Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 9, S. 138-142, 5, 2025

Araştırma Makalesi

IJANSER

https://as-proceeding.com/index.php/ijanser
ISSN:2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 9, pp. 138-142, 5, 2025

Research Article

Artificial Intelligence and the Problem of Energy Consumption: Challenges and Opportunities for a Sustainable Future

Baki Gökgöz

Department of Computer Technologies, Gümüşhane University

bakigokgoz@gumushane.edu.tr

(Received: 30 April 2025, Accepted: 07 May 2025)

(4th International Conference on Scientific and Innovative Sutdies ICSIS 2025, April 29-30, 2025)

ATIF/REFERENCE: Gökgöz, B. (2025). Artificial Intelligence and the Problem of Energy Consumption: Challenges and Opportunities for a Sustainable Future. *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(5), 138-142.

Abstract- In recent years, artificial intelligence (AI) systems have rapidly advanced, offering transformative solutions that significantly enhance efficiency across various sectors such as healthcare, finance, transportation, and education. However, the growing computational demands required to train and deploy state-of-the-art AI models particularly those based on deep learning architectures—have raised serious concerns regarding their environmental sustainability. Training large-scale models often involves the consumption of vast amounts of electricity, resulting in substantial carbon emissions and a considerable ecological footprint. This paper provides a comprehensive examination of the energy consumption associated with AI systems, highlighting the underlying factors contributing to their environmental impact, including model complexity, dataset scale, and infrastructure design. It also surveys recent efforts aimed at mitigating these effects, including the development of energy-efficient algorithms, adoption of low-power hardware architectures, and implementation of carbon-aware computing strategies. Furthermore, the concept of "Green AI" is discussed as a paradigm shift towards sustainability-aware AI development, advocating for the inclusion of energy and environmental metrics as core evaluation criteria in AI research. The paper concludes by emphasizing the need for interdisciplinary collaboration and policy intervention to align the progress of AI technologies with global sustainability goals.

Keywords — Artificial Intelligence, Energy Consumption, Environmental Sustainability, Green AI, Carbon Emissions.

I. INTRODUCTION

Artificial Intelligence (AI) has become an indispensable part of modern technological development, offering significant breakthroughs in a wide range of domains such as healthcare diagnostics, autonomous vehicles, natural language processing (NLP), and financial forecasting. The rise of deep learning, in particular, has enabled machines to achieve near-human performance in various complex tasks, propelling AI into the forefront of scientific and industrial progress. However, the computational cost of training and deploying these large-scale AI models has triggered increasing concern regarding their environmental impact, specifically their substantial energy consumption and associated carbon footprint (Strubell et al., 2019).

Modern AI models are characterized by billions of parameters and require extensive training on massive datasets using high-performance computing infrastructure. For instance, training a state-of-the-art transformer-based language model such as GPT-3 is estimated to consume over 1,000 megawatt-hours (MWh) of electricity, leading to CO_2 emissions on the order of hundreds of tons (Patterson et al., 2021). These figures are not isolated examples; they reflect a growing trend across the industry where the pursuit of performance comes at a significant environmental cost. Moreover, AI model training often involves iterative tuning of hyperparameters, repeated experiments to benchmark results, and fine-tuning for specific tasks all of which further compound energy usage (Schwartz et al., 2020)

The environmental implications of such computational intensity are increasingly difficult to ignore, especially in the context of global climate change efforts and the urgent need to reduce greenhouse gas emissions. While the AI community has traditionally prioritized accuracy, efficiency, and scalability, recent discourse has called for a paradigm shift towards what has been termed "Green AI." Green AI promotes the development and deployment of AI systems that are not only powerful but also resource-efficient and environmentally conscious (Schwartz et al., 2020). This shift necessitates the inclusion of new performance metrics that consider energy efficiency (e.g., accuracy-per-watt) and transparency in reporting carbon emissions associated with AI research (Strubell et al., 2019).

On the hardware side, efforts are being made to mitigate energy use through the design of specialized processors such as Google's Tensor Processing Units (TPUs) and neuromorphic chips that emulate the energy-efficient architecture of the human brain (Horowitz, 2014). Neuromorphic computing, in particular, has shown promise in delivering significant reductions in power consumption by processing information through spiking neural networks, which are more biologically plausible and event driven. Complementary to hardware innovation, software-level improvements such as model pruning, quantization, knowledge distillation, and low-precision arithmetic also offer pathways toward reducing computational demands (Horowitz, 2014).

Despite these emerging strategies, the problem of energy consumption in AI remains largely under-addressed in both academic and industrial practice. There is a growing recognition that if left unchecked, the environmental footprint of AI could contradict the broader goals of sustainability and carbon neutrality set by nations and global organizations. Furthermore, this issue raises questions of equity and environmental justice, as the energy required for AI development is often sourced in regions with limited renewable infrastructure, disproportionately affecting vulnerable communities.

This paper aims to provide a comprehensive overview of the energy consumption challenges posed by modern AI systems, examine the contributing factors, and evaluate ongoing research and innovations aimed at mitigating these effects. It will also explore the policy landscape and advocate for cross-disciplinary collaboration to integrate sustainability as a core principle in AI development. By framing the discussion within the context of Green AI, this paper contributes to the growing body of literature that seeks to harmonize technological advancement with environmental responsibility.

II. RELATED WORK

The energy demands of artificial intelligence systems have garnered increasing attention from both academia and industry, leading to a growing body of research aimed at quantifying and mitigating the environmental impact of AI development. Early concerns were raised by Strubell, who provided one of the first comprehensive studies estimating the carbon footprint of training large-scale natural language processing (NLP) models (Strubell et al., 2019). Their analysis revealed that training a single transformer model with neural architecture search (NAS) could emit more than 626,000 pounds of CO_2 comparable to the lifetime emissions of five American cars bringing the issue of computational sustainability to the forefront of AI ethics and research.

Following this work, introduced the concept of Green AI, advocating for a shift in research priorities toward developing models that are not only accurate but also efficient in terms of computational and energy cost (Schwartz et al., 2019). They proposed the inclusion of energy and carbon metrics in the standard evaluation criteria of AI research, emphasizing that excessive computational scaling without consideration for environmental cost is unsustainable. This framework has since inspired numerous follow-up studies seeking to evaluate the trade-off between model size, performance, and energy usage.

On the hardware front, advancements in AI accelerators have played a central role in improving energy efficiency. Google's Tensor Processing Units (TPUs), designed for optimized matrix multiplication operations, have shown to provide significant performance-per-watt gains over traditional GPUs (Jouppi et al., 2017). Similarly, neuromorphic chips such as Intel's Loihi emulate the architecture of the human brain and use spiking neural networks to achieve event-driven computation with far lower power consumption (Roy et al., 2019) . These technologies highlight the importance of domain-specific architectures in tackling the power-hungry nature of modern deep learning models.

In addition to hardware innovations, several software-level methods have been proposed to reduce AI’s environmental footprint. Model pruning, quantization, and knowledge distillation are widely used techniques to compress neural networks without significantly degrading performance (Han et al., 2015). These methods aim to reduce the number of operations and memory access required during inference, thus lowering energy consumption. Furthermore, the emergence of efficient architectures like MobileNets (Howard et al., 2017) and EfficientNet (Tan & Le, 2019) demonstrate that performance can be achieved with minimal computational overhead through careful model design.

Recent studies have also focused on carbon-aware computing, where training workloads are scheduled based on the availability of renewable energy or the carbon intensity of electricity at a given time (Lacoste et al., 2019). Tools such as CodeCarbon and experiment trackers now allow researchers to monitor energy use and carbon emissions during experiments, promoting transparency and accountability.

Despite these efforts, a standardized methodology for reporting and benchmarking energy usage remains a challenge. Patterson et al. (2021) highlighted the lack of consistent reporting practices across institutions, calling for the adoption of more rigorous and reproducible protocols for energy and emissions tracking (Patterson et al., 2021). Their work also underscored the significant disparities in carbon efficiency between different data centers, noting that location and infrastructure matter as much as the algorithms themselves.

Collectively, these studies underscore the multifaceted nature of the AI energy consumption problem and the diverse strategies being pursued to address it. While progress has been made in both theoretical and practical aspects, further research is essential to integrate sustainability into the core design and deployment principles of AI systems.

III. DISCUSSION

The rapid advancement of AI technologies has undoubtedly yielded significant societal benefits, from enhancing diagnostic accuracy in healthcare to optimizing logistics and transportation systems. However, these achievements have come with a hidden cost: substantial energy consumption and a growing carbon footprint. As shown in prior sections, training large-scale models such as BERT, GPT-3, and PaLM often demands extensive computational resources, resulting in carbon emissions equivalent to those of dozens or even hundreds of automobiles (Strubell et al., 2019; Patterson et al., 2021).

This raises an important question: Can the pace of AI development be sustained without exacerbating global climate challenges?

The answer lies in a collective shift toward sustainable AI practices, but the path forward is not without complexity. For instance, Green AI (Schwartz et al., 2020) promotes the development of efficient models, yet this approach may conflict with commercial incentives that favor larger, more powerful models with higher benchmark scores. Moreover, transparency in reporting energy use and emissions is still not the norm in many research institutions and private companies (Patterson et al., 2021). Without standardization in environmental reporting metrics, it is difficult to assess and compare the true sustainability of AI systems.

Another concern is the global disparity in energy infrastructure. High-resource AI research is often concentrated in regions with access to efficient data centers and renewable energy, while developing regions may lack the infrastructure to adopt such technologies sustainably (Lacoste et al., 2019). This reinforces existing inequalities in access to cutting-edge AI and creates an ethical dilemma regarding who bears the environmental burden of global technological progress.

On the technical side, although model compression techniques like pruning and quantization (Han et al., 2015) show promise, they often require domain-specific expertise and can degrade model performance if not carefully implemented. Similarly, neuromorphic hardware architectures (Roy, Jaiswal, & Panda, 2019) remain largely experimental and are not yet widely adopted for mainstream applications.

Ultimately, the challenge is not simply technical but also institutional and cultural. A sustainable AI ecosystem will require interdisciplinary cooperation spanning computer science, environmental science, public policy, and ethics to design systems that are both performant and ecologically responsible. Policies that incentivize the use of green computing resources, mandate emissions reporting, or fund low-power AI research could catalyze this shift.

IV. CONCLUSION

The environmental cost of artificial intelligence, while often overlooked in mainstream discourse, represents one of the most pressing challenges facing the field today. As AI systems grow increasingly complex and resource-intensive, their energy consumption threatens to undermine global sustainability goals. The literature clearly demonstrates that current practices focused predominantly on performance maximization are not sustainable in the long term.

Encouragingly, a range of solutions is emerging, from efficient model architectures (Gökgöz et al., 2024; Howard et al., 2017; Tan & Le, 2019) and hardware accelerators to frameworks for carbon-aware computing (Lacoste et al., 2019). However, these efforts must be more widely adopted and supported by transparent reporting standards, open-source tools, and appropriate policy interventions. Researchers and practitioners alike must acknowledge that innovation should not come at the expense of planetary health.

Going forward, sustainability must become a first-class consideration in AI development. This will require a rethinking of what constitutes "progress" in the field moving beyond accuracy and throughput to include environmental impact as a core metric of success. Only by integrating these concerns into the heart of AI research can the field continue to evolve responsibly and equitably.

REFERENCES

Gökgöz, B., Aydın, T., & Gül, F. (2024). Optimizing Memristor-Based Synaptic Devices for Enhanced Energy Efficiency and Accuracy in Neuromorphic Machine Learning. *IEEE Access*, *12*, 154401–154417. https://doi.org/10.1109/ACCESS.2024.3482110

Han, S., Pool, J., Tran, J., & Dally, W. J. (2015). *Learning both Weights and Connections for Efficient Neural Networks*. http://arxiv.org/abs/1506.02626

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 (ISSCC)*, 10–14. https://doi.org/10.1109/ISSCC.2014.6757323

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. http://arxiv.org/abs/1704.04861

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). *Quantifying the Carbon Emissions of Machine Learning*. https://doi.org/1910.09700

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). *Carbon Emissions and Large Neural Network Training*. http://arxiv.org/abs/2104.10350

Roy, K., Jaiswal, A., & Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature*, *575*(7784), 607–617. https://doi.org/10.1038/s41586-019-1677-2

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2019). *Green AI*. http://arxiv.org/abs/1907.10597

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54–63.

https://doi.org/10.1145/3381831

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. http://arxiv.org/abs/1906.02243

Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. http://arxiv.org/abs/1905.11946