International Journal of Advance Research in Science, Management, Engineering and Technology ISSN: 3139-1141
Volume 1, Issue 5 | November - December 2026 | www.ijarsmet.com

Artificial Intelligence and Its Growing Environmental Impact: Challenges and Sustainable Solutions

Lilly Michael Daniel[#1], Arpita Ravindra Dedge[#2], Rupali Kalekar [#3]

Postgraduate Research Scholar, Department of Computer Application, Zeal Institute of Business Administration, Computer Application and Research (ZIBACAR), Savitribai Phule Pune University, Pune, Maharashtra, India

Research Guide: Dr. Rupali Kalekar

[1] lillydaniel0012@gmail.com , [2]arpitadedge2002@gmail.com , [3] Rupali.kalekar@zealeducation.com

***Abstract*—Artificial intelligence is being adopted across nearly every sector, yet the computing infrastructure behind it draws growing amounts of electricity and water and depends on resource intensive hardware that eventually becomes electronic waste. At the same time, artificial intelligence tools help forecast weather, balance power grids, and monitor ecosystems. This paper reviews the evidence on both sides of that balance. Using a descriptive and analytical secondary research design, it draws on peer reviewed studies, intergovernmental reports, and official statistics from India, the United States, and international agencies, published mostly between 2019 and 2026. The reviewed sources show that data centre electricity demand is projected to more than double by 2030, that water use occurs both on site and through power generation, that hardware manufacturing contributes embodied emissions that are often left out of reporting, and that electronic waste is growing much faster than documented recycling. The sources also indicate that the environmental benefits of artificial intelligence applications are real but depend on deployment context and are seldom measured with the same rigor as the costs. The paper classifies impacts across the life cycle, compares costs with benefits, and proposes a layered Green AI framework covering governance, infrastructure, models, and applications. Recommendations are given for developers, operators, researchers, and policymakers. No new experiments were conducted, and the findings are limited by the quality and comparability of published data.**

***Keywords*—Artificial Intelligence; Environmental Sustainability; Carbon Footprint; Energy Consumption; Green AI; Data Centres; Water Consumption; Electronic Waste**

I. INTRODUCTION

Machine learning, which has been developed during the past decade, has shifted from research laboratories to commercial applications including search engines, language translators, software for image processing in medicine, credit risk assessment, logistical planning and even more recently, algorithms for generating text and images on demand. This is all enabled through the use of physical infrastructure. The training of models is done in large clusters of specialized processors powered from data centers which run uninterrupted. The amount of electricity consumed by these data centers according to the International Energy Agency (IEA) in 2024 would be about 415 TWh, accounting for 1.5 percent of total electricity consumption, rising up to 945 TWh in 2030 [1].

However, electricity is just one aspect of the problem. Coolants use water, silicon-chip production uses lots of chemicals, and old servers that are out of service are a part of e-waste which is mostly not recorded as recycled [2]. UNEP and OECD have stated that the environmental impact of AI should take into account the whole process, including the full life-cycle of the system rather than the electricity required to run it once [3], [4]. At the same time, the very same system can be employed to predict weather patterns, control electricity networks and track down forests, which means that an analysis should take into account all the costs and benefits, instead of looking at AI solely as either a menace or a boon [5], [6].

This issue has a unique significance to India, as according to official statistics, data center capacities have increased over four times since 2020, with the IndiaAI Mission increasing the shared computing capacity to over 45,000 graphics processing units [7]. According to a government white paper, data centers currently responsible for around 0.5 percent of national electricity consumption can rise to 3 percent of total national electricity consumption by 2030 [8]. Thus, decisions taken in this regard at present would have ramifications for AI in India for years to come.

There are three goals of this study. Firstly, it categorizes the environmental impacts of AI through the whole process of its life cycle and measures them where possible using official or peer-reviewed sources of information. Secondly, it contrasts these impacts with the advantages of AI applications that are either proven or potential. Finally, it formulates a Green AI strategy based on recommendations for all stakeholders. The contribution of the paper is more theoretical than empirical as it critically analyzes and cross-references available sources.

II. BACKGROUND AND RELATED CONCEPTS

A. Artificial Intelligence and Generative Artificial Intelligence

AI is defined as the computer systems that discover patterns from data and use them for classification, prediction, recommendations, or generation. Modern AI models rely on deep neural networks, which have two lives. The first life involves adjusting billions of parameters based on the data and is carried out only once or a few times during the model's development. The second life involves applying the model to new data and is repeated each time the model is used. Inference is carried out many times and, thus, its share in overall energy consumption could surpass that of the training process. This idea is highlighted in life cycle analyses of AI systems [9].

Generative models further amplify this issue. Empirical evidence across several tasks shows that multi-task generative models require substantially more energy per query compared to single-task models, especially in the case of image generation [10].

B. *Environmental Sustainability*

The concept of environmental sustainability as discussed in this context is defined as the ability of the current generation to meet its needs without compromising the climate, water, material reserves, and ecosystems which support future generations' needs. In relation to the digital technology under discussion, the criteria can be narrowed down to just five measurable environmental impacts: greenhouse gas emissions, electricity consumption, freshwater consumption, material extraction and waste creation. It should also be noted that two types of impact are distinguished in this paper: operational and embodied.

C. *Green Artificial Intelligence*

The concept of Green AI was proposed to be in opposition to that of Red AI, which refers to research aimed at improving accuracy through ever-increasing amounts of computation. It is a relatively humble proposal, in terms of content – to simply note the amount of computation used to obtain an outcome in addition to its accuracy and consider efficiency as a primary evaluation criterion to avoid any penalties incurred for lightness [12]. The Green AI movement has broadened its scope to include, amongst other things, carbon-efficient scheduling and hardware.

D. *Artificial Intelligence Infrastructure and Its Life Cycle*

The AI system's life cycle involves the mining and processing of metal ores and minerals, followed by the manufacture of semiconductors, assembling servers and building the data centers, operating through the process of training, inference, and cooling, until the final stages of either decommissioning, reuse, or disposal. OECD characterizes the impacts created as being either direct impacts that arise from the development, deployment, and decommissioning of AI systems and equipment, or indirect impacts from the uses to which these technologies can be put [4]. Table I presents the mapping of these phases against their main pressures.

TABLE I. ENVIRONMENTAL PRESSURES ACROSS THE ARTIFICIAL INTELLIGENCE LIFE CYCLE

Life-cycle stage	Principal environmental pressure	Evidence
Raw materials and mining	Extraction of metals and minerals; ecosystem disturbance	[3], [2]
Chip and server manufacturing	Embodied emissions; process water and chemicals	[11], [3]
Facility construction	Embodied emissions of buildings and equipment; land use	[4]
Model development and training	Electricity use and associated emissions	[13], [14], [15]
Inference and deployment	Continuous electricity demand at scale	[10], [9]
Cooling and power supply	On-site and off-site water consumption	[16], [17]
Retirement	Electronic waste; unrecovered materials	[2], [18]

III. LITERATURE REVIEW

Early carbon accounting. The first widely cited estimate came from a study of natural language processing models, which calculated that training one large transformer model together with a neural architecture search emitted roughly 626,000 lb of carbon dioxide equivalent, about 284 tonnes [13]. A later analysis of large models examined the training of GPT-3 and reported about 1,287 MWh of electricity and about 552 tonnes of carbon dioxide equivalent, while arguing that the choice of model architecture, hardware and data centre location could change emissions by two to three orders of magnitude [14]. The two studies agree that location and hardware matter more than is commonly assumed. They differ in what they count. The first figure includes an architecture search that few practitioners ever run, so it overstates a typical project, whereas the second concerns a single production-scale model and is closer to what developers of very large systems actually do.

Life-cycle studies. Accounting for operational electricity alone understates the burden. For the open BLOOM language model, the authors estimated 24.7 tonnes of carbon dioxide equivalent from dynamic power use but 50.5 tonnes once equipment manufacturing and idle consumption were included [15]. The result echoes earlier work on computing in general, which found that hardware manufacturing and infrastructure make up a major share of lifetime emissions [11]. It should not be read as a universal ratio. BLOOM was trained on a relatively low-carbon electricity supply, so the operational share would be larger on a fossil-heavy grid, and the embodied share smaller.

Aggregate demand. Estimates of sector-wide electricity use rest on different methods. A supply-side projection based on server shipments suggested that AI servers could use between 85 and 134 TWh per year by 2027 [19]. The IEA uses a bottom-up model of data centre capacity and projects global data centre demand of about 945 TWh in 2030 [1], and a United States laboratory study commissioned under the Energy Act of 2020 found that American data centres used 176 TWh in 2023 and could use between 325 and 580 TWh by 2028 [16]. The sources agree on direction and on the role of AI hardware, but their ranges are wide, and the United States study itself presents post-2023 figures as scenarios rather than point estimates because AI hardware has changed so quickly.

Water. Water footprints received attention later than carbon. One study estimated that training GPT-3 in United States data centres consumes about 5.4 million litres of water including off-site use, of which 700,000 litres is evaporated on site, and that a conversation of roughly 10 to 50 responses can correspond to half a litre [17]. The national laboratory study reaches a consistent qualitative conclusion at the sector level, with 66 billion litres consumed directly and about 800 billion litres indirectly through electricity generation in 2023 [16]. Both sources stress that the indirect component is much larger than the direct one and that company disclosures rarely include it.

Benefits and net effects. A broad review catalogued ways in which machine learning can support climate mitigation and adaptation, from power-system forecasting to remote sensing [5]. An expert-elicitation study judged that AI could act as an enabler for 134 of the 169 Sustainable Development Goal targets while potentially inhibiting 59 [6]. These are assessments of potential, not measured outcomes. A more cautious analysis separates the effects of AI into computing-related impacts, immediate application impacts and system-level effects, and warns that some applications, such as those serving fossil fuel extraction or encouraging consumption, can raise emissions [20]. The strongest demonstrated benefits are in tasks such as weather prediction, where a trained model has been shown to outperform a leading operational forecasting system on most verification targets [21].

Measurement and governance. The OECD identified gaps in measurement standards, data collection and transparency, and recommended going beyond operational energy to include embodied and indirect effects [4]. UNEP reached a similar conclusion and called for scientific methods that allow objective measurement of AI's footprint [3]. The European Union has begun to act: its delegated regulation requires operators of data centres above 500 kW of installed information technology power to report harmonised indicators on energy, renewable energy, cooling and water [22].

Position of the present review. The earlier work is strong on individual models, and the official reports are strong on aggregates, but few studies place electricity, water, materials and waste next to the benefits of applications in a single framework, and fewer still use Indian official statistics for that purpose. This review combines the strands, applies one classification across them, and calculates simple derived indicators so that readers can see how the conclusions follow from the data.

IV. Research Gap and Problem Statement

Five gaps emerge from the literature. First, environmental reporting by developers and operators is limited and inconsistent: model cards seldom disclose water use, and operational energy is reported far more often than embodied emissions [17], [4]. Second, energy and water requirements are estimated with divergent methods and boundaries, so published figures cannot be compared directly [19], [1], [16]. Third, the contribution of AI hardware to manufacturing impacts and to electronic waste is poorly separated from that of electronics in general; national and global waste statistics are not broken down in a way that identifies servers and accelerators [2], [18]. Fourth, costs and benefits are usually discussed in separate literatures, with benefits mostly framed as potential [20], [6]. Fifth, practitioners lack a consolidated, practical set of strategies that links technical measures to infrastructure and governance.

The problem addressed in this paper is therefore how the environmental costs of AI across its life cycle can be assessed together with its documented benefits, using the best available public evidence, and what practical measures can reduce the net burden without discarding the benefits.

V. Research Objectives

The study pursues the following objectives. These were drafted from the research direction supplied for this work and should be matched word for word to the approved research synopsis before submission.

- To examine the electricity, greenhouse-gas, water, material and waste pressures associated with the AI life cycle.
- To review published evidence on AI applications that support environmental monitoring, prediction and resource efficiency.
- To identify gaps in measurement and reporting that prevent a consistent comparison of environmental costs and benefits.
- To analyse official global, United States and Indian statistics relevant to AI infrastructure and electronic waste.
- To propose a Green AI framework and practical recommendations for developers, operators, researchers and policymakers.

VI. Research Methodology

A. Research Design

The study is a descriptive and analytical secondary research. It does not report laboratory measurements, surveys or simulations. The analytical element consists of classifying evidence, cross-checking figures between independent sources, and calculating simple derived indicators from published numbers.

B. Literature Identification and Selection

Sources were identified in October 2026 through the publication portals of the IEA, UNEP and the OECD; parliamentary records and white papers of the Government of India; the United States Department of Energy and its national laboratories; and the ACM Digital Library, IEEE Xplore, Springer Nature and Science. Google Scholar was used to discover original studies and to trace citations. Search terms combined artificial intelligence, machine learning, large language model, data centre, energy, carbon, water, e-waste and sustainability. Priority was given to material published from 2019 onward. Studies were included when they reported a method, a dataset or an official statistic relevant to one of the objectives. Vendor marketing, unattributed blog posts and secondary summaries that could not be traced to an original source were excluded, and where press coverage was the only route to a government statement, the underlying official source was located before the figure was used.

C. Source Reliability and Relevance

Evidence was ranked. Official statistics and intergovernmental reports came first, peer reviewed journal and conference papers second, and preprints third and labelled as such. Four of the cited studies circulate as preprints and are flagged in the reference list [14], [17], [23], [24]. Whenever two sources estimated the same quantity, both were retained and the difference was reported rather than reconciled by assumption.

D. Classification and Comparative Analysis

Impacts were classified using the life-cycle scheme in Table I. Benefits were grouped by application domain. Comparisons are qualitative where data are not

commensurable, and quantitative only where the same measure is available at both ends. Derived indicators use the compound annual growth rate, calculated as the ratio of the final to the initial value raised to the power of one over the number of years, minus one. Every derived figure is labelled as the authors' calculation and the inputs are shown in Table III.

E. Limitations of the Method

Because the study depends on published data, it inherits their boundaries, vintage and uncertainty. Section XIV discusses these limitations and the quantitative work that would address them.

VII. Environmental Impact of Artificial Intelligence

Table II collects the principal official and intergovernmental figures used in this section. The remainder of the section interprets them, following the life-cycle classification.

TABLE II. Selected official and intergovernmental statistics used in this review

Indicator	Reported value	Source
Global data centre electricity use, 2024	About 415 TWh (about 1.5% of global use)	[1]
Projected global use, 2030 (Base Case)	About 945 TWh	[1]
United States data centre use, 2023	176 TWh (4.4% of national use)	[16]
United States projection, 2028	325 to 580 TWh (6.7% to 12%)	[16]
United States direct water use, 2023	66 billion litres	[16]
Global e-waste generated, 2022	62 million tonnes; 22.3% documented as formally recycled	[2]
India e-waste, 2023-24	1.75 million tonnes generated; 0.76 million tonnes recycled	[18]
India data centre capacity	375 MW (2020) to about 1,575 MW (2026)	[7]
India data centre share of electricity	About 0.5% now; about 3% projected by 2030	[8]

A. Electricity Consumption

The IEA reports that the United States accounted for 45 percent of global data centre electricity use in 2024, followed by China at 25 percent and Europe at 15 percent, and that consumption has grown by about 12 percent per year since 2017, more than four times faster than total electricity demand [1]. From its 2024 estimate of 415 TWh, the Base Case rises to about 945 TWh in 2030 and about 1,200 TWh in 2035, as shown in Fig. 1. The implied compound growth from 2024 to 2030 is about 14.7 percent per year (authors' calculation). The same report places this growth in perspective: data centres account for around one-tenth of global electricity demand growth to 2030, less than industrial motors, air conditioning or electric vehicles, although in advanced economies the share exceeds 20 percent [1].

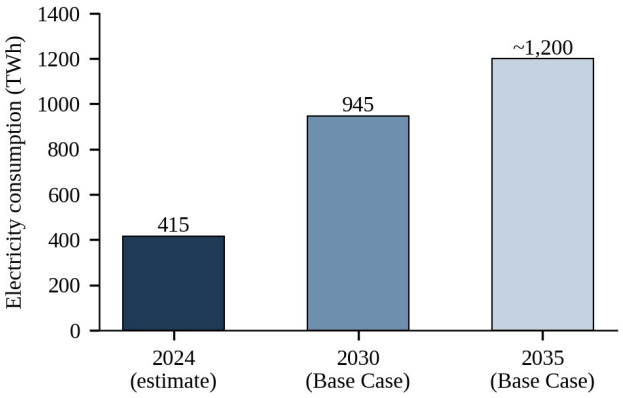


Fig. 1. Global data centre electricity consumption, 2024 estimate and IEA Base Case projections. Source: IEA, Energy and AI [1].

The American record is longer and shows the acceleration clearly. Laboratory estimates put data centre use at 58 TWh in 2014, about 76 TWh in 2018 and 176 TWh in 2023, with 2028 scenarios of 325 to 580 TWh (Fig. 2) [16]. From these values, annual growth was about 13.1 percent over 2014 to 2023 and about 18.3 percent over 2018 to 2023 (authors' calculation), so the recent pace is faster than the decade average. The 2028 range is wide, which reflects uncertainty about how many AI servers will be installed and how intensively they will be used.

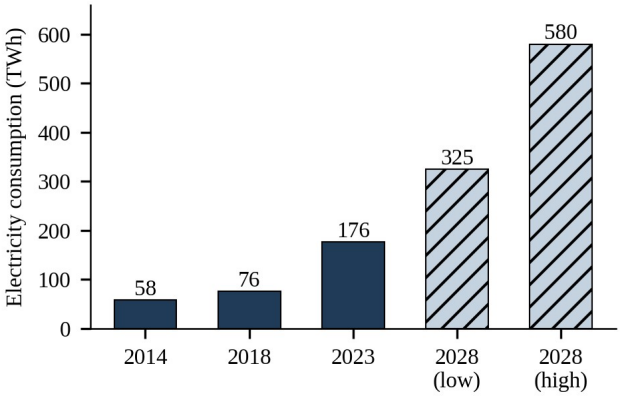


Fig. 2. United States data centre electricity consumption, 2014 to 2023 and 2028 scenario range (hatched bars are projections). Source: Lawrence Berkeley National Laboratory [16].

Indian figures are smaller in absolute terms but are growing quickly. The Ministry of Electronics and Information Technology reports installed capacity rising from 375 MW in 2020 to about 1,575 MW, a factor of about 4.2 [7]. The Principal Scientific Adviser's white paper uses a lower baseline of about 960 MW and projects 9.2 GW by 2030 [8]. The two government sources are therefore not on the same basis, a difference worth noting because capacity statistics are routinely used to plan power supply.

B. Greenhouse-Gas Emissions

Emissions follow electricity use multiplied by the carbon intensity of the supply, so identical computation can have very different footprints in different places and at different times. This is visible in the model-level studies already discussed [13], [14], [15]. At system level, the IEA expects renewables to supply about half of the growth in data centre demand, with natural gas and nuclear power each adding a substantial

amount, which means that the emissions outcome will depend heavily on regional choices about the generation mix [1]. By the same logic, a workload run on a more carbon-intensive grid carries a larger operational footprint unless clean power is procured explicitly; whether this applies to particular Indian facilities has not been measured in the present study.

C. Data Centre Cooling and Water Use

Servers convert almost all of their electrical input into heat, and many facilities remove it by evaporating water. The national laboratory study estimated direct consumption of 66 billion litres in 2023, up from about 21 billion litres in 2014, and a further 800 billion litres consumed indirectly at power plants [16] (Fig. 3). The ratio of indirect to direct use is about 12 to 1 (authors' calculation), and the direct intensity is about 0.38 litres per kWh. At the level of one model, training GPT-3 was estimated to need 5.4 million litres in total, and the water cost of a given amount of computing was found to vary with place and season [17]. National totals can look small beside agriculture, but data centres are clustered, so local stress can be significant. The authors did not locate an official Indian statistic for data centre water consumption, which is itself a finding about the state of public reporting.

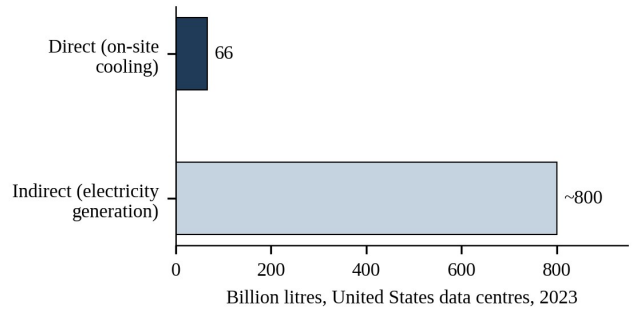


Fig. 3. Direct and indirect water consumption of United States data centres, 2023. Source: Lawrence Berkeley National Laboratory [16].

D. Semiconductor and Hardware Manufacturing

Building the hardware is energy, water and chemical intensive. A study of the carbon footprint of computing concluded that manufacturing and infrastructure account for a large part of lifetime emissions, a conclusion that grows in importance as operational energy shifts to cleaner sources [11]. UNEP lists mineral and energy consumption in chip production among the direct impacts that remain poorly quantified for AI specifically [3]. The BLOOM result, in which the all-inclusive estimate was roughly double the dynamic-power figure, illustrates the scale of what is omitted when only operations are counted [15].

E. Raw-Material Consumption

Servers and accelerators contain copper, aluminium, gold, silicon, rare earth elements and other materials. Global electronic waste in 2022 contained an estimated 31 million tonnes of metals, and recycling met only about 1 percent of rare earth element demand [2]. AI-specific material demand has not been separately quantified in the sources reviewed, so claims about its scale should be treated with caution.

F. Electronic Waste

The Global E-waste Monitor reports that 62 million tonnes of electronic waste were generated worldwide in 2022, 82 percent more than in 2010, and that only 22.3 percent was documented as formally collected and recycled; generation is projected to reach 82 million tonnes by 2030, with the documented rate falling to about 20 percent [2] (Fig. 4). Between 2010 and 2022, generation rose by 28 million tonnes while documented recycling rose by 5.8 million tonnes, so the former increased almost five times as much in absolute terms (authors' calculation). The documented share fell slightly from about 23.5 percent to 22.3 percent.

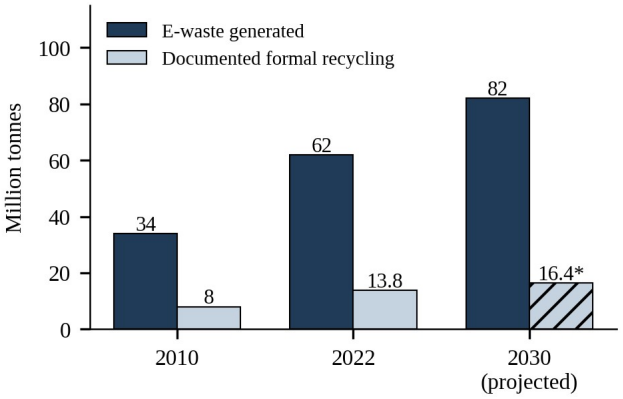


Fig. 4. Global e-waste generated and documented as formally collected and recycled. The 2030 recycled value (marked with an asterisk) is the authors' calculation from the projected 20 percent rate. Source: Global E-waste Monitor 2024 [2].

Government of India data give a national picture. Parliamentary replies from the Ministry of Environment, Forest and Climate Change report generation of 1.60, 1.61 and 1.75 million tonnes in 2021-22, 2022-23 and 2023-24, with 0.53, 0.54 and 0.76 million tonnes recycled respectively [18] (Fig. 5). The recycled share therefore rose from about 33 percent to about 43 percent (authors' calculation). Two cautions apply. The ministry notes that the E-Waste (Management) Rules, 2022 widened the list of covered items from 21 to 106 and came into force in April 2023, so the series is not strictly like-for-like [18], and recycled quantities are reported by the authorities as collected and processed, not as an independent audit. Neither the global nor the Indian statistics isolate data centre equipment.

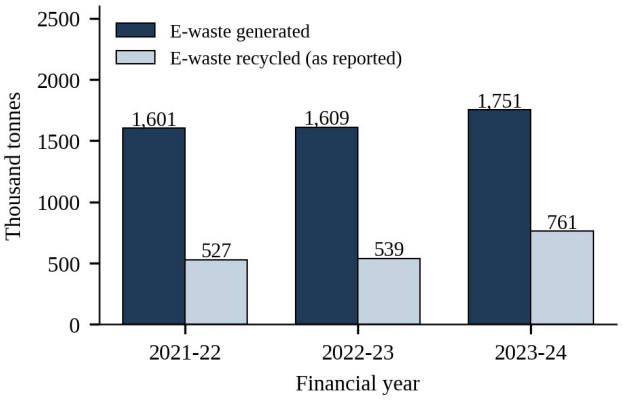


Fig. 5. India e-waste generated and recycled, financial years 2021-22 to 2023-24. Source: Ministry of Environment, Forest and Climate Change, reply in the Rajya Sabha [18].

G. Life-Cycle Environmental Assessment

A life-cycle view reconciles the preceding items. Operational electricity dominates the visible footprint today, but embodied emissions, indirect water and end-of-life waste are each of comparable order in the studies that attempt to measure them [15], [16], [2]. The OECD and UNEP both recommend assessments that cover all stages and distinguish direct from indirect effects [4], [3]. Such assessments remain rare, and the resulting evidence is therefore partial.

TABLE III. DERIVED INDICATORS CALCULATED BY THE AUTHORS FROM PUBLISHED DATA

Indicator	Calculation	Result
Global data centre electricity growth, 2024 to 2030	(945/415)^(1/6) - 1 [1]	About 14.7% per year
United States growth, 2014 to 2023	(176/58)^(1/9) - 1 [16]	About 13.1% per year
United States growth, 2018 to 2023	(176/76)^(1/5) - 1 [16]	About 18.3% per year
Indirect to direct water, United States, 2023	800/66 [16]	About 12 to 1
Documented recycling share of global e-waste	8/34 (2010); 13.8/62 (2022) [2]	23.5% to 22.3%
Increase in e-waste 2010 to 2022: generated vs. documented recycling	62 - 34; 13.8 - 8 [2]	28 vs. 5.8 million tonnes
India recycled share of generated e-waste	Recycled/generated, 2021-22 to 2023-24 [18]	32.9%; 33.5%; 43.4%

VIII. POSITIVE ENVIRONMENTAL APPLICATIONS OF ARTIFICIAL INTELLIGENCE

The same techniques that create the footprint also produce measurable environmental gains in some domains. The evidence is uneven, and for each application below the paper distinguishes demonstrated results from potential.

A. Climate and Weather Prediction

The clearest demonstrated benefit is in medium-range weather forecasting. A graph neural network model trained on decades of reanalysis data was reported to outperform the European Centre for Medium-Range Weather Forecasts deterministic system on most of the verification targets evaluated and to produce a ten-day forecast in under a minute on a single accelerator, which is far less computing than conventional numerical methods require [21]. The training cost is substantial but is incurred once, while the forecasts are repeatable, which is the kind of asymmetry that favours net benefit.

B. Renewable-Energy Forecasting

Wind and solar output varies with weather, and better short-term forecasts allow grid operators to hold less standby capacity and to integrate more renewable generation. Machine learning methods for this purpose are among the application areas catalogued in the climate-and-machine-learning literature [5]. The IEA likewise notes that energy companies already use AI to optimise generation and consumption [1]. Published evidence on realised emissions savings is limited, and results depend on the market and the forecasting baseline.

C. Smart Electricity Grids

Grid operators can use AI for demand forecasting, fault detection, congestion management and flexible operation of loads, including data centres themselves. The IEA identifies flexibility in how data centre servers and on-site generation operate as one of the key options for easing grid constraints [1]. As with forecasting, the magnitude of the benefit is context dependent and has not been established in a single comparable figure.

D. Precision Agriculture and Water Conservation

Sensor and satellite data combined with machine learning can guide irrigation, fertiliser use and pest control, which may reduce water and chemical inputs. These are listed among promising applications in the literature [5], [6]. The review found no robust Indian national statistic that quantifies the water saved, so the benefit is treated here as potential.

E. Air and Water Pollution Monitoring

AI-based analysis of sensor networks and satellite images can detect pollution sources, estimate air quality where monitors are scarce, and flag water quality problems earlier. The literature identifies these uses and their value for environmental management [5], [6], although accuracy depends on the quality and coverage of ground-truth data.

F. Forest and Biodiversity Monitoring

Image and acoustic classification supports the detection of deforestation, illegal activity and changes in species populations at scales that human surveys cannot reach [5]. The footprint of the computing needed is typically modest compared with the monitoring benefit, although that comparison has seldom been quantified.

G. Natural-Disaster Preparedness

Forecasting models also support early warning for storms and floods. The weather model discussed above was evaluated on tropical cyclone tracking as well as general fields [21], and broader reviews list disaster response among the adaptation uses of machine learning [5]. Warnings are valuable only when they are acted upon, which depends on institutions and communications as much as on models.

Evidence status. Taken together, these applications show that benefits are credible, but most evidence is of potential rather than realised, net-of-footprint effect. The expert study that found 134 enabling interactions with the Sustainable Development Goal targets also found 59 inhibiting ones [6], and the IEA notes that AI's own emissions could be offset only if it enables broader reductions elsewhere, and that rebound effects and continued fossil fuel use could negate that outcome [1].

IX. CHALLENGES IN SUSTAINABLE ARTIFICIAL INTELLIGENCE

A. Growing Computational Demand

Demand is rising on every available measure. The IEA expects AI to be the most important driver of data centre growth to 2030 [1], and one projection placed AI server electricity use between 85 and 134 TWh annually by 2027 [19]. India's own plans, with capacity projected to rise about tenfold from the 960 MW baseline by 2030 and shared GPU capacity above 45,000 units, point the same way [8], [7].

B. Lack of Consistent Environmental Metrics

Studies use different system boundaries, grid factors and time frames, and the authors of the BLOOM analysis themselves stress the difficulty of precise estimation [15]. The OECD identified measurement gaps with direct policy consequences [4], and model documentation seldom includes water indicators [17]. Without common metrics, claims of efficiency improvement cannot be verified or compared.

C. Data-Centre Infrastructure Constraints

Data centres cluster geographically, and nearly half of United States capacity sits in five regional clusters [1]. The IEA estimates that around 20 percent of planned projects could be at risk of delay because of grid constraints and long connection queues [1]. India's projected rise to about 9.2 GW by 2030 implies that grid and power planning will need to treat data centres as a distinct category of demand [8]; this is the authors' inference from the projection.

D. Rebound Effects

Efficiency gains lower the cost of computing, and lower cost invites more use. Analyses of AI and climate warn that system-level effects can outweigh per-task savings [20], and the IEA likewise lists rebound effects among the factors that could negate AI's emissions benefits [1]. Efficiency therefore reduces the footprint only when total demand is also managed.

E. Hardware Supply-Chain Impacts

Rapid hardware turnover shifts impact upstream and downstream. Embodied emissions are incurred before a chip runs a single computation [11], and retired equipment enters a waste system that documents under a quarter of global e-waste as recycled [2]. Material recovery is limited by design and by collection systems.

F. Trade-Offs Between Accuracy, Cost and Sustainability

Larger models often score higher on benchmarks, but the gains can be marginal relative to the added computation, which is the reasoning behind proposals to report efficiency alongside accuracy [12], [13]. Distilled models offer a counterexample: one compact version of a widely used language model was reported to be 40 percent smaller and 60 percent faster while retaining 97 percent of the original's language-understanding performance [24]. The trade-off therefore depends on the task and should be assessed case by case.

X. Sustainable Solutions and Green Artificial Intelligence Framework

A. Model and Algorithm Measures

Efficient algorithms, compression and quantisation reduce the computation required per result. Pruning combined with quantisation and coding was shown to cut the storage of well-known image models by a factor of 35 to 49 without loss of accuracy on the tests reported [23], and knowledge distillation produces smaller models that keep most of the capability of larger ones [24]. Appropriate model selection matters as much: because multi-purpose generative models use markedly more energy per query than task-specific ones, routine classification or extraction tasks do not need a general model [10]. Training can also be scheduled by time and place to use cleaner electricity and less water [14], [17].

B. Infrastructure Measures

Operators can procure renewable electricity, improve cooling efficiency, reuse waste heat and site facilities with regard to local water availability. The IEA expects renewables to meet about half of the growth in data centre demand, though that depends on grid and storage capacity [1]. Because indirect water is far larger than direct water, power sourcing is also a water measure [16]. Hardware should be used for longer, reused where possible and routed to authorised recyclers at end of life, which in India falls under the extended producer responsibility provisions of the 2022 rules [18].

C. Governance and Transparency Measures

Environmental reporting should cover energy, emissions, water and hardware, with disclosed boundaries. The European delegated regulation on data centres is a concrete example of mandatory harmonised reporting [22], and the OECD sets out the case for measurement standards and expanded data collection [4]. Responsible AI governance should add a requirement to assess whether a proposed application is likely to produce a net environmental benefit before large resources are committed [20].

D. Proposed Framework

Fig. 6 integrates these measures into four layers. The layers are ordered by the level at which decisions are made, from governance at the top to application use at the bottom, and each constrains the one below it: reporting rules shape infrastructure choices, infrastructure shapes which models are practical, and model choice shapes how applications are designed. Table IV summarises the techniques, their expected advantage and their main caveats. The framework is conceptual; it is derived from the reviewed literature and has not been empirically validated.

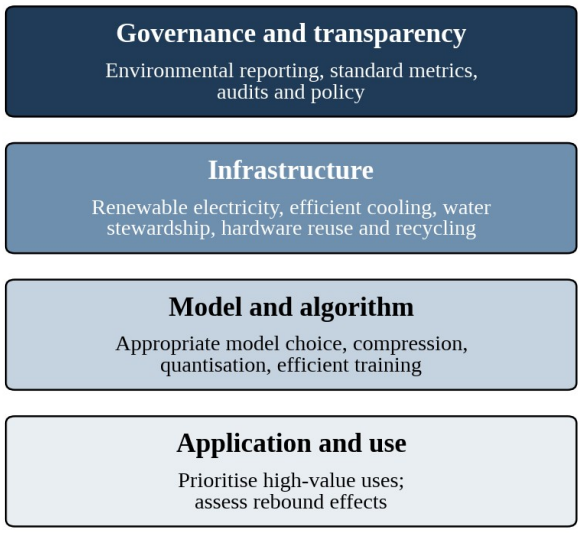


Fig. 6. Proposed layered Green AI framework (conceptual; developed by the authors from the reviewed literature).

TABLE IV. Sustainable Artificial Intelligence Techniques, Expected Advantages and Caveats

Technique	Expected advantage	Main caveat
Pruning, quantisation, compression [23]	Lower storage and computation at inference	Accuracy must be validated per task

Technique	Expected advantage	Main caveat
Distillation [24]	Smaller, faster model keeping most capability	A larger teacher model is still trained
Task-appropriate model choice [10]	Avoids multi-purpose overhead	May need several specialised models
Carbon- and water-aware scheduling [14], [17]	Lower emissions and water use by time and place	Latency and data-location limits
Efficiency reporting [12]	Comparison beyond accuracy	Needs shared standards
Renewable electricity [1]	Lower operational emissions	Grid, storage and additionality limits
Efficient cooling and water stewardship [16]	Lower direct water and energy use	Trade-offs between water and electricity
Hardware reuse and recycling [2]	Recovers materials; delays waste	Collection systems; data security

XI. Comparative Analysis and Discussion

Table V sets the documented costs of AI against the documented or potential benefits across five dimensions. The right-hand column states how strong the evidence is on each side, since the main discipline of this comparison is keeping established findings apart from interpretation.

TABLE V. Comparison of Artificial Intelligence Environmental Costs and Potential Benefits

Dimension	Cost evidence	Benefit evidence	Evidence status
Electricity	Data centres about 415 TWh in 2024; about 945 TWh projected for 2030 [1]	AI used to optimise energy supply and demand [1]	Costs measured and projected; benefits mainly scenario based
Carbon	BLOOM 24.7 to 50.5 t; GPT-3 about 552 t [15], [14]	Mitigation and adaptation uses [5]	Costs model specific; benefits largely potential
Water	US direct 66 billion L; indirect about 800 billion L [16]	Irrigation and water monitoring [5]	Costs estimated; benefits site specific
Materials and waste	62 Mt e-waste in 2022; 22.3% documented recycling [2]	Monitoring and resource-efficiency uses [5]	AI share of waste unknown
Development goal targets	59 targets possibly inhibited [6]	134 targets possibly enabled [6]	Expert judgement, not measured outcomes

Established evidence. Several findings rest on strong sources. Data centre electricity use is large and rising quickly [1], [16]. Indirect water use is an order of magnitude greater than direct use in the United States [16]. Electronic waste is rising much faster than documented recycling, and the global documented share fell slightly between 2010 and 2022 [2]. Accounting only for operations understates the footprint of at least one large model by about half [15]. At least one AI application, global weather prediction, has been shown to outperform an operational benchmark with low per-forecast computation [21].

Interpretation. The following points are the authors' reading of that evidence and are not themselves measured results. First, costs are better measured than benefits. The documented costs appear in meters and national statistics, while most benefits appear as potential in reviews. This asymmetry makes any claim that AI is on balance beneficial or harmful premature. Second, whether an application is net positive depends on counterfactuals: weather forecasting displaces more expensive computation, whereas a recommendation engine may increase consumption [20]. Third, India sits in an early phase where infrastructure is expanding fast while environmental reporting is thin; the 0.5 percent to 3 percent electricity path projected by the government [8] shows that the policy window for building reporting and clean-power procurement into growth is open now. Fourth, the divergence between the two government capacity baselines [8], [7] and between the old and new e-waste coverage [18] indicates that better statistical definitions are a prerequisite for sound policy.

Recommendations versus evidence. The measures in the previous section are supported by studies showing technical feasibility [23], [24], [10], but their system-wide effect on the footprint has not been demonstrated in a controlled national or global setting. They should be adopted as reasoned practice and monitored, not as guaranteed reductions.

XII. Recommendations

AI developers. Publish training and inference energy, emissions and water for released models, with boundaries stated; evaluate whether a smaller or task-specific model meets the requirement before choosing a larger one [10], [12]; use compression and distillation where tests show acceptable accuracy [23], [24].

Organisations that adopt AI. Include environmental criteria in procurement, ask providers for footprint data, and test whether an application plausibly yields a net environmental or operational benefit before scaling it [20].

Data-centre operators. Report energy, renewable share, cooling and water indicators in a harmonised form [22]; prefer power purchase and siting decisions that consider both carbon and water; extend equipment life and use authorised recyclers [2].

Researchers. Measure rather than assume: develop methods for embodied and indirect impacts, break down e-waste statistics so that data centre equipment is visible, and evaluate applications on a net basis [4], [11]. Table VI suggests indicators and measurement approaches.

Policymakers. Align official capacity and electricity statistics, introduce reporting requirements for large facilities, include data centre water in planning, and integrate AI infrastructure with grid and renewable-energy planning [8], [1], [3].

TABLE VI. Suggested Environmental Indicators and Measurement Methods

Indicator	Unit	Suggested measurement method
Electricity use	kWh per training run; kWh per 1,000 inferences	Metered at server or facility level; report power usage effectiveness (PUE)
Operational emissions	kg CO_2e per run	Electricity multiplied by location-based and market-based grid factors [15]
Embodied emissions	kg CO_2e per server	Life-cycle assessment using manufacturer data [11]

Indicator	Unit	Suggested measurement method
Water	Litres per kWh; litres per 1,000 inferences	On-site metering plus grid-based indirect factors [17], [16]
Electronic waste	kg per year; share reused or recycled	Asset registers and extended producer responsibility reporting [2], [18]
Net application benefit	Avoided emissions minus footprint minus rebound	Counterfactual baseline [4], [20]

XIII. Conclusion

The current review concludes that the environmental impacts of AI are indeed true, measurable in the sectoral context, and increasing. Electricity consumption of data centers is anticipated to double by 2030, water usage mainly comes from electricity production, embodied emissions can double the emissions from the model itself, and e-waste generation is happening at an extremely rapid rate compared to recycling. It also recognizes that tangible positive effects of AI are possible in fields like weather prediction, grid management, and environmental monitoring, but they are generally mentioned as a future prospect.

The implication here is the need for balance. The two concepts of technology advancement and environmental conservation are not conflicting, but the latter is not automatic in light of the former. This is dependent on efficient models, clean energy sources, effective water usage, durability of hardware, good accounting, and the question of whether the project is worth the energy consumption at hand. The contribution to knowledge made in this paper include the creation of a classified life cycle, the generation of clear and reliable indicators from both global and Indian official statistics, a comparison of cost and benefits in order to separate facts from interpretations, and a framework of four levels which can be used as a checklist.

XIV. Limitations and Future Research

The research relies upon secondary data, so it uses their definitions, boundaries and period. Some numbers are examples, not statistics; some numbers are taken from government press releases quoted in news agencies, and the original release needs to be seen for citing. Indicators are easy, and the assumption is that there is steady growth between the two numbers. The series of e-waste for India is not consistent from year to year, and none of the sources seen has separated AI devices. The review is only based on English language sources from October 2026.

Subsequent research should cover energy and water measurements of sample AI workloads within Indian data centers, localized life-cycle analyses incorporating embodied emissions, rebound effect studies using quantitative methods, and assessment of the net benefits of particular applications like irrigation advice and flood warnings. Also helpful would be a disaggregated accounting of electronic waste at the national level, distinguishing data center components.

References

[1] International Energy Agency, “Energy and AI,” IEA, Paris, France, Apr. 2025. [Online]. Available: https://www.iea.org/reports/energy-and-ai

[2] C. P. Baldé et al., The Global E-waste Monitor 2024. Geneva, Switzerland, and Bonn, Germany: International Telecommunication Union and United Nations Institute for Training and Research, 2024. [Online]. Available: https://www.itu.int/en/ITU-D/Environment/Pages/Publications/The-Global-E-waste-Monitor-2024.aspx

[3] United Nations Environment Programme, “Artificial intelligence (AI) end-to-end: The environmental impact of the full AI lifecycle needs to be comprehensively assessed,” Issue Note, Nairobi, Kenya, Sep. 2024. [Online]. Available: https://wedocs.unep.org/20.500.11822/46288

[4] OECD, “Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint,” OECD Digital Economy Papers, no. 341, OECD Publishing, Paris, France, 2022, doi: 10.1787/7babf571-en.

[5] D. Rolnick et al., “Tackling climate change with machine learning,” ACM Comput. Surv., vol. 55, no. 2, Art. no. 42, 2022, doi: 10.1145/3485128.

[6] R. Vinuesa et al., “The role of artificial intelligence in achieving the Sustainable Development Goals,” Nature Commun., vol. 11, Art. no. 233, 2020, doi: 10.1038/s41467-019-14108-y.

[7] “India’s data centre capacity grows over four-fold to 1,575 MW since 2020,” The Tribune (ANI), Aug. 14, 2026 (reporting a statement of the Ministry of Electronics and Information Technology, Government of India). [Online]. Available: https://www.tribuneindia.com/news/business/indias-data-centre-capacity-grows-over-four-fold-to-1575-mw-since-2020/

[8] Office of the Principal Scientific Adviser to the Government of India, “Democratising access to AI infrastructure,” White Paper Series, ver. 3.0, Dec. 29, 2025. [Online]. Available: https://psa.gov.in/CMS/web/sites/default/files/publication/WP_Democratising%20Access_V3.0_29122025A.pdf

[9] C.-J. Wu et al., “Sustainable AI: Environmental implications, challenges and opportunities,” in Proc. Conf. Mach. Learn. Syst. (MLSys), vol. 4, 2022, pp. 795–813.

[10] A. S. Luccioni, Y. Jernite, and E. Strubell, “Power hungry processing: Watts driving the cost of AI deployment?” in Proc. ACM Conf. Fairness, Accountability, and Transparency (FAccT), Rio de Janeiro, Brazil, 2024, pp. 85–99, doi: 10.1145/3630106.3658542.

[11] U. Gupta et al., “Chasing carbon: The elusive environmental footprint of computing,” in Proc. IEEE Int. Symp. High-Performance Comput. Archit. (HPCA), 2021, pp. 854–867, doi: 10.1109/HPCA51647.2021.00076.

[12] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, “Green AI,” Commun. ACM, vol. 63, no. 12, pp. 54–63, Dec. 2020, doi: 10.1145/3381831.

[13] E. Strubell, A. Ganesh, and A. McCallum, “Energy and policy considerations for deep learning in NLP,” in Proc. 57th Annu. Meeting Assoc. Comput. Linguistics, Florence, Italy, 2019, pp. 3645–3650, doi: 10.18653/v1/P19-1355.

[14] D. Patterson et al., “Carbon emissions and large neural network training,” arXiv:2104.10350, 2021 (preprint).

[15] A. S. Luccioni, S. Viguier, and A.-L. Ligozat, “Estimating the carbon footprint of BLOOM, a 176B parameter language model,” J. Mach. Learn. Res., vol. 24, no. 253, pp. 1–15, 2023.

[16] A. Shehabi et al., “2024 United States data center energy usage report,” Lawrence Berkeley National Laboratory, Berkeley, CA, USA, Rep. LBNL-2001637, Dec. 2024. [Online]. Available: https://eta.lbl.gov/publications/2024-lbnl-data-center-energy-usage-report

[17] P. Li, J. Yang, M. A. Islam, and S. Ren, “Making AI less ‘thirsty’: Uncovering and addressing the secret water footprint of AI models,” arXiv:2304.03271, rev. Mar. 2025, doi: 10.48550/arXiv.2304.03271 (preprint; accepted by Commun. ACM).

[18] Ministry of Environment, Forest and Climate Change, Government of India, “Effective disposal of e-waste,” Rajya Sabha Unstarred Question No. 1177, answered Aug. 1, 2024. [Online]. Available: https://rsdebate.nic.in/bitstream/123456789/748973/1/PQ_265_01082024_U1177_p244_p245.pdf

[19] A. de Vries, "The growing energy footprint of artificial intelligence," Joule, vol. 7, no. 10, pp. 2191–2194, 2023, doi: 10.1016/j.joule.2023.09.004.

[20] L. H. Kaack, P. L. Donti, E. Strubell, G. Kamiya, F. Creutzig, and D. Rolnick, "Aligning artificial intelligence with climate change mitigation," Nature Climate Change, vol. 12, pp. 518–527, 2022, doi: 10.1038/s41558-022-01377-7.

[21] R. Lam et al., "Learning skillful medium-range global weather forecasting," Science, vol. 382, no. 6677, pp. 1416–1421, 2023, doi: 10.1126/science.adi2336.

[22] European Commission, "Commission Delegated Regulation (EU) 2024/1364 of 14 March 2024 on the first phase of the establishment of a common Union rating scheme for data centres," Off. J. Eur. Union, 2024.

[23] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," in Proc. Int. Conf. Learn. Representations (ICLR), San Juan, Puerto Rico, 2016, arXiv:1510.00149 (preprint version).

[24] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv:1910.01108, 2019 (preprint).