Measure What Matters: Energy Economics of AI Image Generation

Understanding the carbon impact of viral AI image generation trends

Scope3

Generative AI image creation has recently surged in popularity. With a few simple prompts, anyone with access to the internet can create unique images or graphics to meet their needs or whims.

From creative personal projects to streamlined marketing operations, and everything in between, there's tremendous opportunity. However, these on-demand visuals come with a hidden cost: significant energy consumption. As this technology becomes more widespread, and online trends go viral, the carbon footprint expands accordingly.

Inspired by the recent action figure boom on LinkedIn, this analysis explores the various generative AI image models available today, how they work, their energy requirements, and their climate impact.

Which models consume the most resources? How does a viral trend amplify these effects? And what responsibility do we have to consider the environmental impact of AI as we use it more and more?

By quantifying these previously invisible costs, we hope to encourage more mindful approaches to AI deployment as the technology becomes increasingly woven into our lives.

Uncovering what matters in a rapidly evolving landscape

AI technology is evolving at such a rapid pace that measuring its environmental impact demands innovation just as fast. The challenge is compounded by the fact that several leading models, like GPT and Claude, are closed source — leaving the public with little visibility into their energy consumption or climate costs. This lack of transparency makes it difficult for users and organizations to make informed, responsible choices about how they use AI.

At Scope3, we've long seen the environmental toll of digital inefficiency. In the ad ecosystem, we developed the first methodology to measure carbon waste — revealing carbon as a powerful metric for identifying inefficiencies and driving smarter decisions. Now, we're applying these learnings to gen-AI, where many of the same patterns are emerging.

This moment presents a critical opportunity to expand our scope and ensure we don't replicate the same mistakes across the broader digital landscape.

As climate change accelerates, understanding the real-world resource demands of our digital tools is no longer optional. Using our methodology for measuring AI emissions, we analyzed the environmental impact of leading image generation models, including GPT-4o, DALL-E 3, and Stable Diffusion 3.5. Our findings reveal the relative energy costs of each model as well as the broader implications for cultural moments, like the action figure trend, that can cause spikes in energy usage.

Understanding AI image generation: techniques and trade-offs

Diffusion models

Diffusion has long been the backbone of many frontier image generation models, including Stable Diffusion and DALL-E. It is a fairly complicated model architecture and is computationally demanding, though the general process is relatively straightforward to understand:

1. It begins with a random, "white-noise" image and text prompt, similar to a polaroid developing.

2. The model refines the entire image step-by-step, using the prompt as guidance.

3. After a fixed number of steps, it outputs a final, coherent image.

While diffusion excels in producing detailed visuals, it is known for high energy consumption which grows even larger with increased image complexity or resolution.

The energy cost of diffusion models

The trade-off for the quality and detail of diffusion-generated images is their energy consumption. Diffusion models are significantly more energy-intensive than text-based LLMs.

For instance, generating a standard-quality image with DALL-E 3 uses more than nine times the energy required for a typical text response from ChatGPT (using GPT-4o and assuming 750 input and 250 output tokens).

This disparity highlights the energy burden of image synthesis compared to textual output, even within the same AI ecosystem.

GPT-4o: A new approach

OpenAI's GPT-4o marks a shift from traditional diffusion-based image generation. Though closed-source, there are enough public details available about GPT-4o for us to theorize about its architecture.

The "o" stands for omni, reflecting its multimodal design. Unlike diffusion models that generate entire images simultaneously, GPT-4o treats image generation more like text generation through a hybrid approach:

- **Auto-regressive generation:** Images are created patch by patch, with each section influenced by preceding ones - mimicking how LLMs generate text token by token.

- **Visual tokens:** Images are segmented into grids, with each section treated as a sequential token (left-to-right, top-to-bottom).

- **Localized diffusion:** While autoregressive in nature, each token still undergoes diffusion-like processing for local refinement.

With GPT-4o, users can select the quality of image output — low, medium, or high — giving them more control over performance and resource use. That is, if users know to specify.

The model's new hybrid approach enhances its ability to capture fine details with greater contextual awareness, but this also comes with increased computational demands.

While this method combines some elements of both text and diffusion models, its overall energy footprint is the greatest — surpassing the demands of purely text-based responses and even diffusion-generated images.

GPT-4o's computational and environmental cost

GPT-4o's novel image generation approach brings powerful capabilities but at a heightened environmental cost according to calculations made using our methodology:

- Generating a high-quality image with GPT-4o produces approximately 5.6 grams of CO_2e.
- For comparison, this is nearly double the emissions of creating an HD image with DALL-E 3.

high   medium   low

Prompt 1: Differences between image quality settings

These three images demonstrate GPT-4o's quality settings: a garden under starry skies in high, medium, and low quality.

All retain significant detail, but high quality achieves a more realistic look compared to the illustrative appearance of lower settings.

high   auto   medium

Prompt 2: Determining the default image quality

Test prompts showed 'auto' quality matched high-quality images.

Celebrating his recent retirement: Spurs coaching legend Gregg Popovich riding a merry-go-round with Duncan and Ginobili.

Measuring the emissions of image generation models

Knowing that different models use different techniques to generate images, and that even within the same model, output quality settings can vary dramatically, it's clear that not all image generation is created equal.

We looked at four widely used models and compared their image generation capabilities to estimate their carbon emissions, broken down by quality tier where applicable. The models we analyzed included GPT-4o (at high-, medium-, and low-quality), DALL-E 3 (standard and HD), Stable Diffusion XL 1.0, and Stable Diffusion 3.5.

Using Scope3's predictive modeling approach, we estimated the carbon emissions of each task. The results were striking: both the model selected and the requested image quality had a significant effect on energy consumption.

One standout finding was that GPT-4o's 'high quality' setting consumes a whopping 1,300% more energy than its low-quality option, driven by an increase in the number of visual tokens the model produces.
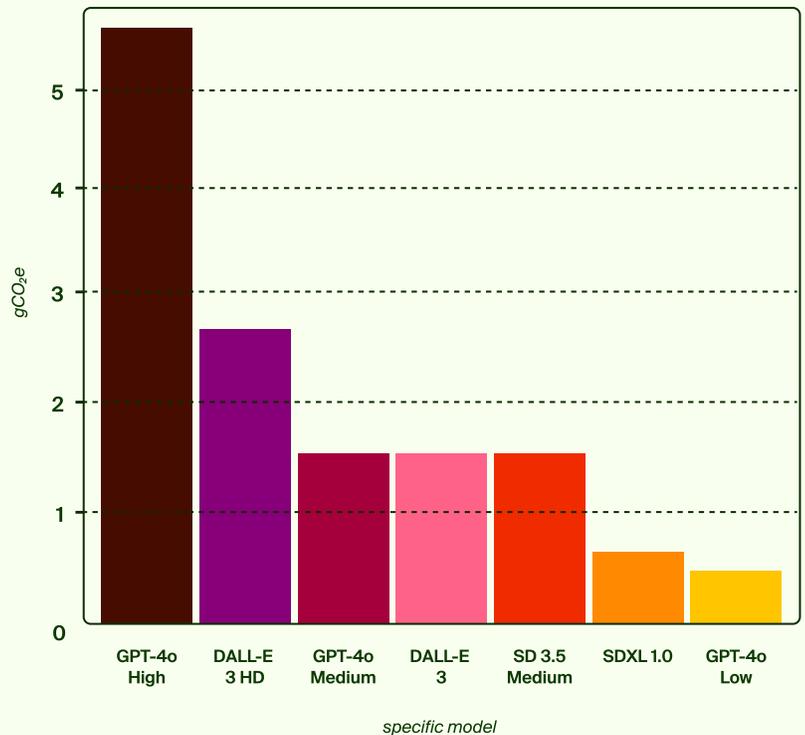
1,300%
more energy burned using GPT-4o's 'high quality' setting v. low

This highlights a key consideration for anyone using AI tools at scale: higher image quality may not always justify the environmental cost. Choosing lower-impact settings can dramatically reduce emissions while still delivering compelling visuals.

Estimated Relative Energy Use by Model

	Model	Emissions per Inference (gCO$_2$e)	Ratio to GPT-4o (text)
Image	GPT 1 (4o) - Low	0.40	2.1
	GPT 1 (4o) - Medium	1.50	8.0
	GPT 1 (4o) - High	5.63	30.3
	DALL-E 3 - Standard	1.40	7.4
	DALL-E 3 - HD	2.74	14.7
	SDXL 1.0	0.55	2.7
	SD 3.5 - Medium	1.37	7.0
Text	**GPT-4o**	0.19	1.0
	GPT-4.1	0.13	0.7
	GPT-4.1 mini	0.03	0.1
	GPT-4.1 nano	0.01	0.0

Carbon Impact of Gen AI Images

specific model (x-axis), *gCO$_2$e* (y-axis)

GPT-4o High, DALL-E 3 HD, GPT-4o Medium, DALL-E 3, SD 3.5 Medium, SDXL 1.0, GPT-4o Low

So, what's the carbon cost of a viral gen-AI image trends?
Let's look at AI 'action figures' to find out.

The 'action figure' trend on LinkedIn (and elsewhere) provides a real-world example of how quickly these costs can accumulate. Each high-quality figure created with GPT-4o generates about 5.6 grams of CO_2e. On its own, that's manageable. But when scaled, the impact is much bigger.

If we assume 1 in 5 active U.S. LinkedIn users participated, roughly 32 million people, the first round of image generation would produce 179.2 metric tons of CO_2e. Knowing it likely takes multiple attempts to get the perfect output, let's average it at two attempts, bringing the emissions output to 358.4 metric tons of CO_2e.

That means, in the month that the trend was traveling across LinkedIn, the CO_2e produced by these images was roughly the same energy usage of 1,000 households' monthly emissions.

While individual contributions seem small, viral trends dramatically magnify energy consumption. Switching to low- or medium-quality outputs can rein in the carbon cost of viral AI image generation trends.

Shaping the environmental impact of gen AI

This is just the beginning. As businesses and individuals rush to experiment with AI image generation, we're entering a pivotal moment.

The models, tools, and habits we adopt today will shape the environmental impact of this technology for years to come. That makes now the ideal time to ensure we don't lock in wasteful defaults or normalize excessive resource consumption.

By making informed choices about the models we use, we can unlock the creative possibilities of AI-driven image generation without sacrificing sustainability. More importantly, we should encourage accountability from tech providers. AI companies must be transparent about the energy demands of each of their models.

- **Model efficiency matters:** Models vary in terms of their energy efficiency. Select the right model and output for the job -- use the model that gets you to the desired end result most efficiently.

- **Measure and optimize:** AI providers should surface emissions data and offer tips for reducing impact. Users can help by choosing lower-emission settings (like "low-quality" images) and refining prompts to minimize trial and error.

- **Balancing performance and sustainability:** As AI continues to reshape media production and creative, there's an urgent need to create models that exist within the tension of performance and sustainability, ensuring the environment doesn't pay the cost of progress.

About Scope3

Scope3 is a technology company transforming how organizations optimize their AI operations. Through real-time monitoring and optimization, Scope3 helps enterprises reduce unnecessary costs and energy consumption while maintaining peak AI performance.

Led by pioneers who scaled algorithmic systems in digital advertising to create a $100B industry, Scope3's leadership team brings proven expertise in optimizing complex machine learning operations for both economic and environmental outcomes.

The company enables organizations to harness AI's transformative power while minimizing their environmental impact, demonstrating that business performance and sustainability can work together.

About Our Methodology

Most of the providers of the generative AI models in this analysis have not disclosed detailed information about their computational or energy requirements. Therefore, to produce these estimates, we relied on Scope3's predictive regression models, as described here.

These regressions are fit to real-world benchmarked data of open source generative AI models. They take input information about the prompt as well as the price charged for inference. It has been shown that the price charged correlates strongly with computational requirements.

This makes sense - the energy is the primary cost to providers while running inference. Therefore we use the price to help us arrive at our best estimate for how much energy is used (and emissions produced) for inference for closed models.