AI Environmental Impact Report

Elie Jerjees, Julia Holt & Michaela Bryant

April 2026

Table of Contents

Introduction

The rise of generative artificial intelligence (generative AI) has prompted students, faculty, and staff across higher education to carefully consider the environmental costs associated with the tools they use. This report seeks to address questions regarding the environmental impacts of three major uses of generative AI across college campuses - text, image, and video generation. While institutional accountability for the environmental impact of AI usage is not yet considered "best" practice, or quintessential to yearly institutional emissions inventories and analysis, growing environmental concerns regarding AI and its rapid expansion could lead to it being included in future institutional accountability commitments. This accountability is especially important with the rising demand for highly commodified natural resources like water, given that AI is a growing sector of energy consumption. Furthermore, with climate change exacerbating resource scarcity across the globe, the proportion at which data centers consume these resources requires careful attention as current usage and cooling procedures can cause strain among local communities where these centers are located.

While the main purpose of our environmental impact report is to inform Davidson College's students, faculty and staff of the broader energy and water usage implications of common uses of generative AI and provide considerations for future use, we hope that our brief report on the topic will also function to inspire readers to remain conscious of how small actions have broader community and global implications. In this report, we will detail the consumption of energy, water, and carbon emissions used per three query types to brief readers on the hidden cost of using generative AI. The report begins with a brief overview of data centers and inferences before transitioning to data on resource consumption by common query type.[1]

[1] The comparisons drawn in this report are limited by the availability of reliable sources for numerical quantification and data. It is important to note that information regarding generative

Data Centers

A common misconception about data centers is that they are solely for AI; however, it is important to recognize that they have been part of our built landscape for over a decade. A data center is “a physical facility that houses computing infrastructure like IT equipment,” meaning that each center can serve a different purpose, depending on its equipment and ownership (Spenrath). This separates their construction and existence from their utility. However, it is still true that many data centers do function to support the computational and energy demands of AI queries and that demand for these centers has skyrocketed (Spenrath). As more data centers are built, it is essential that we do not overgeneralize or villainize data centers more generally, but rather understand the differences between traditional data centers and “hyperscale” data centers, which power AI.

The most crucial difference is that AI servers consume vastly larger quantities of land, water, and electricity. They often require 10 times the power of a conventional data center, hundreds of acres of land, and millions of gallons of water (Gorey). This need for greater electricity and water is due to the computational needs of AI systems and additional necessary powering and cooling for memory storage. In effect, “this rapid proliferation of resource consumption-heavy infrastructure can put an enormous strain on local and regional resources—burdens that many host communities are not fully accounting for or prepared to meet” (Gorey). This strain is in part caused by high consumption needs but also by resource

---

AI’s emissions and consumption is often clouded by misinformation, non-specific quantification, rapidly changing methodology, and an overall lack of transparency from parent companies. Additionally, quantifying energy and water usage is deeply complex due to countless uncontrolled variables regarding model type, location of user, time of query, and more (O’Donnell). As authors of this report, we have worked carefully to ensure the sources and data used in this report are as precise, accurate, and current as possible, though we advise some reader discretion at the time of reading.

acquisition since data centers often tap into municipal water services, lowering the groundwater availability in the area. These location-specific details also mean that these centers can fluctuate in efficiency from one another dependent upon their construction and locality. To add, these implications not only affect the immediate local community but can also cause cascading impacts across other regions. For example, the water supply in Washington, D.C. can be affected by data centers in Northern Virginia if the water is sourced out of the Potomac River because the water is not replenished due to the evaporation that occurs during the cooling process (Gorey).

To put this into perspective, it can be helpful to understand how this relates to our local community in Mecklenburg County. Like many homes in North Carolina, Davidson College's campus is powered by the Duke Energy Corporation. Duke Energy is an investor-owned utility regulated by a state agency, the North Carolina Utilities Commission. Becoming a power monopoly across the whole region, Duke Energy is the provider for millions of customers, as many do not have the flexibility of choice for their energy provider. Because of this, it is recommended for Duke Energy customers to remain updated on the corporation's decision process and relevant updates, such as their carbon plan. Then, customers can remain knowledgeable about the corporation's climate commitments and provide input or pushback when the corporation does not meet public expectations. Relevantly, when a data center was proposed in Mooresville, community members and leaders pushed back on the proposal, effectively preventing the center's construction in the area. This pushback came amid community concerns for strain on the local electricity grid and possible price hikes for local energy consumers. Additionally, the pushback also came amid uncertainty of which company would purchase the center and how it would be utilized once it was built. This is an important example that emphasizes knowledge seeking practices to understand proposed changes to one's community to effectively evaluate the possible consequences those changes might have. This report similarly serves as a mode of informed consumer practices, specifically tied to generative AI usage.

Training

Training is often the first aspect considered when determining the environmental impact of AI. Training a model requires a large amount of energy and water because it involves running massive computations for long periods of time and keeping that computational hardware cooled. The cooling largely relies on water based systems, where water absorbs the heat from the servers and is either evaporated (open-loop system) or circulated with some loss (closed-loop system). In both systems, it is difficult to fully reuse the water.

It is estimated that “training GPT-3 alone consumed 1,287 megawatt-hours” (Bozkurt) and hundreds of thousands of liters of water. However, GPT-3 is now largely outdated, and newer models are significantly larger and more advanced. As MIT Technology Review notes, “more parameters generally means better answers but more energy required for each response.” This indicates that increases in model size are tied to higher energy demands. Although direct comparisons of training costs between newer models and older models remain unclear, larger models require more computational energy during development, suggesting model size can also drive up training energy demands.

When opening ChatGPT, Claude, Gemini, or any other tool, the model has already been trained. That process is done ahead of time to create the product that users interact with. This continues over time to produce newer models. While training absolutely impacts the energy usage that AI causes, it is important to understand that it should not be framed as an individual user’s power usage like queries are. User data may be used for training purposes later, but conversations are not actively training the models. **(AI tools provided by Davidson, like Gemini, do not use user data to train their models)**

What users interact with are inferences. Inferences can be understood as applying the AI's training, where the AI model interprets the user's input and generates a response based on its learned behaviors and patterns. In simple terms, it is the process of AI responding to a prompt.

So, training should be understood as a background context. It explains why AI systems require so much infrastructure and resources. However, when talking about what users can actually control, the focus should shift toward how we use these systems day-to-day.

Inferences

Researchers have identified that "80-90% of computing power for AI is used for the inferences" (O'Donnell). However, the journey of where exactly your inferences go after you hit send is quite a mystery. There is no way to determine which data center your request is processed at, how energy efficient their process is, or which energy sources are used to power that process. Since this information is guarded by AI companies, a common methodology for measuring AI impact is to utilize open-source models. Open source models provided by AI companies allow researchers to run models locally and measure the energy usage themselves. Then they scale their calculations by estimating the other parts of the process using the available information about AI data centers. To contextualize these estimates, we break down the estimations for a single request to a text-model, image-generation model, and video model.

Text-Models

Text models are the form of AI people encounter every day, tools like ChatGPT, Claude, or Gemini, where you type a question and receive a written response. Unlike a simple search engine, these models are conversational, adapting their tone and follow-up responses based on the context of your conversation.

The energy cost of a text query depends heavily on the size of the model being used. MIT estimated the energy demands for the smallest open-source model available, Llama 3.1 8B, finding that the 8-billion-parameter model required about “57 joules per response, or an estimated 114 joules when accounting for cooling, other computations, and other demands” (O’Donnell). The researchers tested the model on a variety of prompts, such as explaining complex topics or planning a travel itinerary. To put that in perspective, MIT notes that this is roughly the energy needed to **ride six feet on an e-bike or run a microwave for one-tenth of a second.**

However, there is a stark difference in the energy demand when using a larger model. The largest open-source model available is Llama 3.1 405B, which is about 50 times larger than the previous model with 405 billion parameters. This model is designed to compete with GPT-4 and Claude 3.5 sonnet (Meta AI), and on average requires 6,706 joules per request, **enough to carry a person about 400 feet on an e-bike or run the microwave for eight seconds.**

These MIT findings are broadly consistent with other research. The EcoLogits Calculator estimates that a GPT-4 prompt generating a short paragraph (roughly 75 words) emits 926 mgCO2eq and consumes about 9.92 mL of water–approximately two tablespoons (Rincé & Banse). Google, meanwhile, reports that the median Gemini text prompt uses 0.24 Wh (roughly 864 joules), along with 0.26 mL of water (about four drops) and 30 mgCO2eq. This lower figure likely reflects that Google may be using data from a smaller or more efficient model, though the company does not disclose the number of parameters for its models or the specific model analyzed.

Image Models

Unlike text models, where energy consumption is driven primarily by prompt complexity, the energy cost of image generation is determined by model size, image resolution, and image quality. MIT estimates that generating a standard-quality image (1024 x 1024 pixels) using the leading open-source image generator, Stable Diffusion 3 Medium, requires 2,282 joules total. Opting for a higher quality image, which increases that figure to about 4,402 joules. This is equivalent to about **250 feet on an e-bike, or around five and a half seconds running a microwave.**

While precise data on per-image CO2 emissions is not reported consistently, existing research shows that emissions scale directly with the intensity of the task, the same way it does with text. The ICT environmental impact study shows that generative tasks like images can be 10 to 100 times more intensive than simpler AI tasks, which implies a proportional increase in carbon emissions since it is a much higher energy demand. Given that water is used primarily for cooling, consumption scales alongside energy demand, suggesting image generation carries a proportionally higher water cost as well.

Video Models

Video generation is by far the most energy-intensive form of AI, yet it is also the hardest to measure precisely. Because proprietary video models are prohibitively expensive to run independently, open-source alternatives lag significantly behind in quality, making direct comparisons difficult. To work around this, MIT cites research by Sasha Luccioni, an AI and climate researcher at Hugging Face, who used the tool Code Carbon to measure energy consumption directly. She identifies that to generate a 5-second video at 16 frames per second requires around 3.4 million joules, more than 700 times the energy required to generate a high-quality image. This is equivalent to **riding 38 miles on an e-bike or running a microwave for over an hour**. Given that commercial models from major AI companies can produce longer, higher-resolution videos, the energy cost of those outputs would likely be staggering by comparison.

Carbon emissions and water usage for video generation are similarly difficult to quantify directly, but the extreme energy demands make clear that both are substantial. Since emissions and water consumption scale with energy use, video generation is the most carbon and water-intensive form of AI usage covered in this report.

Energy Saving

Since users interact mainly with AI through inferences, the biggest impact comes from how we actually use these tools.

One of the main takeaways is that not all prompts are equal. Simple, direct tasks use much less energy than ambiguous or complex ones. In some cases, generative tasks, like writing long texts or creating images, can be "up to 10-100 times more power hungry than basic Q&A or classification tasks." (Rottenberg) So the framing of prompts really matters.

The simplest way to reduce wasted energy is to plan before sending. If the prompt is vague, the model will likely give a broad or unhelpful answer, which then leads to follow-ups, prompt edits, or retries. Being specific upfront about the task, the desired length, and the intended output format will lead to a more accurate final response on the first attempt.
Another large factor is the type of task. Text is generally the least energy-intensive task, while image generation causes a lot more, and video generation is by far the most. If something can be solved with text, then that is the most efficient approach to it.

Finally, the simplest recommendation, not everything needs AI. Using an AI chatbot for simple searches or basic tasks can be inefficient compared to traditional tools. Many times, Google, a calculator, or even just brainstorming with a friend is sufficient. At the end of the day, saving energy with AI does not require avoiding AI completely. It's about using it intentionally: fewer prompts, clearer requests, and only using more complex tools when actually needed.

Work Cited

Bozkurt, Aras. “AI’s Thirst, AI’s Heat, AI’s Waste: Exposing the Hidden Environmental Impact of Every Artificial Intelligence Interaction.” *Open Praxis*, vol. 17, no. 4, 2025, pp. 638–647, https://doi.org/10.55982/openpraxis.17.4.1058.

Dauner, Maximilian, and Gudrun Socher. “Energy Costs of Communicating with AI.” *Frontiers in Communication*, vol. 10, 19 June 2025, https://doi.org/10.3389/fcomm.2025.1572947.

Gorey, Jon. “Data Drain: The Land and Water Impacts of the AI Boom - Lincoln Institute of Land Policy.” *Lincoln Institute of Land Policy*, 17 Oct. 2025, www.lincolninst.edu/publications/land-lines-magazine/articles/land-water-impacts-data-centers/.

Meta. *“Meta Llama 3.1.”* Meta AI, https://ai.meta.com/blog/meta-llama-3-1/. Accessed 19 Mar. 2026.

O’Donnell, James. “We Did the Math on AI’s Energy Footprint. Here’s the Story You Haven’t Heard.” *MIT Technology Review*, 20 May 2025, www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/?gad_source=1&gad_campaignid=20737314952&gbraid=0AAAADgO_mjS-IJxAa3Jo1HHWTyYy-JNB&gclid=Cj0KCQjw9-PNBhDfARIsABHN6-07b2ixCSMrDQBRyklfx8HPfQXKS3GlpDSZ2MS0SOMt16kG_1nPn4saAhF9EALw_wcB. Accessed 17 Mar. 2026.

Spenrath, Micalah. “Powering Texas’ Digital Economy: Data Centers and the Future of the Grid - Houston Advanced Research Center.” *Houston Advanced Research Center*, 21 May 2025, harcresearch.org/research/powering-texas-digital-economy-data-centers-and-the-future-of-the-grid/.

Rincé, Samuel, and Adrien Banse. "*EcoLogits: Evaluating the Environmental Impacts of Generative AI*." *Journal of Open Source Software*, vol. 10, no. 111, 2025, p. 7471. JOSS, https://doi.org/10.21105/joss.07471.

Rottenberg, François, et al. *The Environmental Impact of ICT in the Era of Data and Artificial Intelligence*. 2026, arxiv.org/abs/2601.06174.