AI: Driving Innovation, Draining Sustainability

Author: Ritayush Suchismita Dey
Saturday, 13th June 2026

Behind Every Prompt...

Think about the last time you typed a question into an AI chatbot, asked a voice assistant for directions, or let an algorithm choose your next playlist. These interactions feel effortless, almost weightless. There is no factory chimney, no diesel exhaust, no visible resource being consumed. The interface is clean, the response is instant, and the transaction feels entirely contained within the screen. That frictionless experience is precisely what makes the environmental cost of artificial intelligence so easy to ignore.

Artificial intelligence has become one of the defining technological forces of this decade. From personalised recommendation engines and real-time translation tools to generative systems capable of producing essays, images, and code on demand, AI now touches nearly every domain of modern life. Its proponents frequently highlight its potential to decarbonise industries, reduce logistical waste, and accelerate climate modelling. These benefits are real, and they matter. But the same technology that promises to optimise the world runs on an infrastructure of extraordinary physical scale, one that consumes enormous quantities of electricity, water, and rare materials. The sustainability of artificial intelligence cannot be evaluated through its interface alone. The question worth asking is what it costs to keep the lights on beneath that interface.

The Myth: Artificial intelligence is becoming so computationally efficient that its environmental footprint is naturally diminishing.

The Reality: While individual AI models and processors have become more efficient, these gains are frequently outpaced by explosive growth in model size, training requirements, data-centre expansion, user base and the trillions of inferences performed daily.

The Underlying Infrastructure

Every AI system, regardless of how it presents itself to the user, is physically anchored to a network of data centres. These are not passive storage rooms. They are vast facilities that operate continuously to handle the storage, processing, and retrieval of data generated by training pipelines, user interactions, and model outputs. Importantly, modern data centres have become significantly more efficient over time, with improvements in server utilisation, cooling systems, and power management. Yet efficiency alone has not reduced overall resource consumption because the scale of AI deployment has expanded even faster. In 2024, global data centre electricity consumption reached approximately **415 terawatt-hours**, accounting for nearly **1.5% of total global electricity use.** Projections from the International Energy Agency suggest this figure is on course to more than **double by 2026**, driven primarily by the rapid expansion of generative AI systems. This is the central challenge to the assumption that greater efficiency automatically produces sustainability: even as computing infrastructure becomes more efficient, total electricity demand continues to rise because AI adoption is growing at an extraordinary pace.

© Ritayush Suchismita Dey | Eco-Mythbusters Roundtable

The energy demand alone would be striking enough. But the burden extends further. Servers generate enormous quantities of heat, and cooling them requires water, in quantities that dwarf what most people would imagine. Hyperscale data centres, the kind that power large language models and cloud AI services, can consume hundreds of thousands of gallons of water daily. Some individual facilities use up to **550,000 gallons of water per day** just to maintain operating temperatures. Microsoft's data centre operations consumed approximately **1.7 billion gallons of freshwater in 2022** alone, a **34% increase** from the prior year, largely attributable to AI infrastructure expansion. In regions already under water stress, including parts of the American Southwest and South Asia, siting these facilities raises questions that extend well beyond electricity bills. It is worth acknowledging that data centres also power services that are not AI at all, from email servers to video streaming. The attribution is not clean. But what distinguishes AI workloads is not merely their per-transaction intensity, but also the speed at which demand for them is increasing. The infrastructure was scaling before generative AI arrived; it is now scaling far faster. This distinction matters because sustainability depends not only on how efficiently resources are used, but also on how much total demand is created. AI systems may be becoming more efficient, yet the infrastructure supporting them continues to grow in absolute size and resource consumption.

Creation Comes at a Cost: Training an AI

Before an AI model responds to a single query, it must be trained. Training is the process by which a model learns patterns from enormous datasets, adjusting billions of internal parameters across repeated computational cycles. Although advances in hardware and optimisation techniques have made training significantly more efficient than it was a decade ago, the scale of modern models has expanded just as rapidly. Training a large-scale language model can still take weeks or months, running on thousands of specialised processors simultaneously.

A landmark 2019 study from the University of Massachusetts Amherst estimated that training a single large transformer model could emit approximately **284 tonnes of carbon dioxide equivalent**, comparable to the lifetime emissions of **five average American cars**. More recent models are considerably larger. GPT-3, with 175 billion parameters, is estimated to have consumed around **1,287 megawatt-hours** of electricity during training. The significance of this figure lies not merely in its size, but in what it represents: efficiency gains in computing have often been accompanied by an even faster increase in model complexity and computational ambition. A 2023 study published in Nature estimated that training a single large language model consumed approximately **700,000 litres of freshwater**, primarily for data centre cooling. These are not marginal costs that can be absorbed quietly into a sustainability report.

There is a further dimension that is often omitted from these calculations: models are not trained once. They are retrained, fine-tuned, and updated repeatedly. Organisations running competitive AI products may retrain foundational models multiple times annually. As training becomes faster and more affordable, the incentive to retrain and develop larger models also increases. The cumulative footprint is therefore a multiple of any single training run, compounded across the industry.

© Eco-Mythbusters Roundtable

None of this is an argument against building AI systems. Rather, it is an argument for questioning the assumption that greater efficiency automatically translates into sustainability. If each improvement in training efficiency is accompanied by larger models, more frequent retraining, and greater deployment, total resource consumption may continue to rise despite technological progress.

Every Prompt Comes at a Price: Inference

If training is the foundation, inference is the ongoing operation. Inference refers to the process by which a trained model generates a response: the computation triggered every time a user submits a query. Advances in hardware and model optimisation have made this process significantly faster and more efficient over time, reducing the computational cost of generating any single response. Unlike training, which is periodic, inference is continuous. It happens billions of times daily, across every product built on top of an AI model.

Several studies indicate that inference now accounts for approximately **60% of the total energy consumption** of AI systems, surpassing the training phase, which is often mistakenly assumed to dominate. A single query to a large language model can consume up to **ten times the electricity** of a standard web search. When that figure is scaled across the millions of daily users of platforms like ChatGPT, which reportedly crossed 100 million users within two months of launch and has continued growing since, the cumulative energy demand becomes very large very quickly.

The counterargument often raised here is that AI inference replaces other, more energy-intensive activities. If a chatbot reduces the number of customer service calls handled by physical call centres, or if a logistics algorithm prevents unnecessary truck journeys, the net environmental equation may still favour AI.

This logic holds in specific, well-designed applications. It does not hold universally, and particularly not in the widespread casual use cases that now constitute the majority of AI queries. As AI becomes cheaper, faster, and easier to access, it is increasingly used for tasks that were never previously performed at all, meaning efficiency gains do not necessarily translate into proportional environmental savings.

The Machines Behind the Mind: Hardware

The intangibility of AI output can obscure just how physically demanding its hardware requirements are. AI systems rely on specialised processors, particularly Graphics Processing Units (GPUs) and, increasingly, custom AI accelerator chips. Manufacturing these components is energy-intensive and depends on the extraction of critical minerals, including cobalt, lithium, tantalum, and rare earth elements, many of which are sourced under conditions that carry their own environmental and ethical costs.

Beyond manufacture, there is the question of lifecycle. The pace of advancement in AI hardware is rapid. A GPU generation that is competitive today may be considered obsolete within **two to three years** as newer architectures offer efficiency and performance gains. Paradoxically, the very improvements that make AI hardware more efficient can also shorten replacement cycles,

© Eco-Mythbusters Roundtable

encouraging organisations to retire functional equipment in pursuit of greater computational performance and lower operating costs.

This accelerated obsolescence drives a corresponding acceleration in electronic waste. Efficiency improvements, therefore, do not eliminate environmental costs; they can shift those costs upstream into mineral extraction and downstream into disposal and replacement.

The United Nations Environment Programme estimates that global e-waste generation reached **62 million tonnes** in 2022, a figure projected to rise further as AI hardware demands increase. Unlike some forms of industrial waste, electronic waste contains hazardous materials including lead, mercury, and cadmium, requiring specialised handling that, in practice, is frequently absent.

The hardware lifecycle is therefore not a footnote to AI's environmental story. It is a chapter that begins before any model is trained and continues long after any particular system has been retired.

The Rebound Effect

Perhaps the most structurally important dimension of AI's environmental impact is not any single cost in isolation, but the dynamic by which efficiency gains lead to increased total consumption. This phenomenon, known as the rebound effect, has appeared repeatedly across the history of technological improvement. When fuel efficiency in automobiles improved significantly through the **1980s** and **1990s**, consumers responded in part by driving more, buying larger vehicles, and commuting further. Total fuel consumption did not fall proportionately to the per-kilometre efficiency gain.

The same logic applies to AI. As models become faster, cheaper, and more accessible, they attract more users, more use cases, and more integration into products and workflows. Each marginal improvement in AI efficiency is likely to be met, and then exceeded, by growth in demand. Recent projections from Goldman Sachs Research suggest that AI-related energy consumption could increase electricity demand in the United States by as much as **15% by 2030**, even under conservative adoption scenarios. Global AI-related emissions could reach up to **80 million tonnes of carbon dioxide** annually under high-growth trajectories, with water consumption potentially exceeding global bottled water usage.

This does not make AI development indefensible. It does assume that AI is automatically and inherently sustainable deeply questionable. Efficiency at the system level.

Comparative Data Analysis: AI's Environmental Footprint

The figures below synthesise estimates from peer-reviewed research, IEA energy data, and published corporate sustainability disclosures to contextualise AI's resource demands against benchmarks that are simpler to visualise.

Process	Energy Consumption	Equivalent Impact	Research Basis

© Eco-Mythbusters Roundtable

© Ritayush Suchismita Dey | Eco-Mythbusters Roundtable

Training one large-scale **LLM** (e.g., GPT-3 scale)	~1,287 MWh electricity; ~700,000 L water	~284 tCO_2e (equivalent to the lifetime emissions of 5 cars)	Strubell et al. (2019); Li et al. (2023, Nature)
Single AI chatbot **query**	~0.001–0.01 kWh (approximately 10× a standard web search)	~0.5–4 g CO_2e per query; ~500 mL water consumed at the data centre	Goldman Sachs Research (2024); Luccioni et al. (2023)
1 billion daily AI **queries** (platform scale)	~1,000–10,000 MWh/day electricity	Equivalent to powering ~100,000 average U.S. households for one day	Derived from IEA data and per-query estimates
Global **data centres** (2024, all workloads)	~415 TWh/year electricity	~1.5% of global electricity use; projected to double by 2026	IEA Data Centres Report (2024)
Microsoft data centre water use (2022)	~1.7 billion gallons of freshwater/year	+34% year-over-year increase, largely driven by AI infrastructure expansion	Microsoft Environmental Sustainability Report (2022)
Global **AI-related emissions**	Energy demand +15% in the U.S. by 2030 (conservative estimate)	Up to 80 Mt CO_2e/year; water use potentially exceeding global bottled water consumption	Goldman Sachs Research (2024); UNEP (2023)
Global e-waste from **AI hardware** (2022 baseline)	62 million tonnes of e-waste are generated globally	Contains hazardous materials such as lead (Pb), mercury (Hg), and cadmium (Cd); recycling infrastructure remains inadequate in many regions	UNEP Global E-Waste Monitor (2024)

Table 1: Synthesised environmental footprint estimates for AI systems across the value chain. Sources: Strubell et al. (2019); Li et al. (2023); IEA Data Centres and Data Transmission Networks Report (2024); Goldman Sachs Research (2024); Microsoft Environmental Sustainability Report (2022); UNEP Global E-Waste Monitor (2024); Luccioni et al. (2023).

Conclusion

Using AI is not inherently unsustainable. The technology holds genuine promise in applications where computational intelligence can replace more resource-intensive processes, accelerate scientific discovery, or reduce systemic inefficiency at scale. The challenge is not that AI is becoming less efficient; in many respects, it is becoming dramatically more efficient. The challenge is that improvements in efficiency are frequently accompanied by even greater growth in deployment, infrastructure, and demand. Beneath every query, every generated image, and every auto-completed sentence lies a physical system consuming electricity, water, and hardware resources. The infrastructure supporting these interactions is expanding at a pace that often exceeds the environmental savings generated by technological progress alone.

The most important reframe is this: sustainability is not a property of the technology itself, nor is it an automatic consequence of greater efficiency. It is a property of how the technology is built, powered, deployed, and scaled. The same AI system can represent a net environmental gain in one context and a net environmental loss in another, depending on the energy mix powering it, the necessity of its application, the efficiency discipline of its designers, and whether growth in demand remains proportionate to the efficiency gains being achieved.

When AI Systems Are Sustainable:

- When data centres are powered predominantly by verified renewable energy, directly reducing the carbon intensity of every training run and inference cycle.
- When AI is deployed in high-impact applications, such as grid optimisation, precision agriculture, or drug discovery, where the efficiency gains demonstrably outweigh the computational cost.
- When models are designed with computational efficiency as a deliberate objective, favouring smaller, purpose-specific architectures over unnecessarily large systems, and limiting redundant retraining cycles.
- When hardware lifecycle management includes sustainable sourcing of critical minerals and formal, accessible pathways for electronic waste recycling.

When They're Not:

- When data centres continue to rely on fossil fuel-based electricity, meaning every interaction contributes directly to carbon emissions regardless of how the AI is used.
- When large-scale models are retrained repeatedly without commensurate improvements in capability or application value, generating excessive computational expenditure.
- When the rebound effect takes hold and increased accessibility drives disproportionate growth in trivial or low-value use cases, transforming marginal per-query costs into a substantial collective burden.
- When rapid hardware obsolescence, driven by competitive pressure rather than genuine necessity, accelerates the extraction of finite materials and contributes to growing volumes of hazardous electronic waste.
- When the framing of AI as a digital and therefore resource-free technology goes unchallenged, allowing the infrastructure costs to remain invisible in both public discourse and corporate accountability frameworks.

© Ritayush Suchismita Dey | Eco-Mythbusters Roundtable

Bibliography

1. Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. https://arxiv.org/abs/1906.02629. Accessed 13 June 2026.
2. Li, Pengfei, et al. "Making AI Less Thirsty: Uncovering and Addressing the Secret Water Footprint of AI Models." arXiv, 2023. https://arxiv.org/abs/2304.03271. Accessed 13 June 2026.
3. International Energy Agency. "Data Centres and Data Transmission Networks." IEA, 2024. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks. Accessed 13 June 2026.
4. Goldman Sachs Research. "AI is Poised to Drive 160% Increase in Data Centre Power Demand." Goldman Sachs, 2024. https://www.goldmansachs.com. Accessed 13 June 2026.
5. Luccioni, Alexandra Sasha, Sylvain Viguier, and Anne-Laure Ligozat. "Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model." Journal of Machine Learning Research, 2023. https://arxiv.org/abs/2211.02001. Accessed 13 June 2026.
6. Microsoft Corporation. "Environmental Sustainability Report 2022." Microsoft, 2022. https://blogs.microsoft.com/on-the-issues/2022/05/05/2022-environmental-sustainability-report Accessed 13 June 2026.
7. United Nations Environment Programme. "Global E-Waste Monitor 2024." UNEP, 2024. https://www.unep.org/resources/report/global-e-waste-monitor-2024. Accessed 13 June 2026.
8. Crawford, Kate. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, 2021.

© Eco-Mythbusters Roundtable