

# Meta Has Turned to the Hot AI Circuit

---

HONG KONG, CHINA, April 27, 2023 /EINPresswire.com/ -- In late summer 2022, Meta CEO Mark Zuckerberg convened his key associates for a five-hour analytical discussion of Meta's computing capabilities, focusing on Meta's ability to do advanced artificial intelligence (AI) work, according to internal emails.

According to internal emails, company announcements, and people familiar with the matter, Meta faces a tricky problem: Despite massive investments in AI research, Meta is making slow progress on how to introduce AI-friendly hardware and software systems into its main business. As Meta has increasingly relied on AI to support further growth, this has slowed the pace at which the company is driving overall innovation.

The email, from Meta's new head of infrastructure, Santosh Janardhan, said: "We are significantly behind in tools, workflows, and processes in AI-specific development and need to make significant investments in this area." The email was posted on Meta's Intranet in September and first came to light recently.

The email also said that supporting the AI effort would require Meta to "fundamentally change hardware infrastructure design, software systems, and approach to providing a stable platform."

For more than a year, Meta has been working on a massive project to fill in the gaps in its AI infrastructure. Although Meta has publicly admitted to lagging behind in the development of AI hardware, the details of this, including computing pressure, management changes and an abandoned AI chip project, have not been previously reported.

In response to the email and related restructuring initiatives, Meta spokesman Jon Carvill said that with deep expertise in AI research and engineering development, Meta has a proven track record of building and deploying state-of-the-art large-scale infrastructure".

"As we deliver new AI experiences for apps and consumer products, we are confident that we can continue to expand our infrastructure capabilities to meet short - and long-term needs," he said.

He declined to comment on Meta's decision to abandon the AI chip project.

The restructuring resulted in the company's capital expenditures increasing by about \$4 billion

per quarter, and nearly doubling from 2021, led to the suspension or cancellation of plans to build data centers at four locations separately, according to Meta disclosures.

Meta is also under financial pressure. Since November, Meta has embarked on a scale of layoffs not seen since the dot-com crash of the millennium.

On the other hand, Microsoft-backed OpenAI launched ChatGPT on November 30, and the AI chatbot quickly became the fastest growing consumer app of all time, sparking an AI arms race among tech giants. Big tech companies are rolling out their own generative AI products. In addition to recognizing patterns in data, the AI can generate text and visual content based on input in a human-like manner.

According to [JAK Electronics](#), generative AI consumes much computing power, making it more urgent for Meta to expand its computing infrastructure.

1. Meta haven't invested enough in projects that cost money.

According to sources, a big part of the problem is that Meta was slow to introduce [GPU chips](#) into its AI work. GPUs are well suited for AI computing and can perform a large number of tasks in parallel, greatly reducing the time spent processing massive amounts of data. And, of course, more expensive, with 80% of the market held by Nvidia.

Therefore, Meta largely relies on the CPU to undertake AI computing tasks. CPUs, the workhorses of the computer industry, have been found in data centers around the world for decades, but are not well suited to handling AI computing tasks.

Meta also uses custom chips of its design for AI reasoning, according to two sources. By 2021, however, adopting CPUs and custom chips will prove to be slower and less efficient in AI than GPUs. In addition, GPUs are also more flexible when running different types of AI models than the chips used by Meta.

Meta declined to comment on the performance of its AI chips.

As Zuckerberg pushes Meta toward a metaverse, computational pressures have affected Meta's ability to deploy AI to counter competitive threats, such as the rise of social media rival TikTok and Apple-led changes to AD privacy policies, the sources said.

These setbacks also caught the attention of Peter Thiel, a former Meta board member. He resigned from Meta's board in early 2022 without explanation.

At a board meeting before his resignation, Thiel told Meta executives they were too complacent about Meta's core social media business and too obsessed with the Meta comes, according to people familiar with the matter.

## 2. Meta turns to GPU, but it's behind.

In 2022, after canceling plans for large-scale deployments of custom reasoning chips, Meta executives instead began sourcing multibillion-dollar Nvidia GPUs, one source said. By this time, Meta had fallen significantly behind competitors like Google. Google started deploying custom versions of GPUs, known as TPUs, back in 2015.

In the spring of 2022, Meta executives also moved to restructure Meta's AI division, naming two new engineering heads, including Jonathan, the author of the September email. More than a dozen managers left Meta during a tumultuous months-long period, according to LinkedIn profiles and people familiar with the matter, the MetaAI infrastructure management team was almost replaced.

Next, Meta started replanning the data center infrastructure to accommodate the GPU chips that would be deployed. Compared to CPUs, GPUs consume more power and heat and need to connect a large number of chips through a specially designed network to form a cluster.

According to Janarhan's email and the source, the facilities require 24 to 32 times the network capacity and new water-cooling systems to manage the heat dissipation of the chip clusters, so the facilities need to be "completely redesigned."

As the work progressed, Meta made internal plans to begin developing a new autonomous chip. The chip, which trains AI models and performs reasoning like a GPU, is currently scheduled for completion around 2025.

Cavill, the Meta spokesman, said some data center construction projects, currently on hold and transitioning to new designs, will be restarted later this year. He declined to comment on the chip project within Meta.

## 3. Slow implementation of products

In the process of expanding GPU computing power, Meta currently has almost no new product technologies to show. In contrast, companies like Microsoft and Google are pushing for the public commercialization of generative AI products (Bing Chat, Bard, etc.).

In February, Susan Li, Meta's chief financial officer, admitted that not much computing power is currently invested in generative AI. "All of our AI capabilities go to ads, news streams, and short video Reels," she says.

According to sources, Meta didn't take generative AI products seriously until ChatGPT was launched in November. Facebook's AI lab, FAIR, has been releasing prototypes of the technology since late 2021 but has not turned the research into products, they said.

That is changing as investor interest grows. In February, Zuckerberg announced the creation of a top generative AI team that would "dramatically boost" the company's work in the field.

Meta Chief Technology Officer Andrew Bosworth also said this month that generative AI is the area he and Zuckerberg are spending the most time on, and expects to release a product this year.

Two people familiar with the new team said its work was at an early stage and focused on building a base model as a core that could be adapted for different product needs in the future.

Cavill, the Meta spokesperson, said that many teams at Meta have been working on generative AI products for about a year. He confirms that work has accelerated in the months since ChatGPT arrived.

JAK Electronics  
+852 9140 9162  
it@jakelectronics.com

---

This press release can be viewed online at: <https://www.einpresswire.com/article/630379333>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2023 Newsmatics Inc. All Right Reserved.