

Environmental Sustainability in the Era of Artificial Intelligence: Energy Use, E-Waste and Emerging Challenges

Prasant Kumar¹ & Dr Rana Pratap Yadav²

¹Research Scholar, Department of Geography, VSSD College (C.S.J.M University Kanpur)

²Associate Professor, Department of Geography, VSSD College (C.S.J.M University Kanpur)

Abstract

Artificial intelligence (AI) is rapidly becoming a general-purpose technology, but its environmental footprint is also expanding through electricity demand, data-centre construction, semiconductor production, cooling, water use and accelerated turnover of electronic equipment. This paper evaluates environmental sustainability in the era of AI through a life-cycle framework focusing on energy consumption, electronic waste (e-waste), water demand, critical minerals, rebound effects and emerging governance challenges. The study applies a descriptive and analytical secondary-data methodology using evidence from the International Energy Agency (IEA), International Telecommunication Union (ITU), United Nations Institute for Training and Research (UNITAR), United Nations Conference on Trade and Development (UNCTAD), United Nations Environment Programme (UNEP), Organisation for Economic Co-operation and Development (OECD), and peer-reviewed studies published through 2025. The IEA estimated global data-centre electricity consumption at about 415 TWh in 2024, roughly 1.5% of global electricity use, and projected approximately 945 TWh by 2030 in its base case, with AI as the most important contributor to growth. Global e-waste reached 62 million tonnes in 2022, while only 22.3% was documented as formally collected and environmentally soundly recycled; the total is projected to reach 82 million tonnes by 2030. The analysis shows that efficiency gains alone cannot guarantee sustainability because growing demand, increasingly complex models, short hardware replacement cycles and rebound effects may offset resource savings. A sustainable AI pathway therefore requires transparent life-cycle measurement, low-carbon electricity, efficient and task-appropriate models, water-sensitive infrastructure, extended hardware life, circular-economy systems and environmental standards that distribute benefits and burdens more equitably.

ARTICLE INFORMATION

Corresponding Author: Prasant Kumar

Article History: Received: [03/08/2025] | Accepted: [10/08/2025]

Published: [13/08/2025]

Copyright: © [2025] The Author(s). This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Keywords: artificial intelligence, environmental sustainability, energy consumption, e-waste, data centres, water footprint, circular economy, Green AI

Table 1. Key environmental indicators used in the analysis

Indicator	Observed / projected value	Reference period	Source
Global data-centre electricity use	415 TWh (observed estimate)	2024	IEA, 2025
Global data-centre electricity use	945 TWh (base-case projection)	2030	IEA, 2025
Global e-waste generation	62 million tonnes	2022	ITU & UNITAR, 2024
Formally collected and environmentally soundly recycled e-waste	22.3%	2022	ITU & UNITAR, 2024
Global e-waste generation	82 million tonnes (projection)	2030	ITU & UNITAR, 2024
Generative-AI-related cumulative e-waste	1.2-5.0 million tonnes (scenario range)	2020-2030	Wang et al., 2024
Potential reduction through modelled circular strategies	16-86%	Scenario-dependent	Wang et al., 2024

Introduction

Artificial intelligence has moved from a specialised research field into a core component of contemporary digital economies. Machine-learning systems now support health care, education, finance, scientific discovery, agriculture, logistics, communication, manufacturing and environmental monitoring. Generative AI has accelerated this transformation by making large foundation models accessible to millions of users through text, image, audio, video and software-generation services. Although these services appear immaterial on a computer or mobile screen, they depend on an extensive physical system of semiconductor fabrication plants, data centres, storage devices, high-speed networks, cooling infrastructure, electrical grids and material supply chains. Environmental sustainability in the era of AI therefore cannot be understood by analysing software performance alone; it requires attention to the energy, water and material metabolism that makes digital intelligence possible.

The scale of this infrastructure is becoming environmentally significant. The International Energy Agency (IEA, 2025) estimated that data centres consumed about 415 terawatt-hours (TWh) of electricity globally in 2024, equal to approximately 1.5% of global electricity consumption. This demand had grown by around 12% annually over the preceding five years, substantially faster than total electricity demand. In the IEA base case, data-centre electricity use is projected to reach about 945 TWh by 2030, slightly less than 3% of global electricity consumption. AI is identified as the most important contributor to this growth, particularly

through the deployment of accelerated servers based on graphics-processing units and other specialised processors. These figures refer to the data-centre sector as a whole, not AI alone, but they demonstrate how quickly AI-driven computing is becoming an energy-system planning issue.

Energy use is only one part of the environmental footprint. Advanced processors and servers require minerals, metals, chemicals, high-purity silicon and substantial industrial infrastructure. Semiconductor fabrication is energy- and water-intensive, while data centres may require significant direct water consumption for cooling and indirect water consumption through electricity generation. At the end of the equipment life cycle, rapid technological turnover can generate additional electronic waste. According to ITU and UNITAR (2024), the world generated 62 million tonnes of e-waste in 2022, yet only 22.3% was documented as formally collected and environmentally soundly recycled. Under current trends, e-waste generation is projected to reach 82 million tonnes by 2030. AI hardware remains only one component of this global stream, but its expansion creates a new source of pressure within an already inadequate waste-management system.

This paper therefore treats AI as an environmental paradox rather than as either an inherently green or inherently unsustainable technology. It examines the direct footprint of computing, the embodied footprint of hardware, the environmental effects produced by AI-enabled applications and the broader system changes generated when AI becomes cheap and ubiquitous. The study focuses on three core questions: how is AI changing global electricity and material demand; how may rapid AI hardware expansion interact with the e-waste crisis; and which governance strategies can reduce environmental impacts without preventing socially useful innovation? By integrating evidence on energy, e-waste, water, minerals, rebound effects and environmental justice, the paper aims to provide a balanced framework for sustainable AI policy.

Literature Review, Research Gap and Conceptual Framework

The environmental implications of machine learning initially received attention through studies of model training. Strubell, Ganesh, and McCallum (2019) demonstrated that computationally intensive natural-language-processing research could carry substantial energy and carbon costs, particularly when repeated architecture searches and experimentation were included. Schwartz et al. (2020) subsequently proposed the concept of Green AI, arguing that research culture should value computational efficiency alongside accuracy. Henderson et al. (2020) reinforced this position by calling for systematic reporting of energy use and carbon emissions in machine-learning experiments. These contributions were important because they challenged the assumption that improved model performance should be pursued without accounting for computational resources.

The rapid adoption of generative AI broadened the focus from training to inference. Training a frontier model may require very large but episodic computation, whereas inference occurs every time a deployed system receives a user request. At large scale, cumulative inference can therefore become a dominant part of operational energy demand. Luccioni, Jernite, and Strubell (2024) measured 88 models across different tasks and demonstrated major differences in energy use between task types. In their experimental setting, text classification required very little electricity per 1,000 inferences compared with image generation, which averaged about 2.9 kWh per 1,000 outputs. The result is not a universal per-query value, because actual consumption varies with hardware, model architecture, batching, input length, output length and deployment configuration, but it shows that environmental cost is highly task-dependent.

A second body of literature has shifted attention towards life-cycle and systems analysis. Kaack et al. (2022) distinguish among computing-related emissions, immediate impacts caused by AI applications and system-level impacts that may reshape production and consumption. The OECD (2022) similarly argues that measurement of AI's environmental impact should extend across hardware manufacturing, transport, operations and end-of-life treatment. UNEP (2024) emphasises an end-to-end assessment that includes energy, water, minerals, pollution and waste. This perspective is essential because an AI system powered by low-carbon electricity may still possess substantial embodied impacts, while an application with relatively high compute use may nevertheless produce a net environmental benefit if it displaces much larger emissions elsewhere.

Electronic waste forms a third and less developed strand of the literature. Wang et al. (2024) developed a computational-power-driven material-flow model for generative AI, particularly large language models, and estimated that cumulative AI-related e-waste could reach about 1.2-5.0 million tonnes over 2020-2030 under different development scenarios. Their analysis also suggested that circular-economy strategies could reduce this projected waste by roughly 16-86%, depending on the intervention and scenario. These values are prospective modelled ranges rather than observed quantities. Nevertheless, they highlight a problem often omitted from AI sustainability debates: the environmental benefit of more energy-efficient processors may be partly offset if equipment is replaced so rapidly that large quantities of still-functional hardware enter the waste stream.

The literature therefore reveals a significant research gap in integration. Many studies examine one metric-energy, carbon, water or e-waste-while policy discussions often use a single efficiency indicator to represent sustainability. Yet environmental effects occur across a connected chain from mineral extraction to manufacturing, operation, application and disposal. This paper adopts a life-cycle sustainability framework organised around four stages: production, operation, application and end-of-life. The framework also distinguishes efficiency from sufficiency. Efficiency asks how much resource is required for a unit of computation; sufficiency

asks whether the amount and complexity of computation are proportionate to the social value of the task. This distinction becomes crucial where rebound effects allow total consumption to grow despite continuing improvements in performance per watt.

Data and Methodology

The study applies a descriptive, analytical and comparative secondary-data methodology. Quantitative evidence is drawn primarily from authoritative international sources: the IEA's 2025 Energy and AI assessment for electricity and energy-system projections; the ITU-UNITAR Global E-waste Monitor 2024 for global e-waste statistics; UNCTAD's Digital Economy Report 2024 for the material and developmental dimensions of digitalisation; UNEP's 2024 AI life-cycle issue note; and OECD guidance on environmental measurement of AI compute. Peer-reviewed studies are used to analyse model-level energy differences, carbon accounting, AI-enabled climate applications, generative-AI e-waste and the energy-water-climate nexus. Literature was selected for relevance, methodological transparency and direct connection to AI or digital infrastructure.

Observed quantities are clearly distinguished from projections. For example, 415 TWh represents estimated global data-centre electricity consumption in 2024, whereas 945 TWh is an IEA base-case projection for 2030. Similarly, 62 million tonnes is the measured global e-waste total for 2022, while 82 million tonnes is the projected 2030 level. AI-specific e-waste estimates from Wang et al. (2024) are modelled scenarios rather than administrative waste statistics. The paper avoids converting broad data-centre figures into unsupported AI-only values. This is necessary because companies frequently do not disclose sufficiently detailed information to separate AI workloads from conventional cloud, storage and other computing services.

AI, Data Centres and Rising Energy Demand

The most visible environmental pressure associated with AI is the rapid growth of electricity demand in data centres. A modern data centre combines computing servers, storage devices, networking equipment, power conversion, backup systems and cooling. The IEA (2025) estimates that servers account for roughly 60% of electricity use in modern data centres on average, although the share varies across facility types. Cooling can represent a relatively small share in highly efficient hyperscale facilities but considerably more in older or less efficient enterprise centres. AI changes this profile because accelerated servers use dense arrays of specialised processors that deliver extremely high computing performance but also generate concentrated heat loads, requiring advanced power-delivery and cooling systems.

The IEA's estimate of 415 TWh of global data-centre electricity consumption in 2024 places the sector at about 1.5% of worldwide electricity use. The base-case projection of approximately 945 TWh by 2030 would more than double the total within six years. Accelerated servers, mainly

associated with AI adoption, are projected to show particularly rapid electricity growth and account for almost half of the net increase in global data-centre demand during 2024-2030. This does not mean that AI will consume 945 TWh by itself. Conventional cloud workloads, storage, enterprise computing and other digital services will remain important, but AI is increasingly changing both the scale and composition of new data-centre investment.

Carbon emissions depend on the electricity mix rather than electricity quantity alone. A data centre supplied by renewable, nuclear or other low-carbon sources has a lower operational carbon intensity than a similar facility located on a fossil-fuel-intensive grid. The IEA (2025) projects that renewables will supply a substantial portion of the additional electricity needed by data centres, but natural gas and other dispatchable sources are also expected to contribute. This creates a policy challenge: corporate renewable-energy procurement can reduce market-based emissions accounting, yet the real system impact depends on whether new clean generation is additional, available at the same time as demand and supported by sufficient grid infrastructure. Hourly carbon intensity can therefore be more informative than annual renewable-energy certificates alone.

Hardware and software efficiency can reduce energy per unit of computation. Improvements include more efficient accelerators, lower-precision arithmetic, quantisation, pruning, knowledge distillation, sparse computation, better memory management, batching and higher server utilisation. Smaller, domain-specific models may also perform many tasks adequately while consuming fewer resources than large general-purpose systems. The IEA's high-efficiency scenario shows that aggressive improvements in hardware, software and infrastructure could materially lower future data-centre electricity demand relative to the base case. Environmental efficiency should therefore become a formal design criterion rather than an incidental consequence of reducing cloud costs.

Efficiency, however, creates a rebound problem. When the cost of an inference falls, organisations can afford to make more inferences. When processors become more efficient, developers may use the gains to train larger models, generate longer outputs or add AI features to services that previously used conventional software. Total electricity use may therefore continue rising even while energy per task declines. This does not make efficiency undesirable; it means that efficiency must be combined with measurement of absolute resource consumption and with decisions about whether computational intensity is proportionate to the value of an application. Sustainable AI requires attention to both performance per watt and the overall scale of demand.

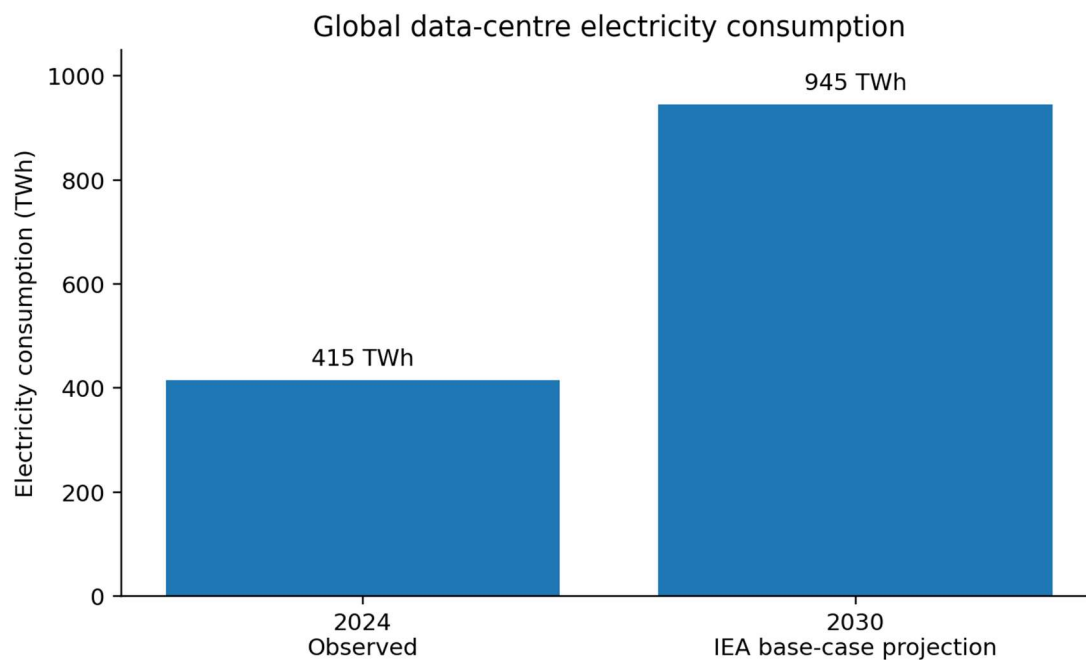


Figure 1. Global data-centre electricity consumption: 2024 observed estimate and 2030 IEA base-case projection. Source: IEA (2025).

E-Waste, Hardware Turnover and the Circular Economy

The material footprint of AI becomes particularly visible at the end of the hardware life cycle. ITU and UNITAR (2024) estimated that global e-waste reached a record 62 million tonnes in 2022, equivalent to about 7.8 kg per person. Only 22.3% of this mass was documented as formally collected and environmentally soundly recycled. If current trends continue, global e-waste is projected to reach 82 million tonnes in 2030, while the documented formal recycling rate could fall because waste generation is increasing faster than collection and treatment. The figures cover all electrical and electronic equipment, not AI infrastructure alone, but they demonstrate that any additional high-turnover computing stream enters a system already struggling to achieve circularity.

AI servers contain high-value and environmentally costly components: specialised processors, memory, storage, printed circuit boards, cooling systems, power electronics and networking equipment. Technological obsolescence may occur before physical failure because newer processors can provide better performance per watt and lower operating cost. A data-centre operator can therefore have a financial incentive to replace still-functional equipment if energy savings and computational throughput justify the capital cost. From a narrow operational perspective this can appear efficient; from a life-cycle perspective it creates a trade-off between lower electricity use and the embodied impacts of producing replacement hardware.

Wang et al. (2024) quantify this emerging risk for generative AI. Their scenarios suggest cumulative large-language-model-related e-waste of approximately 1.2-5.0 million tonnes between 2020 and 2030 under different pathways of AI expansion. The upper estimate is not a forecast that will necessarily occur; it depends on assumptions concerning server demand, hardware lifetimes and technological development. Nevertheless, the study demonstrates that accelerated replacement cycles can create a significant new waste stream. It also shows that circular strategies, including extending server life and improving reuse, could reduce projected generative-AI e-waste by approximately 16-86% across modelled interventions and scenarios.

Poorly managed e-waste carries environmental and health risks because electronic products contain hazardous substances as well as valuable materials. Unsafe dismantling, open burning, acid leaching and uncontrolled disposal can release pollutants into air, soil and water, while workers in informal recycling systems may face direct exposure. At the same time, discarded electronics contain copper, gold, aluminium and strategically important materials that can substitute for some primary extraction if recovered efficiently. The Global E-waste Monitor estimates that formal recycling already avoids substantial primary ore extraction and greenhouse-gas emissions, illustrating that e-waste is both a pollution problem and a secondary-resource opportunity.

A circular AI hardware strategy should therefore prioritise longevity before recycling. The hierarchy begins with avoiding unnecessary replacement, then repair, component reuse, refurbishment and cascading equipment into lower-intensity uses. A processor no longer competitive for frontier model training may remain useful for smaller inference workloads, academic computing or conventional data processing. Only after reuse opportunities are exhausted should equipment enter high-quality material recovery. Extended producer responsibility, manufacturer take-back systems, repairability standards, design for disassembly and digital product passports can support this hierarchy. Such measures reduce both waste volumes and dependence on virgin mineral extraction.

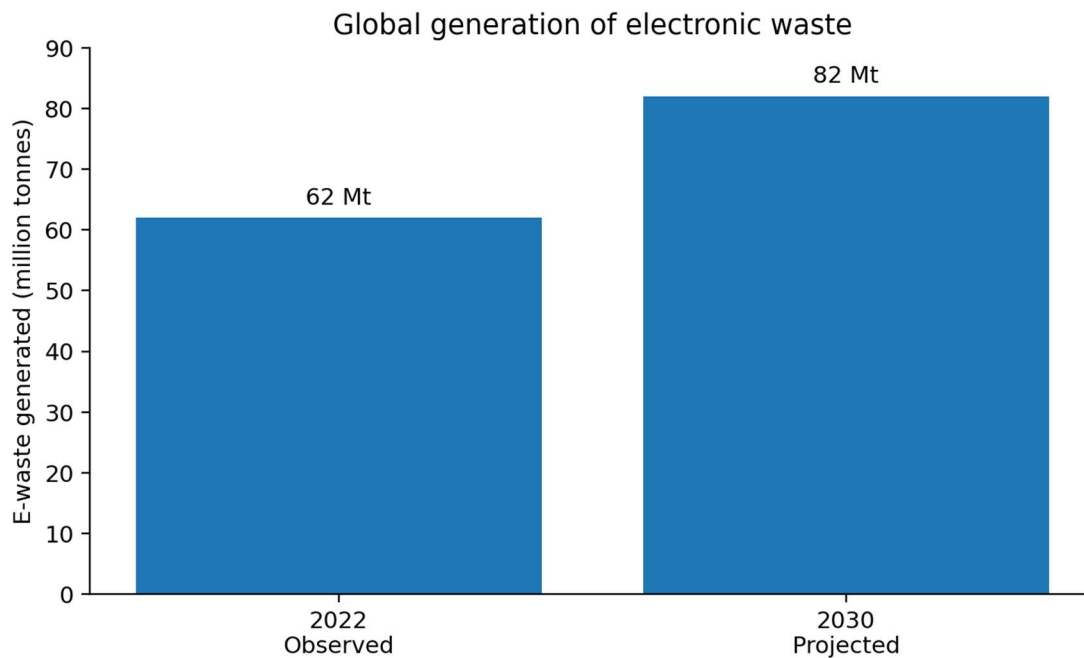


Figure 2. Global e-waste generation: 2022 observed total and 2030 projection. Source: ITU & UNITAR (2024).

Emerging Environmental Challenges Beyond Electricity

Water consumption has become one of the most important emerging issues in AI sustainability. Data centres can consume water directly through evaporative or liquid cooling and indirectly through electricity generation. Semiconductor fabrication also uses large quantities of ultrapure water. UNEP (2024) and OECD (2022) therefore recommend that AI environmental assessments include water withdrawal, consumption and local water stress rather than reporting carbon emissions alone. The significance of water use is highly spatial. A unit of water consumed in a humid, water-abundant basin does not have the same environmental consequence as the same quantity consumed during drought in an arid or heavily allocated watershed.

Recent modelling illustrates this spatial dependency. Xiao et al. (2025) analysed potential expansion of AI servers in the United States and found that location, cooling efficiency, server utilisation and grid development can strongly influence water and carbon footprints. Their results show that improving water-usage effectiveness can substantially reduce direct water consumption, while electricity-source choices affect indirect water use. The broader lesson is global: data-centre siting should consider local hydrology, seasonal water availability, climate risk and competing agricultural and household demands. A facility can have excellent power-use effectiveness and still create environmental stress if it is located in a water-scarce region.

Critical minerals and semiconductor supply chains add another layer of impact. Digital infrastructure depends on copper, aluminium, silicon, gold, rare-earth elements and numerous specialised materials used in processors, memory, power systems and communication equipment. UNCTAD (2024) emphasises that rapid digitalisation increases demand for physical raw materials even when end-user services appear virtual. Mining and refining can cause land disturbance, biodiversity impacts, water use, tailings generation and local pollution. Semiconductor fabrication requires energy-intensive industrial processes and highly controlled chemical inputs. These upstream effects are frequently located far from the data centres and consumers that receive the economic benefits of AI, making them easy to overlook in corporate operational reporting.

The geography of AI therefore raises questions of environmental justice. High-value model development, cloud services and platform ownership are concentrated in a small number of firms and countries, while mineral extraction, manufacturing and e-waste processing are more geographically dispersed. UNCTAD (2024) argues that developing countries can bear disproportionate environmental burdens from digitalisation while capturing a smaller share of the value generated by digital markets. A sustainable AI transition should avoid repeating patterns in which resource-producing regions experience pollution and ecological damage but remain dependent on low-value extraction. Circular supply chains, local value addition, safe recycling and fair environmental standards are consequently development issues as well as technical matters.

A final emerging challenge is the lack of standardised environmental disclosure. Companies may report corporate greenhouse-gas emissions and water consumption without separating AI from conventional computing. Model developers rarely disclose the full energy used for data preparation, experimentation, unsuccessful training runs, fine-tuning and inference. Hardware manufacturers may provide operational efficiency specifications without complete life-cycle inventories. UNEP (2024) therefore stresses the need for comprehensive methods that cover the full AI life cycle. Without comparable disclosure, regulators cannot establish credible benchmarks, researchers cannot validate claims and customers cannot reliably distinguish genuine efficiency improvements from environmental marketing.

AI as an Instrument for Environmental Sustainability

The environmental footprint of AI must be evaluated alongside the environmental services it can provide. Electricity systems are an important example. Machine learning can improve short-term forecasting of wind and solar generation, detect equipment failures, optimise grid operation, support demand response and assist maintenance of transmission assets. Better forecasting can reduce the reserve capacity required to manage variable renewable generation, while intelligent control can shift flexible demand towards periods of abundant clean electricity. AI can also

improve industrial process control, building energy management and logistics, thereby reducing energy waste when deployed carefully.

Environmental applications extend beyond energy. In agriculture, AI can support crop monitoring, irrigation scheduling, pest detection and precision application of fertiliser or pesticides. In conservation, computer vision and remote sensing can assist detection of deforestation, ecosystem change and wildlife populations. Machine-learning models can improve aspects of weather prediction, climate emulation and disaster-risk analysis, helping governments and communities anticipate hazards. Rolnick et al. (2022) catalogue opportunities across electricity, transport, buildings, industry, land use and climate science. These applications demonstrate why the objective of environmental policy should not be to minimise AI use indiscriminately, but to maximise environmental and social benefit relative to resources consumed.

The distinction between gross and net impact is critical. An AI application that consumes additional electricity but prevents a much larger quantity of fuel use, material waste or emissions may be environmentally beneficial. Conversely, an application marketed as efficient may have little net benefit if it stimulates additional consumption. Kaack et al. (2022) therefore emphasise system-level effects beyond the immediate computing footprint. Environmental appraisal should compare an AI-enabled activity with the realistic alternative, include indirect changes in behaviour and identify rebound effects. This approach is more rigorous than simply classifying AI applications as 'green' because they are associated with climate or environmental topics.

Strategies for Environmentally Sustainable AI

The first requirement for sustainable AI is transparent measurement. Developers and data-centre operators should report electricity consumption, carbon intensity, water use, hardware utilisation and relevant life-cycle impacts using comparable boundaries. Training and inference should be distinguished, and where possible companies should disclose environmental performance by major workload category. Location-specific information is essential because the same electricity consumption can generate very different carbon and water impacts in different grids and watersheds. Public disclosure would also enable independent researchers to evaluate whether efficiency improvements translate into lower absolute resource use.

Second, AI systems should adopt efficiency by design. Model developers can choose smaller task-specific models where they provide sufficient accuracy, minimise unnecessary retraining and use quantisation, pruning, distillation and efficient inference. Procurement and research evaluation can include energy or carbon efficiency alongside accuracy and latency. Such criteria would reinforce the Green AI principle proposed by Schwartz et al. (2020). The goal is not to prohibit large models where they provide substantial value, but to make computational intensity an explicit design choice rather than an unexamined default.

Third, data centres should operate on increasingly low-carbon electricity while improving infrastructure efficiency. Renewable energy, nuclear power and other low-carbon sources can reduce operational emissions, but procurement strategies should encourage additional clean generation and grid capacity rather than rely solely on annual accounting instruments. Workloads that are flexible in time can be shifted towards hours of lower carbon intensity, while geographically flexible tasks can be scheduled in regions with cleaner electricity. Cooling design, server utilisation, power conversion and thermal management should also be optimised because facility efficiency reduces both energy costs and environmental pressure.

Fourth, water-sensitive planning must become a formal component of data-centre sustainability. Facilities should assess basin-level water stress, drought risk and competing community needs before construction. Water-use effectiveness should be reported together with the source of water and seasonal conditions. Alternatives such as air cooling, closed-loop systems, treated wastewater and advanced liquid cooling can reduce freshwater demand in suitable contexts, although each technology has trade-offs in energy use, cost and climate suitability. There is no universal best cooling solution; environmental performance must be evaluated locally.

Fifth, circularity should guide hardware policy. Servers and accelerators should be designed and purchased with expectations for durability, repair, refurbishment and secondary use. Replacement decisions should include embodied carbon and material impacts rather than comparing operational electricity costs alone. Extended producer responsibility can require manufacturers to finance take-back and recycling, while procurement rules can require traceable end-of-life management. Wang et al. (2024) show that circular strategies can materially reduce projected AI-related e-waste, making hardware life extension one of the most direct ways to reduce the material footprint of AI expansion.

Finally, sustainability policy must consider sufficiency and social value. Not every digital function requires a frontier generative model. Conventional algorithms, databases, search tools or smaller models may sometimes achieve the same practical outcome using fewer resources. Resource-intensive AI should be justified by commensurate benefits, particularly in publicly funded systems. This principle can coexist with innovation: it simply recognises that computational resources, clean electricity, water and critical materials are finite. The most sustainable AI system is therefore not always the most efficient large model; in many contexts it is the least resource-intensive method that reliably solves the problem.

Major Findings and Discussion

The synthesis produces six principal findings. First, AI is a major driver of rapidly increasing data-centre electricity demand, but data-centre totals should not be misrepresented as AI-only consumption. Second, carbon impacts are determined by both computational demand and the electricity system that supplies it. Third, the environmental footprint extends substantially

beyond electricity to water, semiconductor production, critical minerals, land and infrastructure. Fourth, the existing global e-waste system is already inadequate, making rapid AI hardware turnover a significant emerging risk. Fifth, efficiency gains are essential but may be offset by rebound effects if lower computational costs stimulate much larger volumes of use. Sixth, AI can produce substantial environmental benefits, but these must be evaluated through net system impact rather than through assumptions that digital optimisation is automatically sustainable.

Conclusion

Artificial intelligence is becoming one of the defining technologies of the twenty-first century, but its environmental implications are inseparable from the physical infrastructures that support it. The rapid expansion of data centres, accelerated servers and generative-AI services is increasing electricity demand at a pace that matters for power-system planning. The IEA's estimate of about 415 TWh of global data-centre electricity consumption in 2024 and its base-case projection of roughly 945 TWh by 2030 demonstrate the scale of the transition. These figures do not imply that AI alone consumes the entire amount, but they show that AI is reshaping the trajectory of global digital infrastructure.

The e-waste dimension is equally important. The world generated 62 million tonnes of e-waste in 2022, while only 22.3% was documented as formally collected and environmentally soundly recycled. With global e-waste projected to reach 82 million tonnes by 2030, rapid server replacement could intensify an already serious circular-economy challenge. Modelled estimates of 1.2-5.0 million tonnes of cumulative generative-AI-related e-waste over 2020-2030 further illustrate why hardware longevity, reuse and responsible recycling need to become central components of AI sustainability policy.

Environmental governance must also account for water, minerals and spatial inequality. Data-centre cooling, electricity generation and semiconductor manufacturing can place pressure on water resources, while advanced hardware depends on material supply chains whose environmental costs often occur far from users. The benefits of AI and the burdens of extraction, infrastructure and waste are therefore distributed unevenly across regions. A credible sustainable-AI framework should incorporate local water stress, grid carbon intensity, material circularity, labour and community impacts, and transparent international standards.

AI nevertheless has substantial potential to assist environmental protection. Applications in renewable-energy integration, efficient buildings, transport, agriculture, conservation and climate science can reduce emissions or improve resilience when they replace less-efficient alternatives. The correct objective is therefore neither unrestricted computational growth nor indiscriminate restriction. It is to maximise useful social and environmental outcomes per unit of electricity, water and material consumed while ensuring that absolute resource use remains compatible with climate and sustainability goals.

The pathway forward requires a combination of Green AI and AI for Green. Green AI reduces the footprint of computing through efficient models, low-carbon electricity, responsible siting, improved cooling and circular hardware. AI for Green deploys artificial intelligence where it can produce genuine environmental benefits. Efficiency must be accompanied by transparency and sufficiency so that rebound effects do not erase resource savings. If governments, researchers and technology companies adopt life-cycle measurement, environmental standards and circular-economy principles, AI can be developed in a manner that supports rather than undermines the transition towards a more sustainable global economy.

Table 2. Sustainability challenges and priority responses

Environmental challenge	Main risk	Priority response
Electricity demand	Rapid load growth and grid stress	Efficient models, higher utilisation, low-carbon electricity, flexible scheduling
Carbon emissions	Fossil-intensive electricity and embodied emissions	Additional clean generation, life-cycle carbon reporting, hardware longevity
Water use	Cooling and indirect electricity-related water consumption	Basin-sensitive siting, WUE disclosure, water-efficient cooling and reuse
Critical minerals	Extraction impacts and supply concentration	Material efficiency, responsible sourcing, circular recovery and design for reuse
E-waste	Rapid equipment turnover and inadequate recycling	Repair, refurbishment, cascading reuse, producer responsibility and formal recycling
Rebound effects	Efficiency gains stimulate higher total use	Track absolute consumption and use task-appropriate models
Transparency	Inconsistent boundaries and limited AI-specific disclosure	Harmonised international reporting standards and independent verification

References

- de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191-2194. <https://doi.org/10.1016/j.joule.2023.09.004>
- Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), 1-43.
- International Energy Agency. (2025). *Energy and AI*. IEA, Paris. <https://www.iea.org/reports/energy-and-ai>

- International Telecommunication Union, & United Nations Institute for Training and Research. (2024). The Global E-waste Monitor 2024. ITU/UNITAR. https://www.itu.int/pub/D-GEN-E_WASTE.01-2024
- Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 12, 518-527. <https://doi.org/10.1038/s41558-022-01377-7>
- Luccioni, A. S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 85-99. <https://doi.org/10.1145/3630106.3658542>
- Organisation for Economic Co-operation and Development. (2022). Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint. *OECD Digital Economy Papers*, No. 341. <https://doi.org/10.1787/7babf571-en>
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., et al. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, 55(2), Article 42, 1-96. <https://doi.org/10.1145/3485128>
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63. <https://doi.org/10.1145/3381831>
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645-3650. <https://doi.org/10.18653/v1/P19-1355>
- United Nations Conference on Trade and Development. (2024). Digital Economy Report 2024: Shaping an environmentally sustainable and inclusive digital future. United Nations. <https://unctad.org/publication/digital-economy-report-2024>
- United Nations Environment Programme. (2024). Artificial intelligence (AI) end-to-end: The environmental impact of the full AI lifecycle needs to be comprehensively assessed. UNEP. <https://www.unep.org/resources/report/artificial-intelligence-ai-end-end-environmental-impact-full-ai-lifecycle-needs-be>
- Wang, P., Zhang, L.-Y., Tzachor, A., & Chen, W.-Q. (2024). E-waste challenges of generative artificial intelligence. *Nature Computational Science*, 4, 818-823. <https://doi.org/10.1038/s43588-024-00712-6>
- Xiao, T., Fuso Nerini, F., Matthews, H. D., Tavoni, M., & You, F. (2025). Environmental impact and net-zero pathways for sustainable artificial intelligence servers in the USA. *Nature Sustainability*, 8, 1541-1553. <https://doi.org/10.1038/s41893-025-01681-y>