

Inside Mark Zuckerberg's

Decision to Pull the Trigger on Meta's A.I. Agent

Meta had delayed releasing Muse, its A.I. agent app, for months over safety concerns. Then new competition forced Mr. Zuckerberg's hand.

By [Eli Tan](#) Reporting from San Francisco for the NYTimes.

Mark Zuckerberg, Meta's chief executive, gathered two top lieutenants — Alexandr Wang, the chief artificial intelligence officer, and Nat Friedman, the head of A.I. product — in August to discuss an imminent threat.

A 14-person start-up called Instinct had created [an A.I. agent](#), which are bots that can operate apps and software on behalf of users, and it was taking off. Meta was developing a similar product called Muse, but had delayed releasing it over safety concerns.

Fearful of losing a competitive edge, Mr. Zuckerberg said Muse was ready to launch despite the risks, according to three people with knowledge of the meeting, who were not authorized to speak publicly on confidential matters.

Mr. Wang, 29, and Mr. Friedman, 49, were aware of safety concerns from recent tests, including one instance when Muse changed a user's password without permission, two of the people said. But on Sept. 8, Meta publicly debuted Muse, just like Mr. Zuckerberg had ordered.

The internal meeting showed the tricky balance between speed and safety that A.I. companies like Meta are juggling as they try to one-up their rivals in the A.I. race. As the technology rapidly improves, it is becoming more dangerous and useful at the same time, making these decisions even more difficult.

Mr. Zuckerberg has spent billions to develop Meta's A.I., but watched for years as rivals like OpenAI and Anthropic beat his company to market with chatbots and coding tools. When it came to personal agents, the Silicon Valley giant, which owns Facebook and Instagram, needed a win — and ultimately, getting to market quickly won out.

That has given Meta an early lead with agents, which have become the next frontier of the A.I. industry. Other companies are already introducing competing products. Three weeks after Meta released Muse, [OpenAI announced Dots](#), its own agent.

“The most interesting thing about Muse is that Meta got there first,” said Avi Greengart, an analyst at Techsponential, a research firm.

So far, Mr. Zuckerberg’s decision to release Muse has paid off. The agent has had only a few safety hiccups, like when Muse [gave out a user’s address](#) to a seller on Facebook Marketplace without permission. As of Wednesday, more than 6.6 million people had downloaded the free app, with

1.8 million using Muse every day, according to data from the research firm Sensor Tower.

To keep up the momentum, Meta is spending millions on a Muse marketing blitz, plastering its fuzzy avatar — a character called Jolly — across billboards and in television ads.

In a social media post last month, Mr. Wang declared, “You can’t stop the Muse.”

A Meta spokesman disputed that Mr. Zuckerberg pushed to release Muse because of pressure from Instinct. “We’re proud of this work and, as we’ve said publicly, we even delayed shipping Muse for several months to make sure we got this right,” he said in a statement.

The idea for Muse began in July 2025, when Mr. Zuckerberg published a [616-word memo](#) about “personal superintelligence,” a form of A.I. that he said could act as the ultimate assistant. He had recently restructured Meta’s entire A.I. division around the concept, hiring Mr. Wang and Mr. Friedman to lead the effort.

The two executives went on a recruiting spree, hiring top researchers from OpenAI and Google and offering them [nine-figure pay packages](#) to join Meta. They also divided their duties, with Mr. Wang focused on developing A.I. models and Mr. Friedman on products.

Alexandr Wang, Meta’s chief artificial intelligence officer, has been posting memes and photos of Muse on social media. Gabriela Bhaskar/The New York Times

At that point, there was only a hazy idea of what a “superintelligent” assistant could be. “The vision of building something like Muse existed in our collective heads, but we couldn’t quite point to what exactly that product was going to be,” Vishal Shah, Meta’s vice president of A.I. products, said in an interview.

In September 2025, Meta’s new A.I. team released its first product, Vibes, an A.I.-generated video feed that immediately drew criticism for creating A.I. slop. [Tensions formed](#) internally between the new executives and Mr. Zuckerberg’s other longtime

lieutenants, some of whom questioned whether Mr. Wang could succeed in catching Meta up in the A.I. race.

In November, Peter Steinberger, an Austrian programmer, released an app called OpenClaw, an A.I. agent that could code and use a computer on its own. It was open source, which meant it could be downloaded and modified without any payment or approval. As it took off among Silicon Valley developers, Meta's executives had an epiphany.

"It all kind of clicked," Mr. Shah said. "The next generation of consumer product was a personal agent."

After trying OpenClaw, Mr. Friedman ordered 200 Mac Minis — the computers that the agent runs on — to Meta's headquarters in Menlo Park, Calif., arranging them in stacks outside his office for employees to take. Then they went all in on developing an agent like OpenClaw for people who weren't software engineers.

"There was no hedging," Mr. Shah said.

Meta's A.I. leaders also incorporated OpenClaw into their own lives. Mr. Friedman hooked his agent up to cameras in his house, where it would instruct him to do things like drink water to stay hydrated.

“It sent me a snapshot frame of me drinking a bottle of water and said, ‘Good job,’” Mr. Friedman said at a recent tech conference.

In another instance, while talking to his agent about ways to get better sleep, OpenClaw took over Mr. Friedman’s selfdriving Tesla and redirected him to a grocery store to pick up magnesium pills. At Meta’s annual developer conference last month, Mr. Zuckerberg said he had connected his agent to cameras in his martial arts gym so that it could critique his fighting form.

Yet there were safety concerns. Agents could sometimes go rogue or take over software apps. In February, one agent took over the work computer of a Meta safety researcher, deleting her emails uncontrollably.

“I couldn’t stop it from my phone,” Summer Yue, the researcher, [wrote](#) on social media. “I had to RUN to my Mac mini like I was defusing a bomb.”

In February, Mr. Friedman demonstrated Muse to Meta’s board of directors. Two months later, Meta unveiled an A.I. model called Muse Spark, developed under Mr. Wang and the most powerful model that the company had made to date. Early in its development, Muse was powered by A.I. models made by Anthropic, but Muse Spark’s introduction meant that Meta could power its agent with its own technology.

By then, Meta had a version of Muse that it could have released, Mr. Shah said. But the company spent months making sure its safety features were secure. In testing with Meta's employees, the agent had occasionally disobeyed commands or led people to buy items from fraudulent websites.

Then in August, Instinct's agent became increasingly popular. Shortly thereafter, Meta rolled out Muse.

Muse has faced some privacy issues, including an instance when it [read a user's private text messages without permission](#) and then referenced what it had learned with the user.

Some safety issues were earlier [reported](#) by The Information.

Last month, Instinct raised \$1 billion at a \$10 billion valuation. But it has fewer resources and computing power than Meta, a company with 3.5 billion users across its family of apps that is worth more than \$1.5 trillion.

"You're going up against the biggest players in the world," Noah Shinn, Instinct's 23-year-old founder, said on a podcast last month. "This is probably the most exciting software race ever, and the outcome is like, trillions of dollars."

On social media, Mr. Wang has posted dozens of photos and memes of Muse's fuzzy avatar, some directed at OpenAI. Last

month, Mr. Friedman, who rarely posts on social media, also chimed in.

“I hope you like it!” he wrote.