UDC 0048.8:502.2

THE IMPACT OF AI ON THE ENVIRONMENT

Tsando E.L.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Yushkevich E.V. – Lecturer at the Department of Foreign Languages

Annotation. This paper provides a quantitative and technical analysis of the physical infrastructure supporting large-scale AI models like GPT-4. Rather than focusing on general productivity benefits, the study examines the critical surge in power density within data centers and predicts their emergence among the world's top five electricity consumers by 2026. Special attention is paid to the «water footprint» of computation in resource-stressed regions and the ecological degradation caused by the extraction of rare-earth minerals for GPU production. The author argues for a transition from corporate secrecy to mandatory transparency metrics and algorithmic optimization as fundamental requirements for environmental safety in the AI era.

Keywords: generative AI, environmental impact, data centers, energy consumption, water footprint, greenhouse gas emissions, sustainability, resource intensity, inference phase, hardware production.

Introduction. Amidst the rapid pace of technological progress, generative AI has emerged as a transformative force, rapidly transforming industrial sectors with its promises of increased productivity, groundbreaking scientific advancements, and creative innovations. Tools such as ChatGPT and DALL-E have made sophisticated computing accessible to a broader audience, enabling users to generate text, images, and code with remarkable ease.

However, beneath this surface of advancement lies a concerning reality: the environmental impact of generative AI is escalating at a concerning rate. Recent analyses from institutions like MIT, the United Nations Environment Programme (UNEP), and other experts highlight that the «gold rush» in generative AI is consuming vast quantities of electricity, water, and raw materials, thereby contributing to greenhouse gas emissions, water scarcity, and disruptions to ecosystems.

This article delves into the reasons behind the resource-intensive nature of generative AI, exploring its multifaceted impacts and the pressing need for sustainable practices.

Following a two-part framework akin to MIT's series, we will concentrate on the primary environmental challenges, laying the groundwork for future discussions on strategies for mitigation.

Main part. The computational foundation of generative AI – models comprising billions of parameters, such as OpenAI's GPT-4 – requires extraordinary resources from the outset. Training these models involves processing extensive datasets using complex algorithms, a process that can consume significant electrical power.

For example, estimates from researchers at Google and UC Berkeley indicate that training GPT-3 alone necessitated approximately 1,287 megawatt-hours – sufficient to power around 120 average U.S. households for an entire year – and resulted in the emission of about 552 tons of carbon dioxide.

This energy consumption is not a one-time occurrence; the energy footprint persists through deployment, inference (the process of using the model for queries), and continuous fine-tuning.

As emphasized by Elsa Olivetti, a professor at MIT's Department of Materials Science and Engineering, the environmental implications extend far beyond merely activating a computer. These effects propagate through global systems, spanning energy grids, water supplies, and material supply chains [1].

Central to these demands are data centers, expansive facilities that house the servers, storage, and networking equipment essential for powering AI. Although data centers are not new – the first one was established in 1945 at the University of Pennsylvania for the ENIAC computer – the rise of generative AI has significantly accelerated their growth.

Companies like Amazon operate over 100 data centers globally, each containing tens of thousands of servers. The key differentiator for generative AI workloads is their exceptionally high power density: a training cluster may consume seven to eight times more energy than typical computing tasks, according to Noman Bashir, a postdoctoral researcher at MIT's Computer Science and Artificial Intelligence Laboratory.

In 2022, data centers consumed 460 terawatt-hours of electricity, ranking them as the 11th largest electricity user among nations, positioned between Saudi Arabia and France. Projections are even more alarming: by 2026, this figure could surge to 1,050 terawatt-hours, potentially elevating data centers to the fifth largest electricity consumer, nestled between Japan and Russia.

This increase is primarily driven by generative AI's insatiable demand for computational resources. In North America alone, the power requirements of data centers doubled from 2,688 megawatts at the end of 2022 to 5,341 megawatts by the end of 2023. While not all data center activity is related to AI, this technology serves as a major contributor.

The rapid fluctuations in energy usage during AI training phases introduce additional complexities, often necessitating the use of diesel generators to stabilize power grids – a pollutive backup that worsens emissions.

Bashir notes that the pace of new data center construction is unsustainable, heavily reliant on fossil fuel-based power plants due to the inability of renewable energy infrastructure to keep pace. In the United States, data centers currently account for 4.4% of national energy consumption, while globally, they contribute approximately 1% of energy-related greenhouse gas emissions, a figure expected to rise significantly.

By 2035, the expansion of AI-driven data centers could result in an additional 0.4 to 1.6 gigatonnes of CO2 equivalent emissions, comparable to the annual output of several mid-sized countries [3].

Water consumption represents another significant environmental challenge. Data centers depend on chilled water to dissipate heat generated by high-performance hardware, with estimates indicating two liters of water are required per kilowatt-hour of energy consumed.

On a global scale, AI infrastructure could withdraw between 4.2 and 6.6 billion cubic meters of water by 2027 – quantities that rival the total usage of entire nations.

This situation is particularly problematic in water-scarce regions, where data centers often cluster for logistical reasons.

For instance, in Taiwan, water rationing in 2021 prioritized the semiconductor industry (crucial for AI hardware) over agriculture, highlighting competing demands.

Golestan Radwan of UNEP cautions that with a quarter of the global population lacking access to clean water, AI's substantial water consumption could exacerbate scarcity and disrupt ecosystems.

The physical presence of data centers in the «cloud» belies their tangible impacts: they strain municipal water supplies, disrupt local biodiversity through water diversion, and contribute to broader hydrological imbalances [4].

The hardware that fuels this growth also imposes indirect but significant environmental burdens. Graphics processing units (GPUs), which are essential for managing AI workloads, necessitate complex fabrication processes that consume more energy than standard CPUs.

Market data indicates that shipments of data center GPUs increased from 2.67 million in 2022 to 3.85 million in 2023, with even more pronounced growth anticipated in 2024.

The production of these chips demands rare earth elements and critical minerals such as lithium, silicon, gallium, boron, and graphite – resources frequently extracted through environmentally harmful mining practices.

These operations lead to deforestation, soil erosion, groundwater depletion, and contamination with toxic chemicals, adversely affecting human health and biodiversity.

The processing and transportation of these materials further exacerbate carbon footprints. Additionally, electronic waste generated from outdated hardware poses further risks, as data centers produce hazardous substances like mercury and lead that can leach into environments if not managed properly [2].

Beyond direct infrastructural challenges, the applications of generative AI introduce emergent risks. AI-driven biological design, for instance, functions as a «black box» with opaque processes, potentially introducing untraceable life forms into ecosystems – such as self-spreading viruses for pest control – that could inflict irreversible damage. In military contexts, AI could lower barriers to biological attacks, resulting in lasting ecological harm.

Even in civilian sectors, AI enhances efficiency in polluting industries: the oil and gas sector employs it to optimize exploration and identify new reserves, which could lower fossil fuel costs and encourage overconsumption. Autonomous vehicles powered by AI may shift preferences away from sustainable transportation options such as public transit or cycling, leading to increased emissions. Higher-order effects, including AI-generated misinformation that downplays climate change, could further undermine public support for environmental initiatives.

Marine ecosystems face new threats from innovative yet untested solutions like underwater data centers. Projects such as Microsoft's Natick aim to reduce cooling costs by submerging facilities, but the heat they release could raise local water temperatures, diminishing oxygen levels and disrupting nutrient cycles. This exacerbates climate-induced ocean warming, jeopardizing marine biodiversity and food webs.

The expansion of data centers also competes with other societal needs. Minerals required for AI hardware contend with those needed for renewable energy development, potentially hindering the green transition. In the U.S., companies like xAI are constructing gas-powered turbines for data centers, circumventing regulations and releasing thousands of tons of nitrogen oxides annually – pollutants detrimental to health and ecosystems. On a global scale, the energy demands of AI could redirect public funds away from social and environmental projects toward grid expansions, prolonging reliance on fossil fuels.

Conclusion. Despite these considerable challenges, the environmental trajectory of generative AI is not predetermined. As Olivetti and her colleagues argue in their 2024 paper, a comprehensive evaluation of costs versus benefits is crucial. UNEP's recommendations include standardizing impact metrics, requiring disclosures from companies, optimizing algorithms for energy efficiency, recycling resources, and integrating AI policies into broader environmental frameworks.

Transparency is vital: major providers often withhold energy data for proprietary models, impeding accurate assessments, unlike open-source alternatives.

In summary, while generative AI possesses significant potential for positive environmental impact – such as mapping methane emissions or optimizing resource utilization – its current trajectory remains unsustainable.

The technology's resource intensity, characterized by electricity-hungry data centers, water-dependent cooling systems, and mineral-heavy hardware, underscores the necessity for responsible development.

As the United Nations Environment Assembly prepared to address AI's impacts in December 2025, global stakeholders must prioritize sustainability.

By fostering regulations, promoting green innovations, and encouraging democratic oversight, we can leverage AI's benefits without compromising the health of our planet. The choice lies before humanity: to allow generative AI to serve as a tool for progress rather than a catalyst for environmental degradation.

References

1.Explained: Generative AI's environmental impact [Electronic resource]. – Mode of access: https://news.mit.edu/2025/explained-generative-ai-environmental-impact – Date of access: 25.02.2026

2.AI has an environmental problem. Here's what the world can do about that. [Electronic resource]. – Mode of access: https://www.unep.org/news-and-stories/story/ai-has-an-environemntal-problem-here-what-the-world-can-do-about-that. – Date of access: 26.02.2026.

3.The environmental impact of AI and how to mitigate it [Electronic resource]. – Mode of access https://www.pwc.be/en/news-publications/2025/responsible-ai-environemntal-impact.html. – Date of access: 27.02.2026.

4.Understanding the Environmental Impact of Artificial Intelligence [Electronic resource]. – Mode of access: https://www.snhu.edu/about-us /newsroom/stem/ai-environmental-impact. Date of access: 01.03.2026.