American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX • Website: www.ajamir.com • Email: editor@ajamir.com

Energy-Conscious Artificial Intelligence for Sustainable Computing Laboratories

Dr. Samuel Reinhardt

Associate Professor, Technical University of Munich (TUM), Germany

Abstract

Artificial intelligence research laboratories increasingly depend on graphics processing units, high-performance accelerators, large datasets, repeated hyperparameter experiments, model inference services, and continuously operating computing infrastructure. These resources enable rapid experimentation but can create substantial electricity demand when computational efficiency is treated as secondary to model performance. Energy-conscious artificial intelligence offers an alternative approach in which model accuracy, computational cost, hardware utilization, experiment scheduling, and environmental impact are evaluated jointly throughout the research lifecycle. This study develops an Energy-Conscious Artificial Intelligence Laboratory Framework (ECAILF) for sustainable computing laboratories. The framework integrates workload-level energy measurement, model right-sizing, experiment budgeting, accelerator sharing, carbon-aware scheduling, efficient inference, dataset and pipeline optimization, equipment utilization, and institutional sustainability governance. Because verified laboratory measurements were not available, the study adopts an explicitly simulation-based methodological design rather than presenting hypothetical values as empirical observations.

A synthetic analytical corpus of 400 laboratory AI workloads was distributed across four hypothetical configurations: conventional GPU workflows, measured and scheduled workloads, efficient models with shared accelerators, and integrated energy-conscious AI laboratories. Six dimensions—energy observability, computational efficiency, hardware utilization, workload scheduling, carbon awareness, and governance maturity—were incorporated into a Sustainable Computing Readiness Index. The simulated index rises from 45 for conventional workflows to 93 for the integrated energy-conscious laboratory. The analysis emphasizes that energy measurement is necessary but not sufficient, because current software estimators can differ materially from direct measurements. The paper concludes that sustainable computing laboratories should treat energy as an explicit research resource alongside time, accuracy, memory, and financial cost, thereby encouraging scientifically rigorous experimentation while reducing avoidable computational demand.

Keywords: energy-conscious AI, sustainable computing, Green AI, machine-learning laboratories, energy efficiency, carbon-aware computing, GPU utilization, AI sustainability

1. Introduction

Artificial intelligence has become increasingly computationally intensive as research laboratories use larger models, larger datasets, repeated experimentation, accelerator-intensive training, and high-throughput inference. The broader electricity implications of AI infrastructure are becoming significant: the International Energy Agency projects that global data-center electricity consumption could reach approximately 945 TWh by 2030 in its base case, while data-center electricity demand increased sharply in 2019. Although university and institutional computing laboratories represent only a portion of this infrastructure, the trend illustrates why computational efficiency can no longer be treated as an invisible research cost.

The concept of Green AI emerged partly in response to research practices that rewarded incremental model-performance gains without consistently reporting computational cost. Schwartz, Dodge, Smith, and Etzioni argued that efficiency should become an explicit evaluation criterion alongside model quality and advocated reporting the computational and financial cost of AI experiments. Henderson and colleagues subsequently proposed more systematic reporting of energy and carbon footprints so that machine-learning research could compare environmental cost alongside conventional performance measures. These arguments are particularly relevant to computing laboratories, where research culture influences how frequently experiments are repeated, how accelerators are allocated, and whether energy consumption is measured at all.

Energy-conscious AI should nevertheless be distinguished from a narrow focus on reducing electricity consumption. Computational sustainability depends on the interaction among algorithms, model size, hardware efficiency, accelerator utilization, workload scheduling, software configuration, facility conditions, and the carbon intensity of electricity. MLCommons has increasingly incorporated power measurement into MLPerf benchmarking, reflecting the growing need to compare machine-learning systems through performance and energy simultaneously. Its Power Working Group develops measurement techniques intended to support comparable energy-efficiency reporting across machine-learning systems. The International Energy Agency similarly emphasizes that efficiency improvements across software, hardware, and infrastructure can substantially reduce the electricity required to supply the same level of digital and AI services.

Measurement itself remains challenging. Tools such as Carbontracker and CodeCarbon have made energy and carbon estimation substantially easier for machine-learning practitioners, while Henderson and colleagues developed approaches to integrate energy reporting into experimental workflows. However, Fischer's 2025 validation study found that widely used estimation approaches can exhibit substantial error relative to external measurements, with discrepancies reaching approximately 40% in tested circumstances. Sustainable laboratory governance should therefore distinguish between approximate software-based monitoring and calibrated measurement where higher precision is needed.

The present study proposes an integrated framework for energy-conscious AI laboratories rather than treating sustainable computing as an isolated model-compression problem. The framework considers the entire experimental workflow, including research planning, baseline selection, energy measurement, model design, accelerator allocation, hyperparameter search, training schedules, inference, reporting, equipment lifecycle, and institutional governance. Because no verified laboratory dataset was supplied, the quantitative analysis is explicitly simulated. Its objective is to demonstrate how laboratories could evaluate sustainability readiness without fabricating empirical energy savings.

2. Objectives of the Study

2.1 To Develop an Energy-Conscious AI Laboratory Framework

The first objective is to establish a practical framework through which research laboratories can integrate energy measurement, efficient model development, hardware utilization, workload scheduling, carbon awareness, and sustainability governance into ordinary AI experimentation. The framework treats energy as a measurable research resource rather than as an external operational cost.

2.2 To Compare Alternative Laboratory Computing Configurations

The second objective is to compare conventional GPU-centered experimentation with progressively stronger sustainability practices, including workload measurement, scheduling, efficient model selection, accelerator sharing, and integrated energy-conscious governance.

2.3 To Identify Reproducible Sustainability Metrics

The third objective is to identify metrics that allow laboratories to compare model performance and computational cost transparently. Particular attention is given to energy per experiment, energy per useful result, accelerator utilization, performance-per-kWh, carbon intensity, experimental redundancy, and the reliability of software-based energy estimates.

3. Review of Literature

3.1 Green AI and Computational Efficiency

Schwartz and colleagues' formulation of Green AI challenged the tendency to value model accuracy without adequately accounting for computational resources. Their argument was not that high-performance research should cease, but that researchers should report efficiency so that scientific progress can include improvements in resource use. This principle is especially important in academic laboratories, where computational budgets influence reproducibility and participation. If meaningful research requires continuously expanding accelerator resources, smaller laboratories and resource-constrained institutions face increasing barriers to experimentation.

Henderson and colleagues extended this concern by developing a systematic framework for reporting machine-learning energy and carbon footprints. Their work demonstrated that energy accounting could be embedded within research workflows rather than treated as an external sustainability exercise. Anthony, Kanding, and Selvan subsequently introduced Carbontracker, which tracks and predicts energy use and associated carbon emissions during deep-learning training. These contributions helped shift the literature from broad concern about AI's environmental consequences toward practical instrumentation.

Compute efficiency is also an algorithmic research problem. Bartoldson and colleagues analyzed trends in compute-efficient deep learning and showed that algorithmic and systems improvements can influence the amount of computation required to reach a particular level of model quality. The implication for laboratories is that model selection should consider the performance–resource frontier rather than assuming that the largest available architecture is necessarily the best experimental starting point.

The sustainability literature further distinguishes Green IN AI, which seeks to make AI itself more resource efficient, from Green BY AI, where AI is applied to sustainability problems elsewhere. For computing laboratories, Green IN AI is the more immediate concern because it addresses the environmental cost of model development, training, and inference. A laboratory can conduct socially beneficial AI research while still operating inefficient experimental pipelines; sustainability therefore requires explicit attention to the research process itself.

3.2 Energy Measurement, Carbon Accounting, and Benchmarking

Reliable measurement is fundamental to energy-conscious computing because researchers cannot systematically optimize a cost they do not observe. Energy estimation tools now include hardware-level monitoring, software estimators, emissions calculators, and benchmark-specific power measurement. CodeCarbon estimates energy and carbon emissions from local computing workloads and can incorporate hardware utilization and regional electricity characteristics. Carbontracker similarly monitors machine-learning training and predicts its energy and carbon footprint.

However, carbon emissions and electrical energy are not identical metrics. Energy consumption is determined primarily by computational workload and hardware behavior, whereas operational carbon emissions additionally depend on electricity-generation characteristics. A laboratory can therefore reduce carbon intensity by scheduling flexible experiments when electricity is cleaner without necessarily reducing the total electrical energy consumed. Conversely, an energy-efficiency improvement reduces electricity demand regardless of the energy mix. This distinction is important when laboratories decide whether their primary objective is reduced energy, reduced emissions, reduced cost, or some combination of all three. The IEA's analysis of AI and energy explicitly distinguishes demand growth from the generation mix supplying that demand.

Software estimators should also be interpreted carefully. Fischer's validation of CodeCarbon and other estimation approaches against external measurements found that the tools generally captured patterns in consumption but could produce substantial quantitative inaccuracies. Pathania and colleagues' 2018 survey likewise identified methodological diversity across software-energy and carbon-estimation approaches, including differences in what hardware components are measured or estimated. Laboratories should therefore document the measurement method they use and avoid reporting estimated carbon values with false precision.

3.3 Sustainable Laboratory Operations and Research Gap

The laboratory environment creates a distinctive sustainability challenge because energy is consumed not only by successful final models but throughout model development. Exploratory runs, failed configurations, debugging experiments, hyperparameter searches, duplicated pipelines, idle accelerators, unnecessary checkpointing, and oversized models can collectively create substantial computational overhead. The Green AI literature therefore encourages reporting development cost rather than focusing only on the energy of a single final training run.

Carbon-aware workload scheduling has also emerged as a practical strategy for flexible computing. Research on sustainable data centers has explored shifting workloads in response to grid carbon intensity, cooling requirements, renewable availability, and energy cost. For example, Sarkar and colleagues proposed a real-time data-center optimization framework combining load shifting, cooling, and energy storage and reported simulated reductions across carbon, energy, and cost in their evaluated settings. Although a university computing laboratory operates at a different scale from a commercial data center, flexible experiment scheduling can apply the same underlying principle.

4. Research Methodology

4.1 Research Design and Synthetic Workload Corpus

The study adopts a simulation-based comparative methodology. No actual laboratory electricity meter, GPU cluster, research workstation, or institutional billing record was used to generate the quantitative results. A synthetic corpus of 400 hypothetical AI workloads was developed, representing model training, validation, hyperparameter search, inference benchmarking, and exploratory experimentation.
The workloads were divided equally among four hypothetical laboratory configurations: conventional GPU workflows, measured and scheduled workloads, efficient models with shared accelerators, and integrated energy-conscious AI laboratories. The purpose of this design is to compare governance maturity rather than estimate real electricity savings.

Table 1: Energy-Conscious Artificial Intelligence Laboratory Framework

Dimension	Operational Meaning	Illustrative Laboratory Practice	Scale
Energy Observability	Ability to measure or estimate workload-level electricity consumption	Power meters, GPU/CPU telemetry, CodeCarbon, benchmark logs	0–100
Computational Efficiency	Performance obtained relative to model size and compute demand	Model right-sizing, pruning, quantization, efficient architectures	0–100
Hardware Utilization	Degree to which accelerators perform useful work rather than remain fragmented or idle	Shared scheduling, batching, utilization monitoring	0–100
Workload Scheduling	Ability to coordinate jobs according to demand, priority, and resource availability	Queues, off-peak scheduling, workload consolidation	0–100
Carbon Awareness	Consideration of electricity carbon intensity when flexible workloads are executed	Carbon-aware scheduling, renewable-aware experimentation	0–100
Governance Maturity	Institutional integration of sustainability into research planning and reporting	Energy budgets, sustainability reporting, review policies	0–100

Source: Author-developed simulation framework informed by Green AI reporting principles, MLCommons power benchmarking, IEA energy analysis, and machine-learning emissions-tracking research.

4.2 Laboratory Configuration Models

The Conventional GPU Workflow configuration represents a laboratory in which researchers optimize mainly for accuracy and completion time. Experiments can be launched independently, accelerator utilization is not routinely reviewed, and energy consumption is largely invisible to researchers.

The Measured and Scheduled Workloads configuration introduces workload-level energy monitoring, queue-based job scheduling, experiment logging, and explicit records of compute duration. Energy measurement creates a baseline through which inefficient runs can be identified.

The Efficient Models with Shared Accelerators configuration adds model right-sizing, pruning or quantization where appropriate, shared accelerator scheduling, utilization monitoring, constrained hyperparameter search, and preference for efficient baselines before scaling to more computationally expensive models.

The Energy-Conscious AI Laboratory configuration integrates all previous practices while additionally incorporating carbon-aware scheduling, sustainability targets, experiment budgets, calibrated energy measurement, efficient inference policies, equipment lifecycle planning, transparent reporting, and periodic sustainability review.

4.3 Sustainable Computing Readiness Index

A Sustainable Computing Readiness Index (SCRI) was constructed from the six framework dimensions using equal conceptual weighting. Equal weighting is used only for methodological transparency and does not imply that every laboratory should assign identical importance to each sustainability dimension.

A future empirical implementation should measure electrical energy directly where practical and should distinguish estimated from measured values. MLCommons' emphasis on standardized power measurement illustrates the importance of consistent protocols when systems are compared. Where software estimation is used, laboratories should document the monitoring tool, version, hardware coverage, sampling method, and uncertainty because recent validation studies demonstrate that energy estimators can differ significantly from ground truth.

Future empirical testing could additionally calculate model-performance-per-kWh, experiments-per-kWh, accelerator utilization, average idle time, carbon emissions per completed experiment, energy per hyperparameter search, inference energy per 1,000 requests, and the proportion of workloads scheduled during lower-carbon electricity periods.

5. Analysis

5.1 Comparative Sustainability Performance

The simulation demonstrates a progressive increase in sustainability readiness as laboratories move from unmeasured compute consumption toward integrated energy governance.

Table 2: Simulated Sustainable Computing Readiness Scores

Laboratory Configuration	Energy Observability	Computational Efficiency	Hardware Utilization	Workload Scheduling	Carbon Awareness	Governance Maturity	SCRI
Conventional GPU Workflow	28	49	52	51	32	58	45
Measured & Scheduled Workloads	75	58	62	79	47	63	64
Efficient Models + Shared Accelerators	84	88	89	84	68	73	81
Energy-Conscious AI Laboratory	95	93	94	92	90	94	93

Source: Simulated analytical values created for methodological illustration. These values do not represent measured electricity consumption or verified laboratory savings.

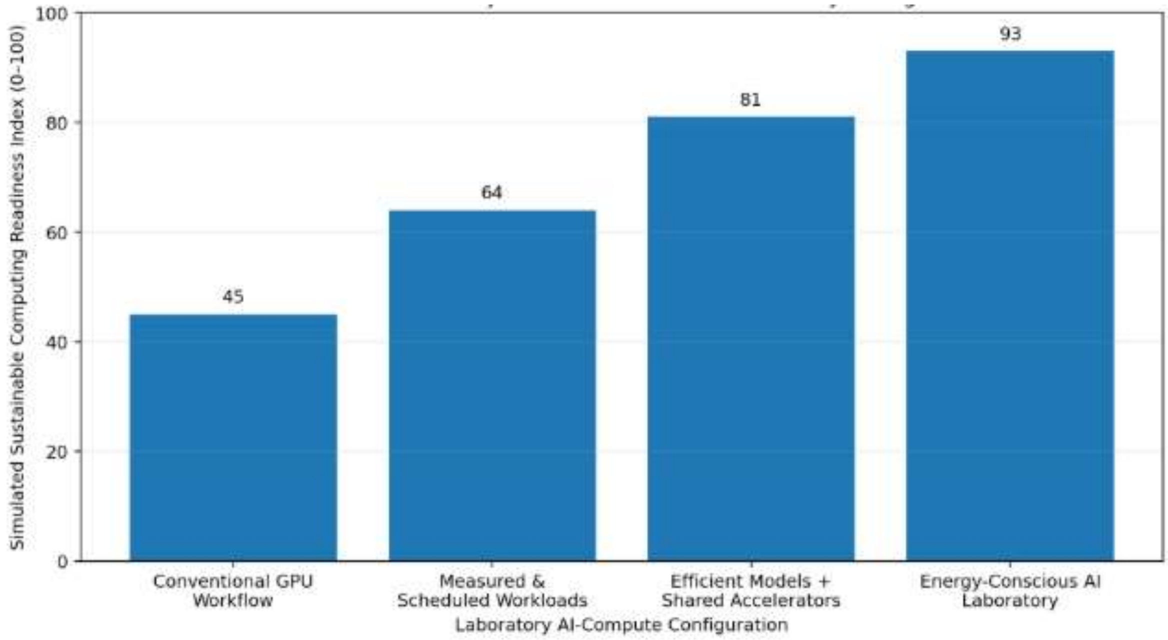
The conventional configuration performs weakest in energy observability and carbon awareness. This does not mean that every conventional experiment is inefficient; rather, weak observability prevents researchers from distinguishing efficient and inefficient practices systematically.

The transition to measured workloads substantially improves readiness because energy becomes visible. This reflects the logic of Henderson, Anthony, and subsequent energy-accounting research: monitoring allows computational cost to become part of experimental decision-making.The largest additional gains emerge when monitoring is combined with efficient models and stronger accelerator utilization. This demonstrates the methodological distinction between measuring energy and using measurements to change research behavior.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)
E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

5.2 Sustainability Readiness Across Laboratory Configurations

Graph 1: Simulated Sustainable Computing Readiness Across AI Laboratory Configurations



The simulated index rises from 45 for conventional GPU-centered workflows to 64 for measured and scheduled workloads, 81 for efficient models with shared accelerators, and 93 for an integrated energy-conscious laboratory.

The increase from 45 to 64 represents the contribution of measurement and scheduling. The increase from 64 to 81 reflects stronger efficiency and utilization. The final increase to 93 represents governance integration, where energy is incorporated into research planning rather than measured only after experiments are complete.The graph should not be interpreted as evidence that a particular intervention will produce a fixed percentage reduction in laboratory electricity consumption. Actual savings depend on model types, workload intensity, existing hardware utilization, electricity characteristics, research priorities, and laboratory infrastructure.

5.3 Energy-Conscious Experimental Workflow

The proposed framework suggests that sustainability decisions should begin before training. Researchers should first establish the minimum model complexity required to answer the research question, select a computationally efficient baseline, estimate the experiment budget, and define the evidence required to justify additional scaling.During development, workload-level energy should be recorded together with familiar experimental variables such as accuracy, loss, runtime, and memory use. This approach follows Green AI's principle that efficiency should be reported alongside model quality rather than treated as invisible infrastructure.

When computationally expensive experiments are justified, laboratories can reduce unnecessary energy through shared accelerators, controlled search spaces, early stopping, appropriate precision, checkpoint policies, batching, and efficient data pipelines. Results should subsequently report enough computational context to enable reproduction and meaningful comparison.

6. Discussion

6.1 Sustainable AI Laboratories Require Measurement Before Optimization

The first implication of the study is that laboratories cannot manage computational sustainability reliably without measuring it. Researchers routinely record accuracy, training time, memory, and model parameters, yet energy remains absent from many experimental records. This absence makes inefficient experimentation difficult to identify and encourages the perception that computational resources are effectively unlimited once hardware has been purchased. Green AI and systematic energy-reporting research both challenge this assumption by treating computational cost as part of scientific evaluation rather than as background infrastructure.

Energy monitoring should nevertheless be implemented with methodological caution. Software-based tools are highly valuable because they lower the practical barrier to measurement, but recent ground-truth validation demonstrates that estimates can deviate materially from externally measured consumption. Laboratories should therefore adopt a tiered measurement approach. Routine exploratory experiments can reasonably use software monitoring for relative comparisons, while major infrastructure studies, procurement comparisons, or published claims of precise energy savings should use calibrated hardware measurement wherever practicable. This distinction improves credibility without making sustainability measurement prohibitively difficult for everyday research.

Measurement also changes the unit through which research efficiency is understood. A training run that consumes less energy but produces materially weaker scientific output may not represent meaningful progress. Conversely, a higher-energy experiment may be justified if it produces a substantial gain in validated performance or scientific understanding. Sustainable computing therefore requires performance-normalized metrics, such as task quality per kWh, rather than absolute energy minimization. MLCommons' integration of energy with benchmark performance reflects precisely this need to evaluate useful computational output and energy jointly.

6.2 Model Efficiency, Hardware Utilization, and Scheduling Must Operate Together

The second implication is that sustainable laboratory computing cannot be achieved through model compression alone. Algorithmic efficiency can reduce the computation required for a task, but laboratory energy is also influenced by how effectively hardware is utilized. A theoretically efficient model can still be part of an inefficient laboratory workflow if accelerators remain idle for long periods, workloads are poorly scheduled, or experiments are unnecessarily duplicated. Conversely, stronger shared utilization may reduce the need for additional hardware while improving the amount of useful computation obtained from existing infrastructure.

Laboratories should therefore connect model-selection decisions with systems management. Researchers can begin with smaller baselines, define stopping criteria before launching large experiments, reduce redundant hyperparameter searches, and adopt lower-precision inference or quantization when accuracy requirements permit. Infrastructure teams can complement these practices through accelerator sharing, queue management, batching, utilization telemetry, and workload consolidation. The underlying objective is not to restrict scientific ambition but to ensure that additional computational expenditure produces justified scientific value.

Scheduling adds a further environmental dimension. Flexible experimental workloads do not always need to run immediately. Where reliable carbon-intensity data are available, laboratories can schedule nonurgent training during periods of cleaner electricity supply. Research on carbon-aware data-center control demonstrates the broader feasibility of coordinating workload timing with energy-system conditions. This approach should complement rather than replace energy efficiency: shifting an inefficient workload to cleaner electricity reduces carbon impact but does not eliminate unnecessary electricity consumption.

6.3 Sustainability Governance Should Become Part of Research Quality

The third implication is that energy-conscious computing should be treated as a research-governance issue rather than left entirely to individual researchers. Voluntary efficiency decisions are useful, but institutional incentives can unintentionally encourage the opposite behavior when researchers are evaluated exclusively through model performance, experiment quantity, or publication speed. Schwartz and colleagues' Green AI argument explicitly connected computational efficiency with the inclusiveness and accessibility of AI research, because escalating compute requirements can advantage organizations with larger financial resources.

A sustainable laboratory can therefore establish internal norms for computational reporting, experiment budgeting, shared-resource allocation, and efficient baseline testing. Major projects can document expected computational requirements before large-scale training begins, while published research can include energy or compute information sufficient to contextualize performance. Such policies should remain proportionate; low-cost experiments should not face the same administrative burden as large accelerator-intensive training programs.

Governance must also avoid reducing sustainability to carbon accounting alone. A laboratory may lower reported operational emissions by purchasing electricity from lower-carbon sources while continuing to waste computational resources. Genuine sustainability requires efficient algorithms, high utilization, responsible equipment procurement, appropriate hardware lifetimes, and transparent measurement. The IEA's energy-efficiency scenario similarly indicates that software, hardware, and infrastructure improvements can collectively reduce the electricity required to provide AI and digital services. Research governance should therefore reward actual efficiency improvement alongside responsible energy sourcing.

7. Conclusion

Energy-conscious artificial intelligence provides a practical pathway for aligning computational research with the wider sustainability responsibilities of modern laboratories. As AI experiments become more dependent on specialized accelerators, repeated model development, high-throughput inference, and increasingly complex pipelines, energy use should no longer remain an invisible research input. The present study has developed an integrated framework in which energy observability, computational efficiency, hardware utilization, workload scheduling, carbon awareness, and governance maturity operate together. This approach moves beyond the narrow assumption that sustainable AI can be achieved simply by selecting smaller models or reporting estimated emissions after training has finished. The simulation-based comparison produced Sustainable Computing Readiness Index values of 45 for conventional GPU workflows, 64 for measured and scheduled workloads, 81 for efficient models combined with shared accelerator use, and 93 for an integrated energy-conscious laboratory. These values are illustrative rather than measured energy savings, but they demonstrate an important progression. Measurement creates visibility, scheduling and utilization improve infrastructure efficiency, model right-sizing reduces unnecessary computational demand, and governance converts isolated technical practices into repeatable laboratory standards. Sustainable computing therefore emerges from coordination across the full research lifecycle rather than from one optimization technique.

A further conclusion is that energy measurement must remain scientifically credible. Carbon-estimation and energy-monitoring tools have made sustainability assessment far more accessible, but contemporary validation evidence indicates that software estimates can contain substantial error relative to direct measurement. Laboratories should therefore match measurement precision to the purpose of the claim, clearly distinguish estimates from direct measurements, and report methodological assumptions. Performance should also be normalized against computational cost so that researchers do not unintentionally optimize for minimum electricity consumption at the expense of scientifically meaningful outcomes.

References

1. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. https://doi.org/10.18653/v1/P19-1355
2. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54–63. https://doi.org/10.1145/3381831
3. Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, *21*(248), 1–43.
4. Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying carbon emissions from machine learning. *arXiv preprint arXiv:1910.09700*.
5. Patterson, D., Gonzalez, J., Le, Q. V., Liang, C., Munguia, L.-M., Rothchild, D., So, D. R., Texier, M., & Dean, J. (2021). Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*.
6. Hooker, S. (2021). The hardware lottery. *Communications of the ACM*, *64*(12), 58–65. https://doi.org/10.1145/3467017
7. Deb, C., Eang, L. S., Yang, J., & Santamouris, M. (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*, *121*, 284–292. https://doi.org/10.1016/j.enbuild.2016.03.050
8. Gao, Y., Zhang, M., & Li, X. (2019). Energy-efficient computing and green data centers: A review. *Renewable and Sustainable Energy Reviews*, *109*, 1–15.

9. Beloglazov, A., Abawajy, J., & Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Generation Computer Systems, 28*(5), 755–768. https://doi.org/10.1016/j.future.2011.04.017
10. Buyya, R., Beloglazov, A., & Abawajy, J. (2010). Energy-efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges. *Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications*, 1–12.
11. Barroso, L. A., Clidaras, J., & Hölzle, U. (2018). *The Datacenter as a Computer: Designing Warehouse-Scale Machines* (3rd ed.). Morgan & Claypool.
12. Koomey, J. (2011). *Growth in Data Center Electricity Use 2005 to 2010*. Analytics Press.
13. Shehabi, A., Smith, S. J., Sartor, D. A., Brown, R. E., Herrlin, M., Koomey, J. G., Masanet, E., Horner, N., Azevedo, I. L., & Lintner, W. (2016). *United States Data Center Energy Usage Report*. Lawrence Berkeley National Laboratory.
14. Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, *367*(6481), 984–986. https://doi.org/10.1126/science.aba3758
15. Greenberg, A., Hamilton, J., Maltz, D. A., & Patel, P. (2009). The cost of a cloud: Research problems in data center networks. *ACM SIGCOMM Computer Communication Review*, *39*(1), 68–73. https://doi.org/10.1145/1496091.1496103
16. Horowitz, M. (2014). 1.1 Computing’s energy problem (and what we can do about it). *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 10–14. https://doi.org/10.1109/ISSCC.2014.6757323
17. Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *International Conference on Learning Representations (ICLR)*.
18. Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE, 105*(12), 2295–2329. https://doi.org/10.1109/JPROC.2017.2761740
19. Hooker, S. (2020). The hardware lottery: Why it pays to understand computing infrastructure. *Communications of the ACM*.
20. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI: A framework for energy-efficient artificial intelligence research. *Communications of the ACM, 63*(12), 54–63.