# Water usage of AI

Creating awareness

Nicolas Wim Landuyt

## Introduction

The demand for AI systems has risen rapidly over the last decade. Companies are investing massive amounts into being able to house these humongous systems [7]. The training and maintenance (inference is the official term) is done through the use of data centers. These data centers have started scaling up in size and power consumption, and thus also in water and electricity usage. Within the scope of this project, we will be considering the water usage of these systems.

## Scope

The idea to focus on sustainability within these ethical frameworks stems from their current lack of attention [16]. However, recently more research has been conducted on the ecological impact of AI, exploring how to make AI more sustainable and address its own environmental impact. Not only has research begun using AI to improve the sustainability of the world (using satellite images for better forecasting, estimating the effects of certain emissions, etc.), but it has also shifted towards making AI itself more sustainable.

Within the broad scope of sustainability, I have narrowed it down to the use of water in AI systems. This focus arises due to the imbalance in the lack of research, measurements, and policies surrounding water usage in AI systems and their immense water consumption. Like energy usage, water is one of the pillars underpinning the AI systems we see around us [5], yet it is almost never considered as a parameter to optimize. Most current efforts within the AI/ML community are focused on the energy consumption and carbon footprint of developing and operating AI systems, leaving the usage of the most valuable resource on the planet undocumented.

The decision also concerns water usage equity. “it is estimated that 20% of datacenter servers’ direct water footprint is sourced from moderately to highly stressed watersheds and 50% of servers are at least partially supplied by power plants in water stressed areas [15].” [18]. This highlights the growing problem of water demand, where the water used is sourced from stressed areas. This could

potentially lead to a resource injustice where water-rich regions consume more AI systems, while water-poor regions bear the cost.

## Water Cooling

Traditional HVAC systems (air-controlled cooling systems) are very costly to maintain. So the industry has turned to water cooling systems. These systems are more efficient and cost-effective [2]. Many different types of water cooling are used in these data centers (e.g., direct water cooling, indirect water cooling, etc.). All of these systems, however, use the evaporation of water to dissipate energy and cool the systems. This is the direct usage of water, however, it is not the only source of water usage. Many of these data centers require enormous amounts of energy in the form of electricity. This electricity is generated by power plants. These power plants require water to cool the systems. This is the indirect usage of water [13].

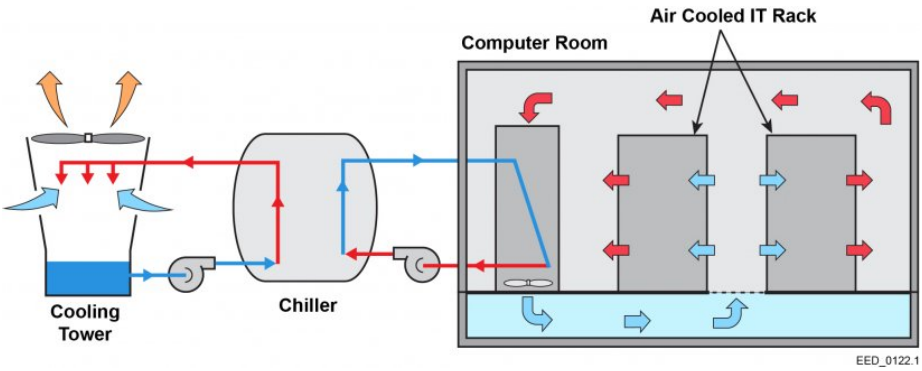


Figure 1: Simpliefied Datacenter Water Cooling [1]

The researchers at *UC Riverside* and *UT Arlington* have highlighted the significant water footprint of AI models, emphasizing the urgent need to address freshwater scarcity and underscored the importance of considering both water and carbon footprints for achieving sustainable AI development [10]. They found that for every 20-50 queries (interactions you have with chatGPT) there is an estimated of 500ml water used (1 drinking bottle of water). These findings are interesting, not only because they expose the scale of water usage and raise questions like how do these systems use water? Cooling water gets recycled, right? These are all questions I found myself and my peers asking and then looking up what was actually going on "under the hood" of these cooling systems exposed the reality of the water evaporation techniques. It is an interesting finding because it shows that AI systems do actually require a lot resources to preform inference.

In the development of AI systems, training and developing the model trump the inference cost. This means developing a model requires much more energy and water than running it. The architectures have been built, the neural nets, their weights and biases have all been learned and assigned, and the hyperparameter tuning has finished; running a model is then just a question of multiple simple calculations, running the input through the systems, and finally acquiring a result at a relatively low cost (in terms of computing power). However, more and more research has started making the case that the widespread use of AI has increased substantially so that the inference costs have not only increased so much that they shouldn't be neglected anymore [6], but also "they account for a large proportion of the data center compute cycles" [9].

"For example, many of Google's billion-user services are empowered by AI and their inference represents 60% of the AI infrastructure emissions [11]; Meta has expanded their infrastructure capacity by 2.5× to meet the ML inference demand [17]; AWS and NVIDIA have estimated that inference accounts for 90% of the ML workloads in HPC and cloud datacenters [3][8]."[9]

## Complications

The complexity of addressing water usage in AI and data centers is compounded by the lack of transparency and data. Currently, companies aren't required to disclose their water usage [13]. Many data centers don't even track their water consumption. Despite various companies committing to water sustainability goals, the absence of robust measures and reporting frameworks undermines these efforts. This lack of data infrastructure hinders the verification of water usage claims, leading to an increased focus on energy consumption and carbon footprints, which are easier to measure by simply consulting with grid operators about specific energy demands.

## Conclusion

Sustainability in AI engineering needs to look beyond mere environmental impacts like carbon footprints and consider the broader ethical, social, and technical landscapes. Simply focusing on carbon numbers might miss the larger ethical questions and the intrinsic value of maintaining a sustainable environment [12]. It's crucial that AI systems are evaluated not in isolation but as part of their wider ecological and socio-technical contexts, which highlights the need to understand the interconnectedness of technology and society [4]. Using a holistic approach with indicators can help in the practical implementation of sustainable AI, ensuring it not only meets environmental standards but also fairly distributes benefits and risks across society [14]. This perspective shows the importance of moving from isolated technical fixes to broader societal changes.

# References


[1] Cooling Water Efficiency Opportunities for Federal Data Centers.

[2] Liquid Cooling vs. Air Cooling in the Data Center.

[3] Amazon EC2 Update – Inf1 Instances with AWS Inferentia Chips for High Performance Cost-Effective Inferencing | AWS News Blog, December 2019. Section: Amazon EC2.

[4] L. Bolte. Conceptual Foundations of Sustainability. A Sustainability Perspective on Artificial Intelligence: Extended Abstract. In *CEUR Workshop Proc.*, volume 3637. CEUR-WS, 2023. Journal Abbreviation: CEUR Workshop Proc.

[5] A.Shaji George, A.S.Hovan George, and A.S.Gabrio Martin. The Environmental Impact of AI: A Case Study of Water Consumption by Chat GPT. April 2023. Publisher: [object Object].

[6] M. Hessenthaler, E. Strubell, D. Hovy, and A. Lauscher. Bridging Fairness and Environmental Sustainability in Natural Language Processing. In *Proc. Conf. Empir. Methods Nat. Lang. Process., EMNLP*, pages 7817–7836. Association for Computational Linguistics (ACL), 2022. Journal Abbreviation: Proc. Conf. Empir. Methods Nat. Lang. Process., EMNLP.

[7] E.A. Isaev, V.V. Kornilov, and A.A. Grigoriev. Data Center Efficiency Model: A New Approach and the Role of Artificial Intelligence. *Mathematical Biology and Bioinformatics*, 18(1):215–227, June 2023.

[8] George Leopold. AWS to Offer Nvidia's T4 GPUs for AI Inferencing, March 2019. awstooffer.

[9] B. Li, S. Samsi, V. Gadepally, and D. Tiwari. Clover: Toward Sustainable AI with Carbon-Aware Machine Learning Inference Service. In *Proc. Int. Conf. for High Perform. Compu., Netw., Storage Anal., SC*. Association for Computing Machinery, Inc, 2023. Journal Abbreviation: Proc. Int. Conf. for High Perform. Compu., Netw., Storage Anal., SC.

[10] Pengfei Li, Jianyi Yang, Mohammad A. Islam, and Shaolei Ren. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models, October 2023. arXiv:2304.03271 [cs].

[11] David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink, April 2022. arXiv:2204.05149 [cs].

[12] C. Richie. Environmentally sustainable development and use of artificial intelligence in health care. *Bioethics*, 36(5):547–555, 2022. Publisher: John Wiley and Sons Inc.

[13] Bora Ristic, Kaveh Madani, and Zen Makuch. The Water Footprint of Data Centers. *Sustainability*, 7(8):11260–11284, August 2015. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[14] F. Rohde, J. Wagner, A. Meyer, P. Reinhard, M. Voss, U. Petschow, and A. Mollen. Broadening the perspective for sustainable artificial intelligence: sustainability criteria and indicators for Artificial Intelligence systems. *Current Opinion in Environmental Sustainability*, 66, 2024. Publisher: Elsevier B.V.

[15] Md Abu Bakar Siddik, Arman Shehabi, and Landon Marston. The environmental footprint of data centers in the United States. *Environmental Research Letters*, 16(6):064017, May 2021. Publisher: IOP Publishing.

[16] Aimee van Wynsberghe. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3):213–218, August 2021. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer International Publishing.

[17] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga Behram, James Huang, Charles Bai, Michael Gschwind, Anurag Gupta, Myle Ott, Anastasia Melnikov, Salvatore Candido, David Brooks, Geeta Chauhan, Benjamin Lee, Hsien-Hsin S. Lee, Bugra Akyildiz, Maximilian Balandat, Joe Spisak, Ravi Jain, Mike Rabbat, and Kim Hazelwood. Sustainable AI: Environmental Implications, Challenges and Opportunities, January 2022. arXiv:2111.00364 [cs].

[18] D. Zhao, N.C. Frey, J. McDonald, M. Hubbell, D. Bestor, M. Jones, A. Prout, V. Gadepally, and S. Samsi. A Green(er) World for A.I. In *Proc. - IEEE Int. Parallel Distrib. Process. Symp. Workshops, IPDPSW*, pages 742–750. Institute of Electrical and Electronics Engineers Inc., 2022. Journal Abbreviation: Proc. - IEEE Int. Parallel Distrib. Process. Symp. Workshops, IPDPSW.