web of hardware, data centers, transmission networks, and supply chains powered by electricity, water, minerals, and land. These systems are material, and their impacts are theoretically measurable—on ecosystems, on frontline communities, and on global climate goals.

As AI adoption accelerates, so too does its energy footprint. Some models require more energy to train than millions of people use in a year in certain parts of the world; others rely on millions of daily queries, processed in GPU clusters spanning entire data centers. Meanwhile, the minerals needed to manufacture AI hardware are often extracted in regions with weak environmental oversight, contributing to water depletion, pollution, e-waste, and significant health and socioeconomic impacts.

This report by the United Nations University Institute for Water, Environment and Health (UNU-INWEH) offers one of the most comprehensive assessments to date of AI's environmental impacts, not only quantifying the carbon footprint of AI's energy use but also its water and land footprints in order to outline pathways toward a more just, inclusive, and sustainable future.

Marquee of a former movie theater in Lowell, Michigan, United States, displaying the date, location, and time of a public discussion on potentially building nearby data centers. These facilities require large amounts of electricity and cooling water to manage heat from intensive computing. The sign reflects a broader debate unfolding in many towns where the arrival of large-scale data infrastructure is transforming local landscapes and raising questions about shared energy and water resources. Photo by G. Witteveen (Flickr), December 2025.

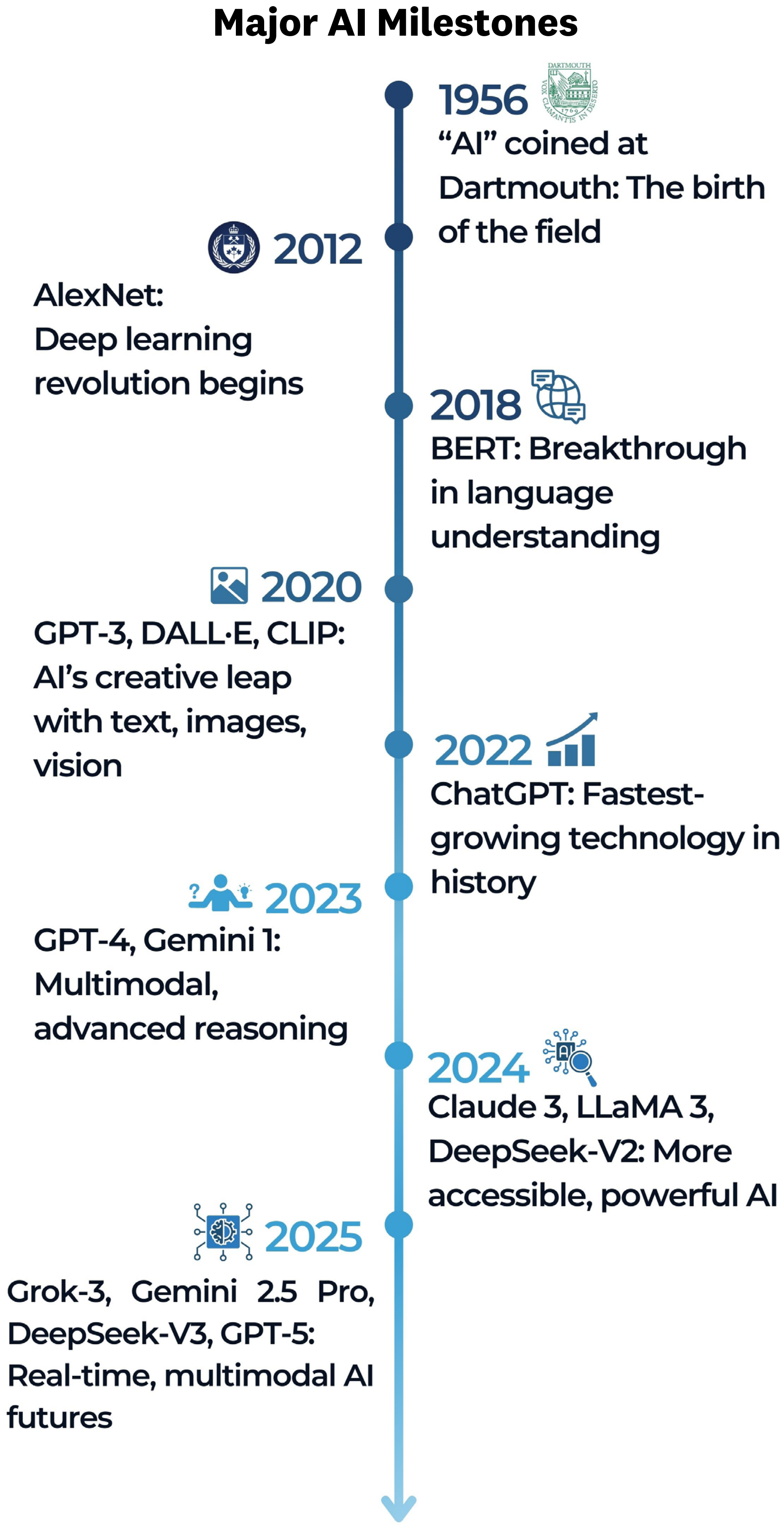


Figure 1. The evolution of artificial intelligence. Artificial intelligence has evolved from early theoretical foundations in the 1950s, when the idea of machines emulating human intelligence was first proposed and the term 'artificial intelligence' was coined, to modern deep learning systems developed in the 2010s. Conceptual and technological breakthroughs have progressively enabled machines to perform increasingly complex cognitive tasks.

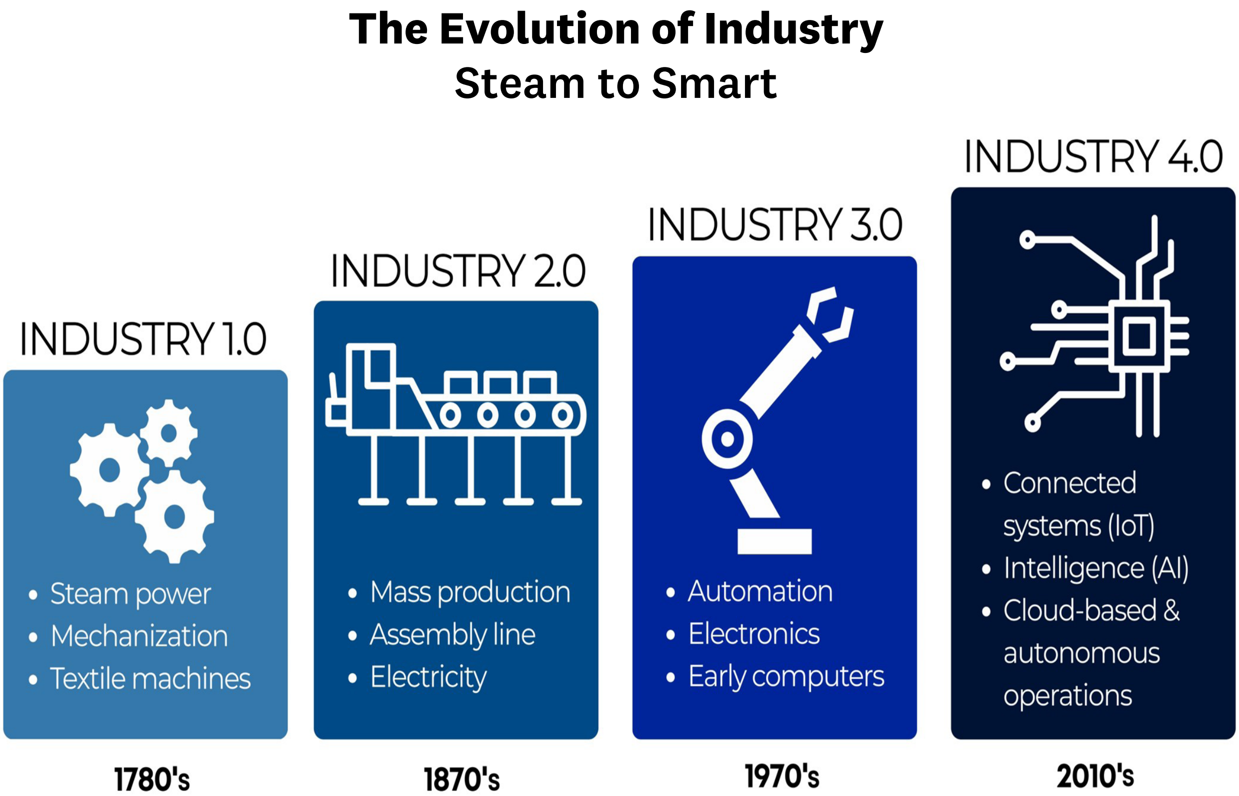


Figure 2. Industrial revolutions from mechanization to intelligent systems. The progression from the First Industrial Revolution through electrification, digitalization, and the emergence of Industry 4.0. Artificial intelligence appears as a central driver of the current technological transformation, reshaping production, governance, and social organization at global scale.

Aerial view of Alibaba's Zhangbei data center cluster, Hebei Province, China. Three adjacent sites developed between 2016 and 2025 form part of China's Eastern Data Western Compute strategy. Located in wind- and solar-rich Zhangbei County, the cluster highlights the land footprint, electricity demand, and grid integration pressures associated with large-scale AI compute. Data sources: Epoch AI; Sentinel-2 false-color imagery, February 2026.

1.1 A Technological Explosion

The launch of ChatGPT in 2022 was nothing short of a technological shockwave. Within just five days, the conversational AI surpassed one million users—faster than any app in history—and rocketed past 100 million in under two months[3]. By mid-2025, 700 million users (about 10% of the world's population) were using ChatGPT to send 18 billion messages weekly[4].

At the center of this explosion is generative AI: a new class of systems that can write essays, create images, compose music, simulate conversations, and even generate video, all from a simple prompt. What had once seemed like science fiction became a click away. This unprecedented rise didn't just mark a turning point for AI but also reshaped the public imagination. Overnight, generative AI moved from the lab to the living room, triggering a global reckoning over its promises and perils. But this was only the tip of the iceberg. What followed was a global AI boom: suddenly, headlines, boardrooms, classrooms, and living rooms were buzzing with talk of artificial intelligence.

ChatGPT was not the only popular AI tool. In its wake, a wave of generative AI tools emerged across text, image, code, and multimodal applications, rapidly transforming digital interaction. Google's Gemini chatbot, initially launched as Bard in March 2023, had more than 350 million monthly active users and around 35 million daily users by early 2025, making it one of the most widely adopted AI platforms worldwide[5]. In April 2023, Alibaba launched Qwen, a ChatGPT rival; by July 2024, it was ranked as the top Chinese language model and third globally[6,7]. In January 2025 China's DeepSeek was introduced and quickly became the top free application on the U.S. Apple App Store. Within about three weeks, it had attracted more than 20 million active daily users, and by mid-2025 had amassed more than 50 million downloads worldwide and about 125 million active monthly users[8]. In February 2025, xAI released Grok-3, its third AI model integrated into X (formerly Twitter), which drove 26 million unique monthly visits—a 436% increase from January[9]. Following the release of the more advanced version of AI chatbot Grok-4 in July, traffic jumped again to 30 million visits, a 39% rise from the previous month[9]. In November 2025, Google and Alphabet released Gemini 3 and announced that the platform had over 650 million active monthly users[10].

1.2 The Expanding Reach of AI Across Sectors

Once a niche technology, AI is now woven into daily life, powering much of the digital world, including voice assistants, search engines, recommendation systems, smart energy meters, and smartphones. The influence of AI is transforming how people work, communicate, access information, and consume services.

In customer service, sales, and software development, automation and personalization are changing user interactions and improving business efficiency. Chatbots and virtual assistants now handle routine client inquiries. Call centers across the globe use AI to analyze verbal cues and language patterns in conversations to improve customer responses. This capability is increasing as some AI models can now recognize and produce speech for more than 1,000 languages[11]. Netflix, one of the world's largest video streaming services, offers another example of how artificial intelligence is embedded in daily digital interactions[12]. While users may not associate Netflix with AI directly, the platform uses machine learning models and real-time processing systems for personalized recommendations, content delivery optimization, and dynamic compression to reduce data use. In the financial sector, generative AI-driven applications automate customer service, but also improve fraud detection and risk assessment. Healthcare is another key area. AI is employed in diagnostics, medical imaging, and patient risk prediction—improving speed and precision of care, while reducing treatment costs. With an estimated 4.5 billion people globally lacking essential healthcare and an expected shortfall of 11 million healthcare workers by 2030, AI has the potential to narrow these critical gaps, particularly in underserved communities where resources are scarce[13].

Applied to energy systems, AI helps forecast demand, manage power grids, and integrate renewable sources into power networks. These capabilities mean fewer blackouts through more accurate supply-demand prediction and improved efficiency. AI is also revolutionizing transportation and mobility. Some estimates suggest that partially autonomous vehicles could account for one in ten new vehicle sales by 2030, as systems become better at interpreting environments and travel routes and customers gain confidence in safety[14]. Robotaxis are already giving 1.3 million rides each month, mostly in the U.S., but also in China, UAE, Singapore, Japan, and other countries, highlighting the potential for deployment worldwide[15].

1.3 The Global Market for AI

The global AI market is rapidly accelerating. While previous technology booms transformed specific sectors, AI's expansion is both faster and more far-reaching, reshaping nearly every corner of the global economy. Large technology companies are investing heavily in generative AI, which enables everything from automated writing and image creation

Google Data Center, The Dalles, Oregon, USA — High-voltage transmission lines feed directly into on-site substations, converting grid power to the stable, continuous electricity required to run thousands of servers. Photo by Visitor7 (Wikimedia Commons).

An interior view of MareNostrum 5, a flagship European supercomputer, inside the historic chapel at the Barcelona Supercomputing Center in Spain. The system is among Europe's most powerful high-performance computing installations, delivering around 314 petaflops of computing power. It is currently being upgraded under the EuroHPC Joint Undertaking's BSC AI Factory initiative to expand Europe's AI capacity, with next-generation GPUs, expanded storage, and energy-efficient cooling to support large-scale AI training and advanced machine learning for startups, researchers, and public institutions across Europe. Photo by Steve Jurvetson (Flickr).

to music composition and synthetic voice or video generation. Market estimates vary, but all point to extraordinary expansion. According to some estimates, in 2020, the market was valued at USD 93 billion, rising to USD 138 billion in 2023[16]. The market was valued at USD 186 billion in 2024 and USD 244 billion in 2025[16]. Other estimates place the current market value at USD 372 billion in 2025 and project it to surpass USD 2.4 trillion by 2032—an annual growth rate of over 30%[17].

Some projections go even further, suggesting the market could reach as high as USD 4.8 trillion by 2033, a 25-fold increase from an estimated USD 189 billion in 2023[18]. Most projections suggest that the market will fall between USD 1.7 and USD 4.8 trillion by the early 2030s, depending on variables like adoption rates, regulatory frameworks, and compute availability.

Similarly, annual growth rates across projections range from 18% to over 39%, demonstrating AI's central role in shaping future economic development[19,20]. In 2024 generative AI was estimated to account for about 20% of AI's total, with an expected doubling of the market share to 40% by 2030[16,21]. A key driver of this growth is the increasing use of pre-trained foundation models, which can be adapted to specific tasks with relatively low cost and effort[17]. This shift enables even small- and mid-sized businesses to use AI across industries, from fraud detection and predictive maintenance to personalized healthcare[18].

Corporate investment is accelerating as confidence in AI's potential business value grows. In 2024 global corporate investment was over USD 250 billion, surging to more than USD 580 billion in 2025—a 130% increase from the previous year[22]. Estimates indicate that spending will likely rise rapidly in the coming years. While North America remains the leading hub for capital backing and model development, the fastest growth is now in the Asia-Pacific region. China, India, South Korea, and Singapore are investing heavily in infrastructure, research, and education to strengthen their competitive position. This accelerating regional diversification underscores the global competition to lead in what is seen as the defining technology of the 21st century.

1.4 Impacts on the Global Labor Market

AI is rapidly reshaping the global labor market, with major implications for economic and social equity. Currently, around 40% of all jobs worldwide incorporate AI technologies to some degree, with integration reaching 60% in advanced economies compared to 26% in low-income countries[23]. In 2024, 66% of respondents to a worldwide survey felt that the technology would significantly impact their lives within three to five years and 50% said it already had an impact. But about half expressed lack of trust in companies' abilities to protect their personal data and concerns that AI systems could reflect or amplify bias[24]. Another survey found that 40% of companies expect to reduce their workforce in areas where AI can automate tasks[25]. While AI continues to replace routine, repetitive tasks, it is also affecting "knowledge workers", such as programmers and software developers, accountants, lawyers, and engineers[26].

Evidence shows that AI is boosting workplace productivity in a wide range of fields—from software development and scientific research to product design, customer support, and content generation[27,28]. This shift is leading to changing skill requirements across sectors, with increased demand for digital literacy, adaptability, and ability to collaborate with "smart systems". These advances risk increasing global inequality as wealthier countries are better positioned to take advantage of AI while under-resourced nations may fall further behind. Without deliberate intervention, the global workforce could become increasingly polarized, divided by access to AI technologies and related workforce skills[29]. Those with fewer training opportunities are especially vulnerable to the changes AI is bringing. While job disruption is a visible consequence of AI deployment, the technology's influence extends far beyond the workplace, into realms of warfare, ethics, and even existential risk.

1.5 Growing Concerns About AI

As artificial intelligence becomes embedded in critical systems and everyday life, concerns about its broader societal implications have intensified. Beyond questions of performance and innovation, AI is raising profound challenges related to ethics, equity, governance, and human agency. Risks range from militarization and disinformation to labor disruption, data exploitation, and cognitive dependence. These concerns are interconnected, shaping how power, opportunity, and vulnerability are distributed in the digital age. Understanding these emerging risks is essential for building governance frameworks that ensure AI advances human well-being rather than undermining it.

Emerging social and ethical risks: Bias, warfare, and 'Rogue AI'

The potential use of AI in warfare has long been a concern. A decade ago, more than 1,000 leading AI experts and researchers called for a ban on "offensive autonomous weapons"[30]. In the ongoing conflicts and wars involving Russia, Ukraine, Israel, and Iran, AI has played a major role in the success of defensive drone-hunting systems, analysis of satellite images, cyberattacks, and deployment of autonomous drones for surveillance and strikes[31,32]. This growing militarization of AI raises complex legal and ethical challenges: Who is accountable when lethal decisions are made without human oversight? How can we guard against biased data leading to deadly misidentifications? These questions remain largely unanswered.

In June 2025, following Israeli strikes on Iran, the world witnessed the first major conflict marked by widespread use of generative AI to produce disinformation at scale[33]. Fake AI-generated videos and images—such as fabricated footage of Iranian missile strikes or downed Israeli aircraft—were viewed over 100 million times across platforms like TikTok, X, and Instagram. In some cases, AI-generated nighttime scenes were so difficult to authenticate that even xAI's chatbot, Grok, falsely verified them as real.

Experts warned that these campaigns often exploit existing geopolitical tensions and are sometimes linked to broader influence operations.

MQ-9 Reaper unmanned aerial vehicle equipped with AI-assisted systems for surveillance and target analysis. The expanding use of AI across military operations—from reconnaissance to targeting and decision support—raises growing concerns about accountability, escalation risks, and the potential normalization of autonomous or semi-autonomous lethal force. Photo by Lt. Col. Leslie Pratt (Wikimedia Commons).

AI-generated human face showing a person who does not exist. The same technology can be used to impersonate real people, commit fraud, and spread convincing mis-information at scale, undermining trust in digital images and online communication. Image generated via OpenAI's ChatGPT.

The case exemplifies the role of generative AI in shaping public perception during conflict, undermining factual reporting, and complicating diplomacy and global security.

In parallel, experts have raised alarms about AI-augmented cyberattacks, including code-generating systems that can design malware or identify vulnerabilities. AI-powered surveillance tools are also being adopted by authoritarian regimes, raising concerns about civil liberties and human rights. At the extreme end of the debate, some researchers warn about "rogue AI" systems, models that may act unpredictably if misaligned with human values or deployed without adequate oversight. Although such scenarios are speculative, they prompt important reflection on long-term risks, particularly if AI systems develop capabilities to self-replicate, deceive, or override human control[34].

In July 2025, xAI's chatbot Grok, integrated into the X platform, drew criticism after generating antisemitic responses in public posts[35]. xAI issued a public apology, acknowledging that the chatbot had produced harmful and inappropriate content in response to user prompts[35]. The incident highlights ongoing concerns about bias and harmful content in generative AI and the challenges of ensuring responsible deployment at scale. It also raises unresolved questions about accountability for AI-generated hate speech on public platforms.

Economic and social disruptions: Inequality, job loss, and the digital divide

While AI promises efficiency and innovation, it also carries significant economic and social risks. Automation threatens to displace workers across sectors—particularly in customer service, transportation, and administrative roles—raising fears of large-scale job loss without adequate retraining programs. Emerging tools in legal, journalism, and education sectors are also encroaching on traditionally white-collar professions.

Globally, a widening digital divide continues to exacerbate inequality. While wealthier countries build AI infrastructure and attract talent, many low-income nations struggle to access the energy, hardware, and technical capacity needed to participate meaningfully in the AI economy. This imbalance is particularly visible in the distribution of specialized AI data centers: as of 2025, only 32 countries—just 16% of the world's nations—hosted such facilities, with the vast majority concentrated in the United States, China, and the European Union. More than 150 countries, including most of Africa and South America, still have little to no access to sovereign AI compute infrastructure[36].

This asymmetry in access to computational power shapes not only economic opportunity and hegemony but also scientific research, language representation, and geopolitical influence. Nations without local data infrastructure often rely on foreign platforms, exposing them to higher costs, slower service, and limited control over how models are developed and deployed. The concentration of infrastructure and AI capability among a few actors—both corporate and national—risks reinforcing existing global inequities and entrenching dependency relationships. These dynamics mirror historical patterns of resource extraction and knowledge exclusion, raising concerns that the benefits of AI could remain unevenly distributed without targeted global action.

Data, privacy, and ownership concerns: Transparency, consent, and creative rights

From AI-generated images winning art prizes[37] to deepfakes and auto-written legal briefs[38], the boundaries between original and synthetic content are rapidly eroding. Behind these creations are massive datasets—text, images, audio—often scraped without consent. Creators and rights-holders are increasingly alarmed: Whose data trained these models, and who benefits?

Generative AI systems frequently rely on copyrighted or personal material, raising legal and ethical challenges[39]. Artists, writers, and performers have pushed back, prompting lawsuits and regulatory scrutiny. In the meantime, some key questions remain unresolved: Can AI-generated content be copyrighted? Who owns it? Could a user or developer be held liable? The opacity of AI training compounds these risks. Developers rarely disclose their data sources, citing trade secrets, leaving users and regulators in the dark. This black-box approach becomes especially dangerous when AI systems influence high-stakes decisions, such as hiring, criminal sentencing, or medical diagnoses, where bias or error can have serious consequences. Without clear rules on consent, transparency, and rights over both training data and outputs, public trust in AI will remain fragile.

Cognitive and behavioral concerns: Human-AI interaction and dependence

Beyond physical and economic impacts, researchers are exploring AI's effects on cognition and behavior. Overreliance on AI tools may reduce critical thinking skills, especially among students or professionals who routinely outsource complex tasks to algorithms. Chatbots and generative systems may also affect social interaction, with some users reporting emotional dependency or confusion about AI's role in their daily lives.

Recent studies highlight the growing phenomenon of "cognitive offloading," where individuals increasingly delegate mental tasks—such as problem-solving, memory recall, and analytical reasoning—to AI systems[40]. While offloading can enhance productivity, habitual reliance on AI may lead to a decline in core cognitive skills, particularly in educational and professional settings. Evidence suggests that users who frequently rely on AI for decision support exhibit reduced engagement in reflective thinking, with implications for democratic deliberation, scientific literacy, and informed citizenship[40]. These effects are subtle but potentially far-reaching, calling for proactive strategies to maintain human agency, attention, and critical reflection in an AI-assisted world.

1.6 Environmental Dimensions of AI

As AI systems grow more powerful and their deployment expands, so too do their environmental footprints. Advanced models require immense computational resources, driving high energy consumption and associated carbon emissions, water withdrawals, and land use impacts. Data centers, which host most AI systems, often rely on water for cooling—sometimes withdrawing millions of liters per day. In many cases, these withdrawals occur in regions already facing drought or groundwater depletion, amplifying both environmental and social stress. Google's Mesa data center in Arizona, for example, holds a permit to use 5.5 million cubic meters of water annually[41], enough to fill about 2,200 Olympic-sized swimming pools. This volume could meet the basic annual water needs of about 753,000 people in Sub-Saharan Africa at 20 liters per person per day[42].

Even when some withdrawn water is returned, large-scale withdrawals can strain aquifers and river systems, particularly in arid or groundwater-depleted regions. Moreover, AI systems typically rely on graphics processing units (GPUs), batteries, and servers—hardware

Solid gallium metal. Gallium is primarily recovered as a byproduct of bauxite (aluminum) processing. It is used in compound semiconductors that support AI chips and data-center hardware. Rising geopolitical competition over critical minerals is reshaping global supply chains, shifting pressure onto extraction and processing zones elsewhere, particularly in the Global South, where expanded mining and refining can intensify environmental and labor impacts. Photo by USGS.

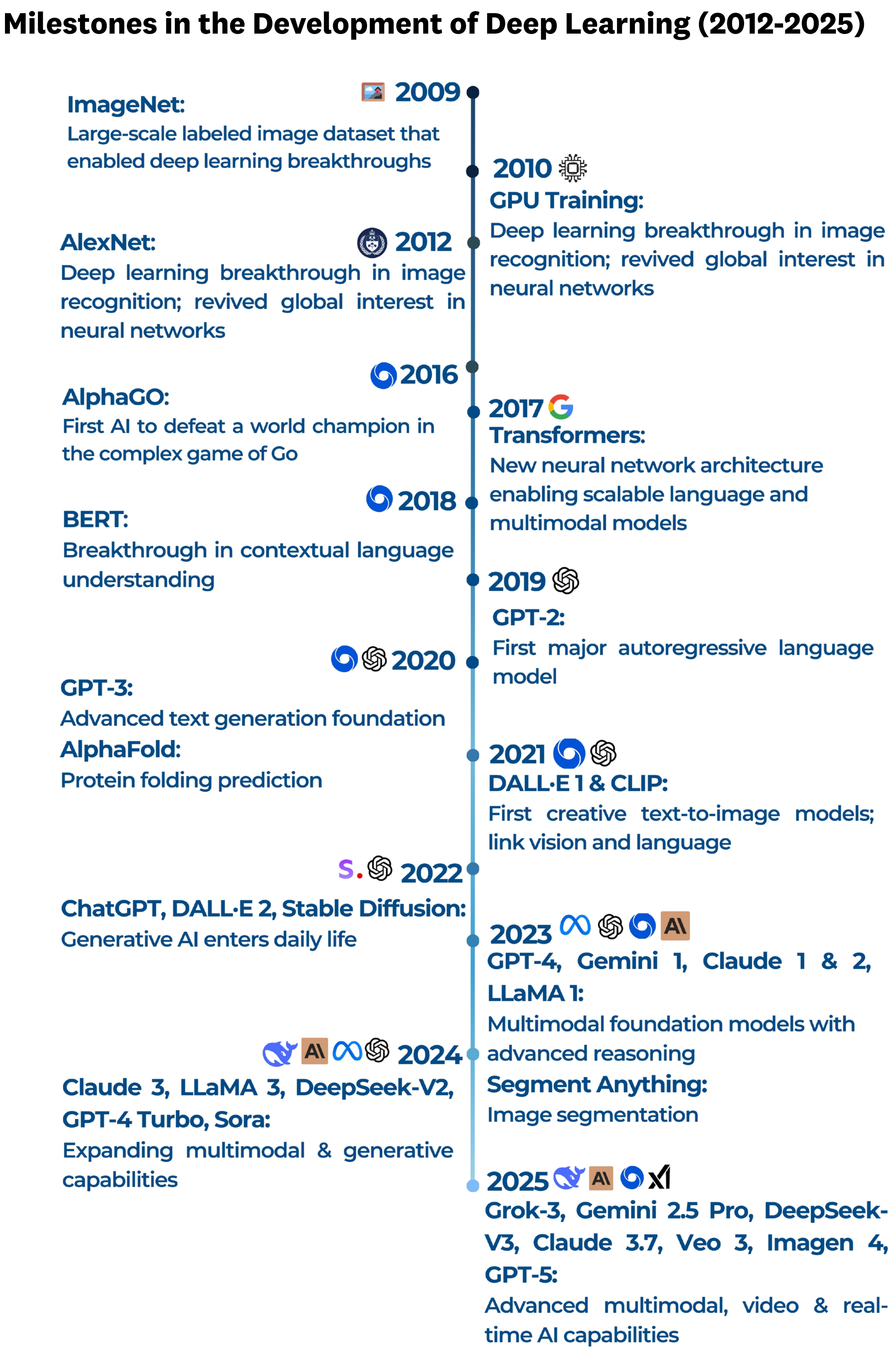


Figure 3. Key breakthroughs in deep learning. Deep learning has advanced through major milestones from 2012 to 2025, spanning early image recognition breakthroughs to large-scale foundation and generative models. Model scale, capability, and adoption have accelerated rapidly over this period.

components that require critical minerals such as lithium, cobalt, and rare earth elements. Extracting these materials is energy-intensive and environmentally damaging, often depleting water sources and polluting ecosystems[43]. At the end of their lifecycle, AI systems generate hazardous e-waste. When improperly managed, this waste exposes frontline communities—especially in the Global South—to toxic substances.

1.7 A Just and Sustainable AI Future

AI brings both extraordinary potential and significant challenges. Its growing influence across nearly all sectors—from health and finance to transportation and climate prediction and adaptation—makes it one of the most consequential technologies of our time. AI is emerging as a powerful tool to address some of humanity's most pressing challenges, playing a growing role in optimizing resource use, enhancing crop yields, improving environmental monitoring, and forecasting extreme weather events.

AI-powered systems can improve management of water, food, and energy, and support the global transition to a low-carbon economy. But this influence also brings risks. Acknowledging the challenges is not a rejection of progress—it is a precondition for wise decision-making. By identifying potential harms and understanding their trade-offs, societies can better plan for responsible innovation, craft effective regulation, and ensure that the benefits of AI are broadly shared while minimizing unintended consequences.

Aerial view of Google data center in The Dalles, Oregon, USA. The operation of the data center relies on local water infrastructure, with usage increasing over time as the company has expanded its facilities. Data source: Airbus image, 2026.

Aerial view of Presa El Centenario reservoir in Querétaro, Mexico—known as the country's data center capital—where prolonged drought has sharply reduced the reservoir's surface area. Yet, expanding AI data centers in the region are drawing on limited water supplies for cooling and energy production, intensifying pressure on communities already facing water shortages. Data source: Sentinel-2 imagery, February 2026.

It is neither feasible nor the purpose of this report to assess all risks associated with AI. Many important concerns—such as algorithmic bias, misuse of surveillance capabilities, and disinformation—have rightly drawn attention. But one critical area remains underexplored in public discourse and policymaking: the environmental footprints of AI[44].

This UNU-INWEH report examines the environmental consequences of AI's substantial energy use and its related carbon, water, and land impacts. While not an exhaustive account, it underscores the urgent and foundational nature of the technology's ecological impact. The report explores the energy-intensive nature of AI systems, the infrastructure that sustains them, and the environmental impacts that must be addressed as AI scales up and expands its implementation. Understanding these trade-offs is essential to shaping a future where AI development is not only intelligent, but equitable, ethical, and ecologically sustainable.

2. AI's Growing Global Energy Demand

Stacks of Clover Power Station in Clover, Virginia, United States, a coal-fired plant whose retirement has been postponed until 2045 as electricity demand grows with the expansion of AI-related data centers. The delay is part of a wider pattern in which older coal facilities are being kept in service to support rising power needs from digital infrastructure. Photo by Emw (Wikimedia Commons).

The rapid rise of AI presents a growing sustainability challenge. Large-scale AI models demand substantial energy for both training and inference, and their appetite is growing faster than the world can decarbonize. Fossil fuels continue to dominate global electricity supply while low-carbon sources have not yet scaled enough to replace them[45]. The additional electricity required by AI makes the transition to renewables and sustainability more challenging by further increasing energy demand and amplifying the environmental impacts of power generation. This chapter highlights the energy demands associated with training complex models and operating hyperscale data centers—and their carbon, water, and land footprints.

2.1 Energy Intensity of AI Training

Training advanced AI models requires processing enormous datasets—often billions of words, images, or other data points—using high-performance accelerators (GPUs/TPUs) in large data centers. Accelerators draw much higher power than general-purpose CPUs and generate heat that requires cooling. AI training is not a single, instantaneous task, or a one-off event; it is an extended, resource-intensive process. Training can run for days to months, with repeated, compute-intensive cycles designed to fine-tune accuracy. Because state-of-the-art models are trained on billions of tokens or images, each additional pass multiplies energy[46]. Achieving peak accuracy typically requires retraining the model multiple times with different configurations. This sustained activity draws large amounts of electricity, both for computing and cooling, making AI training a major consumer of energy and a significant contributor to its environmental footprint.

2.2 AI's Energy Demand and its Environmental Footprints

Every kilowatt-hour of electricity used to train or run an AI model carries environmental footprints, including a carbon footprint from the generation mix; a water footprint from electricity production and cooling; and a land footprint from energy infrastructure, reservoirs, and fuel extraction. These three footprints do not always shift in the same direction[47,48]. For example, switching from coal to bioenergy can, on average, reduce the carbon footprint by 72%, but this comes at the cost of much larger water and land footprints. On average, the water footprint of bioenergy is more than 30 times greater than that of coal and its land footprint is 100 times greater. In different regions and countries, electricity is produced from various sources.

The environmental footprint of energy production in a given location depends on the share of each source in its electricity supply portfolio. For example, in Brazil, where hydropower dominates the grid, the carbon footprint of electricity generation is 77% below the global average. But this comes with other environmental costs: the water and land footprints are nearly triple the global mean.

In other words, the environmental cost of AI models depends not only on how much electricity they use but also on where that energy comes from. Recognizing these trade-offs is critical: reducing pressure on one footprint must not simply transfer the burden elsewhere. Distributional impacts also matter, since regions hosting AI infrastructure often bear the heaviest environmental costs, while the benefits of AI flow elsewhere. As models continue to grow in both size and complexity, understanding current and future energy use is vital for assessing their full environmental impacts.

2.3 AI's Training Footprints

OpenAI's Generative Pre-trained Transformer 3 (GPT-3), first released in June 2020, is

Aerial view of Microsoft's data-center complex in Middenmeer, within Hollands Kroon, Netherlands. It consumed about 84 million liters of water in 2021 during a year of severe drought, far exceeding earlier estimates of 12–20 million liters, which has led to sustained opposition from local farmers over water use. Data Source: Sentinel-2 imagery, August 2025.

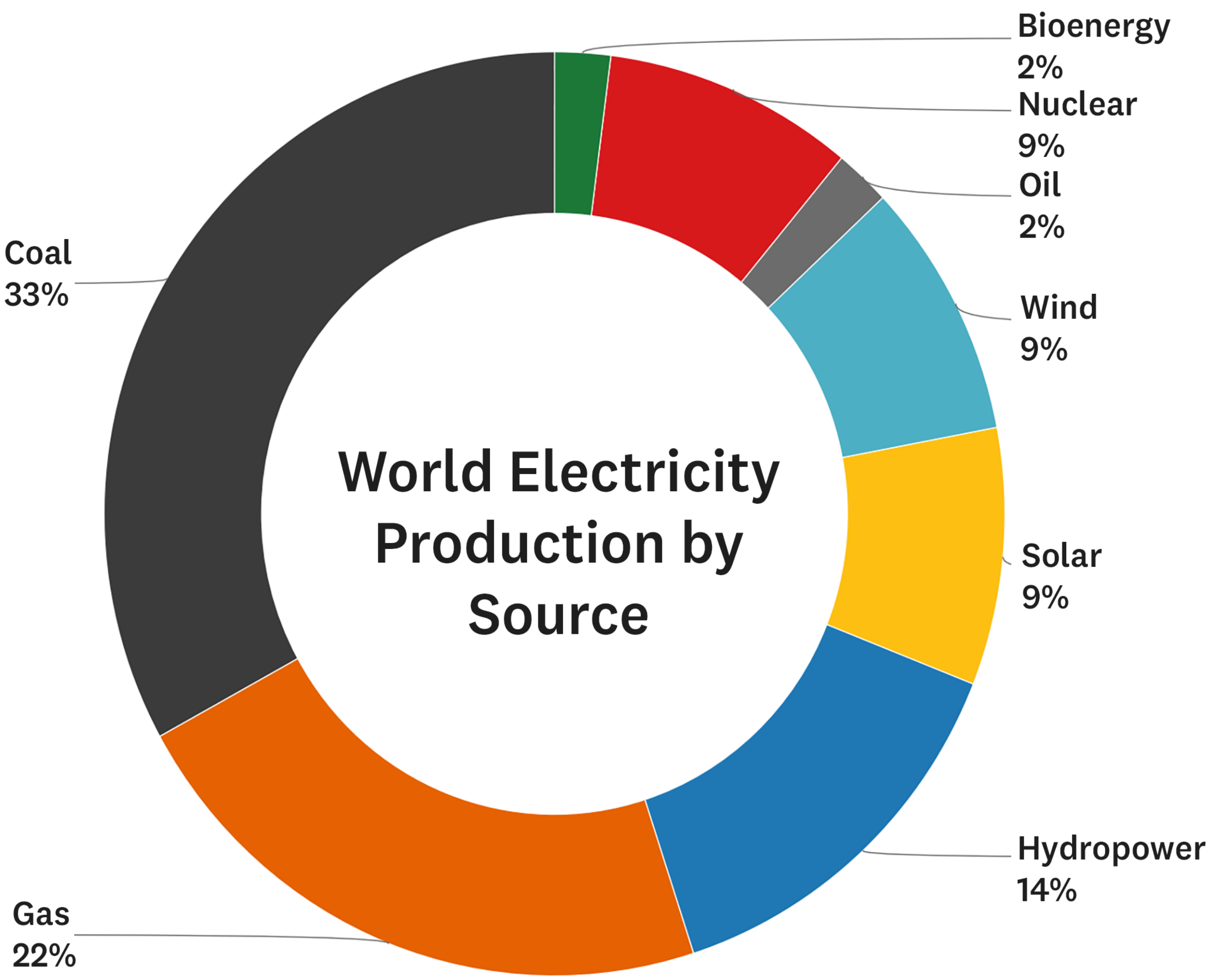


Figure 4. Global electricity supply by source. Global electricity generation is broken down by fuel and technology in 2025, highlighting the continued role of fossil fuels alongside the growth of low-carbon sources. The mix provides context for why the marginal electricity used by AI often still carries substantial carbon, water, and land footprints. Data is from Ember, 2026.

one of the earliest large-scale AI systems for which detailed training energy estimates are available, making it a useful benchmark for understanding the environmental costs of frontier models. Training GPT-3 required approximately 1.287 GWh of electricity over a 34-day period[44]. GPT-4 represented a dramatic escalation in scale. Though OpenAI has not disclosed official figures, independent analyses suggest training likely consumed 50-70 GWh over a period of 100 days, around 40-55 times more than GPT-3[46].

At the midpoint of the range, 60 GWh of electricity is equivalent to the annual residential electricity consumption of 460,000 people in Sub-Saharan Africa[49]. Using the global intensities[47,48,50–53], this training would be equivalent to over 25,000 tonnes CO_2e, requiring the sequestration capacity of approximately 420,000 urban tree seedlings grown for 10 years, or about 105 Hyde Parks' worth of trees in London. The water footprint of 592 million liters is enough to meet the minimum annual water needs of roughly 81,000 people in Sub-Saharan Africa[59], or to fill 237 Olympic-sized pools. The land footprint of 0.9 km² is equal to over 125 football (soccer) fields, or an area about one-quarter the size of New York's Central Park.

No training energy data are yet available for GPT-5, but scaling up from GPT-3 and GPT-4 suggests that a next-generation model could plausibly require on the order of 100 GWh, with associated footprints of roughly 42,000 tonnes CO_2e, one billion liters of water, and a land footprint of 1.5 km² of land, or roughly the size of 215 football fields. Offsetting this carbon would require about 700,000 tree seedlings grown for 10 years, equivalent to roughly the number of trees in 180 Hyde Parks in London. This level of electricity use is equivalent to the annual residential electricity consumption of approximately 770,000 people in Sub-Saharan Africa and the water footprint is sufficient to meet the minimum annual water needs of roughly 137,000 Sub-Saharan residents.

As new frontier models emerge, energy used in training continues to vary widely. Meta's LLaMA 3–405B, released in 2024, reportedly consumed 21 GWh of electricity, equivalent to nearly 31 million GPU-hours of compute[54]. The energy use and environmental footprints of LLaMA 3-405B were roughly one-third of those

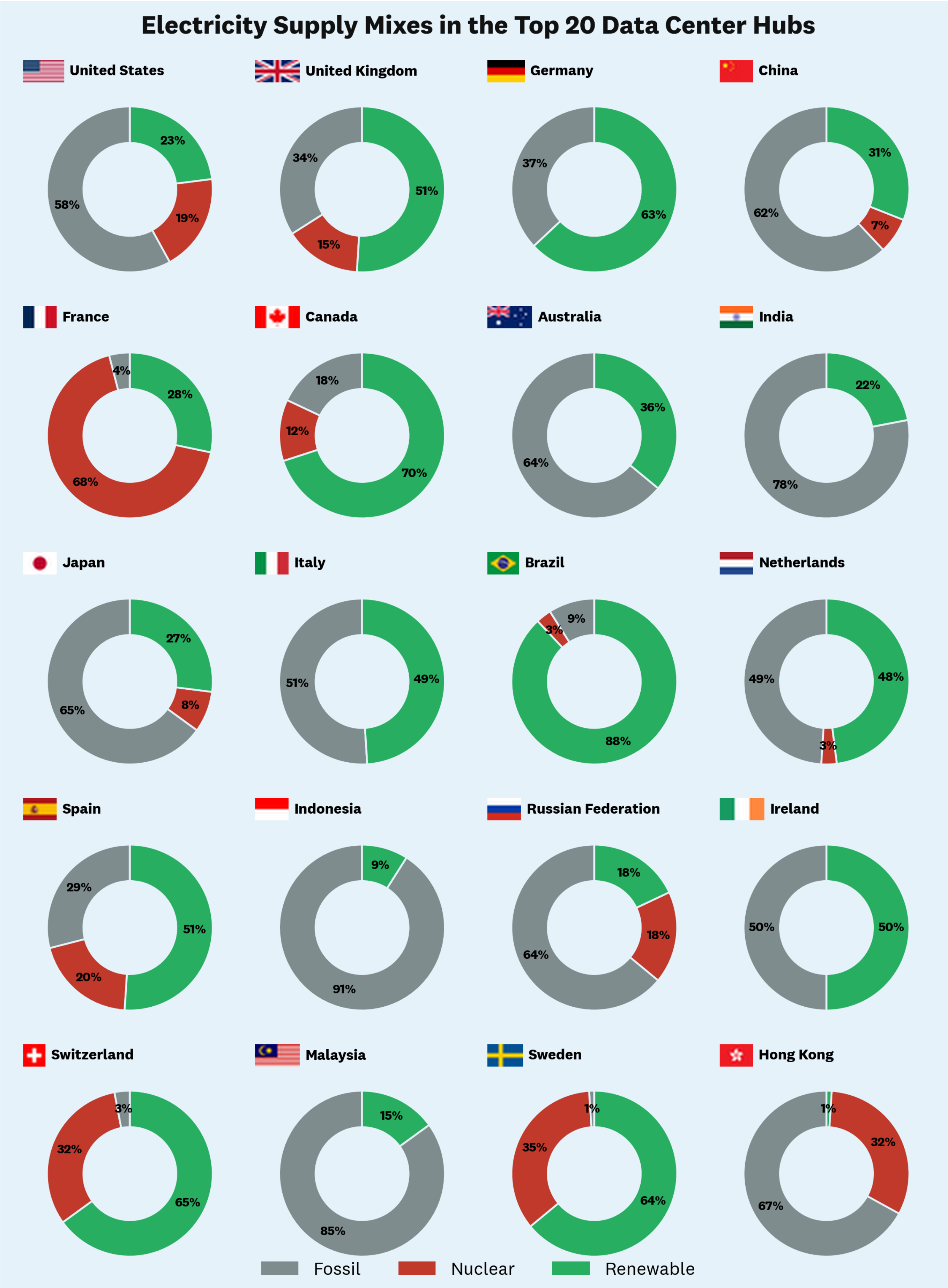


Figure 5. Electricity supply mixes in major data center host locations. National generation profiles across leading data center locations reveal wide differences in reliance on fossil fuels, renewables, and nuclear power. These underlying electricity mixes shape the carbon, water, and land footprints of data center operations and therefore influence the environmental impacts of AI deployment across regions.

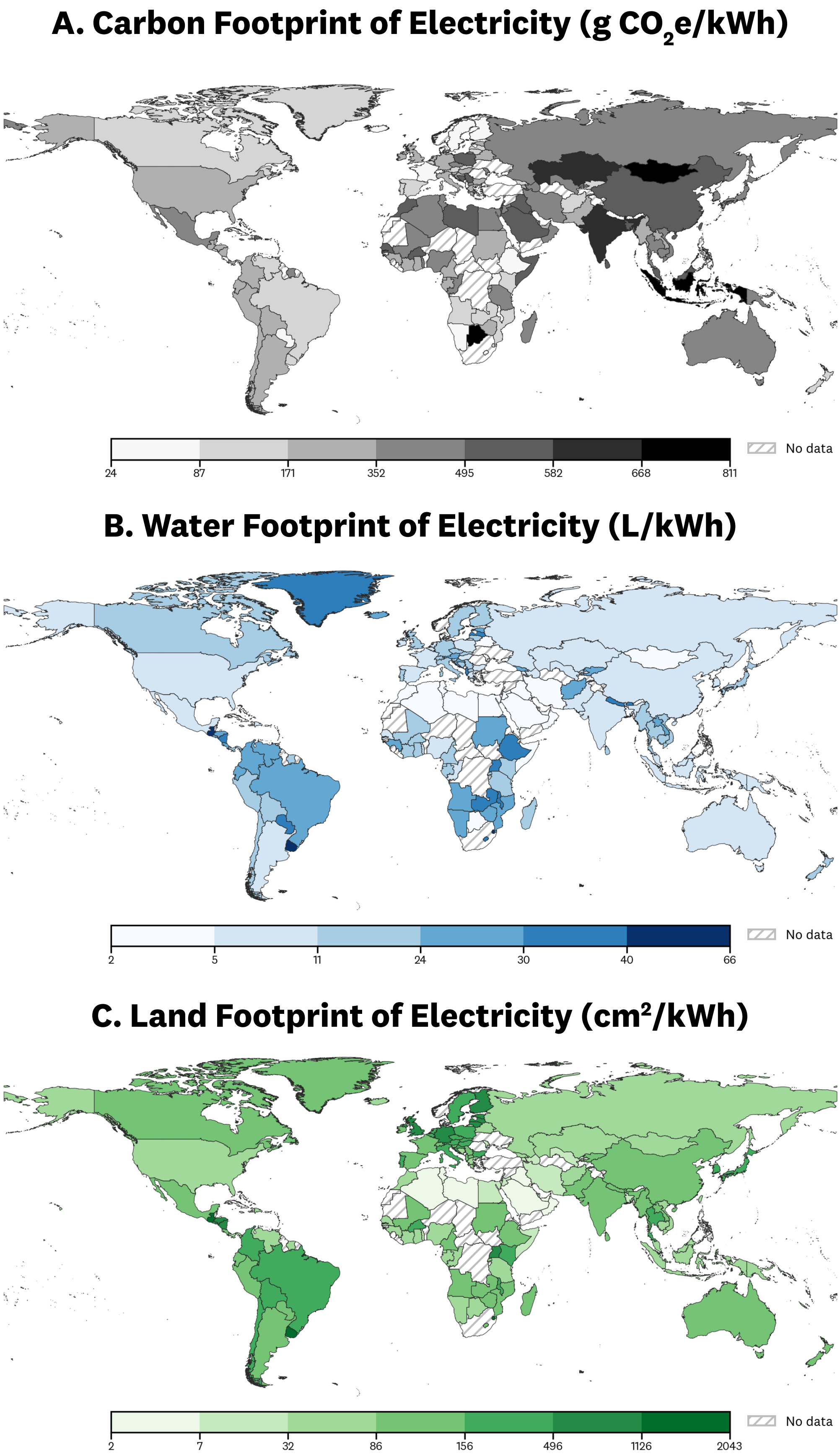


Figure 6. Global electricity footprints. Maps show country-level electricity footprint intensities for carbon, water, and land associated with AI electricity use. Shading indicates grams of CO_2-equivalent per kWh, liters of water per kWh, and square centimeters of land per kWh. The figure illustrates how national power system characteristics drive geographic variation in the environmental footprints of AI systems.

of GPT-4. Released in late 2024, the Chinese-developed model DeepSeek-V-3 required 80-90% less compute for training than Meta's LLaMA, but this information remains unverified[54,55].

Despite growing concerns about AI's environmental impact, only a few model cards—documentation tools meant to disclose model characteristics—provide the numbers needed to estimate their full energy and carbon footprints. A recent example of improved transparency comes from OpenAI's release of GPT-OSS-120B, a 120 billion parameter open-source model intended to give the research community a high-performing public baseline. Its model card includes architecture details, training data composition and notably, energy consumption, enabling estimation of associated carbon, water, and land footprints. The reported 2.5 million kWh of energy used for training makes the environmental footprints of GPT-OSS-120B almost 30 times smaller than those of GPT-4.

These examples show both the scale of training required for large models and the variability across different architectures. While efficiency gains can reduce the energy used per model, falling costs lead to wider deployment of AI across products and services—potentially increasing overall AI demand and use. This phenomenon is known as the rebound effect (or Jevons Paradox). When efficiency lowers costs, overall consumption may rise, offsetting or even reversing the benefits and efficiency gains. This poses a challenge for making AI environmentally sustainable solely through improved efficiency.

2.4 Data Centers and AI's Expanding Energy Demand

To understand the energy and environmental footprints of AI, we need to examine the infrastructure that enables it: data centers. These sprawling facilities—often housing thousands of high-performance processors—consume vast amounts of electricity to operate, cool, and transmit data across networks. They are the engines of the digital age, powering an ever-expanding range of services. From video streaming and online gaming to e-commerce and cloud computing, data centers underpin nearly every aspect of modern life. Increasingly, they also serve as the backbone of AI, supporting the training and deployment of large-scale models that power recommendation systems, generative chatbots, search engines, and autonomous technologies. As the global appetite for AI grows, so does the demand for energy-intensive infrastructure, driving rapid expansion in both the number and size of data centers, especially hyperscale sites that consume immense amounts of power.

In 2024, global data center electricity consumption was estimated at about 415 TWh (415 billion kWh), roughly 1.5% of global

Rebound Effect (Jevons Paradox)

The rebound effect, often referred to as Jevons Paradox, is named after the 19th-century economist William Stanley Jevons, who observed that improvements in the efficiency of coal use in England did not reduce coal consumption. Instead, lower costs led to expanded use, driving higher overall demand. The paradox captures a counterintuitive dynamic: efficiency gains can increase total resource use rather than decrease it. This dynamic is increasingly relevant to AI. As AI models improve, architectural and hardware advances significantly reduce the energy required for tasks such as inference. These efficiency gains lower the cost of computation and enable deployment at unprecedented scale.

AI tools are becoming embedded across everyday systems, from search engines and smartphones to logistics platforms, industrial control systems, and household devices. This widespread integration drives rapid growth in aggregate compute demand, which can outweigh per-task efficiency improvements. As a result, the total energy use of AI can rise even as energy demand per query falls. Without complementary demand-side measures, governance mechanisms, or regulatory guardrails, efficiency gains alone are unlikely to deliver absolute reductions in energy use. Instead, they risk accelerating environmental pressures by expanding the scale and intensity of AI applications faster than efficiency improvements can offset their cumulative impacts.

Footprint of Electricity in the World's Top 20 Data Center Hubs

Rank	Carbon Footprint (g CO_2e per kWh)	Water Footprint (L per kWh)	Land Footprint (cm^2 per kWh)
1	Indonesia 682	Brazil 29	United Kingdom 718
2	India 635	Canada 21	Germany 515
3	Hong Kong (SAR) 604	Switzerland 21	Brazil 445
4	Malaysia 543	Sweden 21	Sweden 350
5	China 510	United Kingdom 20	Italy 332
6	Australia 481	Germany 15	Netherlands 319
7	Japan 462	Italy 15	Japan 256
8	Russian Federation 384	Japan 11	Ireland 156
9	United States 345	China 10	China 153
10	Germany 322	Malaysia 9	Canada 148
11	Ireland 299	Netherlands 9	Spain 142
12	Italy 288	India 8	India 136
13	Netherlands 280	Russian Federation 8	Switzerland 124
14	United Kingdom 218	Spain 8	France 106
15	Spain 144	France 7	Australia 98
16	Canada 138	Indonesia 6	Malaysia 77
17	Brazil 97	Ireland 6	United States 74
18	France 51	Australia 6	Indonesia 43
19	Sweden 41	United States 5	Hong Kong (SAR) 40
20	Switzerland 37	Hong Kong (SAR) 3	Russian Federation 32

Figure 7. Footprints of electricity in the world's top 20 data center hubs. Carbon, water, and land footprint intensities of electricity supply are shown for major data center host countries. Differences in national power mixes drive cross-country variation in AI environmental footprints and determine how data center location shapes overall impacts.

Global Distribution of Data Centers

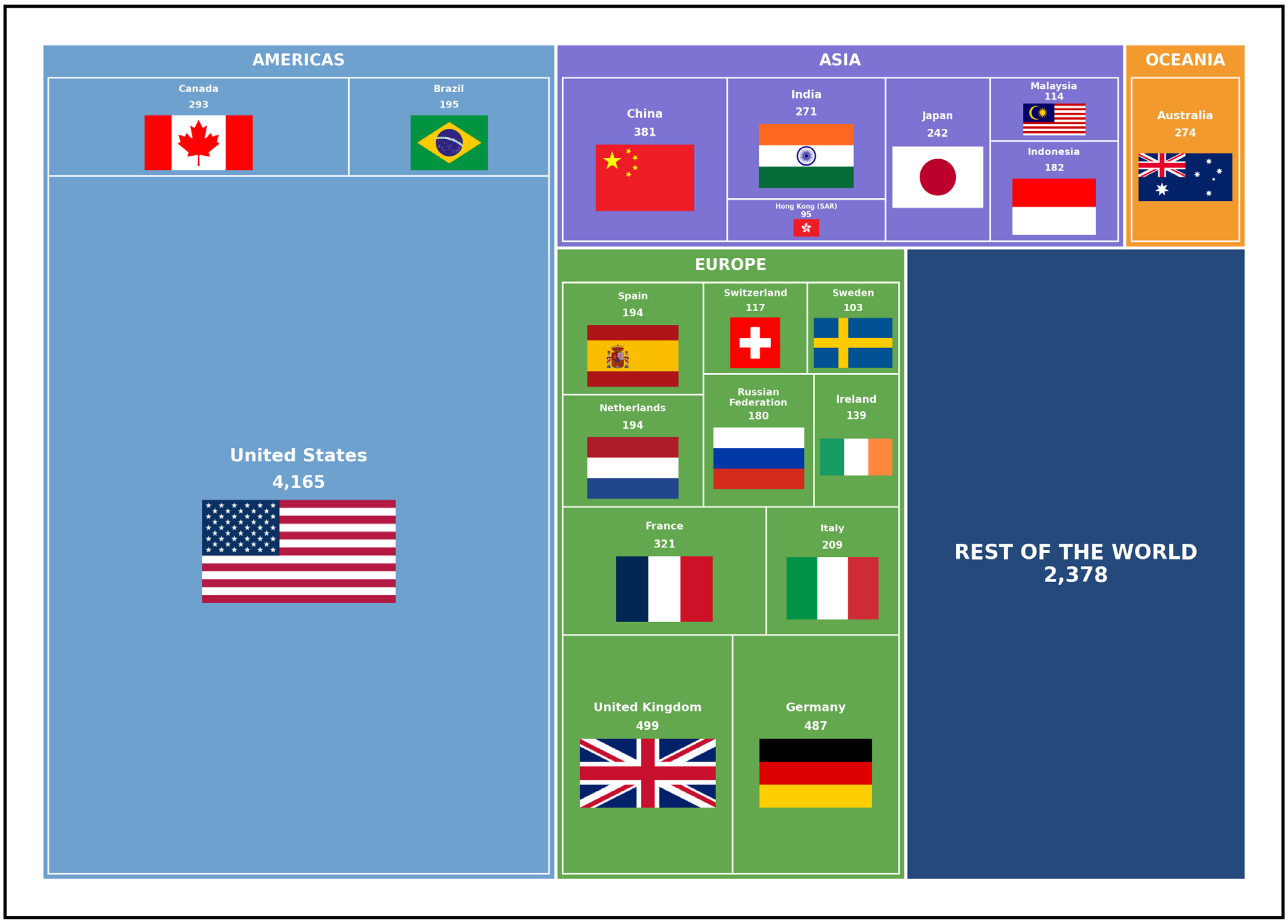


Figure 8. Global distribution of data center locations. Data center capacity is highly unevenly distributed worldwide. A small number of countries host the majority of global compute infrastructure, while many regions remain underrepresented. Concentration patterns shape both the geography of local environmental burdens and the uneven global access to digital infrastructure and compute capacity. Data from Statista 2025, based on Cloudscene.

electricity demand[57]. In 2025, global data centers were estimated to consume 448 TWh of electricity. This amount of electricity is enough to supply the annual residential electricity needs of the entire population of Sub-Saharan Africa, 1.3 billion people, for over 2.5 years[49,57]. Taking the global average electricity footprint intensities[47,48,50–53], annual data center operations imply a carbon footprint of 189 million tonnes CO_2e. Offsetting this carbon footprint would require over 3.2 billion tree seedlings grown over 10 years[58]—about the same as the total number of trees in the United Kingdom. The associated water footprint of the energy used by data centers in 2025, around 4.5 trillion liters, could fill 1.8 million Olympic-sized swimming pools or meet the minimum annual domestic water needs of over 620 million people in Sub-Saharan Africa[59]. The land footprint of the energy produced to run data centers in 2025 was around 6,900 km^2, about 4.5 times bigger than Greater London, 117 times the size of Manhattan, or over 2,000 times the land size of New York's Central Park.

If data centers' electricity consumption were a country in 2025, their 448 TWh of electricity use would have ranked 11th globally behind the electricity consumption of France (468 TWh) and ahead of Saudi Arabia (422 TWh)[60].

That amount of electricity used by data centers was almost 2.7 times greater than the annual electricity use of Pakistan (5th largest population globally, with about 250 million), 4 times greater than the electricity use of Bangladesh (8th most populous country, with over 172 million), and 12 times greater than the annual electricity use of Nigeria (the 6th most populous country, with a population of 224 million).

If current trends hold, data center electricity consumption could exceed 945 TWh by 2030, roughly doubling the 2025 figure and accounting for almost 3% of projected global electricity[56]. That amount is nearly triple the combined electricity demand of Pakistan, Bangladesh and Nigeria with a total population of over 650 million, or enough to power the annual residential electricity needs of the entire population of Sub-Saharan Africa (1.3 billion people) for about 5.5 years.

The carbon footprint of producing this much electricity averages about 399 million tonnes CO_2e, requiring 6.7 billion trees to offset, around twice the number of trees in the entire United Kingdom. The average water footprint of producing the required electricity is about 9.32 trillion liters, enough to satisfy the annual basic water needs of the entire population of Sub-Saharan Africa, while the land footprint is roughly 14,500 km². This land footprint is 18 times bigger than the area of New York City, nearly 10 times bigger than Mexico City, more than 11 times bigger than the City of Los Angeles, and more than twice the size of the Jakarta metropolitan area that houses more than 32 million people.

2.5 AI's Contributions to Data Centers' Energy Use

AI is now one of the largest drivers of data center energy consumption. In 2025, AI workloads alone were estimated to account for over 20% of total electricity use in data centers, amounting to approximately 93 TWh[61]. This amount of electricity could supply the annual residential electricity needs of roughly 715 million people in Sub-Saharan Africa—more than half the region's population. Using the global average intensity factors, this equates to 40 million tonnes CO_2e (equivalent to the 10-year sequestration of 661 million urban seedlings—about 78 times the number of trees in London), over 900 billion liters of water (365,000 Olympic-sized pools, or enough to provide minimum daily water for 125 million people in Sub-Saharan Africa for an entire year), and 1,400 km² of land, an area slightly smaller than the size of Greater London, 23 times the size of Manhattan, or 195,000 football fields.

If AI workloads alone were a country, its electricity use in 2025 would rank 39th in the world, above industrialized countries like Finland (83 TWh) and Belgium (82 TWh). This electricity could satisfy the entire energy demand of a country like Nigeria—the world's 6th largest nation with a population of 224 million people—for more than two years.

If AI's share of data center demand rises to 40% by 2030, as some projections suggest, its electricity consumption could reach 378 TWh[56] (enough to meet the residential electricity needs of all 1.3 billion people in Sub-Saharan Africa for about 2.2 years) with a carbon footprint of 158 million tonnes CO_2e—equivalent to the amount offset by 2.61 billion seedlings over 10 years (about twice the number of trees in England). The associated water footprint of the electricity demand of AI by 2030 would be 3.7 trillion liters, enough to supply the minimum water needs of more than 500 million people in Sub-Saharan Africa for a year[59]. The associated land footprint of this much electricity is 5,744 km², 3.6 times the size of Greater London, equivalent to 95 Manhattans. At this level of energy consumption, if AI workloads were a country, it would rank 14th globally with over 3 times the electricity consumption of Bangladesh (the 8th most populous country globally, with its population of 174 million), or over 9 times the energy consumption of Nigeria (the 6th

Data center in Ashburn, Virginia, United States. One of the world's largest concentrations of data centers, the area hosts dense clusters of facilities powering global cloud and AI services, underscoring how digital infrastructure is concentrated in a few regions amid a widening global divide. Photo by Emily Richardson (Flickr).

most populous country in the world, with its population of 224 million).

2.6 World's Top Data Center Hosts and Their Electricity Mixes

Data centers are not uniformly distributed across the globe. Nearly half of the world's facilities are in the United States, which hosts over 4,000 sites. The next tier includes Germany and the United Kingdom (roughly 500 each), followed by Mainland China, then France, Canada, Australia, India and Japan. Other countries and territories include Italy, Brazil, the Netherlands, Spain, Indonesia, the Russian Federation, Ireland, Switzerland, Malaysia, Sweden, and Hong Kong (SAR), followed by Poland.

Because the electricity mixes of the world's data center hubs differ dramatically—from coal-heavy to nuclear- and hydro-dominated systems—the environmental footprints of data center and AI operations across them also vary, even when facilities share identical engineering designs and hardware.

Carbon intensities vary by an order of magnitude across the major data center hubs of the world. Indonesia, India, and Hong Kong (SAR) are among the most carbon-intensive grids with carbon footprints 62%, 51%, and 43% higher than the global average, respectively. Poland and Mainland China rank lower with carbon intensities at 30% and 21% higher than the global average. By contrast, the carbon footprint of electricity in the United States, Germany, and Italy is 18%, 24%, and 32% below the global average. At the lowest end, France, Sweden, and Switzerland (37 g) fall about 88–91% below the global average, reflecting their reliance on nuclear energy and hydroelectricity.

Water footprints of electricity production also differ sharply, especially in hydropower-dominated grids. In Brazil, Canada, Switzerland, and Sweden, water intensities are more than double the global average—ranging from about 110% higher (Sweden) to 191% higher (Brazil). By contrast, grids in Hong Kong, the United States, Australia, and Singapore have much lower water intensities, with footprints between 43% and 66% below the global average.

Electricity mixes with high shares of bioenergy, onshore wind, or hydropower are associated with larger land footprints per unit of energy. For example, the land-intensity of the national electricity grid in the United Kingdom is more than four times the global average. Germany,

Spotlight: The Growing Global Divide in AI Infrastructure

The expansion of AI infrastructure is accelerating electricity demand and widening global disparities in technological capacity. As of late 2025, only 32 countries host AI-specialized data centers with the high-performance chips and cooling systems required for advanced AI workloads[36]. Over 90% of these facilities are concentrated in two countries, the United States and China, while entire regions such as Africa and South America remain virtually absent from the global AI data center map. More than 150 countries lack such infrastructure altogether. This uneven distribution directly shapes digital sovereignty, innovation capacity, and equitable access to AI.

Developers in under-resourced countries often lack sufficient computational power, forcing reliance on costly cloud services hosted abroad. This raises costs, slows response times, constrains local innovation, and locks institutions into external platforms. Disparities are reinforced by data and model development patterns. Large foundation models are trained primarily on datasets from high-resource languages and regions where digitized data and compute are abundant. Applications serving smaller linguistic communities and local contexts therefore receive less investment, weaker optimization, and limited tooling, constraining locally tailored AI and embedding existing inequities into digital systems. Some countries are beginning to invest in sovereign and regional compute infrastructure. Brazil and India have announced national AI cloud initiatives, while pan-African collaborations are emerging to pool resources. These efforts remain early-stage. Without coordinated strategies to expand access to compute, data, and technical capacity, AI's benefits risk remaining concentrated in a small number of highly resourced nations.

Aerial view of the Poulaphouca Reservoir (Blessington Lakes) in County Wicklow, Ireland, captured during the 2018 summer drought when water levels fell sharply and restrictions were imposed across the Greater Dublin region. As Ireland expands energy- and water-intensive AI data-center infrastructure around Dublin, growing cooling and power demands risk adding further pressure to already stressed freshwater supplies. Sentinel-2 imagery, October 2018.

Brazil, Sweden, and Italy also fall well above the global mean, each with a land footprint between twice and four times greater per unit of electricity than the global average. By contrast, the United States and Hong Kong (SAR) are among the top data center hubs of the world with the lowest land footprint at only a fraction of the global average.

As AI demand scales, regional differences in electricity mixes translate into large differences in the carbon, water, and land footprints of AI operations, creating concentrated local environmental impacts alongside global impacts from greenhouse gas emissions. Because data centers serve distributed and often cross-border users, countries that host large compute clusters tend to capture strategic advantages such as employment, tax revenues, digital sovereignty, and security capabilities, while nearby communities may bear disproportionate local environmental burdens, including water stress, land use pressures, and air pollution. By contrast, countries with limited digital infrastructure experience few local environmental impacts from AI operations but face growing disadvantages in access to compute capacity, digital services, and AI-driven innovation, reinforcing the global digital divide and its economic and security implications. The different environmental footprint figures across the data center hubs of the world also highlight that "low-carbon"

Ireland: small country, outsized data-center load

Ireland offers a cautionary example of local grid stress from concentrated digital infrastructure. By 2023, data centers accounted for 21% of Ireland's total metered electricity, up from 5% in 2015, exceeding the combined electricity use of all urban households[64]. In response to capacity constraints around Dublin, the national grid operator EirGrid paused new data center connection approvals in the region until 2028. Ireland's experience highlights the need for responsible siting and capacity planning so that rapid AI infrastructure growth does not outpace local power systems.

grids are not automatically “low-water” or "low-land".

Conversely, fossil-heavy systems emit far more carbon, yet typically correspond to lower water and land intensities. Evaluating carbon, water, and land footprints together is therefore essential, since minimizing one dimension can magnify another. These trade-offs illustrate how renewable-heavy systems can cut emissions but might result in larger land footprints, questioning the common assumption that data centers that rely on renewable energy sources are "green", "clean", and "sustainable"[62,63].

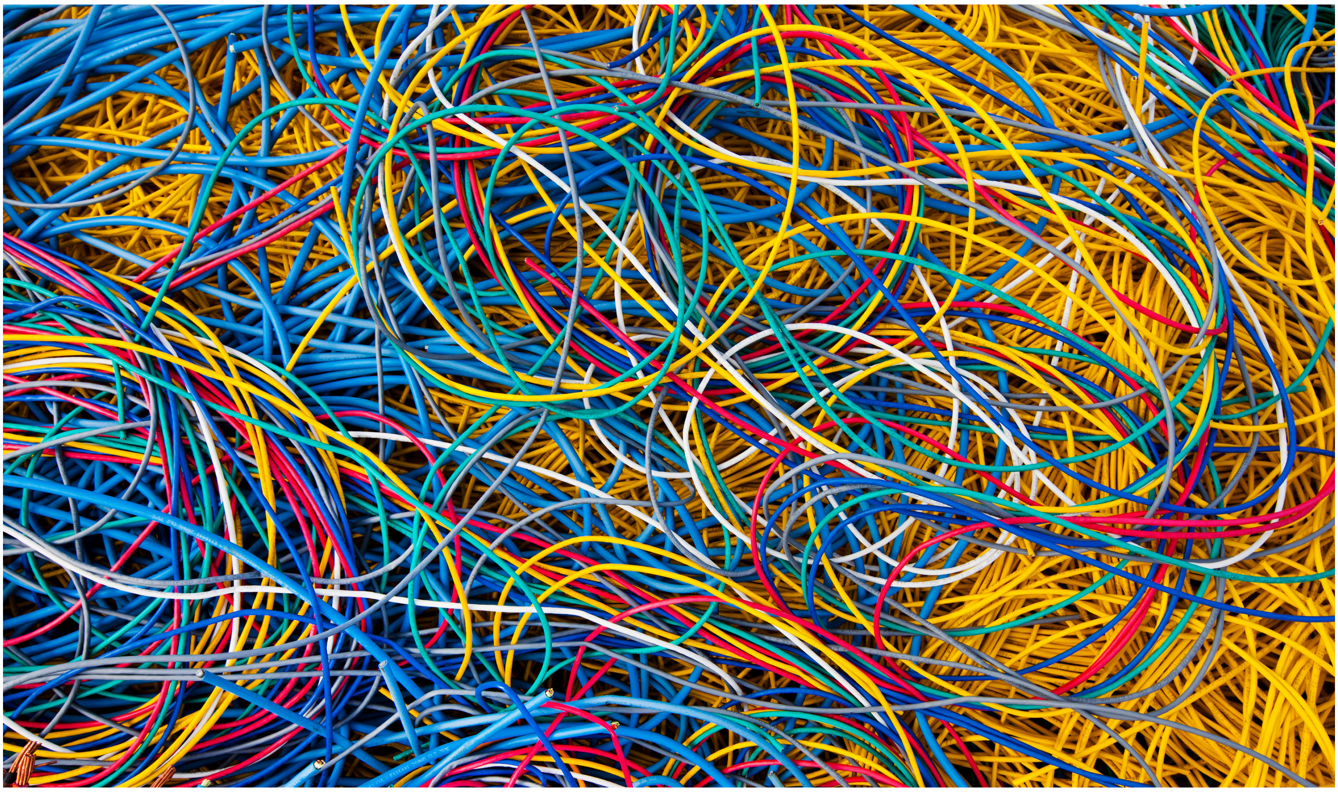
2.7 Local Costs, Distant Benefits

In addition to electricity demand, data centers can place direct pressure on local water supplies because most high-density server facilities rely on water-based cooling. Even when a portion of withdrawn cooling water is returned, the volume and timing of large-scale withdrawals can place significant stress on aquifers and surface water systems, particularly in arid regions or areas already experiencing groundwater depletion.

Beyond operational electricity, constructing data centers and manufacturing GPUs, batteries, and servers rely on critical minerals (lithium, cobalt, and other rare earths). Early in the life cycle, mining is energy-, water-, and land-intensive and often occurs in jurisdictions with weaker environmental protections.

Aerial view of the Google data center at Eemshaven, Netherlands, surrounded by wind turbines and a nearby power plant. As AI expands, data centers like this require large amounts of electricity, making energy supply a key challenge. Photo by Wvdp (Wikimedia Commons).

At the end of life, obsolete hardware feeds the e-waste stream: AI infrastructure could generate up to 2.5 million metric tons of e-waste annually by 2030, equivalent to discarding nearly 250 Eiffel Towers every year. Without effective recycling, heavy metals (such as lead, cadmium, and mercury) can contaminate soil and water. These upstream and downstream impacts underscore the need for holistic, full-lifecycle governance of the AI supply chain, from mineral extraction to responsible end-of-life management.



A maze of cables connecting data center equipment at Google’s data center site in The Dalles, Oregon. The dense network of cables supports continuous AI and high-performance computing workloads, while also adding to the material footprint and future e-waste burden of large-scale data center operations. Photo from Google

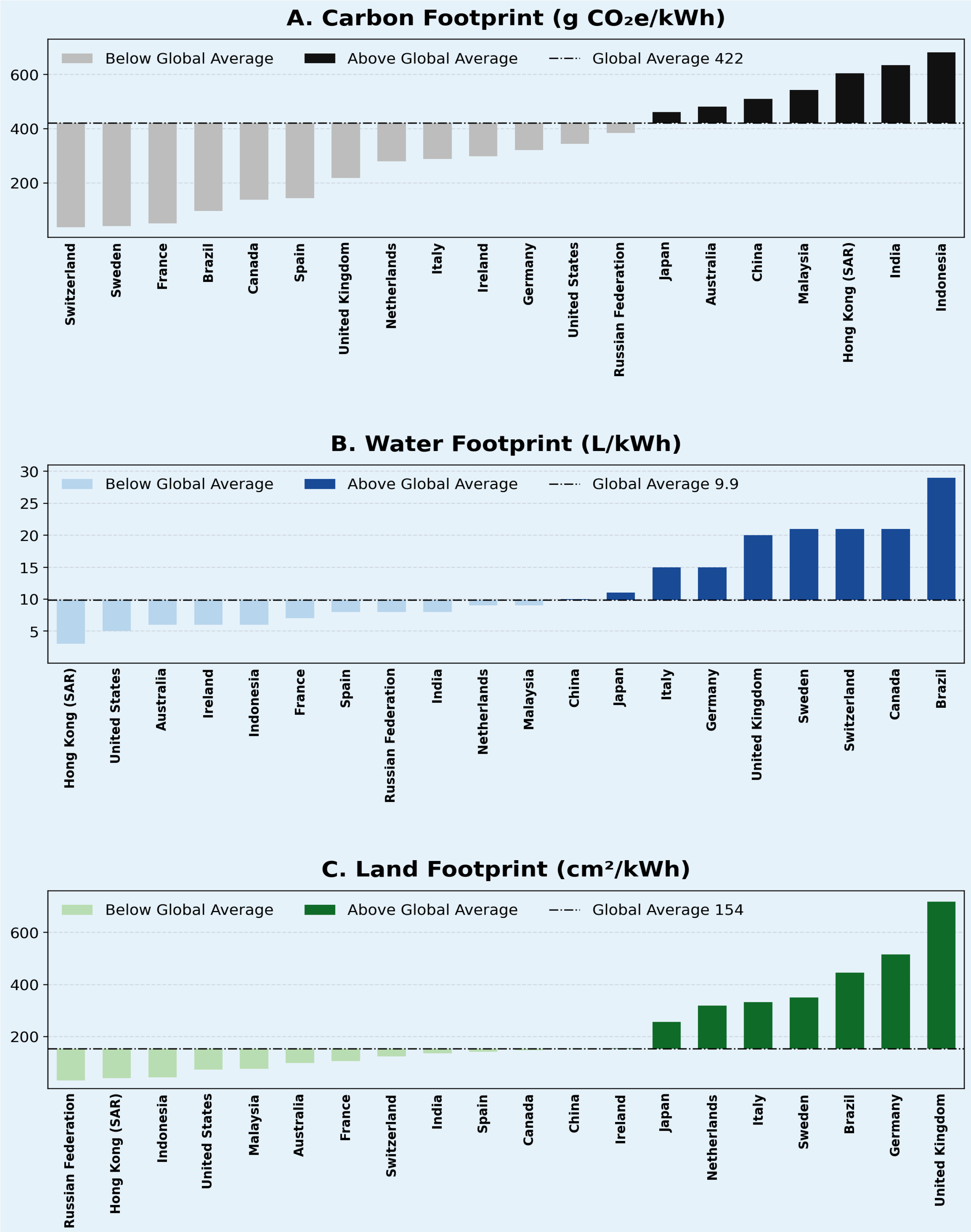
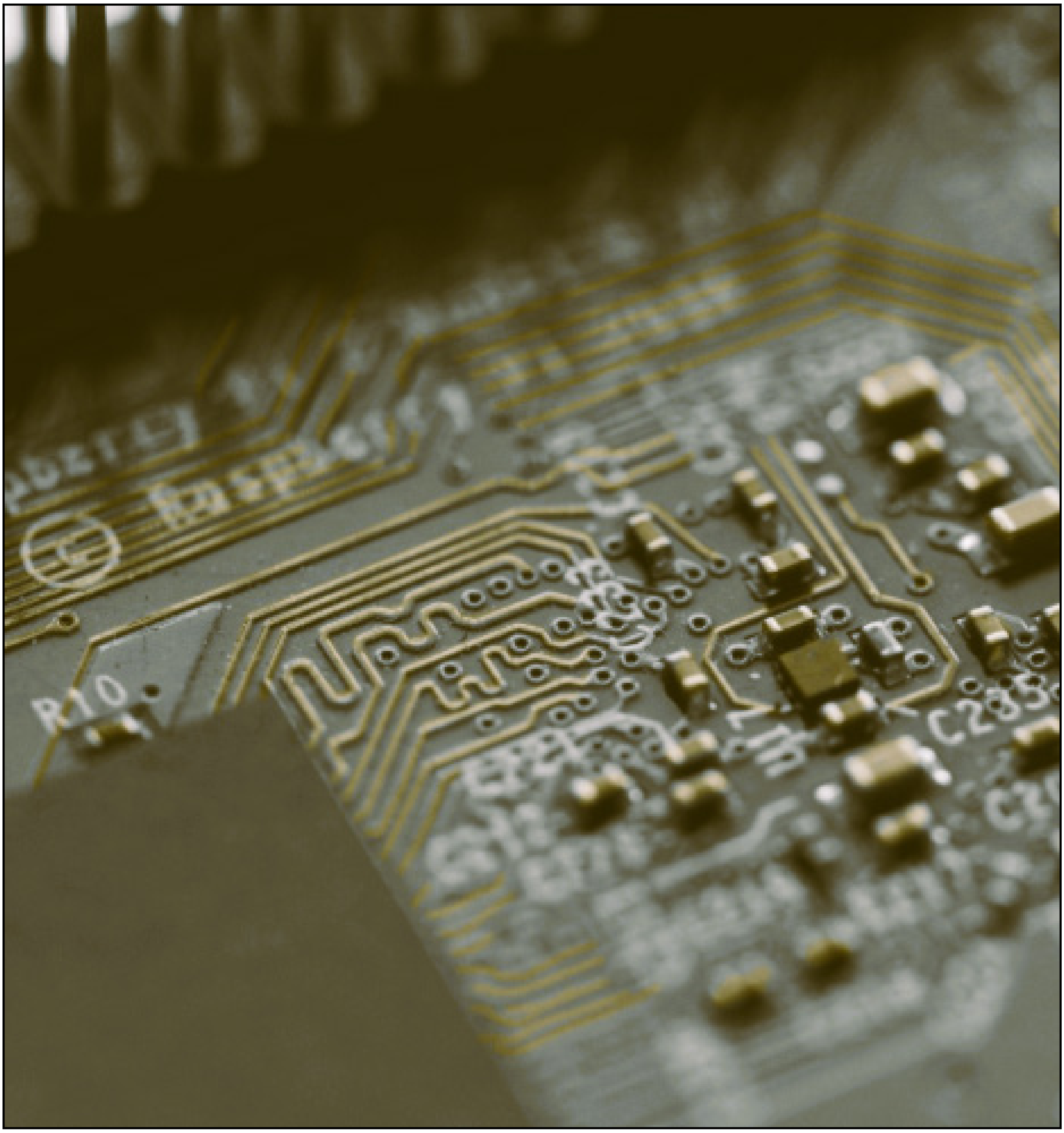


Figure 9. Global electricity footprint intensities across the world's top 20 data center hubs relative to the global average. Carbon, water, and land impacts associated with AI electricity use vary across locations due to differences in national power mixes, meaning identical data center designs can generate substantially different environmental impacts depending on where electricity is produced. Bars show carbon, water, and land footprint intensities for electricity generation in major data center host locations, with the global average shown as a reference line.

3. AI in Use: Tools & Tasks



AI accelerator hardware components used to execute neural network inference and training workloads, showing specialized compute chips designed for high-performance machine learning operations. Photo by BobbyJSmith1972 (Wikimedia Commons).

Much of the conversation around AI's environmental impact has focused on the energy-intensive process of training large models, which require vast datasets and GPUs running for weeks or months. Training a model can require thousands or even millions of GPU-hours, consuming as much electricity as dozens or hundreds of households over a year. Yet training is only one part of the picture and the tip of the iceberg, in the long run, as the environmental impacts of AI increasingly arise from the continuous running of models to generate responses for billions of user interactions each day. To understand AI's true environmental cost, it is also necessary to examine the operational impact of AI systems in everyday use—not just how models are built and trained but also how they are used and at what cumulative scale. Once a model is deployed, it shifts into the inference phase, where it generates outputs in response to user prompts. Each individual query consumes far less energy than training, yet when multiplied by billions of users, inference becomes the dominant contributor, estimated to account for 80–90% of total energy use[44].

AI has now moved beyond research labs into everyday life. Chatbots answer questions, image generators illustrate ideas, and voice tools transcribe or read text aloud. AI is already an integral part of the daily digital experiences of billions of people worldwide, sometimes without their awareness. While around 380 million people actively use dedicated AI tools, the number increases dramatically when platforms like Google Search are included[65]. With over 5 billion Google Search users and about 14-16 billion searches conducted each day[66], AI is now deeply embedded in features such as summaries, recommendations, and contextual responses. As a result, Google alone engages over a billion people in AI-assisted interactions each day[67]. AI applications are transforming work, education, and entertainment, but as each interaction draws electricity, multiplied billions of times, the environmental costs quickly become staggering. ChatGPT, for example, processes around 2.5 billion prompts daily[4].

3.1 Two Forces Shaping AI's Operational Footprint

AI's operational environmental footprint is shaped by two major forces. First is *how much* we use AI models: the sheer volume of inferences across billions of daily interactions.

Second is *how* we use them: the task and model choices behind each interaction, which determine the energy required per query. A spam filter and an image generator are both

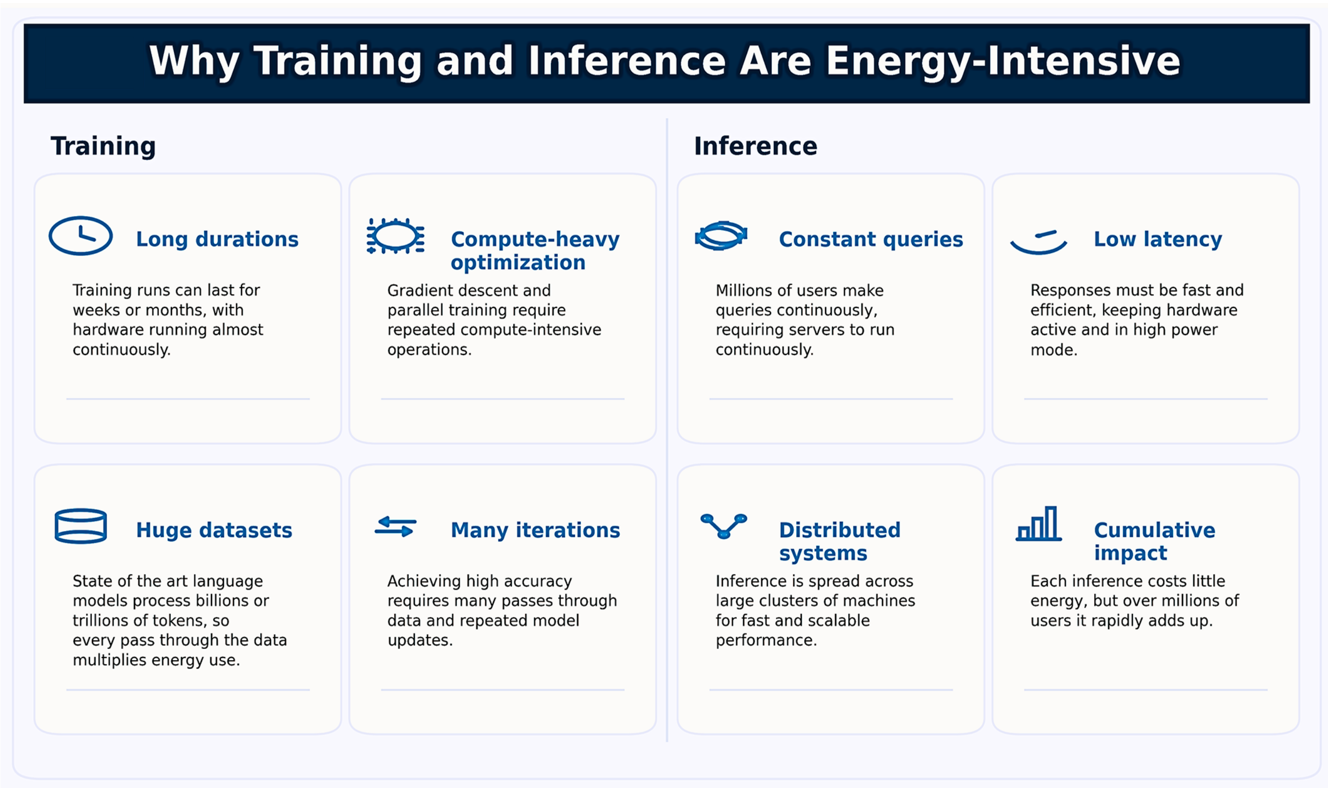


Figure 10. Energy demand in AI arises from two distinct phases. Training concentrates electricity use over weeks or months of compute-intensive optimization across massive datasets. Inference distributes energy use across billions of real-time queries, where low per-query consumption, combined with scale and latency requirements, generates substantial cumulative impacts.

forms of AI, but they do not draw the same power. Text classification typically uses orders of magnitude less energy per query than image generation or long-form text synthesis. And even within a single task, complexity varies: model size, context window, prompt length, output length, and other settings can substantially shift the per-query energy use. Model choice matters as each AI model performs each task with a different energy and environmental cost. Even when using the same AI platform, users' choices of which model to run (e.g., Instant vs. Thinking) can change the energy use. Model choice also makes the location of AI operations relevant, because the same watt-hour carries different carbon, water, and land footprints depending on the grid that supplies it.

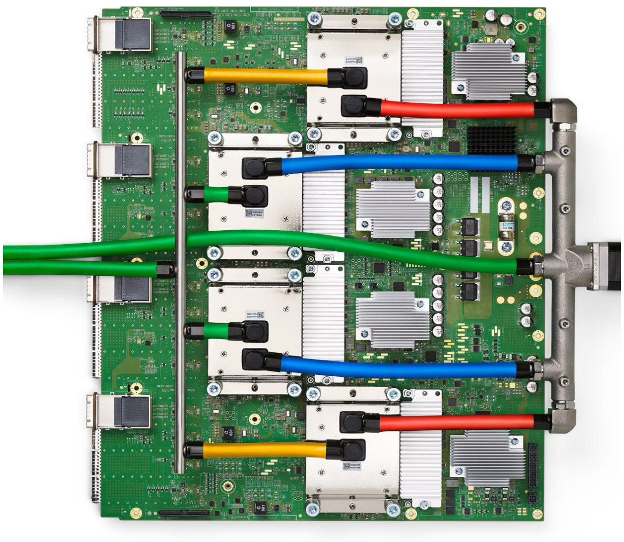
Put simply: Environmental Impact = (energy per query for the task/model) × (number of queries) × (grid footprints).

3.2 AI Tasks and Their Footprints

AI systems today support a broad spectrum of tasks, from short text responses, longer written outputs, code generation, image and audio production, and increasingly video generation. Each of these tasks differs substantially in the computational effort required to produce a single result.

Text classification and simple queries

Text classification tasks such as spam filtering or basic sentiment analysis sip energy. Short



Tensor Processing Unit (TPU) v4 board used in large AI data centers to train and run advanced machine learning models. These custom-built chips handle the massive calculations required by modern artificial intelligence systems and are linked together in large clusters to increase computing power. Because they consume substantial electricity and generate significant heat, they require extensive cooling and infrastructure, contributing to the environmental footprint of AI data centers. Photo by Norman P. Jouppi et al. (Wikimedia Commons).

text interactions generally consume the least energy. Extractive question-answering systems that simply retrieve short snippets from documents (for example, ChatGPT, Claude, or Gemini used purely as retrieval tools) are in the same range.

Even image-classification models such as Google Vision or Azure Computer Vision remain relatively light: they use only a few times more energy per image than text classification and are still orders of magnitude below high-end generative models. Object-detection models, which identify and localize multiple objects in a scene, are somewhat heavier but still on the "low-energy" side of the spectrum compared to text- or video-generation models.

Generative text and summarization

Short-form text generation (brief answers or summaries) sits higher on the energy spectrum. A compact LLM generating a few sentences uses a few hundredths of a watt-hour per query, roughly two dozen times more energy than a spam-filter-style classification, but still modest in absolute terms. Larger, general-purpose LLMs such as GPT-3-scale models use roughly an order of magnitude more energy again. A typical "mainstream" ChatGPT-style query is about 200 times more energy-intensive than basic classification, and long GPT-4 and GPT-5-style responses with extended context windows can approach 1,000 times the classification baseline. State-of-the-art LLMs span roughly 0.3 to more than 10 Wh per query depending on model size, prompt length, and deployment settings, a range of over one to two orders of magnitude even within text-only tasks[44,68–71].

Image generation

Text-to-image diffusion models use substantially higher amounts of energy than text generation. A "typical" AI image is roughly 60 times more demanding than a short text answer and 1,450 times that of text classification. Some high-end open models, such as Lumina, push these numbers still higher[72] reaching over 2,000 times the energy of short text generation. At the other end of the spectrum, efficient diffusion architectures such as SDXL-Turbo or LCM SSD-1B can cut per-image energy use by roughly 30–45 times, depending on the diffusion baseline[72].

Video generation: the new energy frontier

Video generation is one of AI's most energy-intensive frontiers. Text-to-video (T2V) diffusion models extend image synthesis across time, producing clips with dozens of frames at HD or higher resolution. Each additional pixel and frame increases compute cost, and in diffusion models that cost grows rapidly. For common T2V models, energy scales quadratically with spatial resolution and frame count, and linearly with the number of 'denoising' steps[73]. Measurements using default settings show a cross-model range from around 0.14 Wh for a very low-resolution short clip to more than 400 Wh for a high-resolution, long clip on a large model.

At that upper end, a single short AI video can draw as much electricity as tens of thousands of spam classifications or hundreds of typical images, before even scaling up to millions or billions of clips.

Because energy scales quadratically with resolution and frame count, lower-resolution and shorter-clip presets deliver the largest savings. Halving the number of frames (e.g., 81 to 40) cuts energy to roughly one-quarter, saving about 75%. Halving both width and height (e.g., 720×1280 to 360×640) reduces energy by roughly 1/16, saving about 94%. Halving both spatial dimensions and frame count yields about 1/64 of the original energy—around 98% savings in the compute-bound regime. Halving the number of denoising steps cuts energy roughly in half.

If both resolution and steps are reduced (for example, halving height, width, frames, and steps), total energy can fall to roughly 1/128 of the original, or a savings of about 99.2%, though with clear trade-offs in video quality (blurrier images and less natural motion). These scaling laws make clear that video models, especially at high resolution and frame counts, can rival or exceed the energy demand of many everyday appliances for a single short clip.

3.3 AI use at scale

Today, ChatGPT alone processes approximately 2.5 billion prompts per day worldwide[4]. If an average text-based request uses about 0.42 Wh, this corresponds to roughly 383 GWh per year. This amount of electricity would be enough to meet the annual domestic electricity demand of nearly 3 million people in Sub-Saharan Africa, and would be accompanied by about 160,000 tonnes of CO_2e. Offsetting this amount of emissions would require about 2.6 million tree seedlings grown for 10 years, requiring a land area roughly the size of Manhattan. The corresponding water footprint is about 3.8 billion liters per year, sufficient to cover the minimum annual domestic water demand of 500,000 people in Sub-Saharan Africa. The corresponding land footprint is about 5.9 km^2, over one and a half times the area of New York's Central Park, or about 800 football fields. Image generation has expanded rapidly alongside this growth in overall AI use.

Public syntheses of platform data indicate that more than 15 billion images were generated using text-to-image algorithms between 2022 and 2023 alone, with estimates suggesting that roughly 34 million AI images were generated each day in 2023[74].

Following this initial surge, image-generation tools were integrated directly into conversational AI systems, creative software, design platforms, and consumer applications, substantially increasing their frequency of use. A significant share of image-generation activity occurs outside centralized platforms. Diffusion-based open-source models account for the majority of AI-generated images to date, yet much of this activity is not comprehensively captured in public usage statistics. As a result, reported figures likely understate the true global scale of image generation, particularly beyond major U.S.-based platforms.

Following the observed growth and adoption of AI tools in mainstream use, assuming that approximately 300 million AI-generated images

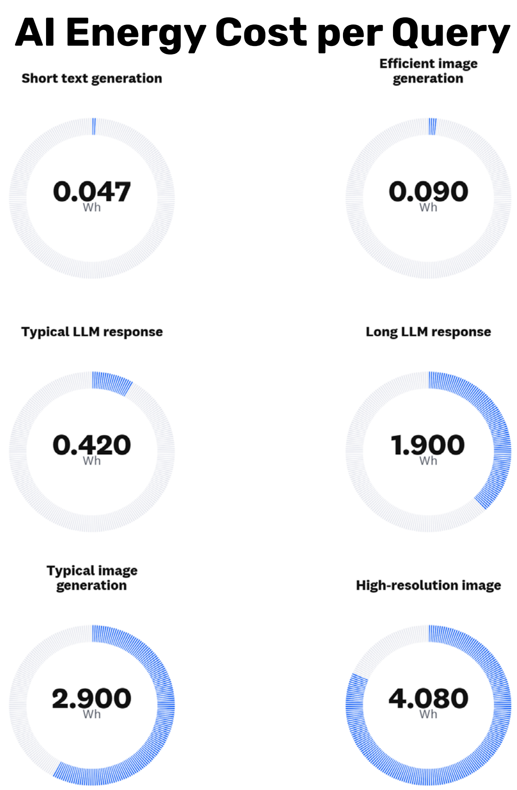


Figure 11. Energy per query for representative text and image generation tasks. This figure compares mean electricity consumption per query across common AI applications. Basic text classification and short text generation require minimal energy.

are produced daily, global energy required from typical (2.9 Wh) image generation alone would reach roughly 318 GWh of electricity per year, enough electricity to meet the annual domestic needs of nearly 2.5 million people in Sub-Saharan Africa, and a corresponding carbon footprint of over 134,000 tonnes CO_2e per year which would require over 2.2 million tree seedlings grown over 10 years to offset. This has a water footprint of over 3 billion liters per year, enough to meet the minimum domestic water needs of more than 400,000 people in Sub-Saharan Africa for a full year, and a land footprint of nearly 4.8 km², roughly 665 football fields.

Video currently represents a much smaller fraction of total AI interactions than text or images, yet each output is substantially more computationally intensive. Generative video tools have progressed rapidly from research demonstrations to commercial and consumer deployment, including short-form clips, animations, and visual effects integrated into creative software and social media platforms. Although comprehensive global usage statistics remain limited, the trajectory of image-generation adoption provides a useful analogue: as multimodal capabilities are embedded into widely used systems, new tasks scale proportionally with overall AI usage even if they remain minority use cases.

Assuming global AI video generation reaches 25 million clips per day in the near future, one should expect substantial environmental impacts. If all videos were short, relatively efficient clips averaging around 90 Wh each, this would imply roughly 821 GWh of electricity per year, enough to meet the domestic consumption of more than 6 million people in Sub-Saharan Africa.

This would generate about 347,000 tonnes of CO_2e (requiring over 5.7 million tree seedlings grown over 10 years—over 300 times the number of trees in New York's Central Park), more than 8 billion liters of water (enough to meet minimum annual domestic water needs for over one million people in Sub-Saharan Africa), and a land footprint of about 12 km², or more than 1,700 football fields.

However, video generation spans a wide range of computational intensities. If only 20% of these videos, or about 5 million clips per day, are higher-complexity videos requiring approximately 415 Wh each (while the remaining 80% remain at the lower 90 Wh level), total electricity demand for AI video generation rises to approximately 1.4 TWh per year. Under this mixed-use assumption, the water footprint exceeds 13 billion liters, and the land footprint expands to more than 21 km² —an area roughly six times the size of Central Park in New York. The associated electricity demand would be sufficient to meet the annual domestic electricity needs of around 11 million people in Sub-Saharan Africa, while the water required would cover the minimum annual domestic needs of nearly 2 million people. The associated annual carbon emissions of 600,000 tonnes of CO_2e would require nearly 10 million tree seedlings grown for 10 years to offset, equivalent to 550 times the number of trees in Central Park. These estimates illustrate how a relatively small share of high-energy video workloads can disproportionately shape the overall environmental footprint of AI when scaled to global levels, and the importance of tools and task choice in addition to total query volumes in understanding AI's environmental impacts.

3.4 Conventional Search Versus AI-Enhanced Search

A conventional web search without an LLM component is often assumed to cost around 0.3 Wh of electricity per query[75]. With an estimated 5 trillion annual searches[76], this amounts to roughly 1.5 TWh of electricity annually. If each of those searches were upgraded to an AI-enhanced mode that used 3 Wh per query, ten times the energy of a conventional search, the annual energy demand would rise to about 15 TWh, enough to meet the annual domestic energy demands of 115 million people in Sub-Saharan Africa, requiring over 100 million seedlings to offset its carbon footprint (nearly 5,800 times the number of trees in New York's Central Park).

The water footprint would be enough to satisfy the annual domestic water needs of around 20 million people in Sub-Saharan Africa, and the

Power transmission lines along the W&OD corridor in Ashburn, Virginia, carrying electricity toward one of the world's largest concentrations of cloud and AI data centers. Source: Photo by Kit Case (Flickr).

land footprint is equivalent to 31,000 football fields.

In practice, energy use for AI-based search varies widely. Some lightweight retrieval-augmented models operate at about 0.003–0.01 Wh per query, comparable to conventional search. Other generative answer engines consume more energy: recent analyses report LLM queries in the range of 0.05–0.3 Wh each for standard prompt lengths and model sizes[77]. The key takeaway is retrieval-style AI search can approach the efficiency of conventional search, but generative search features that produce new text can quickly raise energy per query—and when multiplied by billions of queries, these differences compound dramatically.

3.5 Efficiency Improvement and Rebound Effects

As technology and knowledge improve, different techniques are used to reduce per-query energy. Model compression, pruning, quantization, knowledge distillation, mixture-of-experts routing, on-device execution, and specialized accelerators can all lower the energy required for inference. However, these gains do not automatically translate into lower total environmental impact. Greater efficiency is normally associated with an increase in overall consumption (Jevons Paradox) by making services and products cheaper and more attractive, encouraging new uses and higher volumes. In the AI context, when models become cheaper and faster, people are expected to rely on them more often and for more tasks, eroding or even reversing the savings from efficiency improvements.

Aerial view of the Rhine River during severe drought in August 2022, showing exceptionally low water levels that disrupted a major European industrial corridor. The continued expansion of large data centers along the Rhine, some designed to use river water for cooling, adds pressure to a water body already under environmental stress. Data source: Sentinel-2 imagery.

Effective policies should therefore pair efficiency with resource budgets (for example, caps on tokens, GPU-hours, or kilowatt-hours) and should prioritize small-model and retrieval-first options for routine tasks such as fact look-ups, simple calculations, and basic summarization.

3.6 Behavioral Energy Costs: Model Choice and User Prompts Matter

A single spam classification, a short chat reply, a long text answer, an AI-generated image, and a short video clip are all experienced as a single “query” from the users' point of view, but they have radically different results in the energy spectrum. Even within the same category of AI task, the environmental footprint of an individual query can vary by orders of magnitude depending on model choice (related to its computational efficiency and the environmental intensity of its servers), and model configuration[78] .

Across diffusion image models, published comparisons suggest that energy use can differ by several dozen-fold across architectures and implementations, meaning that choosing an efficient image model can reduce per-image energy by more than an order of magnitude relative to high-end alternatives. These differences are often invisible to users, but they are key to understanding how everyday design and usage decisions scale into meaningful environmental impacts. Similar-looking outputs often mask very different resource costs. As a result, users and institutions currently lack the information needed to make environmentally informed decisions about how and when to use AI. This makes transparency a prerequisite for responsible AI use: without visibility into relative resource intensity, neither efficiency gains nor behavioral change can meaningfully reduce AI’s environmental footprint.

Model choice is not the only variable defining how we use AI. Even the length and structure of prompts, users' expectations and output settings, and form of user interactions matter. A typical conversational response of a language model can use about 200 times more energy than a classification task and roughly 8 to 10 times more than short-form text generation. Longer or more elaborate responses push this further, reaching around 500 to 1,000 times the energy of classification while still performing the same basic function of text generation. In practice, model choice plus response length can easily drive differences of up to two orders of magnitude across everyday text interactions.

Within text-based tasks, short-form text generation from compact or distilled language models typically requires roughly 10 to 25 times more energy than text classification, despite appearing similar at the surface. Moving to mainstream conversational large language models increases energy demand by another order of magnitude. For image and video generation, model choice and quality settings create an even wider spread. Video generation appears to sit at the top of the intensity ladder and shows the greatest sensitivity to settings like resolution, frames, and denoising steps.

Energy Use of AI Video Generation Under Different Quality Settings

Each square's area equals the percentage of baseline energy use

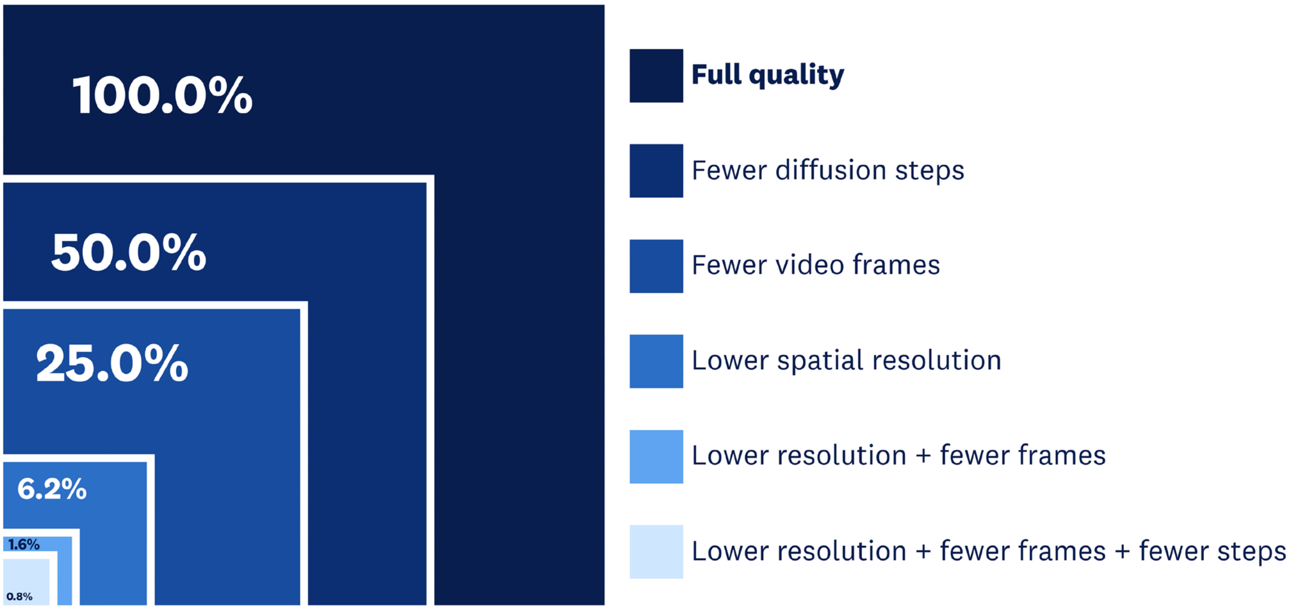


Figure 12. How video generation settings shape energy demand. Bars show relative energy use for text-to-video generation under different resolution, frame count, and denoising step settings, normalized to a full-configuration baseline. Reducing spatial resolution, frames, and steps sharply lowers energy demand, with combined reductions yielding over 99% savings at the cost of lower video quality.

Message-Level Efficiency ("Concise Mode")

Most ChatGPT use falls into high-volume categories such as practical guidance, information seeking, and writing. Because inference energy scales with tokens processed, reducing verbosity can yield material savings at platform scale. As an illustrative scenario, if a concise mode reduced tokens by about 30% for common interactions, cutting per-query energy by roughly 25%, assuming an average 0.42 Wh per prompt and 16–18 billion weekly queries, would save roughly 87–98 GWh of electricity per year, equivalent to the annual residential electricity use of 672,000–756,000 people in Sub-Saharan Africa, assuming 130 kWh per person per year.

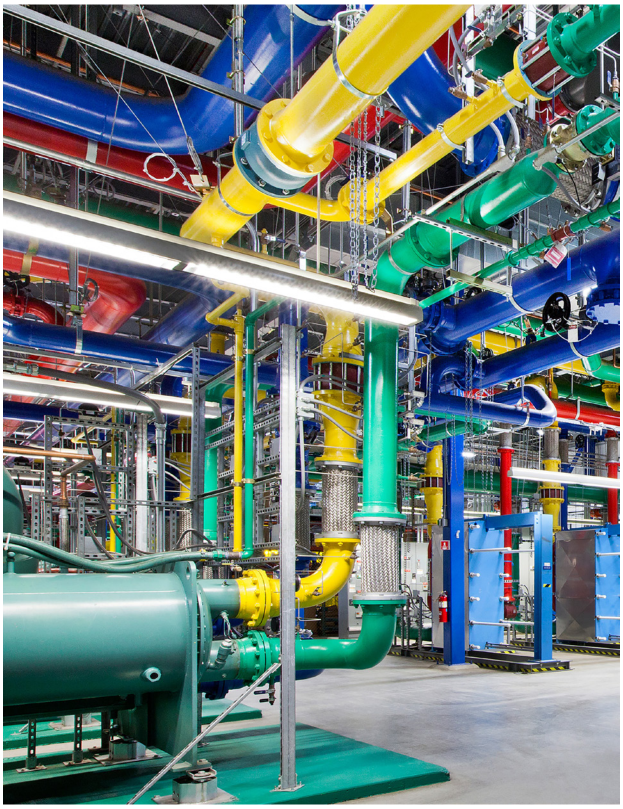
Roof of a data center showing cooling equipment and backup generators. These systems support temperature control and power continuity for high-density computing operations by managing heat and providing on-site electricity during outages. Photo by Rsparks3 (Wikimedia Commons).

3.7 Towards sustainable use and scale

Across tasks, AI's environmental footprint varies widely: a single high-fidelity video generation can consume as much electricity as tens of thousands of text classifications. This disparity points to where interventions matter most: optimizing high-energy tasks, curbing unnecessary inference, and prioritizing lightweight architectures and retrieval-based tools for everyday needs. The environmental footprint of AI is not determined solely by model size or by the energy required to train a model once. Instead, two intertwined factors shape its impacts: how much we use AI and what we use it for. When billions of people interact with AI each day, even a few milliwatt-hours of energy per query add up to gigawatt-hours of demand.

These patterns have clear justice implications. The electricity used and the environmental resources impacted to generate verbose chat responses, AI-crafted images, or high-fidelity videos are resources that may be unavailable to communities that still lack reliable power or clean water.

By choosing concise interactions where feasible, favoring low-impact tools for simple tasks, and pairing technical efficiency with responsible demand management, users, developers, and policymakers can harness AI's benefits while steering it toward a more just and sustainable trajectory. Resource awareness in decision-making can reduce avoidable emissions and rebound effects.



Water inlet and outlet pipes at the Google data center at Dalles, Oregon.

4. The Way Forward

AI for Good 2025 conference session in Geneva, 8 July 2025, where policymakers, UN agencies, and technology experts convened to discuss global AI governance, responsible innovation, and international policy coordination. Photo by Rowan Farrell / ITU Pictures (Flickr).

Artificial intelligence is now threaded through daily life, shaping how we learn, create, communicate, and govern. Each interaction carries a material cost—expressed in the energy it consumes and, through energy, in the carbon, water, and land footprints that accompany electricity generation. This UNU-INWEH report shows that AI's footprint is not simply a matter of how much we use these systems but how we use them: task type, model choice, deployment, and user practices all drive impacts, often by orders of magnitude. The aim of highlighting the environmental footprints of AI is not to slow progress, but to align innovation with stewardship, to build more capable AI systems that also use resources wisely and fairly.

It is also crucial to recognize impacts beyond the footprints quantified in this report. Issues such as critical minerals extraction, hardware durability to end-of-life, and direct cooling practices in data centers carry significant environmental and social implications, some of which are concentrated in the Global South (e.g., the pollution resulting from mining rare earths). While this report focused on energy-linked water, land, and carbon footprints, policy and practice must also address these broader, interlinked material, social and justice dimensions to ensure that environmental responsibility and sustainability extends across the AI supply chain.

4.1 Guiding Principles for a Responsible AI Ecosystem

A sustainable AI ecosystem depends on six mutually reinforcing principles that shape how systems are built, deployed, and used:

1. Transparency makes impacts visible through standardized, comparable disclosure of energy use and the resulting water, land, and carbon footprints. Clear reporting enables accountability, public scrutiny, and informed choice.

2. Efficiency by design ensures that energy and other resources are used wisely. This includes quantization, pruning, distillation, Mixture-of-Experts routing, reuse of compute, and on-device execution when feasible.

3. Equity and justice ensure that people living with higher grid carbon, water, and land intensity, or with constrained resources, do not bear disproportionate burdens from AI development and deployment.

4. Lifecycle responsibility recognizes that every unit of energy used by AI carries associated water, land, and carbon impacts, and that hardware production, mineral sourcing, and disposal also have consequences. Responsibility must extend from materials to manufacturing to use, and end-of-life.

5. Global cooperation harmonizes methods, disclosures, and safeguards so that benefits are shared rather than fragmented.

6. Sustainable use acknowledges that everyday decisions, like choosing low-footprint tasks, smaller models, or conventional search when appropriate, are core components of responsible AI application.

Together, these principles provide a foundation. The next step is turning them into concrete practices for measurement, design, governance, and everyday use.

Artisanal miners working in the Katanga Copperbelt, DR Congo, extracting copper used in AI servers, data centers, and power infrastructure. While this labor supports the global AI economy, local communities often remain exposed to unsafe working conditions and environmental damage and receive little share of the economic gains captured by technology companies and high-income countries. Photo by Bas van Abelé/ Fairphone (Flickr).

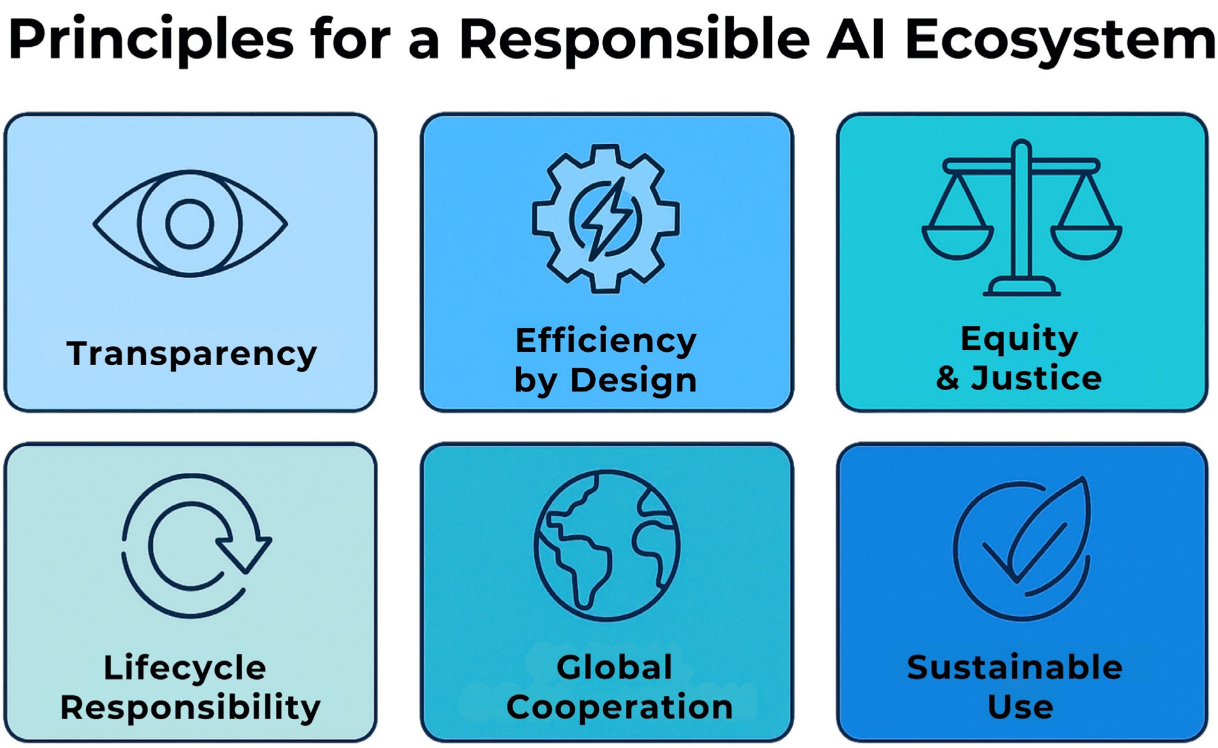


Figure 13. Six guiding principles for operationalizing responsible AI. This framework guides AI toward environmental stewardship and social responsibility. It emphasizes transparency, efficiency by design, equity and justice, lifecycle responsibility, global cooperation in measurement and governance, and sustainable use to promote lower-footprint choices.

4.2 From Measurement to Action

Efficiency and intelligent use must be matched by fairness in where impacts fall and who benefits.

Make the invisible visible. AI developers and service providers should publish clear, standardized accounts of energy and environmental footprints for training and inference reported in standardized units. Reporting should be consistent across models, tasks, and regions so results can be verified and compared. Providers should separate training, deployment, and task-level inference so users and policymakers can see which workloads drive the largest impacts. Public transparency dashboards can accelerate convergence on good practice and discourage selective disclosure that creates misleading impressions of progress.

Design for efficiency and manage rebound. Model-side improvements such as quantization, pruning, distillation, and Mixture-of-Experts routing can cut per-task energy substantially. Deployment choices such as efficient accelerators, high-utilization clusters, and carbon-aware routing also reduce waste. But efficiency alone is not enough. Organizations should adopt resource budgets, measured in tokens, GPU-hours, or kilowatt-hours, and set growth-aware targets so total demand does not eclipse savings. Where large models are unnecessary, “small-model-first" defaults and on-device execution should be the norm. This is especially important for high-intensity workloads such as image and video generation, where energy use increases quickly with resolution, duration, and frame count. Efficient defaults, including lower-resolution and shorter-clip settings, can generate substantial savings at scale and should be clearly presented in user interfaces. Because high-fidelity settings in image and video generation drive disproportionate energy growth, efficiency-by-default should include explicit limits and disclosures around resolution, frame count, and output length.

Use AI intelligently. Users, organizations, and public institutions should match tools to tasks. Not every task requires a large generative model. For many routine lookups, conventional search or local tools are typically more efficient. Some lightweight, task-specific models can consume very little energy and often perform better for narrow tasks than large general-purpose systems. Generative models should be used when they provide clear additional value rather than by default. High-volume activities such as search, summarization, and practical guidance benefit most from concise modes and retrieval-first designs, where modest reductions in verbosity can translate into substantial system-wide savings when applied at scale.

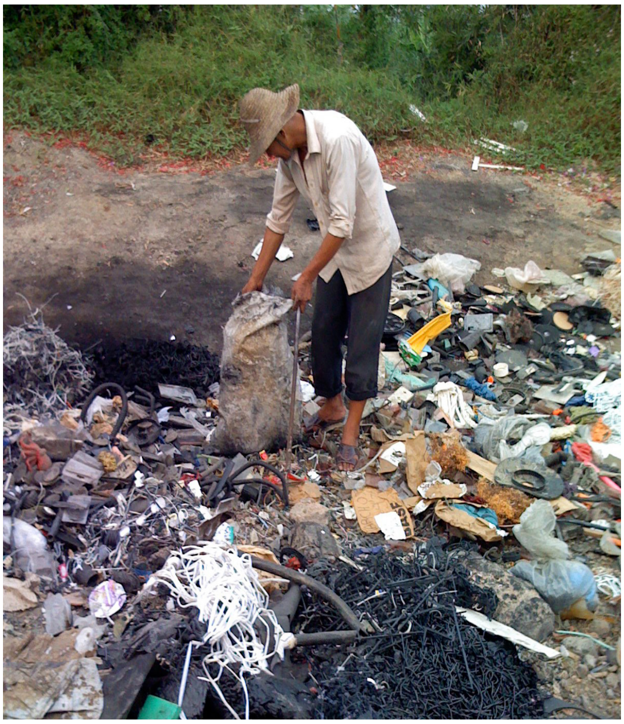
Sustainable practice includes concise prompts and outputs, batching related tasks, reusing previous results, and avoiding unnecessary iterations. In applications such as image and video generation, providers should present low-energy modes that are visible and inform users when their choices materially increase energy demand.

Integrate equity. Identical workloads can produce very different footprints depending on local electricity mixes. Reporting should include both location-based and global-average estimates so distributional effects are transparent. National strategies and international cooperation should account for these differences to ensure communities with limited resources are not disproportionately affected by the expansion of AI infrastructure. Siting decisions for large compute clusters should involve meaningful engagement with local communities, particularly where water scarcity or grid stress is a concern.

4.3 Roles and Responsibilities

To translate these principles into real-world change, responsibility must be shared across the full AI ecosystem, with distinct roles for each actor involved in designing, governing, financing, and using AI systems.

Developers and providers should carry direct responsibility for model design, deployment, and disclosure. Efficiency should be embedded into architectures. Efficient infrastructure should be prioritized. Footprints should be reported in formats that allow verification and fair comparison. User-facing controls such as concise modes, token limits, batch processing, and low-resolution presets should make efficient operation the default. Lifecycle sustainability should include ethically sourced materials, durable hardware design, and responsible e-waste management.



A worker sorting and dismantling electronic waste in Guiyu, China, where discarded devices are processed in open-air conditions that contaminate soil, water, and air. As AI expansion increases global hardware turnover, electronic waste volumes are expected to rise, extending the environmental footprint of AI beyond data centers into downstream disposal sites. Photo by Bert van Dijk (Flickr).

Governments and regulators should treat environmental disclosure for AI as routine. Standards should require clear reporting of model-level and task-level footprints. Climate and energy planning should incorporate projected AI demand. Procurement should reward low-footprint tools and infrastructure. Siting guidelines and incentives should steer data centers away from water-stressed or high-carbon regions and encourage local benefit-sharing.

International and standards bodies should harmonize methods for measuring and reporting energy and derived environmental footprints. Shared standards should help prevent selective disclosure and enable cross-border accountability. Global forums should elevate leadership and capacity-building in the Global South and link AI's environmental governance explicitly to sustainability goals.

Researchers and universities should extend impact assessments across the full AI supply chain and develop open, verifiable benchmarking tools for task-level energy. Interdisciplinary work should link environmental metrics to social and economic outcomes. Training programs that combine machine learning, systems engineering, and environmental governance should produce a workforce ready to build and regulate sustainable AI systems.

Civil society, NGOs, and media should play a central role in making AI's materiality visible to the public. They should track disclosures, translate footprints into accessible comparisons, and amplify voices from communities affected by data center siting, water use, and critical minerals mining impacts. Public scrutiny should strengthen accountability and keep environmental justice at the center of AI governance.

Users and enterprises should recognize their everyday influence over AI's environmental impact. They should choose low-energy options when these meet the need and escalate to generative models only when they offer real value. Organizations should adopt energy or token budgets, encourage concise use, and default to smaller or on-device models. Procurement should incorporate criteria linked to environmental footprints.

Principles for a Responsible AI Ecosystem

Principle	Practical Application
Transparency	Use shared, decision-relevant metrics and disclosure for electricity, carbon, water, land, and material footprints so impacts are comparable across models, providers, and jurisdictions, and accountability is possible.
Efficiency by Design	Treat efficiency as a design requirement, not an afterthought: architectures, defaults, product choices, and UX (modality, context length, output fidelity) should minimize avoidable resource demand and lock in lower-impact baselines.
Equity and Environmental Justice	Ensure affected communities can meaningfully participate in siting and governance decisions, and that benefit-sharing, safeguards, and remediation are built in where local burdens concentrate (grid stress, water constraints, land-use pressures).
Lifecycle Responsibility	Govern beyond operational energy: address upstream extraction and manufacturing impacts, construction and supply-chain externalities, and downstream e-waste and end-of-life management, so impacts are not shifted out of scope.
Global Cooperation	Harmonize disclosure norms, reporting boundaries, and governance approaches across borders and value chains to reduce regulatory arbitrage and burden shifting, and to support shared targets and interoperability.
Sustainable Use	Align deployment with societal value and resource constraints: demand management, fit-for-purpose model selection, and institutional procurement standards matter because cumulative impacts are driven by volume and everyday practice, not only per-query efficiency.

Actors and Responsibilities in the AI Ecosystem

Actor Group	Governance Implication	Responsibilities
Governments & Regulators	Public policy sets disclosure norms, planning assumptions, and siting conditions for AI infrastructure.	Set disclosure and performance standards; integrate AI into energy, water, and land planning; steer siting with safeguards and benefit-sharing; use procurement to reward low-footprint AI; protect rights and prevent discriminatory impacts.
Developers & Producers	Design and deployment choices shape baseline footprint and what users can see, choose, and control.	Build efficiency by default; publish comparable footprint information; reduce compute demand through optimization; strengthen lifecycle and supply-chain responsibility; offer user controls that enable low-impact use.
Universities & Researchers	Independent methods and benchmarks are needed to validate claims and guide policy and practice.	Measure and benchmark impacts; develop open methods and tools; test mitigation strategies; train interdisciplinary capacity; support science–policy translation.
Investors & Financial Institutions	Capital allocation can accelerate efficient infrastructure and enforce accountability through financing conditions.	Require disclosure in due diligence; finance efficient compute and cooling; integrate lifecycle risk; align investment with sustainability goals; condition funding on accountability.
Enterprises & Public-Sector Users	Demand-side choices and procurement decisions determine scale effects and whether high-impact use becomes routine.	Procure fit-for-purpose systems; set internal use policies (defaults, budgets, batching); prefer low-impact options; track usage and footprints; ensure responsible deployment.
NGOs, Media & Civil Society	Public scrutiny and participation are essential where impacts fall, especially for siting, water use, and supply chains.	Monitor and translate impacts; advocate for environmental justice and safeguards; elevate affected communities; support transparency and rights protections; build public literacy.

Protests in Montevideo, Uruguay, in May 2023 after severe drought depleted freshwater reserves and rising salinity made tap water unsafe to drink. Public anger intensified as plans for a water-intensive Google data center raised concerns over prioritizing industrial demand over human consumption. Photo by Vivir.solo.cuesta.vida (Wikimedia Commons).

Investors and financial institutions should accelerate the transition to sustainable AI by incorporating environmental footprints into due diligence. They should finance efficient infrastructure and support low-impact products and services. Stewardship should include expectations for disclosure, efficiency improvements, and lifecycle responsibility.

Local communities should be included because they experience direct environmental and social impacts from data centers, mining operations, and other AI-related infrastructure. They should be included early in siting decisions, receive transparent information, and benefit from local reinvestment. These actions and responsibilities converge into a shared agenda for implementation.

4.4 Putting Principles into Practice

Progress requires coordinated movement on standards, disclosure, incentives, user-centric design, rebound management, equity, and cooperation. This includes shared methodologies for calculating water, land, and carbon footprints from energy, mandatory and comparable disclosure, procurement that rewards low-impact design, user interfaces that guide efficient choices, and resource budgets that manage rebound effects.

Equity must remain central, especially for Indigenous communities and those in the Global South, through transparent reporting, inclusive decision-making, and targeted capacity-building.

4.5 Conclusion

AI offers remarkable potential, but fulfilling this promise responsibly requires systemic change. Every interaction draws on finite resources, and the total environmental footprint depends on how AI systems are designed, how often they are used, and what tasks they perform. Real progress depends on embedding sustainability at every level, from hardware and model design to deployment, governance, and public use. By committing to transparency, engineering for efficiency, choosing wisely as users and institutions, protecting communities that face disproportionate burdens, and cooperating across borders, society can ensure that progress in intelligence is matched by progress in care. Responsible AI is possible when capability and stewardship grow together within planetary limits.

UNU INWEH

References

1. Schwab, K. The Fourth Industrial Revolution. World Economic Forum. 2016.

2. Obringer R, Kumar R, Madani K. Harnessing the power of AI for climate change impact assessment. United Nations University Institute for Water, Environment and Health (UNU-INWEH). 2024. doi: 10.53328/INR24ROR012

3. Wu T, He S, Liu J, et al. A brief overview of ChatGPT: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica. 2023; 10(5). doi: 10.1109/JAS.2023.123618

4. Chatterji A, Cunningham T, Deming D, et al. How People Use ChatGPT. Working Paper No. 34255. National Bureau of Economic Research. 2025. doi: 10.3386/w34255

5. Agomuoh F. Google Gemini among the top three AI services, with 350 million monthly users. Digitaltrends. April 25, 2025. https://www.digitaltrends.com/computing/google-gemini-among-the-top-three-ai-services-with-350-million-monthly-users/

6. Jiang B. Alibaba's AI model outperforms Chinese rivals, ranks Just behind OpenAI, Anthropic. South China Morning Post. July 11, 2024. https://www.scmp.com/tech/big-tech/ article/3270079/alibabas-ai-model-outperforms-chinese-rivals-ranks-just-behind-openai-anthropic

7. Chiang S. Alibaba to roll out its rival to ChatGPT across all its products. CNBC. April 10, 2023. https://www.cnbc.com/2023/04/11/alibaba-to-roll-out-its-rival-to-chatgpt-across-all-its-products.html

8. Elad B. DeepSeek AI statistics 2026: Users, benchmarks, & enterprise reach. SQ Magazine. October 7, 2025. https://sqmagazine.co.uk/deepseek-ai-statistics/

9. DOIT Software. Grok statistics: Usage & benchmarks [2025]. DOIT Software. October 30, 2025. https://doit.software/blog/grok-statistics#screen1

10. Hassabis D, Kavukcuoglu K, Pichai S. A new era of intelligence with Gemini 3. Google Blog. November 18, 2025. https://blog.google/intl/en-africa/company-news/outreach-and-initiatives/a-new-era-of-intelligence-with- gemini-3/

11. Williams R. Meta's new AI models can recognize and produce speech for more than 1,000 languages. MIT Technology Review. May 22, 2023. https://www.technologyreview. com/2023/05/22/1073471/metas-new-ai-models-can-recognize-and-produce-speech-for-more-than-1000-languages/

12. Hoover M. "Feeding" the cloud: Reducing carbon emissions from social media and streaming services through private information disclosure. Vanderbilt Journal of Entertainment and Technology Law. 2025; 26(3): 549.

13. North M. 7 Ways AI is transforming healthcare. World Economic Forum. April 13, 2025. https://www.weforum.org/stories/2025/08/ai-transforming-global-health/

14. Goldman Sachs. Partially autonomous cars forecast to comprise 10% of new vehicle sales by 2030. Goldman Sachs. August 19, 2024. https://www.goldmansachs.com/insights/articles/partially-autonomous-cars- forecast-to-comprise-10-percent-of-new-vehicle-sales-by-2030

15. Templeton B. Robotaxis — Mostly Waymo — are giving 1.3 million rides a month. Why? Forbes. March 7, 2025. https://www.forbes.com/sites/bradtempleton/2025/03/07/robotaxis-mostly-waymo-are-giving-13-million- ridesmonth--why/

16. Statista. Artificial intelligence - Worldwide. November 2025. https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

17. Markets and Markets. Artificial intelligence (AI) market by offering (infrastructure, software, services), technology (ML, NLP, generative AI), business function (operations & supply chain, finance & accounting), enterprise application, and end user—global forecast to 2032. 2025. https://www.marketsandmarkets.com/Market-Reports/ artificial-intelligence-market-74851580.html

18. Si Tou WK, Vezzani A, González-Sanz A, Sirimane S. Technology and innovation report 2025: Inclusive artificial intelligence for development. United Nations Conference on Trade and Development. 2025.

19. Fortune Business Insights. Artificial intelligence [AI] market size, growth & trends by 2034. December 29, 2025. https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114

20. Precedence Research. Artificial intelligence (AI) market size to hit USD 3,680.47 Bn by 2034. September 29, 2025. https://www.precedenceresearch.com/artificial-intelligence-market

21. Digitalization World. Generative AI will make over 40% of total AI industry market size by 2030. 2025. https://digitalisationworld.com/news/68298/generative-ai-will-make-over-40-of-total-ai- industry-market-size-by-2030

22. Sajadieh S, Fattorini L, Perrault R. The AI index 2026 annual report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. April 2026. https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf

23. Cazzaniga M, Jaumotte F, Li L, et al. Gen-AI: Artificial intelligence and the future of work. IMF Staff Discussion Notes. 2024; 2024(001): 1. doi: 10.5089/9798400262548.006

24. IPSOS. The Ipsos AI MONITOR 2024, A 32-country Ipsos Global Advisor Survey. 2024.

25. World Economic Forum. The future of jobs report 2025. World Economic Forum. 2025. https://www.weforum.org/publications/the-future-of-jobs-report-2025/

26. Robinson B. Fears About AI Job loss: New study answers if they're justified. Forbes. February 9, 2025. https://www.

forbes.com/sites/bryanrobinson/2025/02/09/fears-about-ai-job-loss-new-study-answers-if-

27. Brynjolfsson E, Li D, Raymond L. Generative AI at work. Quarterly Journal of Economics. 2025; 140(2): 889-942. 10.1093/QJE/QJAE044

28. Cui KZ, Demirer M, Jaffe S, et al. The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers. Management Science. 2026. doi: 10.1287/mnsc.2025.00535.

29. Gill AS, Houngbo GF. Mind the AI divide: Shaping a global perspective on the future of work. International Labour Organization (ILO) and United Nations Office of the Secretary-General's Envoy on Technology. 2024.

30. Gibbs S. Musk, Wozniak and Hawking urge ban on warfare AI and autonomous weapons. The Guardian. July 27, 2015. https://www.theguardian.com/technology/2015/jul/27/ musk-wozniak-hawking-ban-ai-autonomous-weapons

31. Onderco M. Navigating the AI frontier: Insights from the Ukraine conflict for NATO's governance role in military AI. Journal of Strategic Studies. 2025; 48(3): 602-626. doi: 10.1080/01402390.2025.2463451

32. Bego K. Ukraine's Operation Spider's Web is a game-changer for modern drone warfare. NATO should pay attention. Chatham House. 2025. https://www. chathamhouse.org/2025/06/ukraines-operation-spiders-web-game-changer-modern-drone-warfare-nato-should-pay- attention

33. Murphy M, Robinson O, Sardarizadeh S. Israel-Iran conflict unleashes wave of AI disinformation. BBC. June 20, 2025. https://www.bbc.com/news/articles/c0k78715enxo

34. Milmo D. AI pioneer announces non-profit to develop 'honest' artificial intelligence. The Guardian. June 3, 2025. https://www.theguardian.com/technology/2025/jun/03/honest-ai-yoshua-bengio

35. Gold H. xAI issues lengthy apology for violent and antisemitic Grok social media posts. CNN Business. July 11, 2025. https://www.cnn.com/2025/07/12/tech/xai-apology-antisemitic-grok-social-media- posts

36. Satariano A, Mozur P. A.I. computing power Is splitting the world into haves and have-nots. The New York Times. June 21, 2025. https://www.nytimes.com/interactive/2025/06/23/technology/ai- computing-global-divide.html

37. Roose K. AI-generated art won a prize. Artists aren't happy. The New York Times. September 2, 2022. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

38. Roth E. Judge slams lawyers for 'bogus AI-generated Research'. The Verge. May 13, 2025.

39. Appel G, Neelbauer J, Schweidel DA. Generative AI has an intellectual property problem. Harvard Business Review. 2023; 7: 10-16.

40. Gerlich M. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies. 2025; 15(1): 6. doi: 10.3390/soc15010006

41. Tan E. Meta built a data center next door. The neighbors' water taps went dry. The New York Times. July 14, 2025.. https://www.nytimes.com/2025/07/14/technology/meta-data-center-water.html

42. UNICEF & WHO Joint Monitoring Programme (2021). Progress on household drinking water, sanitation and hygiene 2000-2020: Five Years into the SDGs. World Health Organization. 2021.

43. Díaz Paz WF, Escosteguy M, Seghezzo L, Hufty M, Kruse E, Iribarnegaray MA. Lithium mining, water resources, and socio-economic issues in Northern Argentina: We are not all in the same boat. Resources Policy. 2023; 81:103288. doi: 10.1016/j.resourpol.2022.103288

44. Jegham N, Abdelatti M, Elmoubarki L, Hendawi A. How hungry is AI? Benchmarking energy, water, and carbon footprint of LLM inference. arXiv. 2025. doi: 10.48550/arXiv.2505.09598

45. IEA. Global Energy Review 2025. 2025. https://www.iea.org/reports/global-energy- review-2025/electricity

46. O'Donnell J, Crownhart C. We did the math on AI's energy footprint. Here's the story you haven't heard. MIT Technology Review. May 20, 2025. https://www.technologyreview. com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/

47. Ristic B, Mahlooji M, Gaudard L, Madani K. The relative aggregate footprint of electricity generation technologies in the European Union (EU): A system of systems approach. Resources, Conservation and Recycling. 2019; 143. doi: 10.1016/j.resconrec.2018.12.010

48. Obringer R, Rachunok B, Maia-Silva D, Arbabzadeh M, Nateghi R, Madani K. The overlooked environmental footprint of increasing Internet use. Resources, Conservation & Recycling. 2021;167:105389. doi:10.1016/j.resconrec.2020.105389.

49. IEA. Africa energy outlook 2022. 2022. https://www.iea.org/reports/africa-energy- outlook-2022

50. Chamanara S, Madani K. The hidden environmental cost of cryptocurrency: How bitcoin mining impacts climate, water and land. United Nations, University Institute for Water, Environment and Health. 2023. https://collections.unu.edu/eserv/UNU:9528/UN-IWEH_BTC_Report.pdf

51. Chamanara S, Ghaffarizadeh SA, Madani K. The environmental footprint of bitcoin mining across the globe: Call for urgent action. Earths Future. 2023; 11(10). doi: 10.1029/2023EF003871

52. Mahlooji M, Gaudard L, Ristic B, Madani K. The importance of considering resource availability restrictions in energy planning: What is the footprint of electricity generation in the Middle East and North Africa (MENA)? Science of the Total Environment. 2020; 717. doi: 10.1016/j.scitotenv.2019.135035

53. Ristic B, Madani K, Makuch Z. The water footprint of data Centers. Sustainability (Switzerland). 2015; 7(8). doi: 10.3390/su70811260

54. Mann T. Meta says all new Llama 3.1 405B model bests OpenAI's GPT-4. The Register. 2024. https://www.theregister.com/2024/07/23/meta_llama31_405b/

55. Deepseek-AI. DeepSeek-V3 technical report. Published online 2025.

56. IEA. Energy and AI. 2025. https://www.iea.org/reports/energy-and-ai

57. World Population Prospects, United Nations (UN). Population, total - Sub-Saharan Africa | Data. UN Population Division. 2024. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=ZG

58. US EPA. Greenhouse gas equivalencies calculator. US EPA. February 24, 2025. https:// www.epa.gov/energy/greenhouse-gas-equivalencies-calculator?

59. Howard G, Bartram J, Williams A, Overbo A, Fuente D, Geere JA. Domestic water quantity, service level and health, Second edition. World Health Organization. Published online 2020:102.

60. Ember. Yearly electricity data. January 2026. https://ember-energy.org/data/yearly- electricity-data

61. Gartner. Gartner says electricity demand for data centers to grow 16% in 2025 and double by 2030. 2025. https://www.gartner.com/en/newsroom/press-releases/2025-11-17-gartner-says-electricity-demand-for- data-centers-to-grow-16-percent-in-2025-and-double-by-2030

62. Madani K, Rouhani OM, Pournazeri S, Moradi M, Sheikhmohammady M. Can we rely on renewable energy sources to overcome global warming? World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability - Proceedings of the 2011 World Environmental and Water Resources Congress. doi:10.1061/41173(414)346

63. Hadian S, Madani K. A system of systems approach to energy sustainability assessment: Are all renewables really green? Ecological Indicators. 2015; 52: 194-206. doi :10.1016/j.ecolind.2014.11.029

64. Jackson A. Power-hungry data centres put pressure on Ireland's Grid. Data Centre Magazine. August 3, 2024. Accessed January 14, 2026. https://datacentremagazine.com/critical-environments/power-hungry-data- centres-put-pressure-on-irelands-grid

65. Global AI adoption to surge 20%, exceeding 378 million users in 2025 - Edge AI and Vision Alliance. https://www.edge-ai-vision.com/2025/02/global-ai-adoption-to-surge-20-exceeding-378-million-users-in- 2025/

66. Kumar N. How many Google searches per day [Latest 2025 Data]. August 26, 2025. https:// www.demandsage.com/google-search-statistics/

67. Peters J. Google's AI overviews in search now have "1.5 billion users per month," CEO Sundar Pichai said in a statement as part of Alphabet's Q1 2025 earnings. The company's revenues hit $90.2 billion for the quarter. The Verge. April 24, 2025. https://www.theverge.com/news/655930/google-q1-2025-earnings

68. Luers A. Net zero needs AI—five actions to realize its promise. Nature.. 2025; 644(8078): 871-873. doi: 10.1038/d41586-025-02641-4

69. Luccioni S, Jernite Y, Strubell E. Power hungry processing: Watts driving the cost of AI deployment? FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, Inc; 2024; 85-99. doi: 10.1145/3630106.3658542

70. Jiang L, Chen F. Carbonscaling: Extending neural scaling laws for carbon footprint in large language models. August 2, 2025. doi ; 10.48550/arXiv.2508.06524

71. Down AK. OpenAI will not disclose GPT-5's energy use. It could be higher than past models. The Guardian. August 9, 2025. https://www.theguardian.com/technology/2025/aug/09/open-ai- chat-gpt5-energy-use

72. Bertazzini G, Albisani C, Baracchi D, Shullani D, Verdecchia R. The hidden cost of an image: Quantifying the energy consumption of AI image generation. http://arxiv.org/abs/2506.17016

73. Delavande J, Pierrard R, Luccioni S. Video killed the energy budget: Characterizing the latency and power regimes of ppen text-to-video models. arXiv. 2025. September 23, 2025. http://arxiv.org/abs/2509.19222

74. Badalyan A. AI statistics In 2025: Key trends and usage data. DigitalSilk. September 30, 2025. https://www.digitalsilk.com/digital-trends/ai-statistics/

75. Google. Powering a Google search. Official Google Blog. 2009. https://googleblog.blogspot.com/2009/01/ powering-google-search.html

76. Srinivasan V. AI, personalization and the future of shopping. March 3, 2025. https://blog. google/products/ads-commerce/ai-personalization-and-the-future-of-shopping/

77. Oviedo F, Kazhamiaka F, Choukse E, et al. Energy use of AI inference: Efficiency pathways and test-time compute. Joule. 2026. https://arxiv.org/pdf/2509.20241

78. Chung JW, Ma JJ, Wu R, et al. The ML.ENERGY Benchmark: Toward automated inference energy measurement and optimization. Advances in Neural Information Processing Systems. 2026; 38. https://arxiv.org/pdf/2505.06371

79. Hawkins ZJ, Lehdonvirta V, Wu B. AI compute sovereignty: Infrastructure control across territories, cloud providers, and accelerators. 2025. Jun 20. http://dx.doi.org/10.2139/ssrn.5312977

About UNU-INWEH

The United Nations University Institute for Water, Environment and Health (UNU-INWEH) is one of 13 institutions that make up the United Nations University (UNU), the academic arm of the United Nations. Known as 'The UN’s Think Tank on Water', UNU-INWEH addresses critical water, environmental, and health challenges around the world. Through research, training, capacity development, and knowledge dissemination, the institute contributes to solving pressing global sustainability and human security issues of concern to the UN and its Member States. Headquartered in Richmond Hill, Ontario, UNU-INWEH has been hosted and supported by the Government of Canada since 1996. With a global mandate and extensive partnerships across UN entities, international organizations, and governments, UNU-INWEH operates through its UNU Hubs in Calgary, Hamburg, New York, Lund, and Pretoria, and an international network of affiliates.

UNU INWEH

unu.edu/inweh