HPE GreenLake

Build a more sustainable AI strategy

How? Start with the fundamentals.

Artificial intelligence (AI) is big business: Research firm IDC projects that annual business spending on AI will surpass $500 billion by 2027.[1] In fact, experts predict that it won't be long before technology companies put AI at the core of their services, solutions, and their entire business. In some cases, that's already happening.

"Generative AI, and specifically large language models, have captured the imagination of the general public," says Matt Armstrong-Barnes, an AI specialist and chief technologist at Hewlett Packard Enterprise. "Lots of people are now using generative AI tools every day. AI has moved outside its traditional boundaries of technology, data science, and analytics, and it's part of everyday life."

But the benefit we derive from using generative AI tools to, say, spruce up emails to the boss may not always outweigh the sustainability impacts created when it is extrapolated out on a much larger, enterprise scale. AI's explosive growth is dramatically increasing the need for compute and storage resources, causing a major uptick in carbon emissions created by organizations using AI on a growing scale.

The AI gold rush is already having an impact on the environment. One recent study estimated that by 2030, global AI usage could consume more electricity than a country the size of the Netherlands, while another study posits that AI could be responsible for 14% of the planet's total CO2 emissions by 2040.[2]

"AI is driving exponential growth in demand for computing to train these very large models," says Arti Garg, lead sustainability and edge architect at HPE. "As we move to newer generations of technology in the high-performance computing space, energy consumption has always been a concern. And with AI, the energy demand is growing even faster."

[1] "IDC FutureScape: Artificial Intelligence Will Reshape the IT Industry and the Way Businesses Operate," IDC, Oct 26, 2023

[2] "Technology Untangled," HPE podcast, Feb 2024

Putting sustainability at the heart of your AI strategy

Armstrong-Barnes says the key to addressing AI sustainability is "developing an AI business plan that integrates complete information about the business outcomes you expect, your data needs, the AI models you'll use, requirements for training and tuning those models, the power and cooling required, data center expenses, and so on."

Considering all these components together can help you see the big picture of AI's sustainability impact in your enterprise, which can in turn help you get more from your AI investments.

You need a clear sense of how AI will contribute to your carbon footprint to take steps to reduce it. The more data you're dealing with and the more complex your large language models are, the more your AI-related power needs will grow.

"We're not necessarily saying that AI is bad for the environment," says Armstrong-Barnes. "Yes, it produces CO2. But it is also doing things that no other technology can." For example, there are numerous cases in which AI is being used to devise ways to decarbonize on a scale that will truly make a difference. But AI's carbon emissions are indeed substantial, and they will continue to grow.

In the excitement about generative AI's potential, "organizations are adopting large language models and implementing them in their business without fully understanding the sustainability implications," adds Armstrong-Barnes.

John Frey, an HPE chief technologist focused on sustainable transformation, agrees that it is critical to integrate sustainability into your AI strategy early on, just as you would for any major digital initiative.

"Sustainability should always be part of any transformation strategy, but when it comes to AI, businesses may need help pulling it all together," says Frey. "It has to be an executive-level conversation because many stakeholders — from institutional investors to end customers and employees — often ask boards and CEOs about IT and AI sustainability."

IT fundamentals can help you scale AI sustainably

Here's the good news: Experts say many of the fundamental principles IT leaders use to develop comprehensive IT transformation strategies also apply to AI. "AI should not become a solution in search of a problem," says Frey. As well, Armstrong-Barnes emphasizes the importance of creating a comprehensive AI strategy geared toward long-term results.

"Building and maintaining systems for the long term can help you derive more value from AI investments," Armstrong-Barnes says. "You need to consider the true cost of your AI systems — and when you may be moving in the wrong direction — so you can figure out what to do about it."

To help organizations get grounded in this endeavor, Frey outlines five levers of sustainable transformation — all of which also apply to AI:

- **Data efficiency:** Start with knowing what data you need, where it will come from, how often you'll collect it, the process you'll use to gain insights from it (for example, which AI models you'll use), how data will be moved between systems, and where and how long you'll store it. Can data be consolidated, disposed of, or stored using low-cost techniques such as tape or other backup methods? Data that does not need to be retrieved immediately can often be offloaded to more low-energy media.
- **Software efficiency:** This applies to AI models and any software used in your AI project. Are you employing the software engineering fundamentals of efficiency, repeatability, reusability, and simplicity? Are you integrating architectural efficiency features inherent in compiled code languages, for example?
- **Equipment efficiency:** Is your data center equipment — or the gear your data center provider uses — built using the latest advances in efficiency and sustainability, such as direct liquid cooling, to counteract the heat servers and GPUs generate?
- **Energy efficiency:** This is about getting the most out of every kilowatt consumed by all your workloads. You can team up with your data center equipment provider to determine whether AI and HPC workloads can be run more efficiently using existing resources.

- **Resource efficiency:** Are you using all your resources — including facilities, IT infrastructure, processes, and personnel — as efficiently as possible?

Armstrong-Barnes emphasizes Frey's point about software efficiency and recommends drawing on applicable software engineering principles when planning your AI strategy. "For decades, software engineering has been about building secure, maintainable, robust, environmentally friendly systems," he says. "There are applicable principles we need to borrow from this discipline and apply them when we build AI models. We need to build in sustainability from the very beginning."

Sustainable AI through clean data and tuning AI models

Armstrong-Barnes echoes the importance of the considerations mentioned earlier when it comes to planning your AI strategy, emphasizing the need for smarter software engineering and data management.

"Efficient software engineering and robust data management go hand in hand," he says. One way to ease the pressure on storage systems is to get rid of bad, erroneous, or duplicated data. Data debt, like technical debt, becomes problematic in AI systems because AI results hinge on the data fed into the models.

"If you don't manage and maintain your data, it becomes a big problem. With technical debt, we accrue it a bit more slowly because it's about maintaining and updating code," Armstrong-Barnes says. But with AI, "you're generating such a massive amount of data, including unstructured data. And if you leave your data debt alone, it quickly becomes an enormous mountain to climb."

Along with data management, your AI sustainability strategy should address the need to tune AI models as efficiently as possible. "Unfortunately, this step produces a lot of CO2 because it consumes a lot of power," says Armstrong-Barnes.

But Frey points out that organizations may not need to handle all the AI model training and tuning themselves. There are many prebuilt, pre-tuned AI models available — along with technology vendors and service providers that can help your organization decide how many of these key activities you should do yourself and when to bring in expert help.

Armstrong-Barnes offers this bit of advice: "Make sure you're using generative AI — or any AI application requiring large language models — with purpose. Remember the line in Sun Tzu's "The Art of War": 'Tactics without strategy is the noise before defeat. Strategy without tactics is the slowest route to victory.'"