

Visualizing Freshwater Consumption of LLMs — Draft Report

Abstract

This project visualizes freshwater consumption associated with large language model (LLM) inference and training by mapping water use to a $10 \times 10 \times 10$ m pool (volume = $1,000 \text{ m}^3 = 1,000,000 \text{ L}$). Using benchmark estimates (see “How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference”) the visualization animates how quickly the pool would fill under different usage scenarios. The aim is to provide an intuitive, time-based visual metaphor that helps non-experts and peers grasp the water footprint of current AI workloads.

1. Introduction

Concern over the environmental footprint of AI has mostly focused on energy and carbon. Water is a less-visible but important resource impacted by data center cooling and other infrastructure. To make water use more tangible, a WebGL visualization that simulates filling a $10 \times 10 \times 10$ m pool was built which exposes controls for scenario parameters. The design goal is clarity: viewers should instantly see how quickly a familiar volume of water (a small pool) is consumed.

Goals for this draft:

- Describe the mapping from reported water-use metrics to visual fill rate.
- Provide working example scenarios and step-by-step calculations.
- Explain implementation and how to run the visualization locally.

2. Background and references

The project uses the arXiv paper “How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference” as its empirical starting point. That work benchmarks inference water use across models and datacenter configurations and highlights that water use depends strongly on cooling architecture, PUE/WUE (Power/Water Usage Effectiveness), and geography.

3. Visualization concept and data mapping

Visualization metaphor

- Pool dimensions: $10 \text{ m} \times 10 \text{ m} \times 10 \text{ m} \rightarrow$ pool volume: $V_{\text{pool}} = 10 \times 10 \times 10 = 1000 \text{ m}^3 = 1,000,000 \text{ L}$.
- We animate the water surface rising over time as water is “consumed” by inference/training operations.

What we know from our references

- Global AI water withdrawals (projection for 2027): 4.2–6.6 billion m^3/year (= 4.2–6.6 trillion liters/year). This is the most-cited global figure : [Li et al. 2023](#)
- Inference dominates lifecycle use: the reference paper reports ~70–90% of lifecycle energy/cost sits in inference, not training; we’ll use that as the split for water, with sensitivity bands.

- Of withdrawals, a large fraction is actually consumed (evaporated) in data-center cooling. Expert commentary puts it on the order of ~60–80% being lost to evaporation (vs. returned).
- Single-provider case study (GPT-4o) shows ~1.33–1.58 billion liters/year just for inference on one model.

The Math

Annual withdrawal figures, which refer to freshwater taken out of the ecosystem comes to about 4.2–6.6 trillion L / year. This calculates to

- Per Day:
 - Low: $4.2e12 / 365$ which is roughly equal to 11.5 billion L/day,
 - Per day High: $6.6e12 / 365$ which is roughly equal to 18.1 billion L/day
- Per second:
 - Low: $11.5e9 / 86,400$ which is roughly equal to 133,000 L/s
 - High: $18.1e9 / 86,400$ which is roughly equal to 209,000 L/s

To calculate consumed from the withdrawal figures, where consumed refers to freshwater that is permanently lost to the atmosphere, we do apply 60% to 80% share:

- Consumed Per Day:
 - From 11.5B L/day which is roughly equal to 6.9–9.2B L/day
 - From 18.1B L/day which is roughly equal to 10.9–14.5B L/day
- Consumed per second:
 - Low band: ~80k–106k L/s
 - High band: ~126k–167k L/s

Global totals (4.2–6.6 trillion L/yr) from [Li et al. 2023](#); per-day and per-second are direct conversions

Design affordances in the visualization

- The rate of filling up the pool is presented as a field where calculated figures can be input. Hardcoding the value into the visualization is not opted for, as the water consumption figures are ever growing due to an increase in the use of generative AI.
- This input field accepts numerical values from the user. The default is 209980 Liters / second (this includes both figures from inference as well as training)
- A human model is included in the visualization inside the pool, as this helps the viewer process the sheer scale of freshwater being utilized.
- The visualization shows realtime use of water, which further aids the viewer in processing the scale of things.

The Visualization: How it was made

The visualization is a compact, browser-based application designed to make the freshwater footprint of modern AI tangible: it runs entirely in a web page using standard web technologies (HTML, CSS and JavaScript) and leverages the user's GPU through WebGL to render and animate a realistic 3D pool that fills as water is consumed. The interface is intentionally simple sliders and input fields let viewers change parameters

like water level and rate of fill and immediately see the pool's water level respond, so non-technical audiences can experiment without needing to understand the underlying code.

- Under the hood, a small WebGL helper library and a few focused JavaScript modules handle scene setup, lighting, reflections and the shader-driven water surface.
- A lightweight OBJ loader brings in simple 3D assets (for scale, a human figure sits in the pool) and a polyfill ensures broader browser compatibility for higher-quality water effects.
- Because everything runs client-side, the demo is highly accessible: it can be viewed locally or hosted as a static page with no server or build tools required, which is ideal for classroom demonstrations and peer review.
- Visually, the project pairs an animated, time-scalable pool with clear numeric readouts and preset scenarios so users can compare per-query impacts, single training runs, and annual totals at a glance; for very large totals the interface offers aggregated counters or compressed-time playback so magnitude is perceivable without long waits.
- The design choices prioritize clarity and immediacy: realistic water and reflections make the metaphor compelling, while the minimal, well-documented codebase keeps the project easy to extend.

References and how they helped

The data for this visualization and calculation come primarily from [Li et al. 2023](#) and [Jegham et al. 2025](#) which presents the following facts

- DeepSeek-R1, DeepSeek-V3, o3, and GPT-4.5 exhibit substantially larger environmental footprints across all input sizes. DeepSeek-R1 consistently emits over 14 grams of carbon dioxide and consumes more than 150 milliliters of water per query [Jegham et al. 2025](#)
- Based on scaled inference volumes, GPT-4o's annual water consumption is projected to be between 1,334,991 kiloliters (kL) and 1,579,680 kL. This is just one model. There are many other LLMs available. This consumption refers to evaporated freshwater permanently removed from local ecosystems rather than recycled. GPT-4o alone is responsible for evaporating an amount of freshwater equivalent to the annual drinking needs of almost 1.2 million people.
- [Li et al. 2023](#). analyzed GPT-3's freshwater consumption, estimating over 5 million liters used during training and projecting that AI-related withdrawals could reach 6.6 trillion liters annually by 2027.

These facts, presented by the references helped conceptualize the amount of water, meant for life on earth, be utilized for AI. The figures on paper do not do justice of presenting the magnitude. Visualization like these help get the general viewers process these numbers.

Points to note

While the visualization is accurate with the figures, the impact this has on areas that are prone to drought and face water scarcity is not something that can be understood just from this visualization. South Asia, Southern Africa, Middle East and North Africa face severe water scarcity, and the impact freshwater consumption by AI has on these areas are hard to visualize through such visualization. [Jegham et al. 2025](#) cites AI inference and training consumes water that could have gone to 1.2 million people. To a city like New York, this might not matter much, but for cities across Tunisia, this might be life or death.

