Green AI: Systematic Review and Guidelines for Sustainability in Artificial Intelligence

Hareem Nisar[1*], Austin Tapp [1*], Marius George Linguraru [1,2]
[1]Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, Washington, DC, USA
[2]Departments of Radiology and Pediatrics, George Washington University, Washington, DC, USA

Abstract:

Artificial intelligence (AI) is now a planetary-scale socio-technical infrastructure whose energy demands are rising faster than our capacity to measure or mitigate them. Large-scale commercial applications like Claude and ChatGPT have made AI widely accessible, however, generative AI models are also avid electricity consumers and carbon emissions contributors. Naturally, the question of how we might promote energy-friendly 'green' AI arises. In this systematic review, we synthesize the last seven years (2017–2024) of peer-reviewed work describing how AI's energy use and emissions are measured, how substantial these emissions are across the AI lifecycle, and what techniques can reliably reduce AI-related carbon emissions. We identified five distinct themes that emerge in the context of green AI and have the potential to reduce energy consumption. The observations and analysis of our review suggest a lack of standardization in measuring and reporting energy cost and carbon emissions associated with the AI lifecycle. To address this dearth in reporting standards, we propose eight review-driven, actionable guidelines for researchers, industry, and policymakers to promote environmentally sustainable and green AI as a proactive property of the AI lifecycle.

1. Introduction

Artificial Intelligence (AI) has rapidly evolved from an academic curiosity into everyday technology. Since the creation of Eliza–the world's first chatbot–in 1966 [1], AI technology has grown into complex, sophisticated applications such as the humanoid robot Sophia [2]. In 2017, Google's transformer architecture [3] empowered researchers to create large language models (LLMs) such as BERT [4], leading to the development of intelligent, conversational, and creative tools like ChatGPT [5] and Claude [6]. LLMs have enabled both public access and acclaim of AI. The widespread adoption of AI and constant striving for 'bigger' AI comes with a critical question: what is the environmental impact of AI systems?

The energy requirements for advanced AI models have increased dramatically, raising concerns on the environmental sustainability and cost of AI [7]. While the 2019 BERT model training

*Equal Contribution

**submitted to Nature Machine Intelligence in Dec 2025

consumed around 1.5 MWh of energy [8], the 2023 GPT-4 model was estimated to consume ~3,500 MWh for training [9] - the energy equivalent of powering 331 American homes for a year [10]. In addition to GPT-4, there are over 30 public AI models of the same computation scale [11]. Energy consumption yields a direct impact on the environment in the forms of carbon dioxide (CO_2) and other greenhouse gas emissions. Training of the GPT-3 [12] or Llama 2 [13] models added ~500 tonnes CO_2 to the environment, which is the carbon-equivalent of 294 one-way flights between New York and Tokyo or driving around the world 51 times [14].

Awareness of AI's carbon footprint has grown recently as researchers have designed energy-efficient models. In 2019, Schwartz et al. coined the term 'green AI', defining it as "AI research that is more environmentally friendly and inclusive" [15]. The field of green AI is still in its inception and struggles with metric standardization, fragmented terminology, opaque reporting, and inconsistent study design [16]. The most comprehensive review of green AI by Verdecchia et al. was presented in 2022 [17], which predates the rise of LLMs and generative AI. More recent works on green AI [18, 19] have discussed selective optimization techniques and hardware alternatives, but lack systematic methodology, transparent literature selection, and detailed quantitative synthesis. Therefore, an updated systematic review is required to capture recent advances in green AI and understand environmental sustainability as an emergent property of AI systems.

Environmental sustainability is more than model size or hardware efficiency; it is an emergent property of how AI is conceived, trained, deployed, and used. Every component--from dataset creation and model optimization to the carbon intensity of the regional grid and the behavior of end-users--contributes to AI's environmental footprint. In parallel, a growing body of policy and regulatory efforts has begun to formalize the socio-technical dimensions of AI governance. Notably, the 2024 EU AI Act established the first legal framework for ethical and transparent AI [20], the FUTURE-AI framework proposed principles for trustworthy AI in healthcare [21], the Federal Institute of Technology Zurich proposed ethical principles of fair, non-maleficent, and privacy-aware AI [22], and Gyevnar and Kasirzadeh proposed AI safety guidelines [23]. Yet, environmental sustainability remains largely peripheral, if not absent, within these emerging frameworks. Embedding sustainability as a factor within governance structures is important in mitigating the carbon impact of AI.

Environmental sustainability is relevant in every stage of the AI lifecycle, where distinct computation and energy costs collectively determine the environmental impact of the AI model (Figure 1). The lifecycle of AI typically progresses through four key stages: data management, training, evaluation, and inference. Data management involves curating, pre-processing, and selecting data for training. The training stage of AI is often computationally intensive and demands high energy costs. Evaluation is less structured and entails rigorous testing and validation of the model to ensure safe, generalizable use and performance expectations. The inference stage

constitutes usage of AI for real-world tasks and generally demands the least energy per instance. However, the popular use of AI-assisted web search or LLMs arguably constitutes the biggest energy consumption in the AI lifecycle. For example, querying ChatGPT for one hour consumes the energy equivalent of streaming a video for 73 years [24]. The annual inference energy of ChatGPT (estimated at 1,058.5 GWh) is sufficient to fulfill the annual energy demands of a small country like Barbados [24] or territory like the US Virgin Islands [25]. Data centers supporting this infrastructure already consume ~1.5 % of global electricity and may exceed 3% (i.e., 945 TWh) by 2030. Such growth calls for a shift from ad hoc efficiency measures to carbon-aware AI systems throughout their lifecycle [26].

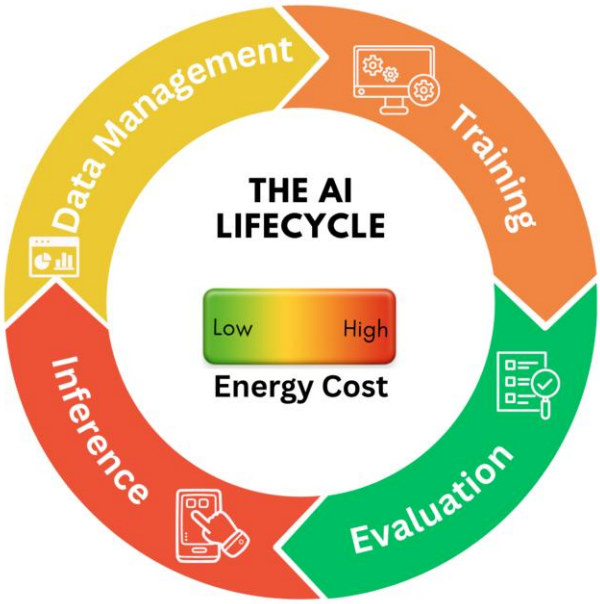


Fig. 1 | Energy Consumption Across the AI Lifecycle. The lifecycle includes four key stages: (1) Data Management, (2) Training, (3) Evaluation, and (4) Inference, each with distinct energy demands. The inference stage can exceed the energy cost of training over time, especially for large-scale models like ChatGPT.

To address the accelerating pace of AI progression and its escalating energy demands, we present a review of literature supporting the environmental sustainability of AI. Our systematic review reflects the post-LLM landscape of approaches for green AI at a software level to advance both scholarly and practical understanding of green AI. Specifically, the review centers on three points of inquiry: how AI's environmental impact is measured during its lifecycle, what is the energy usage for different AI methods, and what approaches can efficiently reduce carbon emissions in the AI lifecycle. Moreover, this review reports sustainable AI metrics and approaches from literature, identifies reporting gaps, and offers practical guidelines for embedding environmental sustainability in AI design and governance.

2. Systematic Review Methodology

We conducted a systematic review of peer-reviewed literature on green AI. All extracted data is provided in the Supplementary file. The review was conducted based on the rigorous workflow by Kitchenham [27]. We complemented the literature search with an additional systematic mapping of selected studies using the snowball method by Wohlin [28].

In this review, we aim to answer three core research questions (RQ):

- RQ1: How can the environmental impact of AI be measured during its lifecycle?
- RQ2: What is the amount of energy consumed by various categories of AI methods?
- RQ3: What methods have been established to reduce AI carbon emissions without compromising model performance?

Search and Review Process:

Relevant publications between January 2017 and December 2024 were initially retrieved from (1) the Association for Computing Machinery (ACM) Digital Library, (2) PubMed, and (3) Web of Science. Given the novelty of green AI, we also searched (4) arXiv to discover peer-reviewed studies accepted for publication but not yet integrated into formal databases.

We designed three search strings incorporating emerging topics from the fields of 'AI' and 'green':

s1. *On AI: {*(artificial intelligence) OR (AI) OR (machine learning) OR (ML) OR (deep learning) OR (DL)}

s2. *On large and trending AI:* {(foundation) OR (generative) OR (genAI) OR (large language) OR (LLM) OR (federated) OR (self supervised) OR (self-supervised) OR (supervised) OR (pretrain) OR (agentic)}

s3. *On green:* {(carbon emission) OR (energy consumption) OR (computation cost) OR (sustainable) OR (green)}

The queries used for databases were q1=[s1 AND s3] and q2=[s2 AND s3] in the title field. The first query aimed to identify research on green AI architectures for traditional machine learning (ML) methods. The second query sought research on more niche and recent AI methods, such as LLMs and federated learning.

Eligibility Criteria and Selection:

The criteria for selecting papers to address our research questions were: (1) must include original empirical, analytical, or simulated experiments addressing the environmental impact and energy consumption of AI methods; (2) must consider software-level strategies; and (3) must be peer-

reviewed. We excluded published abstracts, book chapters, literature reviews, gray literature, social or policy papers, and articles not written in English.

Our search produced 3,179 results from which we identified 64 relevant studies based on the review process [27]. The snowball process yielded 44 additional results, 11 of which met the inclusion criteria. Thus, 75 studies were included and analyzed. The details of the systematic review are reported following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (see Supplementary).

Annotation and Data Extraction:

We employed the online literature review PICO portal [29] to organize the systematic search and divide tasks among multiple reviewers. We extracted information such as the type of study, AI lifecycle stage, type of AI, model learning paradigms, computing infrastructure, choice of hardware and software, model performance metrics, energy consumption tools, metrics reported, and calculated carbon impact. The annotation process also involved extracting the author-defined keywords and determining the primary task/objective of each paper.

3. Empirical Findings

Keyword Co-occurrence:

To track the conceptual evolution of green AI, we performed a co-occurrence analysis of author-defined keywords of the papers in review. Similar words and abbreviations were homogenized before the analysis. We employed the bibliometric analysis webtool VOSViewer [30] to conduct the co-occurrence analysis (Figure 2) showing keywords and index terms found in publications. We observed the frequent keywords to be ‘energy efficiency’, ‘machine learning’, ‘sustainability’ and ‘green AI’. There was also emphasis on a specialized AI paradigm, i.e., ‘federated learning’, while no notable occurrences on generative AI or LLMs were found. The distribution of keywords over time indicated that the terms ‘green AI’ and ‘sustainability’ are more recent and gaining traction.

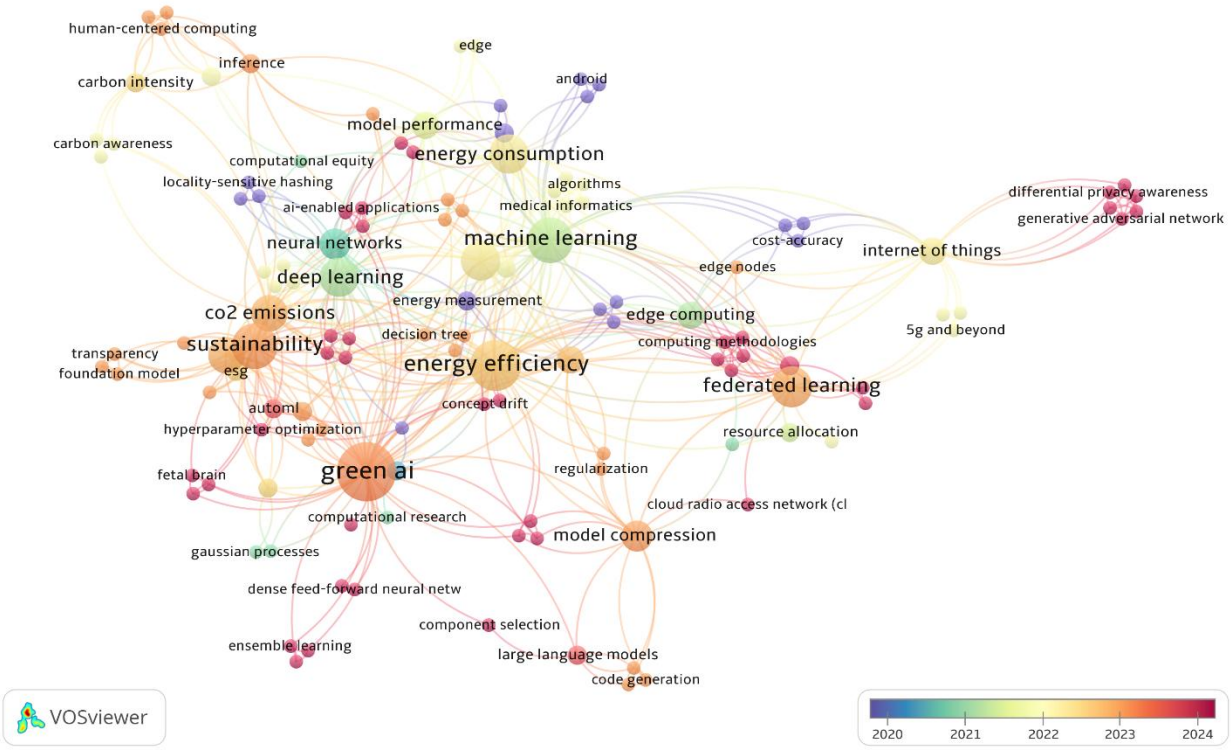


Fig. 2 | Keyword co-occurrence network in green AI literature. Graph nodes are weighted in size by the frequency of occurrence of a keyword. The color-coded legend represents the average year of publication for each keyword. Blue nodes represent keywords most common in earlier publications, while red nodes represent keywords from more recent publications. Only four studies were published between 2017 and 2020 and their keywords are not included for better visualization of color distribution over time.

Citation Analysis:

We performed a high-level bibliometric analysis on co-authorship and citations. No significant connections were observed in the co-authorship network map generated via VOSviewer [30], likely due to the scarcity of work done in green AI. The citation analysis was conducted by Semantic Scholar Academic Graph (S2AG) [31] querying to retrieve paper metadata, and to identify the most influential publications with the highest impact [32]. S2AG calculates 'highly influential citations' using a machine learning classification method based on the location, frequency, context, and structure around the citation [32]. Table 1 shows the summary of the most influential publications in green AI.

Table 1 | Summary of influential works impacting green AI. From the 75 reviewed publications, this table highlights those with an influential citation count greater than 10, along with their overall citations, influential citations, and a brief summary. Publications are listed in descending order of the influential citation count.

Paper	Citations (Overall / Influential)	Brief Summary
Yang et al. 2021 [33]	696 / 76	Energy-efficient transmission and computation resource allocation for federated learning via an iterative optimization algorithm.
Lannelongue et al. 2021 [34]	331 / 30	Generalizable framework and tool for estimating the carbon footprint of any computation task and practical solutions for greener computation.
Strubell et al. 2020 [35]	521 / 20	Energy cost analysis of popular natural language processing models and energy saving recommendations.
Justus et al. 2018 [36]	231 / 16	Execution time prediction for different components of a neural network to facilitate optimal hardware selection and efficient model design.
Yang et al. 2017 [37]	163 / 13	Energy estimation technique for deep neural networks to guide energy-efficient design strategies before training begins.
Dodge et al. 2022 [38]	200 / 11	Framework for measuring the carbon impact of a system, especially considering the geographic location and time of the day.

Environmental Metrics and Tools:

To address RQ1, we studied the 75 publications for their quantitative evaluation of the environmental impact of AI. Several metrics were used (see terminology in Table 2). Energy consumption (EC) is the most direct measure of the power needs of a model. This includes energy consumed by processors, GPUs, and memory. However, the environmental impact of a system is measured by its carbon emissions (CE). In literature, the terms carbon emissions and carbon footprint (CF) are often used interchangeably. CE not only depends on the amount of energy consumed but also on the origin of the energy source as every source has a different emission factor γ (unit: gCO_2/kWh) that dictates the impact of the consumption on the environment. For example, an AI model trained with energy generated from fossil fuels will yield much higher CE than an AI model trained with energy generated by wind turbines. Another factor influencing CE is the Power Usage Effectiveness (PUE) of the system. PUE refers to a unitless measure of how efficiently a system or data center uses energy. An ideal system will have a PUE of 1, however, in practice, systems have an average PUE of 1.56 [39]. Thus, the carbon emissions of a system can be measured as $CE = \gamma * PUE * (\text{Total Energy Consumed})$ [39].

Table 2 | Key metrics for measuring AI's environmental impact with their standard units and definitions.

Energy Consumption (EC): Total amount of energy utilized by a system or process over time. Unit: Joules (J) or kilowatt-hours (kWh).
Power Consumption (PC): The rate at which energy is consumed at a given instant. Unit: Watt (W).
Carbon Emissions (CE): The amount of CO_2 (or equivalent of other greenhouse gas) released into air. Unit: grams of CO_2 equivalent (gCO_2e).
Carbon Footprint (CF): The overall carbon emissions produced by a product or company. It is often used interchangeably with CE. Unit: gCO_2e.
Carbon Intensity (CI): The amount of CO_2 or equivalent gases emitted per unit of activity, like electricity. Example unit: gCO_2e/kWh.

The metrics included in green AI literature included direct measures like EC, PC, CE, and carbon intensity (CI), as well as indirect measures such as training time, number of computations, etc. Only 62/75 (82.67%) of papers reported energy related metrics, with 23/75 (30.67%) of papers reporting carbon impact as CE or CI, and 16 (21.33%) reporting both CE and CI (Figure 3). Furthermore, only 7/75 (9.33%) of papers explicitly incorporated PUE as a metric, even though PUE is crucial to estimating the true carbon impact of a system.

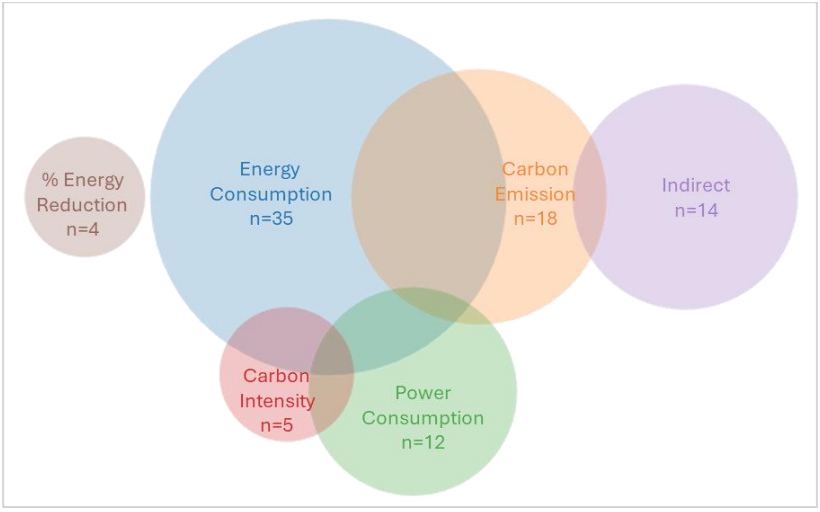


Fig. 3 | Venn diagram illustrating the distribution and overlap of environmental metrics used across reviewed studies. The size of circles and overlapping regions is representative of the number of papers utilizing the metric. Overlap highlights the multi-metric approaches used to evaluate AI's environmental impact.

The reviewed literature mentioned 27 unique tools to quantify the energy consumption of AI through its lifecycle (see Supplementary for citations using specific tools). However, only 29/75 (38.67%) of papers explicitly mentioned tool usage. The most discussed monitoring tool in this review was CodeCarbon (n=7) [40], followed by Experiment Impact Tracker (n=2) [41], Green Algorithms (n=2) [34], and MLCO$_2$ Impact (n=2) [42]. The most discussed performance profiling commands were RAPL (n=4) (Intel, Santa Clara, CA, USA) and NVIDIA-SMI (n=3) (NVIDIA Santa Clara, CA, USA) used to monitor Intel CPU and NVIDIA GPU power usage, respectively.

Geographical Distribution:

We extracted co-author affiliation countries from each paper's metadata and counted affiliations per country. We identified 376 co-author affiliations in 29 countries, with green AI research dominated by the United States (n=99), followed by China (n=39), Spain (n=37), United Kingdom (n=21), and Netherlands (n=21). Representation from the Global South (including Sub-Saharan Africa, South Asia, and Latin America) was sparse (n=13; South Asia 3 and Latin America 10). We rendered a choropleth map (Figure 4) representing the density of co-author affiliations across the globe.

Understanding where green AI research is conducted is important, as geographical context strongly influences the carbon intensity of computation. Regional factors such as climate, energy mix, and temporal variation in grid demand substantially affect the carbon emissions of identical workloads. For example, Dodge et al. [38] demonstrated that EC for equivalent computation tasks can vary by several orders of magnitude, from as little as 0.02 kWh in Norway and France to nearly ~13,000 kWh in the United States and Australia. Similarly, Shaikh et al. [43] reported carbon emissions associated with the same energy use may differ by sevenfold across U.S. regions due to variations in electricity grid intensity. These disparities highlight the need for guidelines that integrate geographic and infrastructural variability into the assessment and governance of green AI.

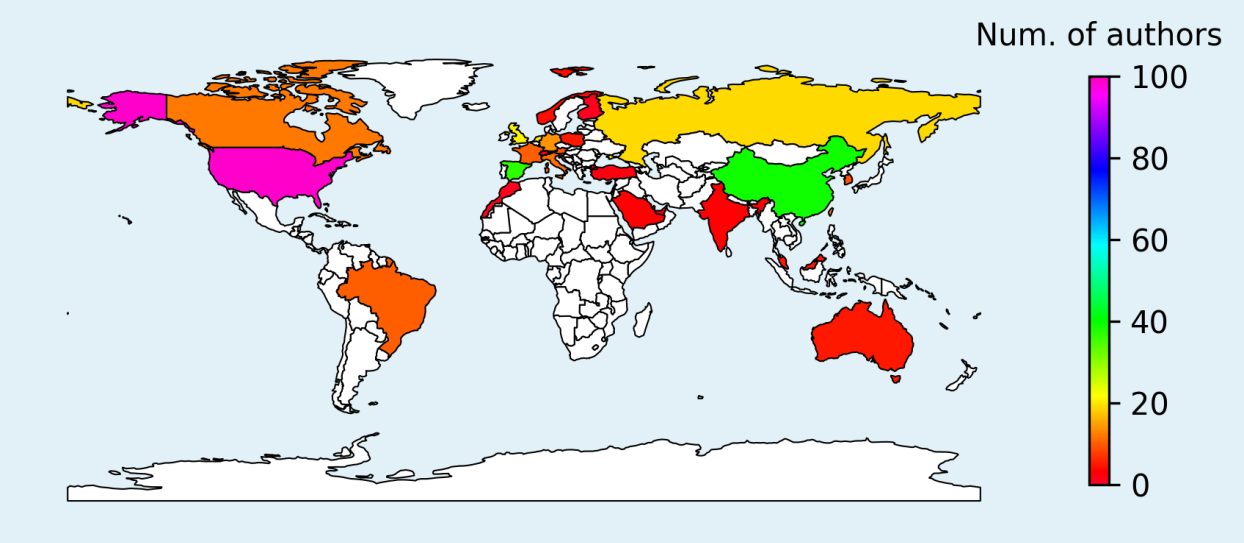


Fig. 4 | Geographic distribution of co-authors contributing to green AI research. The choropleth map shows dispersed but limited geographical distribution. United States has the highest density of 99 co-authors, while the remaining 28 countries each have fewer than 40 co-author affiliations.

Theme Identification and Analysis:

We identified five principal themes in literature by grouping the primary tasks of the reviewed studies:

T1. Energy Consumption Tracking Tools
T2. Carbon Evaluation of Existing AI
T3. Green Frameworks in Federated Learning
T4. Novel Techniques for Green AI
T5. New Carbon-aware AI Frameworks

The themes relate well to our research questions with T1 and T2 addressing research questions RQ1 and RQ2, respectively, while T3, T4, and T5 collectively address RQ3 by detailing novel techniques and frameworks to minimize the carbon impact of AI. We report empirical findings, trends, limitations, and energy improvements from each theme below.

Table 3 | Summary of five principal themes in green AI. Commonly occurring features are noted for each category of extracted data. Features of the highest frequency are in bold format.

Paper Count	Years	Frequently Occurring				
		Tools	**AI Model**	**AI Lifecycle Stage**	**Computing Infrastructure**	**Performance Metric**
Theme 1 – Energy Consumption Tracking Tools						
13	2017 – 2024	Green Algorithms [34] CodeCarbon [40]	**CNN** LLM	**Training** Inference	Local Cloud HPC	**Accuracy** RMSE
Theme 2 – Evaluates Environmental Impact of AI methods						
23	2019 – 2024	**CodeCarbon** [40] **NVIDIA-SMI** (NVIDIA) perf tool (LINUX)	**CNN** ML LLM MLP	**Training** Inference Optimization Data-Management	Local	**Accuracy** F1-score
Theme 3 – Green Frameworks in Federated Learning						
13	2021 – 2024	Mathematical models	CNN	**Training** Optimization Data-Management	Distributed	**Accuracy** SSIM
Theme 4 – Novel Techniques for Green AI						
17	2017 – 2024	$MLCO_2$ Impact [42] Torch profiler (LINUX)	**CNN** LLM	**Training** Inference Data-Management	Local	**Accuracy** RMSE
Theme 5 – New Carbon-aware AI frameworks						
9	2020 – 2024	CodeCarbon [40] Carbon Tracker [44] Intel Power Gadget (INTEL)	**CNN** ML	**Training** Inference Optimization	Local Cloud HPC Distributed	Accuracy Latency BLEU score

CNN - convolutional neural network; LLM - large language model; HPC - high-performance computing; ML -machine learning; MLP - multilayer perceptron; SSIM - structural similarity index measure; RMSE - root mean square error; BLEU - bilingual evaluation understudy.

T1. Energy Consumption Tracking Tools

The papers in this theme present tracking tools designed to estimate and track energy consumed during various AI lifecycle stages, particularly during training and inference. Diverse approaches towards green AI include optimized task scheduling on low carbon intensity days [45], visual comparisons of various ML configurations [44,46], and considering geographical location

[38,34,44,45] in calculating carbon impact of an AI system. Table 4 presents a complete list of methods and tools in this theme. These tools were implemented equally across local setups, on the cloud, and across high-performance computing (HPC) and data centers. Importantly, these tools were developed for single-site AI and underscore the need for scalable carbon tracking frameworks that extend to distributed and federated settings.

Table 4 | Summary of publications proposing standalone tools for supporting green AI. Tools can be categorized as (1) Estimation tools that can be used to gauge carbon footprint prior to training, (2) Monitoring tools that can track energy consumption in real-time, and (3) Governance tools for green AI awareness. DL-based tools indicate methods for measuring energy consumption or carbon footprint of AI systems and are designed using a deep learning (DL) model.

Tool	Feature	Tool Description
Pre-training estimation tools		
Execution-Time Predictor [36]	DL-based tool	Prediction of execution time for neural network training and inference.
GreenRunner [47]	DL-based tool Open source	Preselection of energy-efficient ML configurations in application-specific contexts.
TO model [48]		Estimation of carbon footprint based on model configurations via nonlinear computational modeling.
Green Algorithms Calculator [34]		Estimation of carbon emissions based on data center efficiency and geographic location.
eco4Cast [45]	Open source	Prediction of daily carbon intensity in geographical regions and energy-efficient scheduling of ML tasks.
Runtime monitoring tools		
eCO_2AI [49]	Open source	Estimation of carbon emissions during model training.
Carbon Framework [38]		Measurement of carbon footprint, considering location and time of the day.
Compressed Framework [50]	DL-based tool	Evaluation of energy consumption, especially in cloud data centers.
Benchmark-Tracker [44]		Benchmarking and evaluation of AI system's speed, performance, energy use, and carbon footprint.
EnergyVis [46]	Open source	Tracking, visualization, and comparison of carbon emissions based on hardware or geographical location.
MIEV [37]		Tracking, visualization and comparison of energy-performance tradeoffs in ML models
Online Resource Estimator [37]	Open source	Estimation of energy consumption of deep neural networks considering both computation energy and data movement energy.
Governance		
Explainable Sustainability [52]		Embedding sustainability feedback within AI systems.

T2. Carbon Evaluation of Existing AI

Theme 2 has the greatest breadth by evaluating multiple AI lifecycle stages and concluding that large AI models incur high carbon emissions [35]. In addition to reporting energy consumption, this theme reveals several practical solutions and considerations for AI engineers when designing energy-efficient systems:

- Use simpler model architecture when possible [53].
- Employ CPUs for non-imaging inference tasks [54].
- Add power caps to GPUs (**10–24% energy efficiency**) [55].
- Pair lightweight models with local devices and heavier models with cloud GPUs [56].
- Perform hyperparameter tuning (**50–160% energy efficiency**) [57].
- Reduce data and choose appropriate hardware (**80% computational efficiency**) [58].
- Deliberately select optimized techniques tailored to the model such as caching for data loading (**400% faster training**), choosing adaptive optimizer like Novograd, and using lightweight architecture. [59].
- Consider Racetrack Memory device instead of GPUs for energy-efficient inference [60].
- Consider ONNX Runtime for energy-efficient deployment of edge AI models [61].
- Use lower-precision computation e.g. float16 instead of float64 (~**24% energy efficiency**) [62].
- Employ energy-aware hyperparameter optimization techniques e.g. Bayesian optimization HyperBand (BOHB), HyperBand, population-based training (PBT), and the asynchronous successive halving algorithm (ASHA) (**~29% energy efficiency**) [63].
- Reduce size of training data (**50% energy efficiency**) [64].
- Reduce network complexity by reducing the number of convolutional layers [65].

T3. Green Frameworks in Federated Learning

Within this theme, studies highlighted efforts to reduce both data transmission and computation energy in federated learning systems. Studies incorporated edge AI, quantized neural networks, device scheduling, and resource allocation to lower energy costs. Table 5 summarizes the quantitative improvements achieved through novel federated learning methods. Studies in this theme only reported total energy usage but omitted CF estimates. This gap likely stems from inconsistent and unclear emission factors across clients, making carbon impact calculations difficult. Altogether, these studies demonstrate promising advances in energy-aware federated learning but expose a gap in green AI, i.e., without standardized carbon measurement across distributed nodes, the true environmental footprint of federated systems remains uncertain and lacks metrics for governance.

Table 5 | Summary of publications proposing green frameworks in federated learning with reported quantitative improvement metrics. Maximum percentage improvement derived from either direct or indirect measures reported in publications, along with concise descriptions of the corresponding methods employed to achieve these improvements are summarized. Publications are organized in descending order of maximum reported improvement within each measurement category.

Paper	Maximum reported improvement	Method in Federated Learning
Using direct measures (energy consumption, power consumption)		
Kim et al. 2022 [66]	70%	Quantized neural networks to optimize convergence rate and energy cost.
Hsu et al. 2022 [67]	69%	Two-stage iterative optimization for optimal data sampling and communication energy resources.
Li et al. 2021 [68]	57%	Fine-grained gradient compression technique that adapts sparsity/precision per network layer.
Wang et al. 2023 [69]	50%	Quantized models combined with optimized joint fronthaul-rate and power allocation.
Driss et al. 2024 [70]	43%	Grey-Wolf-Optimizer–based multi-attribute client selection.
Albaseer et al. 2024 [71]	28%	Fairness-aware client scheduling that weights devices by historical participation.
Kuswiradyo et al. 2024 [72]	27%	Energy-aware omnidirectional workload offloading across fog, edge, and cloud computing.
Using indirect measures (time, communication overhead)		
Fontenla-Romero et al. 2023 [73]	87%	One-shot federation for a network for efficient single round communication.
Qi et al. 2024 [74]	80%	One-shot federated learning with generative learning to achieve outcomes in a single communication round.

T4. Novel Techniques for Green AI

Like T3, this theme also answers RQ3 by proposing new algorithms that result in reduced carbon impact of AI technology. Table 6 presents a summary of energy-saving techniques and their quantitative impact. T4 demonstrates the most inconsistent tool usage and lack of standardization across all themes, as energy costs were often reported as percentage improvements [75,76] without details on how they were calculated. Other approaches relied on indirect computation indicators like FLOPs and training time [77-83], which offer limited insight into the true environmental impact of the system. Collectively, these studies illustrate rapid innovation toward green AI, yet the absence of standardized measurement protocols limits direct comparison in energy efficiency and hinders the translation of greener AI methods into verifiable environmental gains.

Table 6 | Summary of publications proposing novel techniques in green AI with reported quantitative improvement metrics. Maximum percentage improvement derived from either direct or indirect measures reported in publications, along with concise descriptions of the corresponding methods employed to achieve these improvements are summarized. Publications are organized in descending order of maximum reported improvement within each measurement category.

Paper	Maximum reported improvement	Method
Using direct measures (energy consumption, carbon emissions)		
Shi et al. 2024 [84]	99%	Compression and energy-focused usage of LLMs.
Pirnat et al. 2022 [85]	94%	Compact feature and model design for low-power applications
Liu et al. 2024 [86]	87.6%	Green-aware benchmarking framework using only-encode-once (OLEO) training.
Xiangyu et al. 2020 [87]	67%	Optimized task selection and structured sparsity for efficient edge inference.
Yin et al. 2024 [83]	62%	Dynamic coreset selection to train on a compact subset of data.
Nijkamp et al. 2024 [81]	57%	Optimal ensemble selection to minimize computing costs.
Reguero et al. 2024 [88]	56%	Reduce training compute using progressive layer freezing, model quantization, and early stopping.
Wei et al. 2023 [89]	55%	Post-training quantization method for LLMs.
Lazzaro et al. 2023 [76]	31%	Novel training loss function guided by measured energy to optimize consumption.
Gille et al. 2022 [75]	25%	Structured sparsity-regularized convolutional autoencoders for low-energy image compression.
Using indirect measures (compute cost, FLOPs)		
Huang et al. 2023 [79]	64%	Adaptive backpropagation in LLM finetuning to reduce computing costs.
Spring et al. 2017 [82]	95%	Randomized hashing technique to reduce memory and computing during training.
Using indirect measures (time, model size)		
Balderas et al. 2024 [80]	20%	Lightweight deep learning architecture with efficient training for time-series prediction.
Acmali et al. 2024 [77]	50%	Model pruning technique for reducing the number of parameters.
Yang et al. 2023 [90]	60%	Dynamic data subset selection during training to reduce redundant samples.
Candelieri et al. 2021 [78]	66%	Augmented Gaussian processes for efficient hyperparameter optimization.
Bird et al. 2020 [91]	94%	Binary-activated neural networks for generative models to limit computation.

T5. New Carbon-aware AI frameworks

Unlike other themes, literature on new carbon-aware frameworks focused on both single-site and distributed learning paradigms. These frameworks highlighted the trade-offs between model performance and energy cost across the AI lifecycle [86-88]. The emergence of such frameworks suggests the need for standardized policies and practices for embedding environmental sustainability into AI systems. Some of these carbon-aware frameworks are:

- CEMAI, a carbon-aware machine learning pipeline that measures and reduces emissions across the entire ML lifecycle using CodeCarbon and smart caching [92].
- Clover, a carbon-aware inference system that dynamically selects models, allocates GPU resources, and schedules inference based on grid carbon intensity (**80% reduced carbon emissions**) [93].
- FREEDOM, a privacy-aware, energy-efficient resource scheduling framework to optimize batch sizes and privacy budgets in federated learning environments using Deep Q-learning and GANs (**10.9-11.5% energy efficiency**) [94].

Extracted Data Analysis:

Green AI is a rapidly evolving domain, reflecting a growing awareness of the AI's environmental impact. Of the selected papers, 66/75 (88%) were published in or after 2021, marking a decisive acceleration in research activity following the public release of Open AI's GPT-3 in June 2020. The temporal evolution of publications (Figure 5) reveals a clear thematic transition: early work (pre-2019) concentrated on optimizing conventional machine learning and convolutional neural network (CNN) architectures, while research between 2020 and 2022 shifted toward transformer-based models and LLMs. Moreover, since 2022, tool-centric publications have emerged, advocating for energy measurement practices in AI systems. This trajectory reflects a maturing field, moving from conceptual discussions to data-driven approaches for AI sustainability assessment.

We further synthesized and illustrated the distribution of extracted data features (Figure 6), the relationship between learning paradigms, lifecycle stages, and model architectures (Figure 7), and the potential of energy-saving strategies across the AI lifecycle (Figure 8). The reviewed literature prioritized reducing training costs, with only 18/75 (24%) addressing inference-stage efficiency. The Sankey diagram (Figure 8) reveals that the reviewed literature on green AI is primarily dominated by centralized, single-site learning pipelines focused on training and inference of CNN-based models. While federated learning is present, it is prevalent in training stage only. LLMs are trending but generative or other complex vision models appear only in niche cases. We also identified that improvements in energy are not confined to any single AI lifecycle stage. This suggests that meaningful energy savings can be pursued throughout the lifecycle of an AI system.

It is noteworthy that the evaluation phase of the AI lifecycle was absent from all studies included in this review (Figures 6, 7, and 8).

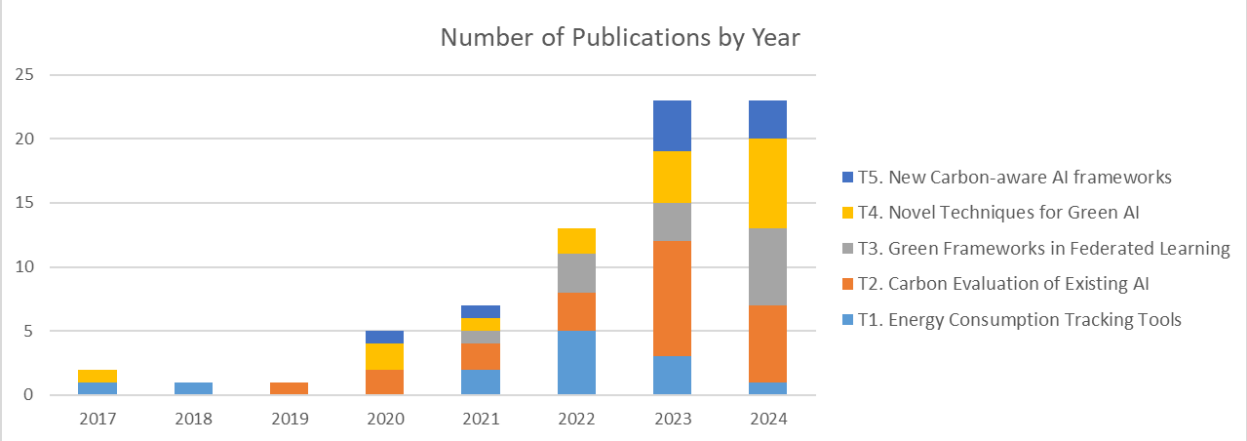


Fig. 5 | Yearly trends in green AI publications (2017 – 2024). Distribution of publication volume and dominant thematic focus across years.

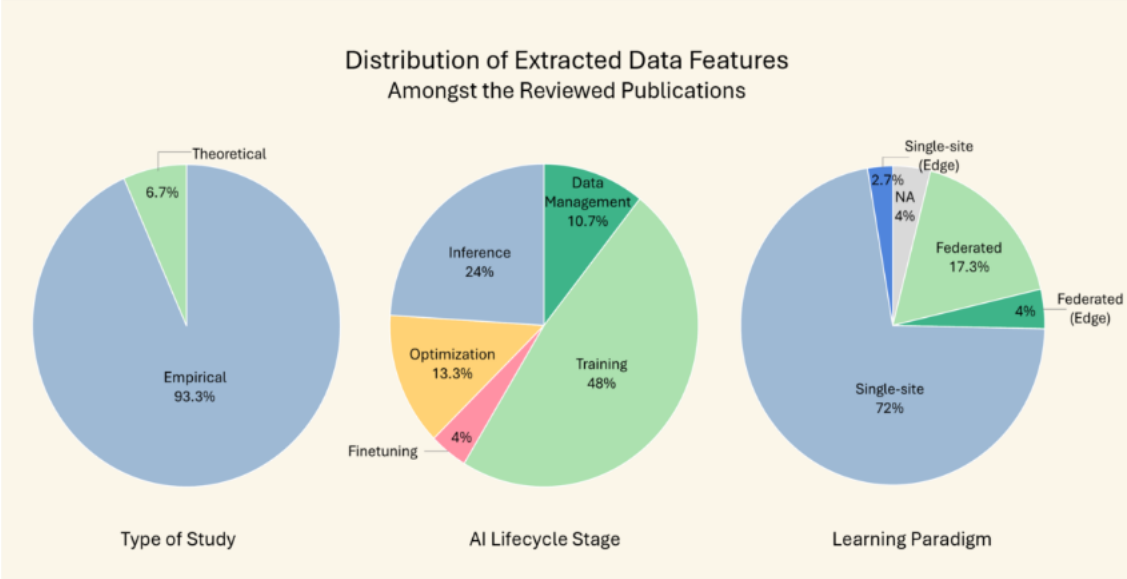


Fig. 6 | Distribution of extracted data features across reviewed studies. Visualization of quantitative distribution of reviewed literature across study types, AI lifecycle stages, learning paradigms.

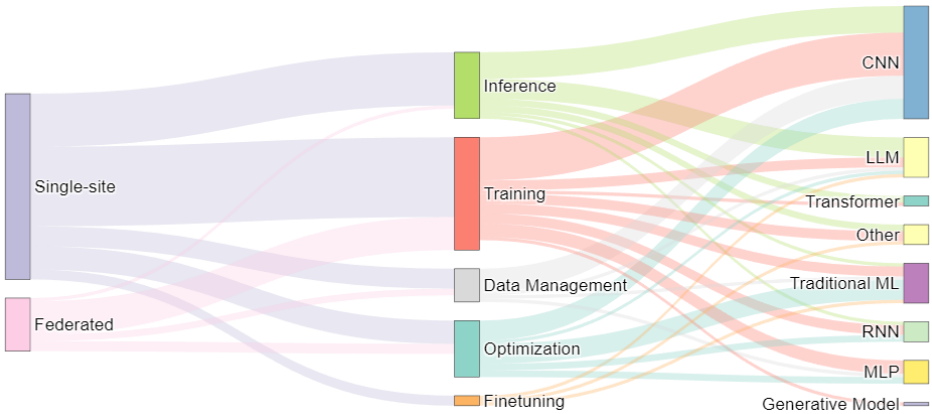


Fig. 7 | Sankey diagram of AI attributes in the literature. Visualization of quantitative flow between AI learning paradigms, lifecycle stages, and model architectures in reviewed research. The width of each stream represents the relative prevalence of that combination in the reviewed studies. CNN - convolutional neural network; LLM - large language model; RNN - recurrent neural network; ML - machine learning; MLP - multilayer perceptron.

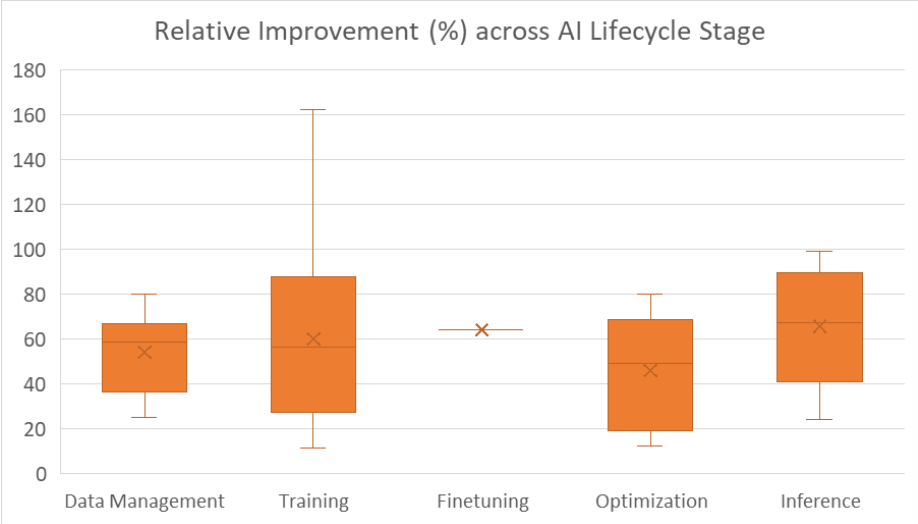


Fig. 8 | Percentage improvement in energy consumption reported across AI lifecycle stages. The box plot shows the median, interquartile range, and minimum and maximum values of reported relative energy improvements.

4. Guidelines for Green AI

We propose eight literature-backed guidelines for green AI. Figure 9 summarizes the energy-intensive stages of the AI lifecycle and highlights techniques to reduce their environmental impact. Each guideline is aligned with the recommendation where evidence suggests it delivers the greatest impact, though all guidelines are relevant to each of the AI lifecycle.

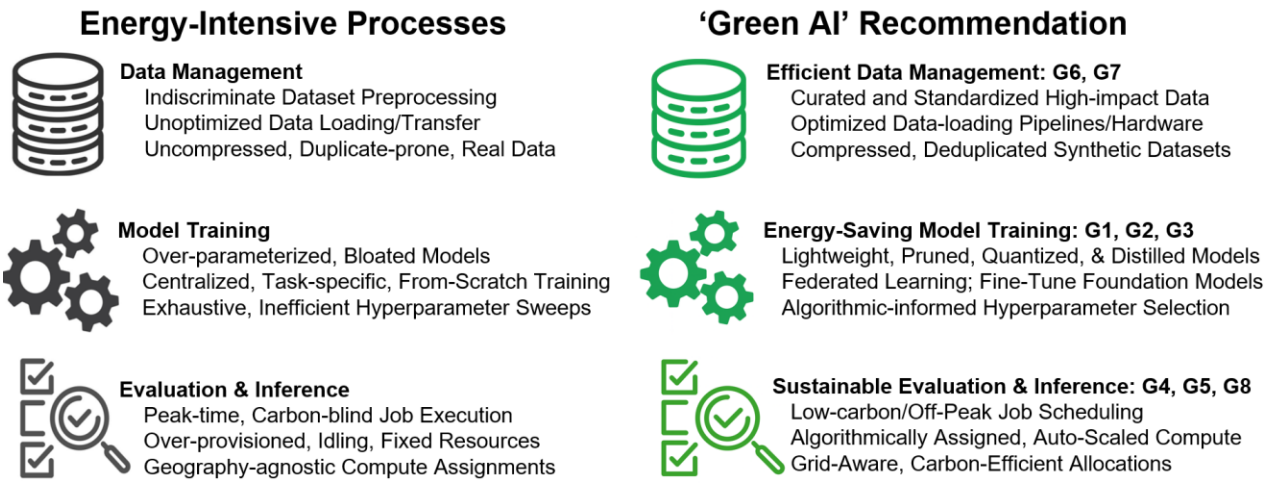


Fig. 9 | Broad categorization of energy-intensive AI subprocesses with corresponding techniques for reducing environmental impact. Green techniques can be employed at all stages to reduce energy costs.

G1. Adopt Emissions-Aware Scheduling and Resource Optimization

Use emissions-aware scheduling for compute-heavy jobs so they run when grid carbon intensity and data center load are low, reducing indirect emissions from cooling and supporting infrastructure [34, 95]. Consolidating workloads into hyperscale facilities can cut energy use by 25% [34]. Complement emissions-aware scheduling with hardware resource allocation systems that provide only the minimum necessary computing and storage resources, to avoid over-provisioning and idle energy usage [95].

G2. Design Energy-Efficient Models and Datasets

Design AI architectures with energy efficiency as a core objective and train models with high-impact, optimized data. Apply pruning, quantization, distillation, early stopping, and lightweight ensembling to reduce model size and training time while maintaining competitive performance [35,44,50,55,59,63,64,78,79,85,88,91,96]. Reported savings vary by method and setting: energy-aware training has been shown to cut training operations by 8 to 27% [76]; power capping can reduce energy use by 20% with minimal slowdown in reported configurations [55]; sustainable network design reduces carbon footprints by 42% [85]; energy-efficiency model training with early stopping and quantization save 56% of energy costs. Additionally, reduce system-level overhead by scaling datasets when appropriate and optimizing data-loading [53, 56, 68, 73, 83,]; smart data selection can decrease computation time by 41% [83].

G3. Leverage Pretrained Models and Efficient Learning Paradigms

Avoid training large models from scratch when possible. Instead, apply transfer learning, fine-tune foundation models, or use federated learning to reduce redundant computation and associated

emissions [48,58,97,60,66,69,73,80,97]. Reported results reduce energy by 84% with energy-efficient federated learning [98], 70% with quantized federated learning [66], 50% with joint optimization [69], and 32% with gradient compression [68]. Further, sharing open model weights and providing transparent methods prevents unnecessary re-training and duplicated emissions.

G4. Track Carbon Emissions Throughout the AI Lifecycle

Adopt standardized carbon-tracking tools (e.g., CodeCarbon, eCO₂AI) to consistently report energy use and emissions across the AI lifecycle [45,50,52,60,65,76,77,82,93,99,100,101]. Tracking alone does not cut emissions, but tracking software supports accuracy-carbon savings tradeoffs, enabling up to 75% carbon saving with minimal accuracy loss [93]. Prefer tools that support controls such as pausing, resuming and early stopping, allow for scheduled starts [36,46,54,], and that integrate geographic and time-varying grid intensity data, which can reduce carbon costs by up to 90% via multi-zone workload scheduling [45]. Figure 10 displays a carbon intensity map from July 2025 extracted from the Electricity Maps platform (Copenhagen, Denmark) and demonstrates the vast ranges of carbon intensity from grids worldwide.

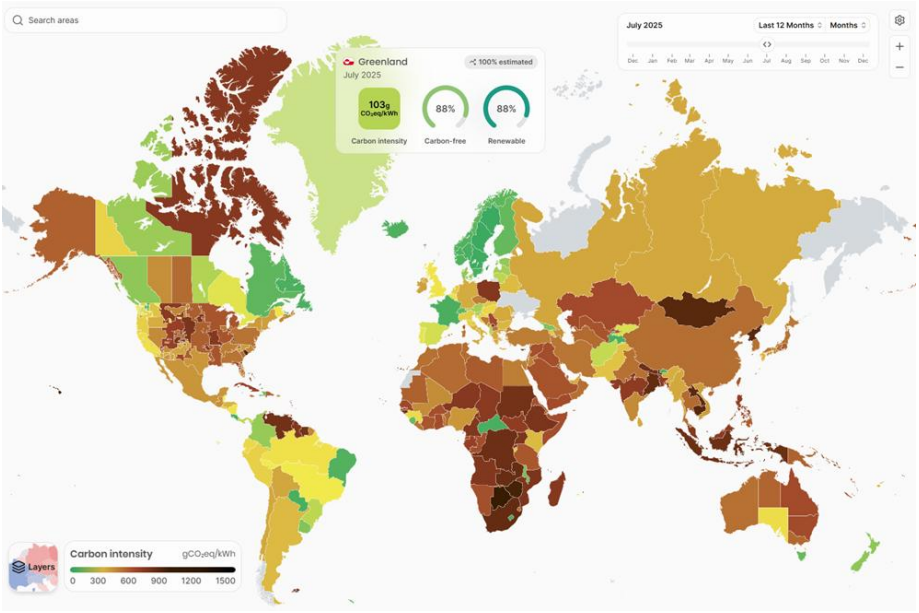


Fig. 10 | Example carbon-intensity map from the Electricity Maps platform illustrating wide geographic variation in grid emissions (gCO₂e/kWh). Such variability motivates carbon-aware scheduling and multi-zone workload placement to reduce the emissions of AI training and inference.

G5. Benchmark Environmental Metrics Alongside Performance

In addition to standard performance metrics (e.g., accuracy, F1-score), it is recommended to report environmental indicators such as carbon footprint, training time, memory usage, and energy consumed using standard PUE measures. This practice facilitates meaningful comparisons between models and fosters a culture of environmentally sustainable evaluation [36, 50, 57].

G6. Establish Benchmark Datasets for Energy Efficiency

Develop and maintain community-driven standard datasets dedicated to benchmarking energy efficiency across diverse AI tasks. A centralized repository for evaluating green AI methods on standardized tasks with embedded energy consumption metrics ensures fair comparisons and guides future innovation in green AI [37, 47].

G7. Encourage Policy and Institutional Reform

Encourage the adoption of institutional and governmental policies that prioritize funding, publication, and implementation of green AI systems. Reform requires aligning review criteria and research funding calls with environmental sustainability goals and advocating for national AI strategies that integrate environmental responsibility [38, 54, 62].

G8. Promote Explainable Sustainability in AI

Make energy and emissions impacts easy for non-experts to understand by providing clear dashboards and visual interfaces that translate technical metrics (i.e., energy, runtime, utilization, and carbon-intensity–based emissions estimates) into intuitive summaries and trends [37, 52]. Tools can surface per-query emissions and a user's cumulative footprint to encourage lower-impact choices (e.g., smaller models, batching, or deferring non-urgent runs to cleaner grid periods) and strengthen accountability [37, 52].

5. Discussion

We present a systematic review addressing sustainable AI strategies and guidelines synthesized from our findings to improve the environmental sustainability of AI. Data analysis revealed a rapidly expanding interest in green AI, yet the field remains methodologically fragmented, with inconsistent carbon reporting and a disproportionate focus on the training phase of AI models. Particularly, the inference stage, which may constitute the largest share of AI's energy footprint, has received limited attention. For instance, internal estimates from Google suggest that inference accounts for nearly 60% of its AI-related energy use [102]. Similarly, the evaluation stage is virtually absent from current literature, and its unstructured character means its environmental implications remain unexplored and poorly understood.

The sparse distribution of environmental metrics in the review reveals the lack of reporting standards and hinders reproducibility across studies. Carbon emissions, incorporating the carbon intensity of the power source, offer a more accurate reflection of AI's environmental impact than energy consumption alone. Indirect metrics such as runtime, number of computations, and model size require further evaluation to determine their validity as proxies for green AI metrics. Similarly, in federated learning studies, the absence of carbon reporting despite calculating energy cost

reflects unaligned practices in the AI community and highlights the need for standardized reporting protocols.

The observed geographical skew of publications toward high-income countries may reflect unequal access to hardware resources, renewable energy, and measurement infrastructure. Without equitable participation, AI can risk deepening the global divides in both sustainability and innovation. Establishing standardized reporting frameworks and lifecycle-aware benchmarks is therefore not only a scientific consideration but also a matter of governance and fairness. For researchers, these findings suggest embedding environmental transparency and regional context into every stage of the AI lifecycle. The guidelines proposed in this paper offer a pathway to harmonize research practices and act as building blocks for green AI inclusive policy frameworks.

Altogether, three main conclusions emerge from our synthesis: (1) Lifecycle Blind Spots: Research on green AI remains heavily skewed toward the training stage, underrepresenting the long-term energy costs of evaluation and inference; (2) Quantifiable Metrics: The development of carbon-tracking tools and model optimization techniques reflects a growing emphasis on measurable sustainability practices; and (3) Reporting Gaps : The lack of standardized reporting frameworks in AI covering elements such as environmental metrics, hardware, software details, and geographic context, continues to hinder meaningful cross-study comparisons and real-world relevance.

The review underscores the timely need to develop global standardization protocols, design carbon-aware methods, and conceptualize environmental sustainability as a socio-technical property of AI systems. Environmental sustainability could become a guiding principle of AI systems, not a secondary engineering concern. AI's energy profile is a structural characteristic of its development ecosystem, shaped by decisions across data curation, model architecture, hardware allocation, temporal scheduling, and geographic context. In January 2025, the World Economic Forum released a white paper highlighting the challenges and potential opportunities for AI sustainability through efficient lifecycle design and optimized data centers [103]. In this framing, green AI is not a matter of just minimizing wattage, but of rethinking how AI is designed, implemented, and scaled within a global computational infrastructure. This perspective aligns with broader agendas in responsible AI and climate science, where transparency, equity, and long-term impact are central.

The review has several limitations. To preserve scientific rigor, the review process excluded industrial white papers, non-peer-reviewed publications, and papers in other languages than English, thus omitting potentially informative real-world data. Our focus on software-level strategies means that embodied emissions from hardware manufacturing, cooling systems, and data center construction were excluded. Finally, heterogenous reporting in literature, such as missing context information and non-standard units, precluded a formal meta-analysis. These limitations reflect the nascency of the field and further underscore the need for standardized carbon accounting protocols.

In addition to energy-aware AI systems, we call for increased awareness of AI's carbon impact as aligned with UN Sustainable Development Goal 13 [104], which emphasizes education and institutional readiness for climate change mitigation. Awareness of AI's environmental footprint can begin with public AI chatbots and tools displaying the carbon emissions generated by individual user queries and their cumulative impact over time.

The environmental impact of AI is an emerging challenge. The findings and guidelines in this review have the potential to connect state-of-the-art technical innovation and nascent environmental governance through actionable and explainable frameworks for environmentally sustainable AI. The systematic review emphasizes the need for integrated policies and collaborative standards that embed sustainability alongside safety, ethics, and trustworthiness in AI regulation. The choices our community of researchers, policymakers, and industry make today will determine not only the capabilities of AI, but also its footprint on our environment tomorrow.

Acknowledgement:

We would like to acknowledge the support of Faisal Al Munajjed at Children's National Hospital, DC during the review.

References

1. Weizenbaum, Joseph. Computer power and human reason: from judgment to calculation. *W. H. Freeman and Company* (1976).
2. Sophia. Hansen Robotics https://www.hansonrobotics.com/sophia/ (2016).
3. Vaswani, A. et al. Attention is all you need. Preprint at https://doi.org/10.48550/arXiv.1706.03762 (2017).
4. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Preprint at https://doi.org/10.48550/arXiv.1810.04805 (2019).
5. ChatGPT. *OpenAI* https://chat.openai.com/chat (2023).
6. Claude 3 Opus, *Anthropic* https://claude.ai/ (2024).
7. Dhar, P. The carbon impact of artificial intelligence. *Nature Machine Intelligence* 2, 423–425 (2020).
8. Strubell, E., Ganesh, A. & McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 34, 13693-13696 (2019).
9. Mei, L. Energy Considerations for Large Pre-trained Neural Networks (San José State University, 2025).
10. EIA releases consumption and expenditures data from the Residential Energy Consumption Survey. *U.S. Energy Information Administration (EIA)* https://www.eia.gov/pressroom/releases/press530.php (2023).

11. Rahman, R., Heindrich, L., Owen, D. & Emberson, L. Over 30 AI models have been trained at the scale of GPT-4. *Epoch AI* https://epoch.ai/data-insights/models-over-1e25-flop (2025).
12. Luccioni, A. S., Viguier, S. & Ligozat, A.-L. Estimating the carbon footprint of BLOOM, a 176B parameter language model. *The Journal of Machine Learning Research* 24, 1–15 (2023).
13. Touvron, H. et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. Preprint at https://doi.org/10.48550/arXiv.2307.09288 (2023).
14. Flight Carbon Footprint Calculator. *Calculator.now* https://calculator.now/flight-carbon-footprint-calculator/ (2025).
15. Schwartz, R., Dodge, J., Smith, N. A. & Etzioni, O. Green AI. *Communications of the ACM* 63, 54–63 (2020).
16. Debus, C., Piraud, M., Streit, A., Theis, F. & Götz, M. Reporting electricity consumption is for sustainable AI. *Nature Machine Intelligence* 5, 1176–1178 (2023).
17. Verdecchia, R., Sallou, J. & Cruz, L. A systematic review of Green AI. *WIREs Data Mining and Knowledge Discovery* 13 (2023).
18. Bolón-Canedo, V., Morán-Fernández, L., Cancela, B. & Alonso-Betanzos, A. A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing* 599, 128096 (2024).
19. Tabbakh, A. et al. Towards sustainable AI: a comprehensive framework for Green AI. *Discover Sustainability* 5, 1–14 (2024).
20. EU Artificial Intelligence Act. *European Union* https://artificialintelligenceact.eu/ (2024).
21. Lekadir K, Frangi A F, Porras A R, Glocker B, Cintas C, Langlotz C P et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* 388 (2025).
22. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 389–399 (2019).
23. Gyevnár, B. & Kasirzadeh, A. AI safety for everyone. *Nature Machine Intelligence* 7, 531–542 (2025).
24. Hoffman, P. AI's Power Demand: Calculating ChatGPT's electricity consumption for handling over 365 billion user queries every year. *BestBrokers* https://www.bestbrokers.com/forex-brokers/ais-power-demand-calculating-chatgpts-electricity-consumption-for-handling-over-78-billion-user-queries-every-year (2025).
25. International, Total Energy Production. *U.S. Energy Information Administration (EIA)*. https://www.eia.gov/international/rankings/world.
26. Kaack, L. H. et al. Aligning artificial intelligence with climate change mitigation. *Nature Climate Change* 12, 518–527 (2022).
27. Kitchenham, B. Procedures for performing systematic reviews. *Keele University* 33, 1–26 (2004).
28. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)* 38, 1–10 (2014).

29. Minion J.T. et al. PICO Portal. *The Journal of the Canadian Health Libraries Association* 42, 181–183 (2021).
30. Bukar, U.A. et al. A method for analyzing text using VOSviewer. *MethodsX* 11 (2023).
31. Wade, A. D. The Semantic Scholar Academic Graph (S2AG). In *Companion Proceedings of the Web Conference 2022* 739–739 (2022).
32. Valenzuela-Escarcega, M. A., Ha, V. A. & Etzioni, O. Identifying Meaningful Citations. *AAAI Workshop: Scholarly Big Data* (2015).
33. Yang, Z., Chen, M., Saad, W., Hong, C. S. & Shikh-Bahaei, M. Energy Efficient Federated Learning Over Wireless Communication Networks. *IEEE Transactions on Wireless Communications* 20, 1935–1949 (2021).
34. Lannelongue, L., Grealey, J. & Inouye, M. Green Algorithms: Quantifying the Carbon Footprint of Computation. *Advanced Science* 8 (2021).
35. Strubell, E., Ganesh, A. & McCallum, A. Energy and Policy Considerations for Modern Deep Learning Research. In *Proceedings of the AAAI Conference on Artificial Intelligence* 34 13693–13696 (2020).
36. Justus, D., Brennan, J., Bonner, S. & McGough, A. S. Predicting the Computational Cost of Deep Learning Models. In *IEEE International Conference on Big Data* 3873–3882 (2018).
37. Yang, T.-J., Chen, Y.-H., Emer, J. & Sze, V. A method to estimate the energy consumption of deep neural networks. In *51st Asilomar Conference on Signals, Systems, and Computers* 1916–1920 (2017).
38. Dodge, J. et al. Measuring the Carbon Intensity of AI in Cloud Instances. In *ACM Conference on Fairness Accountability and Transparency* 1877–1894 (2022).
39. Taylor, P. Data center average annual PUE worldwide 2024. *Statista*. https://www.statista.com/statistics/1229367/data-center-average-annual-pue-worldwide/ (2025).
40. Courty, B. et al. mlCO₂/codecarbon: v2.4.1. Zenodo https://doi.org/10.5281/zenodo.11171501 (2024)
41. Henderson, P. et al. Towards the systematic reporting of the energy and carbon footprints of machine learning. arXiv 2002.05651 (2020).
42. Bannour, N., Ghannay, S., Névéol, A. & Ligozat, A.-L. Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. In *Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing* 11–21 (2021).
43. Lacoste, A., Luccioni, A., Schmidt, V. & Dandres, T. Quantifying the carbon emissions of machine learning. arXiv 1910.09700 (2019).
44. Shaikh, O. et al. EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, ACM* 1–7 (2021).
45. Tiutiulnikov, M. et al. eco4cast: Bridging Predictive Scheduling and Cloud Computing for Reduction of Carbon Emissions for ML Models Training. *Doklady Mathematics* 108, S443–S455 (2023).

46. Yoo, T., Lee, H., Oh, S., Kwon, H. & Jung, H. Visualizing the Carbon Intensity of Machine Learning Inference for Image Analysis on TensorFlow Hub. In *Computer Supported Cooperative Work and Social Computing, ACM* 206–211 (2023).
47. Kannan, J., Barnett, S., Simmons, A., Selvi, T. & Cruz, L. Green Runner: A Tool for Efficient Deep Learning Component Selection. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI* 112–117 (2024).
48. Li, C., Tsourdos, A. & Guo, W. A Transistor Operations Model for Deep Learning Energy Consumption Scaling Law. *IEEE Transactions on Artificial Intelligence* 5, 192–204 (2024).
49. Budennyy, S. A. et al. eCO$_2$AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. *Doklady Mathematics* 106, S118–S128 (2022).
50. Ma, H. & Ding, A. Method for evaluation on energy consumption of cloud computing data center based on deep reinforcement learning. *Electric Power Systems Research* 208, 107899 (2022).
51. Carastan-Santos, D. & Pham, T. H. T. Understanding the Energy Consumption of HPC Scale Artificial Intelligence. In *CARLA 2022- Latin America High Performance Computing Conference* 1660 131–144 (2022).
52. Jääskeläinen, P. Explainable Sustainability for AI in the Arts. In *The 1st International Workshop on Explainable AI for the Arts, ACM Creativity and Cognition Conference* (2023).
53. Castanyer, R. C., Martínez-Fernández, S. & Franch, X. Which design decisions in AI-enabled mobile applications contribute to greener AI? *Empirical Software Engineering* 29, 2 (2024).
54. Caspart, R. et al. Precise Energy Consumption Measurements of Heterogeneous Artificial Intelligence Workloads. In *International Conference on High Performance Computing* 108–121 (2022).
55. Zhao, D. et al. Sustainable Supercomputing for AI. In *Proceedings of the 2023 ACM Symposium on Cloud Computing* 588–596 (2023).
56. del Rey, S., Martínez-Fernández, S., Cruz, L. & Franch, X. Do DL models and training environments have an impact on energy consumption? In *49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), IEEE* 150–158 (2023).
57. Ferro, M., Silva, G. D., de Paula, F. B., Vieira, V. & Schulze, B. Towards a sustainable artificial intelligence: A case study of energy efficiency in decision tree algorithms. *Concurrency and Computation: Practice and Experience* 35 (2023).
58. Gómez-Carmona, O., Casado-Mansilla, D., Kraemer, F. A., López-de-Ipiña, D. & García-Zubia, J. Exploring the computational cost of machine learning at the edge for human-centric Internet of Things. *Future Generation Computer Systems* 112, 670–683 (2020).
59. Mazurek, S., Pytlarz, M., Malec, S. & Crimi, A. Investigation of Energy-Efficient AI Model Architectures and Compression Techniques for “Green” Fetal Brain Segmentation. In *Computational Science – ICCS 2024. Lecture Notes in Computer Science* 14835 61–74 (2024).
60. Ollivier, S. et al. Sustainable AI Processing at the Edge. *IEEE Micro* 43, 19–28 (2023).
61. Tomlinson, B., Black, R. W., Patterson, D. J. & Torrance, A. W. The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports* 14, 3732 (2024).

62. Yokoyama, A. M., Ferro, M., de Paula, F. B., Vieira, V. G. & Schulze, B. Investigating hardware and software aspects in the energy consumption of machine learning: A green AI-centric analysis. *Concurrency and Computation: Practice and Experience* 35 (2023).
63. Castellanos-Nieves, D. & García-Forte, L. Strategies of Automated Machine Learning for Energy Sustainability in Green Artificial Intelligence. *Applied Sciences* 14, 6196 (2024).
64. Perera-Lago, J. et al. An in-depth analysis of data reduction methods for sustainable deep learning. *Open Research Europe* 4, 101 (2024).
65. Yarally, T., Cruz, L., Feitosa, D., Sallou, J. & van Deursen, A. Uncovering Energy-Efficient Practices in Deep Learning Training: Preliminary Steps Towards Green AI. In *IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN)* 25–36 (2023).
66. Kim, M., Saad, W., Mozaffari, M. & Debbah, M. Green, Quantized Federated Learning Over Wireless Networks: An Energy-Efficient Design. *IEEE Transactions on Wireless Communications* 23, 1386–1402 (2024).
67. Hsu, Y.-L., Liu, C.-F., Wei, H.-Y. & Bennis, M. Optimized Data Sampling and Energy Consumption in IIoT: A Federated Learning Approach. *IEEE Transactions on Communications* 70, 7915–7931 (2022).
68. Li, P., Huang, X., Pan, M. & Yu, R. FedGreen: Federated Learning with Fine-Grained Gradient Compression for Green Mobile Edge Computing. In *IEEE Global Communications Conference (GLOBECOM)* 1–6 (2021).
69. Wang, J., Mao, Y., Wang, T. & Shi, Y. Green Federated Learning Over Cloud-RAN With Limited Fronthaul Capacity and Quantized Neural Networks. *IEEE Transactions on Wireless Communications* 23, 4300–4314 (2024).
70. Driss, M. B., Sabir, E., Elbiaze, H., Diallo, A. & Sadik, M. A Green Multi-Attribute Client Selection for Over-The-Air Federated Learning: A Grey-Wolf-Optimizer Approach. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* 10, 1–24 (2025).
71. Albaseer, A., Seid, A. M., Abdallah, M., Al-Fuqaha, A. & Erbad, A. Novel Approach for Curbing Unfair Energy Consumption and Biased Model in Federated Edge Learning. *IEEE Transactions on Green Communications and Networking* 8, 865–877 (2024).
72. Kuswiradyo, P., Kar, B. & Shen, S.-H. Optimizing the energy consumption in three-tier cloud–edge–fog federated systems with omnidirectional offloading. *Computer Networks* 250, 110578 (2024).
73. Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernandez-Pereira, E. & Perez-Sanchez, B. An effective and efficient green federated learning method for one-layer neural networks. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* 1050–1052 (2024).
74. Qi, Y. & Hossain, M. S. Harnessing federated generative learning for green and sustainable Internet of Things. *Journal of Network and Computer Applications* 222, 103812 (2024).
75. Gille, C., Guyard, F., Antonini, M. & Barlaud, M. Learning Sparse auto-Encoders for Green AI image coding. In *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1–5 (2023).

76. Lazzaro, D. et al. Minimizing Energy Consumption of Deep Learning Models by Energy-Aware Training. In *Image Analysis and Processing – ICIAP 2023. Lecture Notes in Computer Science* 14234 515–526 (2023).
77. Acmali, S. S., Ortakci, Y. & Seker, H. Green AI-Driven Concept for the Development of Cost-Effective and Energy-Efficient Deep Learning Method: Application in the Detection of Eimeria Parasites as a Case Study. *Advanced Intelligent Systems* 6, (2024).
78. Candelieri, A., Perego, R. & Archetti, F. Green machine learning via augmented Gaussian processes and multi-information source optimization. *Soft Computing* 25, 12591–12603 (2021).
79. Huang, K., Yin, H., Huang, H. & Gao, W. Towards Green AI in Fine-tuning Large Language Models via Adaptive Backpropagation. In *The 12th International Conference on Learning Representations* (2024).
80. Balderas, L., Lastra, M. & Benítez, J. M. An Efficient Green AI Approach to Time Series Forecasting Based on Deep Learning. *Big Data and Cognitive Computing* 8, 120 (2024).
81. Nijkamp, N., Sallou, J., van der Heijden, N. & Cruz, L. Green AI in Action: Strategic Model Selection for Ensembles in Production. In *Proceedings of the 1st ACM International Conference on AI-Powered Software* 50–58 (2024).
82. Spring, R. & Shrivastava, A. Scalable and Sustainable Deep Learning via Randomized Hashing. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Part F129685, 445–454 (2017).
83. Yin, Z., Pu, J., Wan, R. & Xue, X. Embrace sustainable AI: Dynamic data subset selection for image classification. *Pattern Recognition* 151, 110392 (2024).
84. Shi, J. et al. Greening Large Language Models of Code. In *Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Society, ACM* 142–153 (2024).
85. Pirnat, A. et al. Towards Sustainable Deep Learning for Wireless Fingerprinting Localization. In *IEEE International Conference on Communications (ICC)* 3208–3213 (2022).
86. Liu, Q., Zhu, J., Dai, Q. & Wu, X.-M. Benchmarking News Recommendation in the Era of Green AI. In *Companion Proceedings of the ACM Web Conference* 971–974 (2024).
87. Yang, X. et al. Sparse Optimization for Green Edge AI Inference. *Journal of Communications and Information Networks* 5, 1–15 (2020).
88. Reguero, Á. D., Martínez-Fernández, S. & Verdecchia, R. Energy-efficient neural network training through runtime layer freezing, model quantization, and early stopping. *Computer Standards & Interfaces* 92, 103906 (2025).
89. Wei, X. et al. Towards Greener Yet Powerful Code Generation via Quantization: An Empirical Study. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* 224–236 (2023).
90. Yang, Y., Kang, H. & Mirzasoleiman, B. Towards Sustainable Learning: Coresets for Data-efficient Deep Learning. In *Proceedings of the 40th International Conference on Machine Learning (PMLR)* 202 39314–39330 (2023).

91. Bird, T., Kingma, F. H. & Barber, D. Reducing the Computational Cost of Deep Generative Models with Binary Neural Networks. In *International Conference on Learning Representations (ICLR)* (2021).
92. Husom, E. J., Sen, S. & Goknil, A. Engineering Carbon Emission-aware Machine Learning Pipelines. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI* 118–128 (2024).
93. Li, B., Samsi, S., Gadepally, V. & Tiwari, D. Clover: Toward Sustainable AI with Carbon-Aware Machine Learning Inference Service. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, ACM* 1–15 (2023).
94. Zhang, S. et al. Differential Privacy-Aware Generative Adversarial Network-Assisted Resource Scheduling for Green Multi-Mode Power IoT. *IEEE Transactions on Green Communications and Networking* 8, 956–967 (2024).
95. Pagliari, R. et al. A Comprehensive Sustainable Framework for Machine Learning and Artificial Intelligence. In *27th European Conference on Artificial Intelligence* 392, 834–841 (2024).
96. McIntosh, A., Hassan, S. & Hindle, A. What can Android mobile app developers do about the energy consumption of machine learning? *Empirical Software Engineering* 24, 562–601 (2019).
97. Hu, Y., Huang, H. & Yu, N. Resource Optimization and Device Scheduling for Flexible Federated Edge Learning with Tradeoff Between Energy Consumption and Model Performance. *Mobile Networks and Applications* 27, 2118–2137 (2022).
98. Salh, A. et al. Energy-Efficient Federated Learning With Resource Allocation for Green IoT Edge Intelligence in B5G. *IEEE Access* 11, 16353–16367 (2023).
99. Jean-Quartier, C. et al. The Cost of Understanding—XAI Algorithms towards Sustainable ML in the View of Computational Cost. *Computation* 11, 92 (2023).
100. Shi, M. et al. Thinking Geographically about AI Sustainability. *AGILE: GIScience Series* 4, 1–7 (2023).
101. Henderson, P. et al. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *The Journal of Machine Learning Research* 21, 10039–10081 (2020).
102. D. Patterson et al. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. *Computer* 55, 18-28 (2022).
103. Artificial Intelligence's Energy Paradox: Balancing Challenges and Opportunities. *World Economic Forum – AI Governance Alliance* https://reports.weforum.org/docs/WEF_Artificial_Intelligences_Energy_Paradox_2025.pdf (2025).
104. The Sustainable Development Goals Report 2025. *United Nations* https://unstats.un.org/sdgs/report/2025/The-Sustainable-Development-Goals-Report-2025.pdf (2025).

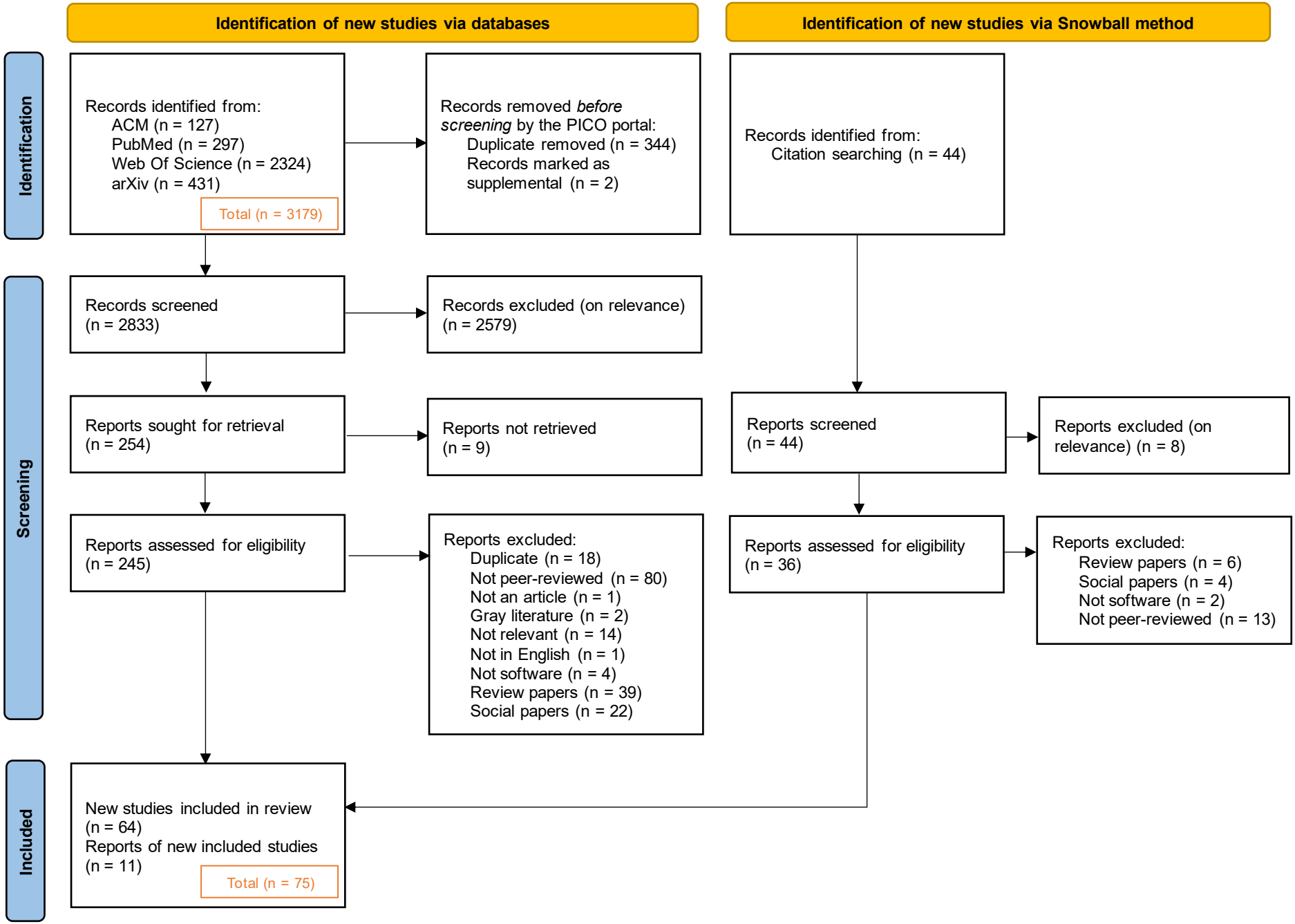
Supplementary Infromation:

1) PRISMA chart

2) Extracted Data for the review

3) List of tools

PRISMA flow diagram for "Green AI: Systematic Review and Guidelines for Sustainability in Artificial Intelligence"



“Green AI: Systematic Review and Guidelines for Sustainability in Artificial Intelligence”

Data Extracted from 75 Publications:

Paper	Type of Study	AI Lifecycle Stage	Type of AI Model	Learning Paradigm	Computing Infrastructure	Hardware Used	Software Libraries & Framework	Consumption-related Measuring Tools	Environmental Metrics	Performance Metrics	Energy Saving
Theme 1: Energy Consumption Tracking Tools											
Budennyy et al. 2022 [1]	Empirical	Training	Transformer *(Malevich, Kandinsky)*	Single-site	Cloud; HPC	CPU *(AMD EPYC)*; GPU *(NVIDIA A100)*	PyTorch	PyNVML; psutil	Energy (*kWh)*; Carbon (*gCO2e)*	Accuracy *(via loss function)*	17%
Jääskeläinen et al. 2023 [2]	Theoretical	Inference	-	-	Local	GPU	-	-	-	-	-
Jesse et al. 2022 [3]	Empirical	Inference	LLM *(BERT)*	Single-site	Cloud	GPU *(NVIDIA TITAN)*	-	-	Energy (*kWh)*; Carbon (*gCO2e)*	-	-
Justus et al. 2018 [4]	Empirical	Training	CNN	Single-site	Cloud; Local	CPU; GPU *(NVIDIA V100, P100, K80, M60, GTX 1080 Ti)*	TensorFlow	-	Indirect *(time)*	RMSE	-
Kannan et al. 2024 [5]	Empirical	Inference	-	Single-site	-	-	PyTorch	-	Indirect *(num. of computations)*	Accuracy	-
Li et al. 2022 [6]	Empirical	Training	CNN	Single-site	Local	CPU *(Intel Core i9, AMD Ryzen 9)*; GPU *(NVIDIA RTX 3080)*	PyTorch	-	Energy (*J)*	RMSE; R-squared	-
Lannelongue et al. 2021 [7]	Empirical	Inference	LLM *(BERT)*	Single-site	HPC	CPU *(Intel Xeon)*; GPU *(NVIDIA V100)*; TPU *(TPU-v3 Pod)*	-	Green Algorithms	Energy (*kWh)*; Carbon (*gCO2e)*	-	-
Ma et al. 2022 [8]	Empirical	Training	-	Single-site	HPC; Local	-	-	-	Power (*W)*	Mean relative deviation	-
Santos et al. 2022 [9]	Empirical	Inference	CNN *(MobileNet, Inception,*	Single-site	HPC	CPU *(Intel Xeon)*; GPU *(NVIDIA GTX*	TensorFlow	AI Benchmark Alpha;	Energy (*kWh)*; Carbon (*gCO2e)*	Accuracy	-

			ResNet, Inception-ResNet, VGG-16)			*1080 Ti)*; HPC *(Dell PowerEdge R730)*		Experiment Impact Tracker			
Shaikh et al. 2021 [10]	Empirical	Training	-	-	-	-	-	Intel RAPL; NVIDIA-SMI	-	-	-
Taewon et al. 2023 [11]	Empirical	Inference	CNN	Single-site	-	CPU; GPU	TensorFlow	CodeCarbon	Carbon (*gCO2e)*	-	-
Tien-Ju et al. 2017 [12]	Empirical	Training	CNN	Single-site	-	-	-	-	-	Accuracy	-
Tiutiulnikov et al. 2023 [13]	Empirical	Training	CNN	Single-site	Cloud	CPU; GPU *(NVIDIA V100)*	PyTorch	-	Carbon intensity *(gCO2e/kWh)*	-	17.8% (CE)
Theme 2: Evaluates Environmental Impact of AI methods											
Caspart et al. 2022 [14]	Empirical	Training	RNN *(LSTM)*	Single-site	HPC	CPU; GPU *(NVIDIA A100)*	PyTorch	NVIDIA Management Library *(NVML)*	Energy (*J*); Power (*W)*	Accuracy; Mean absolute percentage error *(MAPE)*	-
Castanyer et al. 2024 [15]	Empirical	Inference	CNN *(ResNet)*; RNN *(GRU, LSTM)*; Transformer	Single-site	-	CPU; GPU *(NVIDIA P100)*; Mobile *(Xiaomi Poco X3 NFC)*	PyTorch; TensorFlow; ONNX; TensorFlow Lite; Unity3D; Android Studio	Android Profiler; Unity Profiler	Indirect *(time)*	Accuracy	-
Dan et al. 2023 [16]	Empirical	Inference	LLM *(BERT, LLaMA 65B)*; CNN *(ResNet)*; Graph Neural Network *(DimeNet)*	Single-site	HPC	GPU *(NVIDIA A100, V100)*	Slurm	NVIDIA-SMI	Power (*W)*	training time; inference time*(latency)*	10-24%
del Rey et al. 2023 [17]	Empirical	Training	CNN *(MobileNetV2, NASNet Mobile, Xception, ResNet50, VGG16)*	Single-site	Cloud; Local	CPU; GPU *(NVIDIA GTX 750 Ti, RTX 3070, RTX 3090)*	TensorFlow	NVIDIA-SMI	Power (*W)*	F1 score	81%

Desislavov et al. 2023 [18]	Empirical	Inference	CNN *(AlexNet, VGG, ResNet, EfficientNet)*; Transformer*(BERT, MobileBERT)*	Single-site	-	GPU	PyTorch; TensorFlow	ptflops	Energy (*J)*	Accuracy	-
Emma et al. 2020 [19]	Empirical	Training	LLM *(BERT, GPT-2)*; RNN*(ELMo)*	Single-site	Cloud; HPC	-	TensorFlow ; Hyperopt	NVIDIA-SMI; Intel RAPL	Energy (*kWh)*; Power (*W)*; Carbon intensity *(gCO2e/kWh)*	-	-
Ferro et al. 2023 [20]	Empirical	Training	Traditional ML *(Decision Tree, Random Forest, XGBoost)*	Single-site	Local	CPU *(Intel Core i7)*; GPU *(NVIDIA RTX 2080 Ti)*	scikit-learn; pandas; NumPy	perf; PyRAPL	Energy (*J)*	Accuracy; MSE	53%, 62%, 135%, 162%
Gomez-Carmona et al. 2020 [21]	Empirical	Optimization	Traditional ML *(Logistic Regression, Random Forest, KNN, Naive Bayes, Linear SVM, Decision Trees)*; MLP	Single-site *(Edge)*	Local	CPU *(Intel Core i7, Rasberry Pi 3, Rasberry i Zero W)*	scikit-learn	-	Indirect *(time)*	F1 score; Precision; Recall; Accuracy	80% (time)
Jean-Quartier et al. 2023 [22]	Empirical	Optimization	CNN *(YOLO)*; Traditional ML *(Random Forest, Decision Tree)*	Single-site	Local	CPU	PyTorch; scikit-learn; SHAP; SciPy	CodeCarbon	Energy (*kWh)*	Accuracy; MSE; R-squared	-
Yu et al. 2022 [23]	Empirical	Inference	Traditional ML *(LR, KNN, SVM, RF, XGB)*; Neural Networks *(NN1, NN5, Pruned NN, Quantized NN)*	Single-site	Local	CPU *(Intel Core i5)*	PyTorch; scikit-learn	Intel Power Gadget	Power (*W)*	Accuracy; AUROC	-

Mazurek et al. 2024 [24]	Empirical	Data Management	CNN *(UNet)*	Single-site	HPC	CPU *(AMD EPYC 7742)*; GPU *(NVIDIA A100)*	PyTorch; MONAI; Lightning-Bagua; Weights & Biases	CodeCarbon	Energy (*J)*	DICE	400% (time)
McIntosh et al. 2019 [25]	Empirical	Training	Traditional ML *(Naive Bayes, J48, SMO, Logistic Regression, Random Forest, KNN, ZeroR)*; MLP	Single-site	-	Mobile *(Galaxy Nexus Android 4.2.2)*	WEKA;Neuroph; Android SDK; Dalvik VM	-	Energy (*J)*; Power (*W)*	Accuracy; RMSE	-
Nesrine et al. 2021 [26]	Empirical	Inference	LLM	Single-site	HPC; Local	CPU; GPU *(GTX 1080 Ti, V100)*	Hugging Face Transformers	CarbonTracker; Green Algorithms; ML CO2 Impact; Experiment Impact Tracker; Cumulator; Energy Usage	Energy (*kWh)*; Carbon (*gCO2e)*	-	-
Ollivier et al. 2023 [27]	Empirical	Inference	CNN *(AlexNet, VGG-16)*	Single-site *(Edge)*	Local	CPU; GPU; Supercomputer *(NVIDIA Jetson NX)*	PyTorch; TensorFlow	-	Power (*W)*; Carbon (*gCO2e)*	Throughput	-
Omar et al. 2024 [28]	Empirical	Data Management	Traditional ML *(SVM, Naive Bayes, Decision Tree, Hoeffding Tree, KNN, Random Forest, AdaBoost, Bagging Classifier)*	Single-site	Local	CPU *(Intel Xeon)*	scikit-multiflow; scikit-learn	CodeCarbon	Energy (*J)*	Drift detection closeness; True alarm rate	-

Rakshit et al. 2021 [29]	Empirical	Training	LLM *(BERT)*	Single-site	Local	CPU; GPU	PyTorch	CodeCarbon	Energy (*kWh)*; Carbon (*gCO2e)*	Accuracy	-
Raluca Maria et al. 2022 [30]	Empirical	Finetuning	-	Single-site	Local	CPU *(Rasberry Pi 3 and 4)*	ONNX; Docker; WebAssembly *(Wasm)*	High Voltage Power Monitor *(HVPM)* hardware by Monsoon	Energy (*J)*	Execution time	-
Tomlinson et al. 2024 [31]	Theoretical	Training	LLM	Single-site	-	GPU *(NVIDIA A100)*	-	-	Carbon (*gCO2e)* per task	-	-
Yokoyama et al. 2023 [32]	Empirical	Optimization	CNN; Traditional ML *(XGBoost, Random Forest, Decision Trees, K-means)*; MLP	Single-site	Local	CPU; GPU *(NVIDIA RTX)*; TPU; Supercomputer *(NVIDIA Jetson TX-2)*	Dask; scikit-learn; pandas; NumPy	perf; NVIDIA-SMI	Energy (*J)*; Carbon (*gCO2e)*	Accuracy	24%
Castellanos-Nieves et al. 2023 [33]	Empirical	Training	Traditional ML *(SGD Classifier, Logistic Regression, KNN, Random Forest, Gradient Boosting, Decision Tree, Linear & Non-Linear SVM, Gaussian Naive Bayes)*	Single-site	Local	CPU *(Intel Core i7)*; GPU	scikit-learn	mathematical model; psutil	Energy (*kWh)*; Carbon intesity *(gCO2e/kWh)*	Accuracy; F1 score	-
Dagoberto et al. 2024 [34]	Empirical	Training	CNN *(EfficientNet)*; RNN *(BiLSTM)*	Single-site	Cloud; Local	CPU; GPU	Keras; PyTorch; Ray Tune	-	Power (*W)*; Carbon (*gCO2e/kWh)*	Accuracy	28%

Toscano-Durán et al. 2024 [35]	Empirical	Data Management	CNN *(YOLO)*; MLP	Single-site	Local	CPU *(Intel Xeon Silver 4210)*; GPU *(NVIDIA Quadro RTX 4000)*	PyTorch	CodeCarbon	Indirect *(time)*; Carbon (*gCO2e)*	Accuracy; Precision; recall; F1 score; Mean Average Precision *(MAP)*	40% (CE)
Yarally et al. 2023 [36]	Empirical	Training	-	Single-site	Local	GPU *(NVIDIA GTX 1080)*	PyTorch	-	Energy (*J)*	Accuracy	-
Theme 3: Green Frameworks in Federated Learning											
Albaseer et al. 2024 [37]	Empirical	Optimization	CNN	Federated	Distributed	CPU; GPU	PyTorch	-	Energy (*J)*	-	17%
Fontenla-Romero et al. 2023 [38]	Empirical	Training	Neural Network *(Single-Layer Perceptron)*	Federated	Distributed	CPU *(Intel Core i7)*	-	-	Power (*Wh)*	Accuracy	87% (time)
Hsu et al. 2022 [39]	Empirical	Optimization	-	Federated	Distributed	CPU	-	Mathematical model	Energy (*J)*	RMSE	69%
Hu et al. 2022 [40]	Empirical	Training	CNN; MLP	Federated *(Edge)*	Distributed	CPU	-	Mathematical model	Energy (*J)*	Accuracy	-
Kim et al. 2024 [41]	Empirical	Training	CNN	Federated	Distributed	-	-	Energy model of processing chip	Energy (*J)*	Accuracy; Precision	70%
Kuswiradyo et al. 2024 [42]	Empirical	Training	-	Federated	Distributed	CPU *(Intel Core i5)*	IBM ILOG CPLEX Optimization Studio	-	Reduction in Energy (*%)*	Latency	11-27%
Maryam Ben et al. 2024 [43]	Empirical	Optimization	CNN	Federated	Distributed	GPU *(NVIDIA T4)*	Keras; TensorFlow	-	Energy (*J)*	Accuracy	43%
Peichun et al. 2021 [44]	Empirical	Data Management	CNN *(VGG-9)*	Federated *(Edge)*	Distributed	CPU	-	Mathematical model	Energy (*J)*	Accuracy	32%, 57%
Wang et al. 2023 [45]	Empirical	Training	Quantized Neural Network *(QNN)*	Federated	Distributed	-	MATLAB	Mathematical model	Energy (*J)*	Fronthaul Capacity *(Gbps)*; Level of precision *(bit)*	50%
Zhou et al. 2022 [46]	Empirical	Inference	-	Federated	Distributed	CPU	MATLAB	-	Energy (*J)*	-	-
Lee et al. 2023 [47]	Empirical	Training	CNN	Federated *(Edge)*	Distributed	CPU; GPU	PyTorch	-	Energy (*J)*	Structural Similarity	-

										Index Measure *(SSIM)*; Membership inference attack accuracy	
Qi et al. 2024 [48]	Empirical	Data Management	CNN	Federated	Distributed	CPU	PyTorch	-	Indirect *(overhead, latency)*	Accuracy	80% (comm. overhead)
Xia et al. 2024 [49]	Empirical	Training	CNN *(LeNet-5, AlexNet, MobileNetV)*	Federated	Distributed	CPU; GPU	-	-	Indirect *(computational complexity and overhead)*	Accuracy	-
Theme 4: Novel Techniques for Green AI											
Acmali et al. 2024 [50]	Empirical	Training	CNN	Single-site	Local	CPU; GPU	-	-	Indirect *(FLOPs)*	Accuracy	50% (model param)
Álvaro Domingo et al. 2024 [51]	Empirical	Training	CNN; Auto-Regressive Prediction Model	Single-site	Local	CPU; GPU	TensorFlow	NVIDIA-SMI	Energy (*kWh)*	Accuracy	56%
Candelieri et al. 2021 [52]	Empirical	Optimization	Traditional ML *(SVM)*	Single-site	HPC	CPU	-	-	Indirect *(time)*	Accuracy	66% (time)
Gille et al. 2022 [53]	Empirical	Data Management	CNN *(Convolutional AutoEncoder)*	Single-site	HPC	CPU; GPU *(NVIDIA A100)*	PyTorch	-	Reduction in Energy (*%)*	Peak SNR *(PSNR)*; Structural Similarity Index Measure *(SSIM)*	25%
Huang et al. 2023 [54]	Empirical	Finetuning	Transformer-Based LLM *(OPT, BLOOMZ, FLAN-T5)*	Single-site	-	GPU *(NVIDIA A100)*	PyTorch; TensorFlow	Torch Profiler	Indirect *(FLOPs)*	ROGUE scores	64% (FLOPs)
Lazzaro et al. 2023 [55]	Empirical	Training	-	Single-site	Local	CPU; GPU	-	-	Reduction in Energy (*%)*	-	7-31%
Liu et al. 2024 [56]	Empirical	Training	-	Single-site	Local	GPU *(NVIDIA RTX 3090)*	-	-	Carbon (*gCO2e)*	AUC; Mean Reciprocal Rank *(MRR)*; Normalized Discounted Cumulative Gain *(NDCG)*	87.6% (CO2)

Luis et al. 2024 [57]	Empirical	Training	MLP	Single-site	Local	-	sktime	-	Indirect *(num. of parameters)*	Mean Absolute Error *(MAE)*; RMSE; Mean Absolute Percentage Error *(MAPE)*; R-squared; Mean Squared Logarithmic Error *(MSLE)*; Median Absolute Percentage Error *(MdAPE)*; DIstortion Loss including shApe and TimE *(DILATE)*	20% (number of model parameters)
Nijkamp et al. 2024 [58]	Empirical	Inference	-	Single-site	Cloud	CPU	Datadog	-	Indirect *(CPU usage)*	Precision; Recall; F1 score	14%, 57%
Pirnat et al. 2022 [59]	Empirical	Training	CNN *(ResNet18)*	Single-site	Cloud	GPU	-	-	Energy (*J*); Carbon (*gCO2e)*	Mean Distance Error *(MDE)*; RMSE	94%
Ryan et al. 2017 [60]	Empirical	Training	MLP	Single-site	Local	CPU *(Intel Core i7, Xeon)*	-	-	Indirect *(num. of computations)*	Accuracy	95% (compute cost)
Shi et al. 2024 [61]	Empirical	Inference	LLM *(CodeBERT, GraphCodeBERT)*	Single-site	Local	CPU; GPU	Hugging Face Transformers	MLCO2 Impact	Energy (*kWh)*; Carbon (*gCO2e)*	Accuracy; latency	99%
Thomas et al. 2020 [62]	Empirical	Training	Generative Models *(VAE, Flow)*	Single-site	Local	-	PyTorch	-	-	Bits per dimension *(BPD)*	90-94% (model size) del
Xiangyu et al. 2020 [63]	Theoretical	Inference	CNN *(GoogLeNet)*	Single-site	-	-	-	-	Power (*W)*	Signal-to-Interference-plus-Noise Ratio *(SINR)*	67%
Xiaokai et al. 2023 [64]	Empirical	Optimization	LLM	Single-site	Cloud	CPU; GPU	PyTorch	-	Carbon (*gCO2e)*	BLEU *(BiLingual Evaluation Understudy)*	28.8%, 48.4%, 55% (CO2)

										score; Accuracy	
Yang et al. 2023 [65]	Empirical	Data Management	CNN *(ResNet-20, ResNet-18, ResNet-50)*; LLM *(RoBERTa)*	Single-site	Local	GPU *(NVIDIA RTX A6000)*	-	-	-	Accuracy	60% (time)
Yin et al. 2024 [66]	Empirical	Data Management	CNN *(ResNet, DenseNet)*	Single-site	HPC; Local	GPU *(NVIDIA A100, RTX 3080)*	-	-	Indirect *(time)*	Accuracy	47%, 52%, 62%
Theme 5: New Carbon-aware AI frameworks											
Erik Johannes et al. 2024 [67]	Empirical	Optimization	CNN; RNN *(LSTM)*; Traditional ML *(Decision Tree, Random Forest, XGBoost, Gradient Boost)*	Single-site	Cloud; Local	CPU; GPU	PyTorch; TensorFlow	CodeCarbon	Energy (*kWh)*; Carbon (*gCO2e)*	F1 score; R-squared	-
Gutierrez et al. 2023 [68]	Empirical	Optimization	CNN; RNN *(LSTM)*; Traditional ML *(Decision Tree, Random Forest, XGBoost, Gradient Boost)*	Single-site	Local	CPU	Scikit	Energy Efficiency Tester *(hardware)*	Power (*W/s)*	-	12%
Li et al. 2023 [69]	Empirical	Inference	LLM *(ALBERT)*; CNN *(EfficientNet, YOLOv5)*	Single-site	Cloud; HPC	CPU *(ADM EPYC)*; GPU *(NVIDIA A100)*	PyTorch; Flask; NetworkX	Carbon Tracker	Carbon (*gCO2e)* per day	Accuracy; Latency	80%
Meilin et al. 2023 [70]	Theoretical	Training	-	-	-	-	-	-	-	-	-
Henderson et al. 2021 [71]	Empirical	Inference	-	Single-site	HPC	CPU *(Intel Xeon)*; GPU *(NVIDIA A100, V100)*	PyTorch; TensorFlow	RAPL; Intel Power Gadget; NVIDIA-SMI; psutil	Energy (*kWh)*; Carbon intesity *(gCO2e/kWh)*	BLEU *(BiLingual Evaluation Understudy)* score	-

Roberto et al. 2024 [72]	Empirical	Finetuning	Traditional ML *(Linear Regression, Decision Tree, Random Forest, XGBoost)*; Neural Network	Single-site	Local	CPU	-	-	Carbon (*gCO2e)*	Accuracy	-
Salh et al. 2023 [73]	Empirical	Training	-	Federated	Distributed	CPU	-	-	Energy (*J)*	Time	-
Sunxuan et al. 2024 [74]	Empirical	Training	-	Federated	Distributed	CPU	-	-	Energy (*J)*	Accuracy *(via loss function)*; Differential privacy cost	11%
Zhaohui et al. 2020 [75]	Theoretical	Training	-	Federated	Distributed	-	-	-	Reduction in Energy (*%)*	Maximum average transmit power in dB	59%

References:

[1] Budennyy, S. A. et al. eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. Doklady Mathematics 106, S118–S128 (2022).

[2] Jääskeläinen, P. Explainable Sustainability for AI in the Arts. in The 1st International Workshop on Explainable AI for the Arts, ACM Creativity and Cognition Conference (2023).

[3] Dodge, J. et al. Measuring the Carbon Intensity of AI in Cloud Instances. in ACM Conference on Fairness Accountability and Transparency 1877–1894 (ACM, 2022). doi:10.1145/3531146.3533234.

[4] Justus, D., Brennan, J., Bonner, S. & McGough, A. S. Predicting the Computational Cost of Deep Learning Models. in IEEE International Conference on Big Data (Big Data) 3873–3882 (IEEE, 2018). doi:10.1109/BigData.2018.8622396.

[5] Kannan, J., Barnett, S., Simmons, A., Selvi, T. & Cruz, L. Green Runner: A Tool for Efficient Deep Learning Component Selection. in Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI 112–117 (ACM, 2024). doi:10.1145/3644815.3644942.

[6] Li, C., Tsourdos, A. & Guo, W. A Transistor Operations Model for Deep Learning Energy Consumption Scaling Law. IEEE Transactions on Artificial Intelligence 5, 192–204 (2024).

[7] Lannelongue, L., Grealey, J. & Inouye, M. Green Algorithms: Quantifying the Carbon Footprint of Computation. Advanced Science 8, (2021).

[8] Ma, H. & Ding, A. Method for evaluation on energy consumption of cloud computing data center based on deep reinforcement learning. Electric Power Systems Research 208, 107899 (2022).

[9] Carastan-Santos, D. & Pham, T. H. T. Understanding the Energy Consumption of HPC Scale Artificial Intelligence. in High Performance Co mputing. CARLA 2022. Communications in Computer and Information Science, vol 1660 131–144 (Springer, Cham, 2022). doi:10.1007/978-3-031-23821-5_10.

[10] Shaikh, O. et al. EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models. in Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems 1–7 (ACM, 2021). doi:10.1145/3411763.3451780.

[11] Yoo, T., Lee, H., Oh, S., Kwon, H. & Jung, H. Visualizing the Carbon Intensity of Machine Learning Inference for Image Analys is on TensorFlow Hub. in Computer Supported Cooperative Work and Social Computing 206–211 (ACM, 2023). doi:10.1145/3584931.3606959.

[12] Yang, T.-J., Chen, Y.-H., Emer, J. & Sze, V. A method to estimate the energy consumption of deep neural networks. in 51st Asilomar Conference on Si gnals, Systems, and Computers 1916–1920 (IEEE, 2017). doi:10.1109/ACSSC.2017.8335698.

[13] Tiutiulnikov, M. et al. eco4cast: Bridging Predictive Scheduling and Cloud Computing for Reduction of Carbon Emissions for ML Models Training. Doklady Mathematics 108, S443–S455 (2023).

[14] Caspart, R. et al. Precise Energy Consumption Measurements of Heterogeneous Artificial Intelligence Workloads. in Internation al Conference on High Performance Computing 108–121 (2022). doi:10.1007/978-3-031-23220-6_8.

[15] Castanyer, R. C., Martínez-Fernández, S. & Franch, X. Which design decisions in AI-enabled mobile applications contribute to greener AI? Empirical Software Engineering 29, 2 (2024).

[16] Zhao, D. et al. Sustainable Supercomputing for AI. in Proceedings of the 2023 ACM Symposium on Cloud Computing 588–596 (ACM, 2023). doi:10.1145/3620678.3624793.

[17] del Rey, S., Martínez-Fernández, S., Cruz, L. & Franch, X. Do DL models and training environments have an impact on energy consumption? in 49th Eur omicro Conference on Software Engineering and Advanced Applications (SEAA) 150–158 (IEEE, 2023). doi:10.1109/SEAA60479.2023.00031.

[18] Desislavov, R., Martínez-Plumed, F. & Hernández-Orallo, J. Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. Sustainable Computing: Informatics and Systems 38, 100857 (2023).

[19] Strubell, E., Ganesh, A. & McCallum, A. Energy and Policy Considerations for Modern Deep Learning Research. in Proceedings of the AAAI Conference on Artificial Intelligence vol. 34 13693–13696 (2020).

[20] Ferro, M., Silva, G. D., de Paula, F. B., Vieira, V. & Schulze, B. Towards a sustainable artificial intelligence: A case study of energy efficiency in decision tree algorithms. Concurrency and Computation: Practice and Experience 35, (2023).

[21] Gómez-Carmona, O., Casado-Mansilla, D., Kraemer, F. A., López-de-Ipiña, D. & García-Zubia, J. Exploring the computational cost of machine learning at the edge for human-centric Internet of Things. Future Generation Computer Systems 112, 670–683 (2020).

[22] Jean-Quartier, C. et al. The Cost of Understanding—XAI Algorithms towards Sustainable ML in the View of Computational Cost. Computation 11, 92 (2023).

[23] Yu, J.-R. et al. Energy Efficiency of Inference Algorithms for Clinical Laboratory Data Sets: Green Artificial Intelligence Study. J ournal of Medical Internet Research 24, e28036 (2022).

[24] Mazurek, S., Pytlarz, M., Malec, S. & Crimi, A. Investigation of Energy-Efficient AI Model Architectures and Compression Techniques for “Green” Fetal Brain Segmentation. in Computational Science – ICCS 2024. ICCS 2024. Lecture Notes in Computer Science, vol 14835 61–74 (Springer, Cham, 2024). doi:10.1007/978-3-031-63772-8_5.

[25] McIntosh, A., Hassan, S. & Hindle, A. What can Android mobile app developers do about the energy consumption of machine learning? Empirical Software Engineering 24, 562–601 (2019).

[26] Bannour, N., Ghannay, S., Névéol, A. & Ligozat, A.-L. Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. in Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing 11–21 (Association for Computational Linguistics, 2021). doi:10.18653/v1/2021.sustainlp-1.2.

[27] Ollivier, S. et al. Sustainable AI Processing at the Edge. IEEE Micro 43, 19–28 (2023).

[28] Omar, R. et al. How to Sustainably Monitor ML-Enabled Systems? Accuracy and Energy Efficiency Tradeoffs in Concept Drift Detection. in 10th International Conference on ICT for Sustainability (ICT4S) 172–182 (IEEE, 2024). doi:10.1109/ICT4S64576.2024.00026.

[29] Naidu, R. et al. Towards Quantifying the Carbon Emissions of Differentially Private Machine Learning. in Workshop on Socially Responsible Machine Learning, The International Conference on Machine Learning (ICML) (2021). doi:10.48550/arxiv.2107.06946.

[30] Hampau, R. M., Kaptein, M., van Emden, R., Rost, T. & Malavolta, I. An empirical study on the Performance and Energy Consumption of AI Containerization Strategies for Computer-Vision Tasks on the Edge. in The International Conference on Evaluation and Assessment in Software Engineering 2022 50–59 (ACM, 2022). doi:10.1145/3530019.3530025.

[31] Tomlinson, B., Black, R. W., Patterson, D. J. & Torrance, A. W. The carbon emissions of writing and illustrating are lower for AI than for humans. Scientific Reports 14, 3732 (2024).

[32] Yokoyama, A. M., Ferro, M., de Paula, F. B., Vieira, V. G. & Schulze, B. Investigating hardware and software aspects in the energy consumption of machine learning: A green AI-centric analysis. Concurrency and Computation: Practice and Experience 35, (2023).

[33] Castellanos-Nieves, D. & García-Forte, L. Improving Automated Machine-Learning Systems through Green AI. Applied Sciences 13, 11583 (2023).

[34] Castellanos-Nieves, D. & García-Forte, L. Strategies of Automated Machine Learning for Energy Sustainability in Green Artificial Intelligence. Applied Sciences 14, 6196 (2024).

[35] Perera-Lago, J. et al. An in-depth analysis of data reduction methods for sustainable deep learning. Open Research Europe 4, 101 (2024).

[36] Yarally, T., Cruz, L., Feitosa, D., Sallou, J. & van Deursen, A. Uncovering Energy-Efficient Practices in Deep Learning Training: Preliminary Steps Towards Green AI. in IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN) 25–36 (IEEE, 2023). doi:10.1109/CAIN58948.2023.00012.

[37] Albaseer, A., Seid, A. M., Abdallah, M., Al-Fuqaha, A. & Erbad, A. Novel Approach for Curbing Unfair Energy Consumption and Biased Model in Federated Edge Learning. IEEE Transactions on Green Communications and Networking 8, 865–877 (2024).

[38] Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernandez-Pereira, E. & Perez-Sanchez, B. An effective and efficient green federated learning method for one-layer neural networks. in Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing 1050–1052 (ACM, 2024). doi:10.1145/3605098.3636144.

[39] Hsu, Y.-L., Liu, C.-F., Wei, H.-Y. & Bennis, M. Optimized Data Sampling and Energy Consumption in IIoT: A Federated Learning Approach. IEEE Transactions on Communications 70, 7915–7931 (2022).

[40] Hu, Y., Huang, H. & Yu, N. Resource Optimization and Device Scheduling for Flexible Federated Edge Learning with Tradeoff Between Energy Consumption and Model Performance. Mobile Networks and Applications 27, 2118–2137 (2022).

[41] Kim, M., Saad, W., Mozaffari, M. & Debbah, M. Green, Quantized Federated Learning Over Wireless Networks: An Energy-Efficient Design. IEEE Transactions on Wireless Communications 23, 1386–1402 (2024).

[42] Kuswiradyo, P., Kar, B. & Shen, S.-H. Optimizing the energy consumption in three-tier cloud–edge–fog federated systems with omnidirectional offloading. Computer Networks 250, 110578 (2024).

[43] Driss, M. B., Sabir, E., Elbiaze, H., Diallo, A. & Sadik, M. A Green Multi-Attribute Client Selection for Over-The-Air Federated Learning: A Grey-Wolf-Optimizer Approach. ACM Transactions on Modeling and Performance Evaluation of Computing Systems 10, 1–24 (2025).

[44] Li, P., Huang, X., Pan, M. & Yu, R. FedGreen: Federated Learning with Fine-Grained Gradient Compression for Green Mobile Edge Computing. in IEEE Global Communications Conference (GLOBECOM) 1–6 (IEEE, 2021). doi:10.1109/GLOBECOM46510.2021.9685582.

[45] Wang, J., Mao, Y., Wang, T. & Shi, Y. Green Federated Learning Over Cloud-RAN With Limited Fronthaul Capacity and Quantized Neural Networks. IEEE Transactions on Wireless Communications 23, 4300–4314 (2024).

[46] Zhou, X., Zhao, J., Han, H. & Guet, C. Joint Optimization of Energy Consumption and Completion Time in Federated Learning. in IEEE 42nd International Conference on Distributed Computing Systems (ICDCS) 1005–1017 (IEEE, 2022). doi:10.1109/ICDCS54860.2022.00101.

[47] Lee, J., Seif, M., Cho, J. & Poor, H. V. Exploring the Privacy-Energy Consumption Tradeoff for Split Federated Learning. IEEE Network 38, 388–395 (2024).

[48] Qi, Y. & Hossain, M. S. Harnessing federated generative learning for green and sustainable Internet of Things. Journal of Network and Computer Applications 222, 103812 (2024).
[49] Xia, C., Tsang, D. H. K. & Lau, V. K. N. Structured Bayesian Federated Learning for Green AI: A Decentralized Model Compression Using Turbo-VBI-Based Approach. IEEE Internet of Things Journal 11, 12783–12798 (2024).
[50] Acmali, S. S., Ortakci, Y. & Seker, H. Green AI-Driven Concept for the Development of Cost-Effective and Energy-Efficient Deep Learning Method: Application in the Detection of Eimeria Parasites as a Case Study. Advanced Intelligent Systems 6, (2024).
[51] Reguero, Á. D., Martínez-Fernández, S. & Verdecchia, R. Energy-efficient neural network training through runtime layer freezing, model quantization, and early stopping. Computer Standards & Interfaces 92, 103906 (2025).
[52] Candelieri, A., Perego, R. & Archetti, F. Green machine learning via augmented Gaussian processes and multi-information source optimization. Soft Computing 25, 12591–12603 (2021).
[53] Gille, C., Guyard, F., Antonini, M. & Barlaud, M. Learning Sparse auto-Encoders for Green AI image coding. in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 1–5 (IEEE, 2023). doi:10.1109/ICASSP49357.2023.10096638.
[54] Huang, K., Yin, H., Huang, H. & Gao, W. Towards Green AI in Fine-tuning Large Language Models via Adaptive Backpropagation. in The Twelfth International Conference on Learning Representations (2024). doi:10.48550/arXiv.2309.13192.
[55] Lazzaro, D. et al. Minimizing Energy Consumption of Deep Learning Models by Energy-Aware Training. in Image Analysis and Processing – ICIAP 2023. Lecture Notes in Computer Science, vol 14234 515–526 (Springer, Cham, 2023). doi:10.1007/978-3-031-43153-1_43.
[56] Liu, Q., Zhu, J., Dai, Q. & Wu, X.-M. Benchmarking News Recommendation in the Era of Green AI. in Companion Proceedings of the ACM Web Conference 2024 971–974 (ACM, 2024). doi:10.1145/3589335.3651472.
[57] Balderas, L., Lastra, M. & Benítez, J. M. An Efficient Green AI Approach to Time Series Forecasting Based on Deep Learning. Big Data and Cognitive Computing 8, 120 (2024).
[58] Nijkamp, N., Sallou, J., van der Heijden, N. & Cruz, L. Green AI in Action: Strategic Model Selection for Ensembles in Production. in Proceedings of the 1st ACM International Conference on AI-Powered Software 50–58 (ACM, 2024). doi:10.1145/3664646.3664763.
[59] Pirnat, A. et al. Towards Sustainable Deep Learning for Wireless Fingerprinting Localization. in ICC 2022 - IEEE International Conference on Communications 3208–3213 (IEEE, 2022). doi:10.1109/ICC45855.2022.9838464.
[60] Spring, R. & Shrivastava, A. Scalable and Sustainable Deep Learning via Randomized Hashing. in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining vol. Part F129685 445–454 (ACM, 2017).
[61] Shi, J. et al. Greening Large Language Models of Code. in Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Society 142–153 (ACM, 2024). doi:10.1145/3639475.3640097.
[62] Bird, T., Kingma, F. H. & Barber, D. Reducing the Computational Cost of Deep Generative Models with Binary Neural Networks. in International Conference on Learning Representations (ICLR) (2021). doi:10.48550/arXiv.2010.13476.
[63] Yang, X. et al. Sparse Optimization for Green Edge AI Inference. Journal of Communications and Information Networks 5, 1–15 (2020).
[64] Wei, X. et al. Towards Greener Yet Powerful Code Generation via Quantization: An Empirical Study. in Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering 224–236 (ACM, 2023). doi:10.1145/3611643.3616302.
[65] Yang, Y., Kang, H. & Mirzasoleiman, B. Towards Sustainable Learning: Coresets for Data-efficient Deep Learning. in Proceedings of the 40th International Conference on Machine Learning, PMLR 202 39314–39330 (2023). doi:10.48550/arXiv.2306.01244.
[66] Yin, Z., Pu, J., Wan, R. & Xue, X. Embrace sustainable AI: Dynamic data subset selection for image classification. Pattern Recognition 151, 110392 (2024).
[67] Husom, E. J., Sen, S. & Goknil, A. Engineering Carbon Emission-aware Machine Learning Pipelines. in Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI 118–128 (ACM, 2024). doi:10.1145/3644815.3644943.

[68] Gutiérrez, M., Moraga, M. Á., García, F. & Calero, C. Green-IN Machine Learning at a Glance. Computer 56, 35–43 (2023).

[69] Li, B., Samsi, S., Gadepally, V. & Tiwari, D. Clover: Toward Sustainable AI with Carbon-Aware Machine Learning Inference Service. in Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis 1–15 (ACM, 2023). doi:10.1145/3581784.3607034.

[70] Shi, M. et al. Thinking Geographically about AI Sustainability. AGILE: GIScience Series 4, 1–7 (2023).

[71] Henderson, P. et al. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. The Journal of Machine Learning Research 21, 10039–10081 (2020).

[72] Pagliari, R. et al. A Comprehensive Sustainable Framework for Machine Learning and Artificial Intelligence. in 27th European Conference on Artificial Intelligence (volume 392) 834–841 (2024). doi:10.48550/arXiv.2407.12445.

[73] Salh, A. et al. Energy-Efficient Federated Learning With Resource Allocation for Green IoT Edge Intelligence in B5G. IEEE Access 11, 16353–16367 (2023).

[74] Zhang, S. et al. Differential Privacy-Aware Generative Adversarial Network-Assisted Resource Scheduling for Green Multi-Mode Power IoT. IEEE Transactions on Green Communications and Networking 8, 956–967 (2024).

[75] Yang, Z., Chen, M., Saad, W., Hong, C. S. & Shikh-Bahaei, M. Energy Efficient Federated Learning Over Wireless Communication Networks. IEEE Transactions on Wireless Communications 20, 1935–1949 (2021).

Tools mentioned in the Systematic Review

Tool	#Occurrences	Function	
Experiment Tracking & Monitoring Tools			
Ai Benchmark Alpha	1	Benchmarking tool (python-based) for AI algorithms in HPC setting to compare energy and accuracy outcomes.	
Code Carbon	7	Python package that computes carbon emissions based on time, power (GPU + CPU + RAM), and carbon intensity.	
ML CO2 Impact	2	Online tool that computes carbon emissions based on *user-provided* information such as runtime, hardware, cloud provider and location.	
Carbon Tracker	2	Python package. (same formula as MLCO2)	
Experiment Impact Tracker	2	Python package. ((same formula as MLCO2)	
Green Algorithms	2	Convenient online tool. (same formula as MLCO2)	
Cumulator	1	Python package that estimates carbon emissions based on runtime, GPU load and carbon intensity as well as communication energy.	
Energy usage tool	1	Python package (possibly buggy)	
Profiling & Performance Tools			
NVIDIA-SMI	7	command-line tool	Monitor NVIDIA GPUs
NVML	1	C-based API	
pyNVML	1	Python library	
Intel Power Gadget	2	Software app	Monitor Intel CPU power usage
Intel RAPL	3	Hardware feature	
pyRAPL	1	Python library	
perf	2	Command-line profiler tool in Linux to monitor CPU performance and system events.	
psutil	3	Python library for retrieving information on system utilization (CPU, memory, disks etc.)	
ptflops	1	Python library for calculating the number of FLOPs/computations during inference only.	
PyTorch Profiler	1	Python library for profiling time and memory usage by CPU/GPU.	
Android Profiler	1	Built-in Android Studio tool for monitoring CPU usage, memory, device power consumption, etc.	
Unity Profiler	1	Built-in Unity engine tool to get performance information on app's CPU, GPU, memory, and rendering.	