NATIONAL INSTITUTE ON ARTIFICIAL INTELLIGENCE IN SOCIETY

http://csus.edu/ai

NIAIS Briefs Series *AI impact on Society*

Environmental Impact of Generative AI: Carbon and Water Footprint

Alexander M. Sidorkin, California State University Sacramento

Summary

Generative AI technologies, such as ChatGPT, have measurable environmental impacts primarily from electricity consumption and freshwater usage at data centers. An individual AI query emits roughly 4.3 grams of CO_2 and uses around 10 milliliters of freshwater. In comparison to common everyday tasks, **AI's carbon footprint is small, significantly lower than driving or showering but higher than simple digital activities like web browsing**.

At scale, however, generative AI contributes notably to global energy and water demands. AI data centers consume tens of terawatt-hours (TWh) of electricity annually, with major companies reporting rapid increases in energy use due to expanding AI capabilities. Freshwater use at these centers is similarly substantial, reaching billions of gallons annually.

Though concerns regarding AI's environmental impacts are supported by data, significant mitigation is achievable through energy-efficient designs, renewable energy sourcing, and enhanced operational transparency.

Carbon Footprint and Freshwater Use per AI Interaction

Each interaction with a large language model (LLM) like ChatGPT involves energy-intensive computations. A single query to ChatGPT emits approximately 4.3 grams of CO_2 (Patterson et al., 2021), significantly higher than simple digital searches (around 0.2 grams per query) (Strubell et al., 2019). Energy per GPT-4 query ranges from 0.001 to 0.01 kWh, substantially more than typical web searches (Patterson et al., 2021).

Water usage arises primarily from cooling systems in data centers. GPT-3 queries, for instance, consume roughly 500 milliliters of water per 10–50 interactions (Li et al., 2023). Another estimation indicates about 2 liters of water are needed per kWh consumed by data centers (Mytton, 2021), translating to around 10 milliliters per AI interaction.

Comparison with Everyday Activities

Carbon Footprint

Activity (typical use)	CO_2 Emissions (grams)	Comparison to AI Queries
Generative AI Query	~4.3	Baseline
Driving a gasoline car (1 mile)	~404	Significantly higher than AI queries
Hot shower (10 minutes)	~2,000	Substantially higher than AI queries
Clothes dryer (1 load, ~1 hour)	~1,000	Substantially higher than AI queries
Electric oven (1 hour)	~675	Substantially higher than AI queries
Boiling water in kettle (~1 liter)	~50–70	Significantly higher than AI queries
Microwave (5 minutes)	~40–50	Significantly higher than AI queries
Video Conferencing (per minute)	2.5–16.7	Higher or comparable to AI queries
Streaming Video (per minute)	0.6–1.7	Slightly lower or comparable to AI queries
Online Gaming (per minute)	0.3–1.0	Lower or comparable to AI queries
Social Media (per minute)	0.3–0.5	Lower than AI queries
Web Browsing (per minute)	0.1–0.2	Notably lower than AI queries
Email Usage (per email)	0.02–0.17	Significantly lower than AI queries

Freshwater Usage

Activity	Water Usage per Minute (milliliters)	Comparison to AI Queries
Generative AI Query	~10	Baseline
Video Conferencing	~10–40	Comparable or higher than AI queries
Streaming Video	~1–5	Lower than AI queries
Online Gaming	~2–6	Lower than AI queries
Social Media	~0.5–1	Lower than AI queries
Web Browsing	~0.1–0.5	Significantly lower than AI queries
Email Usage	~0.1–0.2 (per email)	Significantly lower than AI queries

Aggregate Energy and Water Use of Large-Scale AI Operations

AI's **aggregate** environmental impact is considerable:

- **Energy Demand:** Global data centers used around 460 TWh of electricity in 2022, with AI servers alone projected to reach 85–134 TWh annually by 2027 (Masanet et al., 2020). Google's operational emissions increased nearly 50% between 2019 and 2023 due to expanded AI use, and Microsoft reported a 13% increase linked directly to AI infrastructure growth (Google Sustainability Report, 2023; Microsoft Sustainability Report, 2023).

- **Water Consumption:** Google's data centers consumed about 5.6 billion gallons (21 billion liters) in 2022, a 20% increase attributed largely to AI workloads. Microsoft similarly reported consuming 1.7 billion gallons (6.4 billion liters), a 34% increase primarily due to AI-driven expansion (Google Sustainability Report, 2023; Microsoft Sustainability Report, 2023).

Assessing the Validity of Concerns

Environmental concerns regarding generative AI are well-founded based on observable data. However, exact impacts vary due to limited transparency and reliance on estimates. While some claims (e.g., each ChatGPT session equals "a bottle of water") come from conservative extrapolations, actual environmental footprints depend significantly on operational practices and energy sourcing. Advances in efficiency and renewable energy could significantly mitigate AI's environmental impact (Patterson et al., 2021).

Future

Generative AI systems have showcased striking efficiency leaps since their debut. From 2018 to 2023, computational efficiency in training advanced language models surged by roughly 30-40% annually, trimming carbon emissions and water use per inference (Kaplan et al., arXiv:2001.08361; ScienceDirect, 2023). These strides spring from innovations like sparse attention, knowledge distillation, and quantization, slashing resource demands without sacrificing output quality. Meanwhile, AI deployment fuels productivity gains that temper environmental tolls through sharper processes. For instance, AI-driven smart building systems cut energy use by 10-30% (Nature Communications, 2024), and logistics optimization trims transportation emissions by 10-20% (Springer, 2023). Beyond this, AI bolsters climate science, renewable energy tuning, and precision farming, directly aiding carbon cuts. Thus, the environmental ledger must weigh both upfront resource gulps and these ripple-effect savings, where AI's initial footprint could pay off in spades—optimizing resources, curbing waste, and streamlining energy across swathes of the economy.

References

Aslan, J., Mayers, K., Koomey, J. G., & France, C. (2018). Electricity Intensity of Internet Data Transmission. *Journal of Industrial Ecology*, 22(4), 785-798.

Google. (2023). *Environmental Report 2023*. Retrieved from https://sustainability.google/reports

International Energy Agency (IEA). (2020). *Data Centres and Data Transmission Networks*. Retrieved from https://www.iea.org/reports/data-centres-and-data-transmission-networks

Kaplan, J., et al., "Scaling Laws for Neural Language Models," *arXiv*:2001.08361, 2021.

Li, P., Bashir, M., & Ren, H. (2023). The Hidden Water Footprint of Generative AI Models. *Environmental Science & Technology*, 57(10), 4155-4165.

Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984-986.

Microsoft. (2023). *2023 Environmental Sustainability Report*. Retrieved from https://www.microsoft.com/sustainability/report

Mytton, D. (2021). Data centre water consumption. *Nature Sustainability*, 4(6), 428-430.

Nature Communications, "Potential of artificial intelligence in reducing energy and carbon emissions of commercial buildings," 2024.

Netflix. (2020). *Environmental Social Governance Report 2020*. Retrieved from https://ir.netflix.net/governance/esg/default.aspx

Obringer, R., Rachunok, B., Maia-Silva, D., Arbabzadeh, M., Nateghi, R., & Madani, K. (2021). The overlooked environmental footprint of increasing internet use. *Resources, Conservation & Recycling*, 167, 105389.

Patterson, D., Gonzalez, J., Le, Q. V., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). Carbon Emissions and Large Neural Network Training. *Communications of the ACM*, 64(4), 58-67.

ScienceDirect, "Trends in AI inference energy consumption," 2023.

Springer, "Artificial intelligence-based solutions for climate change," 2023.

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*, 3645-3650.