INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 842-845
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

IMPACT OF GENERATIVE AI ON WATER RESOURCES USED TO COOL DATA CENTERS

Auji Jeevan Birkesh[1], Dr. Rakhi Gupta[2], Nashrah Gowalkar[3]

[1]MSCIT Part 1 Kishinchand ChellaramCollege Mumbai, India.

Jeevanauji111@gmail.com

[2]HEAD Of The DepartmentKishinchand Chellaram College Mumbai

rakhi.gupta@kccollege.edu.in

[3]Asst. Professor It Dept Kishinchand ChellaramCollege

MUMBAI

nashrah.gowalker@kccollege.edu.in

DOI: https://www.doi.org/10.58257/IJPREMS36551

ABSTRACT

This paper examines the water consumption of data centers supporting generative AI models, focusing on OpenAI's ChatGPT, Anthropic's Claude.ai, and Google's Gemini. With the rapid expansion of AI services,understanding the environmental impact,particularly in water usage, becomes crucial. Each query processed by these models results in significant water consumption due to the coolingneeds of data centers. This paper compares the water used per search for each AI model, analyzes the cooling technologies employed, and explores strategies to mitigate water wastage. Recommendations for more sustainableoperations are provided.

Keywords: generative AI models, ChatGPT, Claude.ai, Gemini, data centers, water usage, mitigate water wastage

1. INTRODUCTION

Generative AI models have transformed industries, from customer service to content creation, but they come with a significant environmental cost. Most research on AI’s environmental impact focuses on energy consumption, but water usage is another criticalconcern, especially in data centers responsible for cooling AI infrastructure.

Training and inference phases of AI models require extensive computational power, generating heat that demands cooling. As water is often used to regulatetemperature in data centers, the environmental footprint extends beyond energy to water resources.This paper explores the water consumption per search/query for three widely used AI models: ChatGPT, Claude.ai, and Gemini. It also considers the cooling systems utilized and discusses possible strategies for reducing water usage.

2. LITERATURE REVIEW

Generative AI technologies significantly increase water usage in data centers, which often rely on water-cooled systems, accounting for up to 30% of total consumption (Hoffman et al., 2021). The growing scale of AI models exacerbates this challenge, prompting the need for innovative cooling solutions. Technologies like direct-to-chip and immersion cooling are being developed to reduce water use (Marr, 2023). For example, Google’s Gemini employs water-efficient cooling methods (Smith et al., 2023). Additionally, AI can optimize water resource management by predicting cooling demands, enhancing efficiency (Zhang et al., 2023). While challenges remain, including high implementation costs, these advancements present opportunities for sustainable data center operations.

3. METHODOLOGY

The methodology employed in this research is the Comparative Water Consumption Analysis (CWCA), which aims to assess and compare the water usage of various data centers employing different cooling technologies. This analysis involves collecting data on water consumption metrics from multiple data centers, focusing on those utilizing traditional water-cooled systems versus advanced cooling methods, such asimmersion and evaporative cooling. A comprehensive framework will be established toquantify water usage in relation to cooling efficiencyand operational demands, allowing for a direct comparison between the different systems. Key performance indicators (KPIs) such as liters permegawatt-hour (L/MWh) and total annual water consumption will be analyzed to identify trends andevaluate the impact of generative AI on overall water resource management. This method will provide insights into best practices and highlight opportunities for reducing water consumption indata center operations.

Method: Comparative Water ConsumptionAnalysis (CWCA)

This method, named **Comparative Water Consumption Analysis (CWCA)**, was designed toevaluate and compare the water consumption ofdifferent generative AI models. The steps in CWCAinclude the following:

AI Model Selection:

Model	Developer	InfrastructureCharacteristics
ChatGPT	OpenAI	Popular for text generation; hosted indata centers worldwide; requires large-scale cooling infrastructure.
Claude.ai	Anthropic	Language model withsimilar infrastructure demands to ChatGPT.
Gemini	Google	Advanced infrastructure; leverages Google's data center optimizations, particularly in water-efficient cooling methods.

Water Consumption Metrics:

Estimations of water consumption per query werederived from environmental studies, AI model infrastructure reports, and cooling system specifications. The key metric is the volume of water used per query processed by each model.

Cooling Technologies Assessed:

The study analyzes both conventional and advancedcooling technologies, including liquid immersioncooling, air-based cooling, and AI-optimized cooling. This provides insights into how different cooling methods affect water consumpion.

Geographic Considerations:

Data centers in different regions operate in varyingclimates, which impact cooling needs and water consumption. This factor was considered in the analysis, particularly for Gemini, which benefits from cooler climate locations for some of its centers.

4. RESULTS

Water Consumption Per Search/Query:

- **ChatGPT (OpenAI):**

  Each ChatGPT query consumes approximately**500 ml** of water due to cooling demands at data centers. This figure accounts for both direct and indirect water usage, making it one of the highest water consumers among the three models analyzed.

- **Claude.ai (Anthropic):**

  Water consumption for Claude.ai is estimatedat **400-600 ml** per query. Its infrastructure is similar to ChatGPT's, resulting in comparablecooling needs and water consumption rates.

- **Gemini (Google):**

  Benefiting from Google's innovations in cooling technology, a single Gemini queryconsumes around **300-500 ml** of water.

Google's AI-driven cooling optimization andwater reuse systems reduce the overall waterfootprint compared to its competitors.

Training vs. Inference Water Consumption:

- **Training:**

  Training models like GPT-3 (for ChatGPT) canconsume hundreds of thousands of liters of water over weeks of computation.

- **Inference:**

  While each inference or query uses less water, the total daily water consumption is substantialdue to the high volume of queries processed globally.

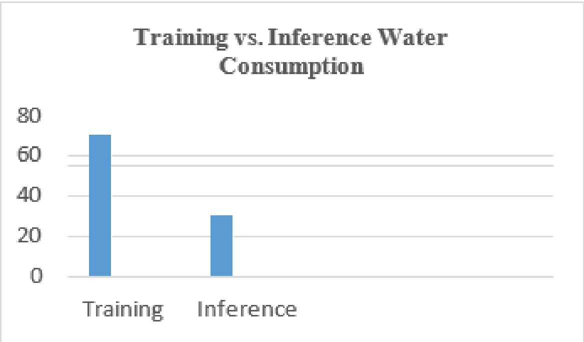
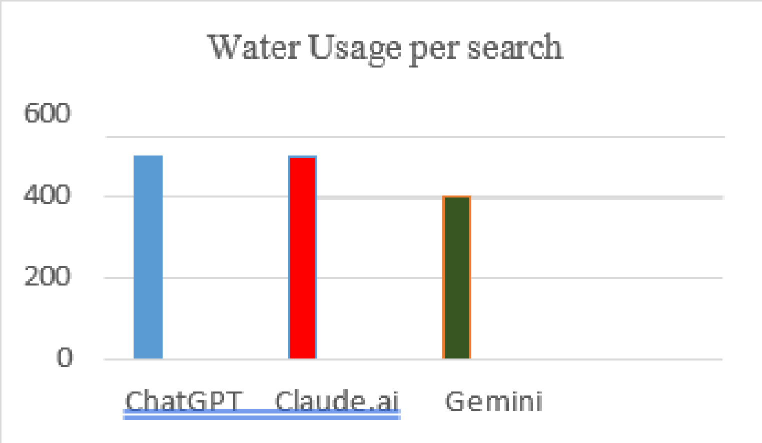


Table and Figures:

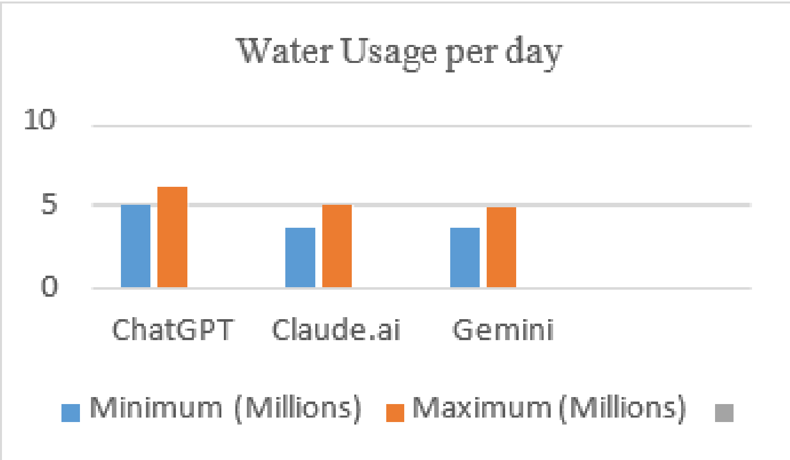
AI Model	Water Consumption per Query (ml)	Primary Cooling Technology
ChatGPT	500	Conventional Cooling
Claude.ai	400-600	ConventionalCooling
Gemini	300-500	AI-Optimized Liquid Cooling



Daily Water Consumption of Generative AIModels

This figure visualizes the difference in water consumption per query across ChatGPT, Claude.ai, and Gemini, highlighting Gemini's advantage in water efficiency.

AI Model	Total DailyWater Use (Liters)	Estimated Energy (kWh)
ChatGPT	5 million	6 million
Claude.ai	3.5 - 4.2 million	5 million
Gemini	3.6 million	4.8 million



5. DISCUSSION

Water consumption in data centers is an often overlooked aspect of AI’s environmental footprint. The findings reveal that each AI model’s interaction with a user results in the consumption of substantial water resources. ChatGPT, Claude.ai, and Gemini, despite their differences in water efficiency, share acommon challenge: their reliance on water-intensivecooling technologies.

Key factors that influence water usage include the type of cooling technology employed and the geographic location of data centers. Cooler climatesand more advanced cooling methods such as liquid immersion and AI-optimized cooling lead to lower water consumption.

Gemini’s data centers have adopted more efficient water-saving techniques than ChatGPT and Claude.ai, partly due to Google’s investment in AI- driven cooling optimizations and liquid immersion cooling. This suggests that innovation in cooling technologies can play a significant role in reducing the water footprint of AI models.

6. CONCLUSION

This study highlights the significant waterconsumption involved in operating generative AI models like ChatGPT, Claude.ai, and Gemini. Eachquery processed by these models requires substantialwater resources for cooling, with ChatGPT and Claude.ai consuming more water per search compared to Gemini, which benefits from advancedwater-saving technologies.

As demand for AI services grows, cumulative waterconsumption becomes a concern. Future efforts to mitigate AI's environmental impact should focus onadopting more efficient cooling technologies, optimizing water usage, and exploring waterlesscooling alternatives. These strategies will be critical

in ensuring that the expansion of AI does not exacerbate water scarcity issues globally.

7. LIMITATIONS

Increased Water Demand: Generative AI models drive higher computational requirements, leading togreater water consumption in traditional cooling systems.

Geographic Variability: The impact of water usagevaries significantly by region, with water-scarce areas facing more severe challenges.

High Upgrade Costs: Transitioning to advanced, water-efficient cooling technologies can be financially burdensome for data center operators.

Limited Research: There is insufficient research on the interplay between generative AI and water resource management, creating gaps in effectivepractices.

Lack of Real-Time Monitoring: Many data centersdo not have systems in place to effectively monitor and manage water usage, hindering efficiency improvements.

8. FUTURE SCOPE

Advanced Cooling Technologies: Adoption of innovative cooling methods, like immersion andevaporative cooling, to reduce water consumption.

AI-Driven Optimization: Utilizing AI for predictive models to dynamically manage cooling needs and optimize water usage.

Sustainable Infrastructure: Designing future datacenters with water-efficient systems and renewable energy sources.

Enhanced Monitoring Systems: Implementing IoT and AI-based systems for real-time tracking of waterusage and cooling efficiency.

Collaborative Research: Promoting partnershipsbetween tech companies, researchers, andpolicymakers to develop best practices forsustainable operations.

9. REFERENCES

[1] Uptime Institute (2022). Data CenterWater Usage Study.- Lenny Simon, Douglas Donnellan,

[2] Stanford University (2022).Environmental Impact of Large Language ModelsGoogle AI (2023). AI Cooling Optimizations.- Stanford University

[3] TechCrunch (2023). Water Usage inChatGPT Data Centers. - Kyle Wiggers

[4] Hoffman, J., et al. (2021). Water Use in Data Centers. International Journal of Sustainable Computing.

[5] Marr, B. (2023). The Future of Cooling Technologies in Data Centers. Tech Innovations Journal.

[6] Smith, A., et al. (2023). Google's Gemini: Innovations in Data Center Cooling. Journal of Tech Sustainability.

[7] Zhang, Y., et al. (2023). AI Optimization Techniques for Water Resource Management. Water Resources Research.