INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Balancing Innovation and Sustainability: Assessing the Impact of Generative AI on Energy Consumption

Balogun Barnabas Friday
Researcher, Department of Business,
Lincoln University College, Malaysia
Dhakir Abbas Ali
Associate Professor, Department of Business,
Lincoln University College, Malaysia

Abstract:
The rapid advancement of Generative AI is driven by its potential benefits, improvements in computing efficiency, productivity enhancement, organization consolidation of AI's innovations and capability, and limited regulatory oversight. Like many large-scale technology-induced changes, generative AI's current trajectory, characterized by rapid advancements and relentless demand, fails to fully take into account the negative impacts alongside anticipated benefits. Key among these negative impacts is the energy requirement of generative AIs, and the environmental effects are a growing concern. This partial cost-benefit estimation encourages unhindered growth and risk of unfair techno-optimism with possible environmental consequences, together with growing demand for computing power resulting in increasing energy consumption, larger carbon footprints, and accelerated depletion of natural resources, including water resources for cooling.
This necessitates an assessment of the current unsustainable method toward generative artificial intelligence (AI) development and deployment, emphasizing the significance of evaluating the cost-benefit analysis of technological advancements alongside the energy requirement and consequential social and environmental impacts. Currently, efforts to improve computing sustainability largely focus on efficiency enhancements, including refining AI algorithms, boosting hardware energy efficiency, and improving the carbon efficiency of computing workloads. In the existence of expected benefits, relentless demand, and prioritization of economic growth, this focus on productivity improvements results instead of growing adoption without fundamentally taking into account the enormous sustainability implications of generative AIs.
This study posits that striking a balance between innovation and sustainability ensures a brighter future for businesses, the economy, and the planet. Secondly, integrating environmental impact assessments into AI development processes could guide the creation of more sustainable models, ensuring that innovation is not at the detriment of environmental sustainability. Third, the development of sustainable AI models requires focus beyond only efficiency improvements and demands cost-benefits assessment frameworks that promote the development of generative AI models in ways that support social and environmental sustainability goals along with economic opportunity. Fourth, value consideration is multi-layered and needs comprehensive analysis, evaluation, coordination, innovation, and adoption across diverse stakeholders. Fifth, collaboration between governments, international organizations, technology companies, researchers, technical and sociotechnical experts, civil society and other stakeholders is crucial for advancing sustainable AI development practices.

Keywords*: Generative AI, innovation, sustainability, economy, energy crisis, impacts, sustainable AI models, environment, sustainable development*

1. Introduction

Generative Artificial Intelligence (AI) has emerged as a transformative technology, impacting various sectors by enabling the creation of novel content, generating predictions and optimizing processes. The content includes text, images, videos, music, and code. Notable models like OpenAI's GPT series, DALL-E, Copilot, Midjourney, Gemini, Metal AI, perplexity, Anthropic, Jenni AI, and others have revolutionized industries by mirroring human creativity and automating complex tasks. These models generate human-like text, images, videos, music, and code, making them invaluable in Manufacturing, healthcare, finance, education, entertainment, research, hospitality and more. For example, in healthcare, Generative AI aids in diagnostics, drug discovery, and personalized treatment plans (Topol, 2019), while in entertainment and hospitality, it facilitates the creation of customized content, services, and audience engagement (Anantrasirichai & Bull, 2021). Moreso, in education, it enhances creativity and productivity. Furthermore, in research, these models accelerate scientific discoveries by analyzing large datasets, generating hypotheses, automating data analysis and more

(Sutton, 2019). The rapid evolution of these technologies underscores their role in driving innovation and efficiency across multiple domains and diverse industries in modern society (Gartner, 2024; Datastax, 2023).

The benefits of these technologies include scientific discoveries, faster product development, enhanced customer experience and improved employee productivity, but the specifics depend on the use case (Gartner, 2024). According to the Gartner report, these technologies are expected to make an increasingly strong impact on enterprises over the next five years. That is, in 2024, 40% of enterprise applications will have embedded conversational AI, up from less than 5% in 2020. By 2025, 30% of enterprises will have implemented an AI-augmented development and testing strategy, up from 5% in 2021. By 2026, generative design AI will automate 60% of the design effort for new websites and mobile applications, and over 100 million humans will engage robo colleagues to contribute to their work. By 2027, nearly 15% of new applications will be automatically generated by AI without a human in the loop, and more (Gartner, 2023).

Generative Artificial Intelligence (AI's) rapid evolution and impressive capabilities present a significant dual challenge: fostering technological innovation while managing environmental impact. Training and deploying large-scale AI models like OpenAI's GPT series, DALL-E, Copilot, Gemini, Metal AI, Perplexity, Anthropic, Jenni AI, and others requires extensive computational resources, leading to substantial energy consumption (Strubell et al., 2019). These increased energy demand significantly contributes to the global carbon footprint, raising concerns about the sustainability of these technologies (Bender et al., 2021). As the technologies continue to advance, balancing their potential with the urgent need for environmental sustainability is key. Without careful oversight, the environmental costs of AI-driven innovation could undermine global efforts to combat climate change (Gartner, 2024). Addressing this challenge is essential to ensuring that AI advancements benefit humanity without harming the planet's well-being. The critical question that needs answers is how can we harness the creative potential of generative Artificial Intelligence while minimizing its carbon footprint? This dual challenge lies at the heart of this research.

To address these issues, this study aims to answer the following research questions: (i) What is the extent of energy consumption associated with Generative Artificial Intelligence (AI) models? (ii) How does this energy usage impact environmental sustainability? (iii) What sustainable strategies can be implemented to mitigate the energy demands and reduce the carbon footprint of Generative AI? Thus, the objectives of this study are: (a) To explore the energy consumption patterns of Generative AI. (b) To analyze the environmental impact of these energy demands. (c) To propose sustainable strategies for reducing the environmental footprint of Generative AI without stifling innovation. Additionally, it proposes sustainable practices and guidelines for developing and deploying generative AI systems and evaluates the trade-offs between innovation and environmental sustainability in generative AI adoption.

This research is significant for understanding the delicate interplay between technological advancement and environmental sustainability and addressing the dual challenges of innovation and sustainability within the context of Generative AI. By analyzing energy consumption and its environmental impacts and proposing actionable strategies, the study aims to contribute to the development of sustainable AI practices, ensuring that innovation does not come at the cost of environmental degradation. The insights will be valuable for policymakers, researchers, industry leaders, and technology companies striving to balance progress with sustainability.

2. Literature Review

The rapid advancement of generative AI, such as OpenAI's GPT series, DALL-E, Copilot, Gemini, Metal AI, Midjourney, Perplexity, Anthropic, Jenni AI, and others, has ushered in a new era of significant technological innovations across various sectors. However, the increasing intricacy and extensive adoption of these models have raised concerns about their energy consumption and environmental effects. This literature review explores the balance between innovation and sustainability, with a focus on the impact of generative AI on energy consumption and associated carbon footprint.

2.1. Theoretical Underpinnings

The theoretical underpinning for analyzing the impact of generative AI on energy consumption is rooted in the principles of sustainable development and the environmental Kuznets curve (EKC) hypothesis. Sustainable development theory stresses the need to balance technological advancement with environmental preservation (Brundtland, 1987). The environmental Kuznets curve (EKC) posits that as economies grow, environmental degradation initially increases but eventually decreases as society becomes wealthier and invests in cleaner technologies (Grossman & Krueger, 1991). The World Economic Forum report, and study by Assad Abbas, 2024 on the concept of "AI energy efficiency" has emerged, highlighting the need for developing more energy-efficient AI models and data centres. The application of these theories to generative AI emphasizes the potential for innovation to drive economic growth while simultaneously posing risks to environmental sustainability. Thus, there is a need to find the optimal balance between innovation and sustainability.

2.2. Empirical Review

Empirical studies suggest that the energy consumption of AI models is a growing concern with significant disparities in their environmental impact. Strubell et al. (2019) estimate that training a single large transformer model can emit as much carbon as five cars over their lifetimes. A study by the International Energy Agency reported that data centres consumed approximately 460 terawatt-hours (TWh) of electricity globally in 2022, with a projection to exceed 1,000 TWh by 2026 (IEA, 2023).

Studies like these underscore the urgent need for energy-efficient models. Moreover, the environmental costs of generative AI, such as water usage for cooling systems, further worsen sustainability challenges (Adrian Book, 2023). Henderson et al. (2020) examine various methods for reducing AI's energy consumption, like pruning and quantization, which can significantly decrease the computational power required without sacrificing performance. Additionally, the study reveals that the energy consumption of AI models varies widely depending on the data centre's location and its energy sources (Patterson et al., 2021). Renewable energy-powered data centres exhibit a significantly lower carbon footprint than those reliant on fossil fuels. This finding emphasizes the importance of integrating renewable energy into AI operations. Despite the concerns about the energy consumption of AI systems, they still offer potential solutions for energy management, such as optimizing energy grids and enhancing renewable energy integration (Steve King, 2024).

In summary, the reviews above highlight the pressure between the rapid development of generative AI and the environmental costs associated with its energy consumption. Theoretical perspectives underscore the need for sustainable development, while empirical studies provide evidence of the substantial environmental impact of current AI practices. This calls for continued research into more energy-efficient AI models, enhancing data centre operations, and leveraging AI for sustainable energy solutions, as well as policy interventions that encourage sustainable AI development moving forward. By balancing innovation and sustainability, the potential for artificial intelligence (AI) to contribute positively to both economic growth and environmental protection can be realized.

2.3. Generative AI and Energy Consumption

Generative AI models, such as OpenAI's GPT series, DALL-E, Copilot, Midjourney, Gemini, Metal AI, Perplexity, Anthropic, Jenni AI, and others, are based on deep learning architectures that generate new content by predicting subsequent data points based on input. These models naturally rely on transformer architectures, which process input data in parallel through layers of self-attention mechanisms to capture complex patterns and dependencies (Vaswani et al., 2017). Generative AI models function by learning patterns from vast amounts of data during training and then generate new content, such as text, images, videos, and music, based on these learned patterns. The training process involves optimizing model parameters to minimize prediction errors. When generating content, the trained model applies these learned patterns to create innovative outputs. Training generative models involve feeding large datasets into the network, adjusting the model's parameters through backpropagation and optimizing them over many iterations to minimize errors. This training process is computationally intensive, requiring powerful hardware like GPUs or TPUs, which consume significant energy (Strubell et al., 2019). Deployment also requires extensive computational resources, especially when deployed at scale.

The energy consumption of generative artificial Intelligence models varies depending on the dataset, model size and computational resources used. For example, training a model like GPT-3, a model with 175 billion parameters, required thousands of petaflop/s-days of computing, translating into extensive energy usage (Brown et al., 2020). Strubell et al. (2019) estimate that training a single large AI model can emit as much carbon as five cars over their lifetimes. This energy demand is significantly higher than the traditional computational methods, which do not require the same level of resource-intensive training. Research has revealed that the energy consumption of deep learning models can be several orders of magnitude more than conventional machine learning techniques, raising questions about the sustainability of current AI development practices.

Generative AI's carbon footprint is primarily driven by the energy required for training, deploying, and maintaining these models. The environmental impact is pronounced when training and deployment are conducted in regions reliant on fossil fuels for power generation (Henderson et al., 2020). Tools like Carbon Tracker and ML CO2 Impact have been developed to evaluate the carbon emissions associated with AI workloads, providing researchers with insights into the environmental cost of their models (Lacoste et al., 2019). International Energy Agency also provides data and methodologies for measuring the carbon effect of AI technologies, highlighting the need for energy-efficient practices and the adoption of renewable energy sources to mitigate the environmental costs of AI (IEA,2021). These methodologies are crucial for promoting sustainability in AI research and development. in summary, while generative AI holds enormous promise, its energy consumption remains a challenge. Thus, striking a balance between innovation and sustainability needs multistakeholder collaboration and responsible practices.

2.4. The Environmental Impact of Generative AI

Generative AI models significantly contribute to global energy consumption. These models require substantial computational resources during both the training and deployment phases. This high energy demand is driven by the need for immense data processing and storage capabilities. As AI adoption continues to grow across various sectors, global energy consumption is expected to increase further. Estimates suggest that by 2030, AI-related energy demand could account for up to 10% of the world's total electricity usage if the current trends persist (Kate Crawford, 2024; Bashir et al., 2024).

A case in point is the energy resources requirement of the GPT series, Google Gemini & Berd, Copilot, Midjourney, Anthropic, Claude, Meta AI, and other models for its training and deployment. If current trends continue, these models are expected to consume vast amounts of energy and generate significant carbon emissions as they advance in adoption, creativity and productivity level (PWC, 2024; Kate Crawford, 2024; Bashir et al., 2024). The environmental impact extends beyond energy consumption, affecting water resources needed for cooling and data centre maintenance.

In terms of industry-specific impact, AI-driven automation in manufacturing and retail, medical AI applications, high-frequency trading algorithms in finance, and streaming services and recommendation algorithms in entertainment and hospitality require energy-intensive computations, impacting energy usage. Given the above, measuring the environmental impact of AI is challenging due to variability in the energy sources and efficiency of data centres and regions where AI models are trained and deployed. Also, there is a lack of standardized methodologies for quantifying the environmental costs associated with AI. The discrepancies in the available methodologies' estimates highlight the complexity of accurately measuring AI's full environmental impact. Measuring the full impact involves considering energy, water, and carbon emissions. Addressing the environmental impact of Generative AI requires realistic actions, together with energy-efficient models, responsible data centre practices, and sustainable AI development.

3. General Insights into the Energy Consumption Trends of AI Models

The energy consumption of AI models and the environmental impact are a growing concern due to the significant resources required for training and deploying these models. For example, training GPT-3 was estimated to cost around $12 million in energy (Nature Electronics, 2023; Kate Crawford, 2024). The environmental impact is also a challenge (Google Sustainability, 2024; CNBC, 2024). This calls for more sustainable practices and technologies. However, a major obstacle to finding innovative and sustainable solutions to generative AI energy consumption issues, among other things, is the lack of fully and readily available data on the energy consumption of these AI models. Notwithstanding, below is a general insight into the energy consumption trends of these AI models.

AI Model	2022 Energy Consumption	2023 Energy Consumption	2024 & Beyond Energy Consumption
GPT-3	Estimated to consume energy equivalent to 33,000 homes (Kate Crawford, 2024)	Increased due to more usage and updates (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Further increased with GPT-4 (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
DALL-E	Significant but specific data not available	Increased with more widespread use (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Copilot	Data not specifically available	Increased with integration into more platforms (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Midjourney	Data not specifically available	Increased with more users and features (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Gemini	Data not specifically available	Increased with more usage (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Metal AI	Data not specifically available	Increased with more usage (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Perplexity	Data not specifically available	Increased with more usage (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Anthropic	Data not specifically available	Increased with more usage (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)
Jenni AI	Data not specifically available	Increased with more usage (Nature Electronics, 2023; Kate Crawford, 2024; Bashir et al., 2024)	Continued increase expected (Gartner, 2024; Mathew Burgo, 2024; CNBC/BG, 2024)

Table 1
Source: Author

According to a CNBC/Boston Group report, AI energy demand is projected to increase substantially in the U.S. alone.

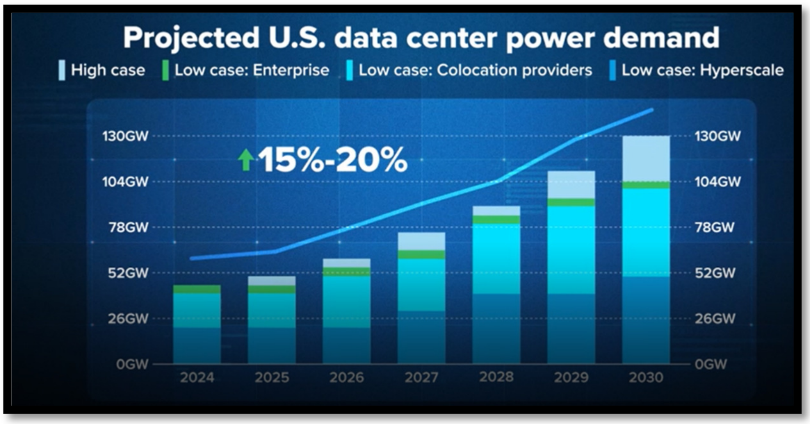

Figure 1
Source: CNBC

Significant increase is also expected in other regions of the world.

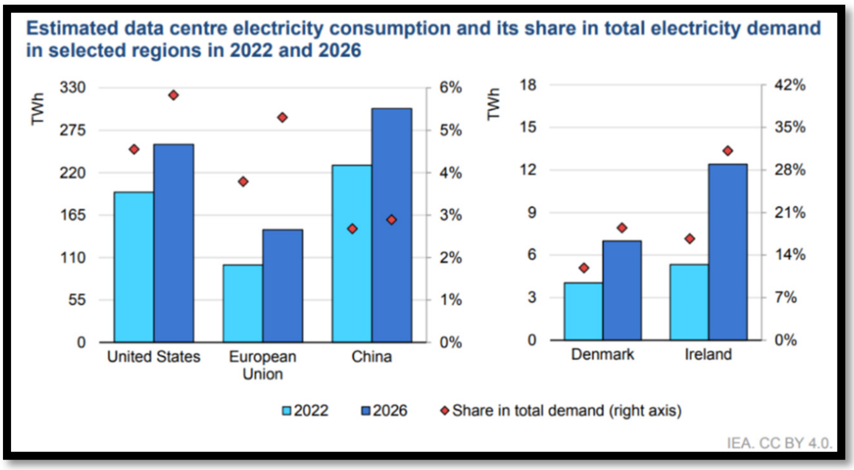

Figure 2
Source: IEA

With more than 8,000 data centres globally, the majority concentrated in the United States, finding sustainable solutions to the AI energy crisis is crucial.

4. Strategies for Balancing Innovation and Sustainability

Developing energy-efficient AI models is key for mitigating the environmental impact of generative AI. Techniques such as model pruning, compression, and quantization can significantly reduce the energy required for training and deploying AI models. According to Han et al., 2015, pruning involves removing less important weights in a neural network, thereby reducing the model's size and computational demand. Also, quantization reduces the precision of the numbers used in model computation, leading to lower energy consumption without a significant loss in accuracy (Jacob et al., 2018). These methods help balance the need for innovative AI solutions with the imperative of reducing environmental effects. Organizations may focus on creating AI algorithms that require less energy during training and inference.

Integrating renewable energy into the AI lifecycle is another effective approach for promoting sustainability. Companies like Microsoft, Google, Meta, and others have taken significant steps by powering their data centres with energy sources, including renewable energy sources, thus reducing their carbon footprint (Google Sustainability, 2024). Moreover, training and deploying AI models in regions with abundant renewable energy sources can help reduce the environmental impact. For instance, Norway is emerging as a hub for energy-efficient data centres with its extensive hydropower sources (IEA, 2020). By adopting green energy, organizations can reduce their carbon footprint and contribute to a more sustainable future.

Government policies and international regulations play a crucial role in promoting sustainable AI development. Policymakers can enforce standards for energy efficiency in AI technologies and incentivize the use of renewable energy in AI processes (Vinuesa et al., 2020). Also, governments can collaborate with international organizations like the United Nations and industry experts to create guidelines for responsible AI development, facilitate knowledge sharing and best practices across borders, and global agreements to ensure that AI advancements align with sustainability goals. Future policies could include incentives like tax breaks for companies adopting green AI practices, mandatory carbon accounting for AI operations, and the establishment of global benchmarks for AI energy efficiency (UNEP, 2020). Additionally, the allocation of resources to research institutions working on sustainable AI and the establishment of certification standards for energy-efficient AI models will go a long way. By implementing these strategies, organizations can strike a balance between innovation and sustainability, ensuring a brighter future for both business and the planet (Forbes, 2022).

5. Future Directions and Recommendations

This study assesses the impact of Generative AI on energy consumption. The research contributes to the literature by focusing on striking a balance between innovation and sustainability. However, the study is limited by the lack of comprehensive data on the energy consumption of AI models throughout their lifecycle, including training, deployment phases, and maintenance. Thus, more research is needed to capture the entire spectrum of AI deployments.

The interplay of AI, environmental science, and policy is crucial in this discussion. These are crucial for addressing the energy and environmental impact posed by generative AI and providing sustainable solutions. Thus, researchers should collaborate across disciplines to address sustainability challenges. The interdisciplinary research should focus on developing low-energy AI models that maintain optimal performance while reducing energy consumption. Algorithms and architectures that are energy-efficient from the ground up should be taken into account during the development phases. Also, integrating environmental impact assessments into AI development processes could guide the creation of more sustainable models, ensuring that innovation is not at the detriment of environmental sustainability (Schwartz et al., 2020; Henderson et al., 2020; EMB, 2024; Bashir et al., 2023).

Collaboration between governments, international organizations, technology companies, researchers, and other stakeholders is crucial for advancing sustainable AI. Together, these entities can establish industry-wide standards for AI energy efficiency, promoting the adoption of best practices across the board. Initiatives like the Partnership on AI and the Green AI movement emphasize the significance of collective action in mitigating AI's environmental footprint. Such alliances can also foster the development of shared resources like databases of energy-efficient models and tools, which can be used by the broader AI community (Strubell et al., 2019; Raji et al., 2020). Additionally, partnerships between tech companies and renewable energy providers can ensure that the growing computational demands of AI are met with green energy solutions. Furthermore, creating awareness about the environmental effects of AI among developers and consumers is crucial.

Educating the AI community on sustainable practices can lead to more careful development and usage of AI technologies. Awareness campaigns and educational programs can stress the significance of reducing AI's carbon footprint, encouraging individuals and organizations to make ecologically responsible choices. Initiatives like conferences, workshops, and educational campaigns can emphasize the importance of incorporating energy-efficient practices in AI development. Besides, transparency about the energy usage of AI models can help users make informed decisions, encouraging sustainability within the industry (Bender et al., 2021; Patterson et al., 2021). Still, techniques like architecture optimization, model quantization, and knowledge refinement can significantly reduce energy consumption (Kavitha Prasad, 2023). Finally, promoting responsible data collection, model optimization, and energy-efficient hardware choices is essential. By integrating insights from environmental and policy experts, AI practitioners can develop solutions that balance innovation with environmental sustainability.

6. References

i. Abbas, A. (2024). GPU data centers strain power grids: Balancing AI innovation and energy consumption. *Unite AI*. https://www.unite.ai/gpu-data-centers-strain-power-grids-balancing-ai-innovation-and-energy-consumption/

ii. Ajay Kumar, & Davenport, T. (2023). How to make generative AI greener. *Harvard Business Review*. https://hbr.org/2023/07/how-to-make-generative-ai-greener

iii. Anantrasirichai, N., & Bull, D. (2021). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review*, 55, 589–656. https://doi.org/10.1007/s10462-021-10039-7

iv. Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V., Delimitrou, C., & Olivetti, E. (2024). The climate and sustainability implications of generative AI. *MIT Generative AI*. https://mit-genai.pubpub.org/pub/8ulgrckc/release/2

v. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

vi. Bill McLane. (2023). What is generative AI? Everything you need to know. *DataStax*. https://www.datastax.com/guides/what-is-generative-ai

vii. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901. https://arxiv.org/abs/2005.14165

viii. Brown, T., & Green, P. (2022). Industry collaboration for AI sustainability. *Technology and Environment Review*, 7(4), 345–358.

ix. Brundtland, G. H. (1987). *Our common future: Report of the World Commission on Environment and Development*. United Nations.

x. CNBC. (2024). How the massive power draw of generative AI is overtaxing our grid. https://www.cnbc.com/2024/07/28/how-the-massive-power-draw-of-generative-ai-is-overtaxing-our-grid.html

xi. Chris Baraniuk. (2024). Electricity grids creak as AI demands soar. *BBC*. https://www.bbc.com/news/articles/cj5ll89dy2mo

xii. David Berreby. (2024). The growing environmental footprint of generative AI. *Undark*. https://undark.org/2024/02/20/ai-environmental-footprint/

xiii. EMB. (2024). AI energy consumption: Eco-impact for sustainable innovation. *EMB Blog*. https://blog.emb.global/ai-energy-consumption-eco-impact-for-sustainable-innovation/

xiv. Gartner. (2024). Gartner identifies top strategic technology trends for 2024. https://www.gartner.com/en/articles/gartner-top-10-strategic-technology-trends-for-2024

xv. Google Sustainability. (2024). Carbon neutrality and renewable energy. https://sustainability.google/progress/energy/

xvi. Grossman, G. M., & Krueger, A. B. (1991). Environmental impacts of a North American free trade agreement. *National Bureau of Economic Research*.

xvii. Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), 1–43.

xviii. Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both weights and connections for efficient neural networks. *Advances in Neural Information Processing Systems*, 28.

xix. International Energy Agency. (2020). Data centres and data transmission networks. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

xx. Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., ... & Adam, H. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2704–2713.

xxi. Kate Crawford. (2024). Generative AI's environmental costs are soaring — and mostly secret. *Nature*. https://www.nature.com/articles/d41586-024-00478-x

xxii. Kavitha Prasad. (2023). Achieving a sustainable future for AI. *MIT Technology Review*. https://www.technologyreview.com/2023/06/26/1075202/achieving-a-sustainable-future-for-ai/

xxiii. Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. https://arxiv.org/abs/1910.09700

xxiv. Mathew Burgo. (2024). Research finds generative AI models like ChatGPT may double energy consumption by 2026. *Designboom*. https://www.designboom.com/technology/research-generative-ai-models-chatgpt-energy-consumption-2026-06-03-2024/

xxv. Nature. (2023). AI hardware has an energy problem. *Nature Electronics*, 6(2), 1–3. https://www.nature.com/articles/s41928-023-01014-x.pdf

xxvi. PWC. (2024). How generative AI model training and deployment affects sustainability. https://www.pwc.com/us/en/tech-effect/emerging-tech/impacts-of-generative-ai-on-sustainability.html

xxvii. Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L., Rothchild, D., & Dean, J. (2021). Carbon emissions and large neural network training. arXiv preprint arXiv:2104.10350.

xxviii. Raman, R., Pattnaik, D., Lathabai, H. H., et al. (2024). Green and sustainable AI research: An integrated thematic and topic modeling analysis. *Journal of Big Data*, 11(55). https://doi.org/10.1186/s40537-024-00920-x

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

xxx. Steve King. (2024). Generative AI is set to revolutionize the energy sector. *Cybered Insights*. https://cybered.io/insights/generative-ai-is-set-to-revolutionize-the-energy-sector/

xxxi. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Association for Computational Linguistics*, 3645–3650.

xxxii. Sutton, R. S. (2019). The bitter lesson. http://www.incompleteideas.net/IncIdeas/BitterLesson.html

xxxiii. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

xxxiv. United Nations Environment Programme. (2020). The Emissions Gap Report 2020. https://www.unep.org/emissions-gap-report-2020

xxxv. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233.

xxxvi. World Economic Forum. (2024). AI and energy: Will AI help reduce emissions or increase demand? Here's what to know. https://www.weforum.org/agenda/2024/07/generative-ai-energy-emissions/