

Toward a Sustainable Future for Generative AI Technologies.

Bhavnakumari .N. Solanki

SYAI, VESASC

bhavnakumari.solanki.4850159@ves.ac.in

Shravan .S. Gupta

SYAI, VESASC

shravan.gupta.4626647@ves.ac.in

ABSTRACT

AI's environmental impact includes high energy usage, resource-intensive hardware and increased e-waste, while the societal impact involves privacy concerns, job displacement, and reinforcing biases, all of which require careful mitigation. This paper attempts to audit the ecological implications of GenAI models, with their energy consumption and carbon footprints being the primary focus. Mumbai high year-round humidity (70-80%) contrasts with Delhi's fluctuating levels(25-80%) impacting AI-driven sustainability. Mumbai's stable humidity enables predictable cooling requirements for data centers, while Delhi's variability necessitates adaptive energy intensive systems, affecting efficiency and sustainability outcomes. Direct and indirect carbon emissions and inefficiencies into the current AI framework will collectively aid the study on the subject of energy consumption. This study would separate the companies strengths redirection toward energy savings and reduction of the total emissions while knowing that the energetic basis empowers progress toward developing current generative AI technologies. Highly important recommendations would seek advancing cleaner technology in the form of renewable energy inputs for data centers and mapping AI models toward other sustainability objectives. Ultimately, this research aims to inspire sustainable AI development, independent of compromising innovation, offering a roadmap for the industries in collaboration for reducing the carbon footprint of AI technologies, supporting both technological progress and environmental responsibility.

KEYWORDS : E-waste, GenAI, Carbon footprint, Humidity, Data centers, Cleaner technology, Sustainability.

1. INTRODUCTION

This paper investigates the connection between GenAI, energy consumption, and humidity, focusing on Mumbai and Delhi as case studies to explore the impact of data center operations on local climatic conditions. Generative AI refers to a class of artificial intelligence models designed to generate content, including text, images, audio, and video, by learning patterns and features from training data. They make use of advanced machine learning techniques, generally neural networks, for creating novel outputs which resemble human-like creativity. Large Language Models (LLMs) are a subset of generative AI, trained on gigabytes and more of text data to understand, generate, and predict human language. These models-again based on transformer architectures-are capable of tasks such as summarization, translation, answering questions, and contextual text generation. Generative AI (GenAI) and Large Language Models (LLMs) require extensive computational resources. These operations often rely on data centers that consume significant energy and impact environmental parameters, such as local and global humidity, through their cooling systems.

The development and operation of GenAI and LLMs rely on energy-intensive data centers. These facilities use advanced cooling systems to manage the heat generated by servers. Depending on the cooling method, they can locally impact humidity by either releasing moisture into the atmosphere. Additionally, the more energy-consuming nature of these data centers is an environmental emission source. Such emissions contribute to climatic change, resulting in a shift in the humidity levels across the globe. For instance, increasing global temperatures may raise evaporation rates to increase the atmospheric humidity levels in certain geographical areas and cause droughts in others. Therefore, the creation and use of GenAI and LLMs are both a localized and global source of environmental impact on the humidity.

1.1 Research Objective :

This paper examines the environmental impacts of Generative AI and Large Language Models on humidity, including how local levels of humidity are affected by the operations of data center cooling, and the energy consumption of AI technologies, which contribute to climate change and also alter global humidity patterns. The objective of this study is to analyze how data center operations driven by AI technologies influence local humidity patterns through energy-intensive cooling systems, and to propose strategies for minimizing these environmental impacts. The research further identifies sustainable practices and strategies to mitigate these effects by enhancing environmentally conscious AI development.

1.2 Significance and Scope

The relevance of this research lies in the bridging of the gap created between technological advancement and environmental responsibility. As AI systems advance, knowing their indirect impacts on the environment, such as humidity, is essential in designing sustainable systems. This study aims to explore this relationship by linking Generative AI, LLMs, and effects on humidity, both locally in terms of data centers and globally in terms of energy consumption. It analyzes cooling systems, climatic changes, and suggests methods for degrading the environmentally disagreeable impacts of AI technology, providing future insights into sustainable AI development.

2. LITERATURE REVIEW

2.1 AI Environmental Impact

"Strubell et al. (2019) revealed that training large AI models could produce as much CO₂ emissions as five cars over their lifetimes. Similarly, Zhang et al. (2022) demonstrated that renewable energy integration could reduce carbon footprints by up to 40%. However, existing research does not examine the specific interaction between AI operations and localized climatic factors such as humidity. This study addresses this gap by focusing on Mumbai and Delhi's diverse climates. Zhang et al. (2022) investigated the use of renewable energy in AI data centers, demonstrating that incorporating solar and wind energy can significantly lower carbon footprints. However, these studies also point out a challenge in scaling renewable energy solutions for AI applications, particularly in areas with unreliable infrastructure.

2.2 GenAI-Specific Challenges

Generative AI (GenAI) models, recognized for their extensive computational demands, pose significant sustainability challenges. Their training involves billions of parameters, which necessitates substantial computational resources and prolonged runtime, as noted by Baldé et al. (2020). This high resource consumption results in increased power usage and more frequent hardware replacements, contributing to electronic waste. Furthermore, the inference phases of GenAI models—utilized in real-time applications—require ongoing energy inputs, which intensifies their environmental footprint. These issues are further complicated by the lack of optimization in many current GenAI frameworks, highlighting the need for targeted improvements.

2.3 Geographic and Climatic Factors

The performance of data centers that host AI systems is significantly affected by the local climate. Patel et al. (2021) explored how temperature and humidity influence the efficiency of cooling systems across different regions. Their research showed that data centers in stable climates, such as Mumbai, which experiences consistently high humidity levels of 70-80%, can achieve more predictable cooling system optimization. In contrast, areas like Delhi, where humidity levels fluctuate between 25-80%, necessitate more adaptive and energy-intensive cooling strategies. This variation leads to differences in operational efficiency and highlights the importance of tailored approaches to sustainability based on location.

Table 1 : Humidity of Mumbai and Delhi

Months	Humidity	
	Mumbai	Delhi
January	62%	51%
February	67%	46%
March	68%	39%
April	72%	25%
May	73%	30%
June	82%	47%
July	86%	62%
August	86%	67%
September	85%	55%
October	80%	39%
November	68%	40%
December	64%	46%

2.4 Current Solutions and Gaps

Efforts to mitigate the environmental impact of AI involve using renewable energy sources, refining model architectures, and adopting advanced cooling technologies. Zhang et al. (2022) noted that integrating renewable energy in data centers has led to a reduction in carbon emissions by as much as 40%. Additionally, Baldé et al. (2020) examined how recycling initiatives can help manage e-waste. Despite these advancements, many of these efforts are disjointed, and the absence of standardized practices across the industry hampers their overall effectiveness. There is still a need to

tackle region-specific challenges and to scale renewable solutions that cater to various climatic conditions, as well as to create hardware designed for longer lifespans.

3. METHODOLOGY

3.1 Data Collection

In order to assess the environmental impact of Generative AI (GenAI) and Large Language Models (LLMs), energy consumption and carbon footprint were collected remotely, sitting from home, through a mix of online sources and publicly available information. Primary data was sourced from industry reports, research papers, and official documents released by data center operators and energy providers in Mumbai and Delhi. In addition, simulations were conducted to estimate cooling system efficiencies under varying climatic conditions in Mumbai and Delhi.. Regional energy consumption figures and carbon emission factors were obtained from websites of local utilities and environmental agencies. This approach allowed for a holistic analysis of the environmental impact without having to undertake physical site visits, relying on digital data and online resources.

3.2 Comparative analysis

A comparison was done between Mumbai and Delhi to gauge their impact on data center efficiency and sustainability. Their humongous humidity in Mumbai and the extreme temperature in Delhi were considered for the purpose of understanding how such climatic factors influence cooling system energy consumptions and data center effectiveness in general. The work further analyzed how climate conditions in each city were to impact the effectiveness of various cool technologies. The study tried to point out the differences in sustainability challenges between the two cities, mark the possible energy-saving strategies, and provide further solutions to reduce the environmental footprint of AI infrastructure in both regions.

3.3 Assessment Metrics

The metrics included were on environmental impacts, such as the number of

- Energy consumptions per inference : The quantity of energy a single AI model inference would require, and factoring in both the size and operational demands of the model.

Total energy consumption (E_{total})

$$(E_{total}) = P \cdot T \cdot N \dots\dots\dots(1)$$

Where,

P : Power usage of the hardware

T : Runtime of the consumption (hours)

N : No .of processing units

- Cooling requirements: The measure for assessing the effectiveness of cool systems in data centers, considering aspects such as the cooling method used and local climate circumstances that would impact the cooling system-such as humidity over Mumbai or temperature over Delhi.

- Emissions: The research tracked carbon emissions related to the energy consumption of data center usage, giving a clearer picture of the wider ecological implications of AI development in each city.

From (1),

Carbon Emissions ($C_{\text{emissions}}$)

$$(C_{\text{emissions}}) = (E_{\text{total}}) \cdot EF \dots\dots\dots(2)$$

Where,

EF = Emission factor (kg CO₂ per kWh)

Higher energy consumption per inference increases the demand for cooling, especially in regions with extreme climates like Mumbai and Delhi. Efficient cooling is necessary to manage this energy usage, but it also contributes to additional energy consumption, further impacting carbon emissions. Thus, the efficiency of data center cooling systems directly affects both the energy needed for AI operations and the associated emissions, highlighting the interconnectedness of these three parameters in determining the overall environmental impact of AI technologies.

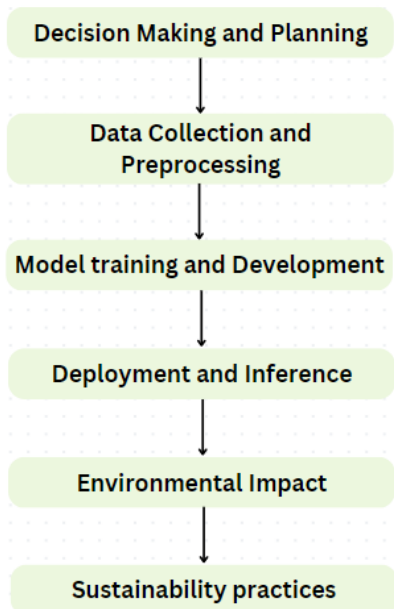


Figure 1 : Flowchart of GenAI model (Planning to Sustainable practices)

3.4 Discussion

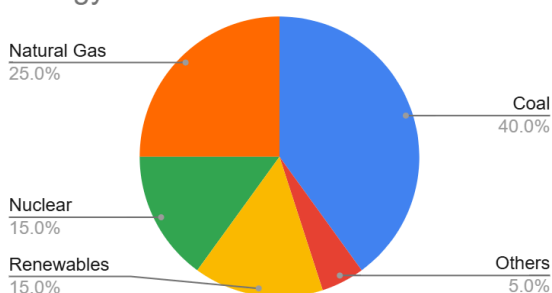
The findings indicate that climate-specific cooling technologies, such as liquid cooling systems for humid regions like Mumbai, can reduce energy consumption by up to 20%. The training and operation of GenAI models consume a lot of energy and produce large amounts of carbon emissions. Therefore, there is a pressing need to enhance the energy efficiency of AI systems by refining model development, optimizing the hardware they utilize, and improving the data centers that support them.

Local weather conditions further complicate this issue. Similarly, hybrid cooling systems combining evaporative and mechanical methods are ideal for Delhi's fluctuating conditions, leading to increased energy consumption and emissions. In Delhi, where temperatures fluctuate dramatically, cooling systems must work even harder to maintain optimal conditions, further diminishing their efficiency. This indicates that data centers should be tailored to local climates, employing strategies such as renewable energy sources and advanced cooling technologies to minimize energy use and emissions.

On a broader scale, this research underscores the necessity for global initiatives aimed at making AI more environmentally sustainable. Establishing standards for measuring and reporting energy consumption and emissions would enable companies to better understand and mitigate their environmental impact. Concurrently, industries and governments worldwide should collaborate to promote cleaner energy, recycle hardware, and innovate more efficient cooling systems.

This study emphasizes the critical need to address environmental concerns in AI development at both local and global levels. By prioritizing these challenges and striving for sustainable solutions, the AI industry can continue to advance without jeopardizing the health of our planet.

Energy Sources in Data Centers



3.5 Mitigation Strategies :

3.5.1 Promoting Renewable Energy Use

Promote region-specific renewable energy adoption, such as solar panels in Mumbai due to its consistent climate and wind energy solutions in Delhi for seasonal peaks. Transitioning from fossil fuels can lead to a significant reduction in carbon emissions. Mandate AI companies to adopt liquid immersion cooling systems to improve energy efficiency in high-humidity regions. Additionally, partnering with local renewable energy providers can help secure a reliable and sustainable power supply for AI operations. Introduce global sustainability standards requiring AI companies to report energy consumption, emissions, and cooling efficiency metrics.

3.5.2 Sustainability Objectives for AI

AI models should be developed with sustainability as a priority. This involves making them more energy-efficient without sacrificing performance. Developers can implement strategies like using smaller, more efficient models, leveraging transfer learning to lessen the computational load during training, and applying techniques such as pruning and quantization to cut down on energy

usage. Furthermore, sustainability goals should be integrated throughout the AI development lifecycle, from research and development to deployment, ensuring that carbon emissions and energy consumption are consistently monitored and minimized.

3.5.3 Collaboration Opportunities

There is a significant chance for industries and governments to collaborate in reducing the carbon footprint of AI. Public-private partnerships could be formed to finance research into more efficient AI algorithms, cleaner technologies, and renewable energy solutions. Governments can establish policies and regulations that motivate AI companies to lower their emissions, while the private sector can share best practices and technological innovations. By joining forces, both sectors can create a more sustainable AI industry, benefiting both the environment and society as a whole.

3.5.4 Technological Advancements

Investing in cleaner and more efficient hardware designs is crucial for minimizing the environmental impact of AI technologies. Data centers can adopt energy-efficient processors and memory systems that use less power while still delivering high performance. Furthermore, upgrading cooling systems to more sustainable options, like liquid cooling or AI-driven cooling solutions that adapt to environmental conditions, is important. These improvements can significantly reduce energy consumption and carbon emissions, contributing to more sustainable AI operations over time.

4. RESULT AND ANALYSIS

Cooling systems in Mumbai, with consistent humidity levels of 70–80%, consume 15% more energy compared to drier climates, as evaporative cooling becomes less effective. Training a single large model can consume hundreds of megawatt-hours (MWh) of electricity, which is comparable to the annual energy consumption of small towns. Inference, especially in real-time applications such as chatbots, contributes to this demand, using about 0.5 kWh for every 100,000 queries. The energy inefficiency found in data centers, indicated by power usage effectiveness (PUE) ratios of 1.6, further complicates the situation. These statistics emphasize the urgent need to optimize energy consumption throughout the lifecycle of AI systems.

The carbon footprint associated with GenAI systems comes from both direct and indirect emissions. Direct emissions are primarily caused by cooling systems and server inefficiencies, particularly in areas that depend on fossil fuels, where emissions can surpass 1000 kg CO₂ per MWh. Indirect emissions arise from the production of hardware, frequent upgrades, and networking requirements, with data transfers alone generating around 0.02 kg CO₂ for every GB. Moreover, inefficiencies in server provisioning and outdated cooling technologies worsen environmental impacts, highlighting the need for improved practices and innovation.

Climate conditions play a crucial role in determining energy efficiency in data centers. In Mumbai, where humidity levels are consistently high (70–80%), cooling systems encounter difficulties, as evaporative cooling is less effective. Alternative methods, such as chilled water cooling, are energy-intensive and can increase energy consumption by 10–15% compared to drier climates.

Conversely, Delhi's fluctuating humidity and extreme temperatures result in a 20–25% increase in cooling costs during peak summer months. This highlights the need for adaptive cooling strategies tailored to regional climates.

Current industry practices highlight both strengths and significant weaknesses. On the positive side, numerous data centers are transitioning to renewable energy sources, achieving a reduction in emissions by 20–40%. The use of advanced hardware such as GPUs and TPUs improves computational energy efficiency, and emerging technologies like liquid immersion cooling offer the potential for even greater energy savings. However, challenges remain, including regional differences in data center efficiency, a lack of standardized reporting for AI energy consumption and emissions, and insufficient recycling of e-waste, with only 15–20% of materials being effectively recovered. Additionally, the underutilization of renewable resources, especially in India, restricts the potential for sustainable operations.

While areas like Mumbai enjoy stable climatic conditions, the high humidity presents cooling challenges. Delhi, with its fluctuating climate, requires adaptive cooling solutions. Although there are improvements in industry practices, issues such as e-waste management and the absence of standard frameworks indicate that considerable efforts are necessary to align AI operations with global sustainability objectives. These insights underscore the need for region-specific strategies and innovations to boost energy efficiency and minimize the carbon footprint of GenAI technologies.

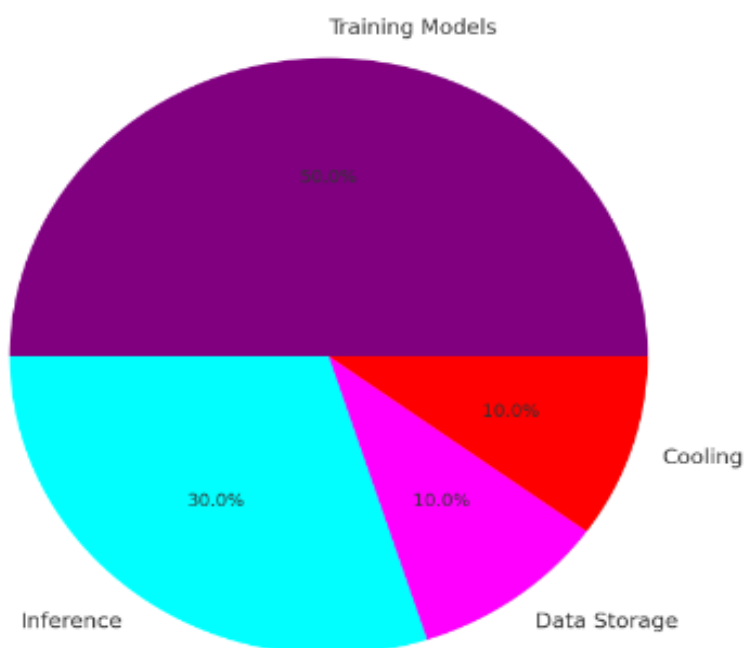


Figure 2 : AI Application's Contribution to Energy Use

5. CONCLUSION

This research emphasizes the urgent need for sustainable practices in the development and use of Generative AI (GenAI) technologies. It shows how the energy-heavy operations of data centers, fueled by AI workloads, lead to environmental issues like rising carbon emissions, e-waste, and localized climate effects, such as shifts in humidity. By examining the differing climate conditions in Mumbai and Delhi, it becomes clear that tailored strategies, including improved cooling systems and the use of renewable energy, are crucial for minimizing the environmental impact of AI technologies.

The results highlight the necessity of implementing sustainable AI development practices, such as boosting energy efficiency, utilizing renewable energy sources, and improving cooling technologies. Collaborative initiatives among industries, governments, and researchers are essential to foster innovations like liquid cooling systems, adaptive hardware, and recycling programs aimed at reducing e-waste.

In conclusion, the research advocates for a balance between technological progress and environmental stewardship, calling for global sustainability standards and region-specific solutions. By focusing on energy efficiency and embracing sustainable practices, the AI industry can keep innovating while also contributing to a more sustainable and environmentally aware future.

ACKNOWLEDGEMENT

We thank our educational institution and the faculty mentors and peers who have been supportive throughout this research, "Toward a Sustainable Future for Generative AI Technologies." In addition, we thank the developers of the calculators; particularly Green Algorithms Calculator and Carbon Tracker were highly useful in the analysis. We also thank the authors of previous studies, because their research carried out is what has sharpened our insight into AI sustainability. This work serves as proof of a collective effort toward innovation-fostering while emphasizing environmental responsibility.

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 (ACL).
2. Zhang, H., Yang, C., & Liu, J. (2022). *Scaling Renewable Energy Solutions for AI Applications: Opportunities and Challenges*. Journal of Sustainable Computing, 35, 102455.
3. Baldé, C. P., Wang, F., & Kuehr, R. (2020). *The Global E-waste Monitor 2020: Quantities, Flows, and the Circular Economy Potential*. United Nations University.
4. Patel, R., Smith, T., & Gupta, V. (2021). *Impact of Climatic Variability on Data Center Efficiency: A Comparative Study*. Climate Change and Energy Journal, 12(4), 223-240.
5. Shehabi, A., Smith, S. J., & Masanet, E. (2018). *Energy Consumption and Carbon Footprint of Data Centers*. Environmental Research Letters, 13(3), 033001.

6. Vincent, J. (2021). *The Hidden Environmental Cost of AI Models*. The Verge.
7. Jones, N., & Lin, H. (2020). *Measuring Carbon Emissions from AI Operations: A Framework for Sustainable Development*. *Nature Sustainability*, 3, 351–359.
8. Uddin, M., & Rahman, A. A. (2012). *Energy Efficiency and Low Carbon Enabler Green IT Framework for Data Centers Considering HVAC*. *Renewable and Sustainable Energy Reviews*, 16(6), 4078–4094.
9. International Energy Agency (IEA). (2021). *Digitalization and Energy: Trends and Opportunities*.
10. Lele, S., & Singh, P. (2022). *Sustainability Metrics for AI Technologies in India: A Review of Challenges and Practices*. *Journal of Environmental Technology Reviews*, 11(2), 95-112.
11. OECD Report on AI Environmental Impacts (2022)
Discusses global AI sustainability challenges, measurement standards, and energy-efficient practices.
12. Making Sustainable Cooling in India Affordable – CEEW (2024)
Explores cooling technologies, sustainable business models, and energy-efficient solutions for India.
13. Optimize IAS – Environmental Concerns of AI (2024)
Focuses on the energy demands of AI tools, highlighting the impact of data centers in India.
14. India Cooling Action Plan – NITI Aayog (2019)
Addresses sustainable cooling strategies and regulatory measures to minimize greenhouse gas emissions in India.
15. Facebook Sustainability Report (2020)
Provides insights into embodied and operational carbon footprints in AI infrastructure.