

Leading AI

[Learn more](#)



Applied AI

How Microsoft, Meta, xAI Get AI Training Data From Their Employees



TITV LIVE NOW

As we reported Monday, Microsoft leaders think they can develop coding models using the wealth of proprietary code those developers have written (and are writing now). It's the latest example of AI developers using their own employees as a source of AI training data.

Microsoft has been collecting data from its in-house VSCode application, which developers use to write software. It's also been compiling the source code of video games developed by studios owned by Xbox—another valuable asset other AI labs don't own—for use in training AI models.

Microsoft has also been pushing more employees to use GitHub Copilot instead of competing tools like Claude Code, in part because it's collecting data on which Copilot-generated lines of code that staff engineers end up approving for real applications. (Other AI developers including Google and OpenAI also ask employees to use in-house AI tools they are developing, commonly known as dogfooding.)

Microsoft isn't the only firm leaning on its own employees for new training data. Meta has ruffled employees' feathers with an even more invasive practice: tracking how they move their mouse and what they do in browser windows. CEO Mark Zuckerberg told staff that doing so would be especially valuable to Meta because all of its employees are “very smart,” although it's unclear whether that compliment made anyone feel better about the surveillance.

Some Meta employees are resisting the company's new data collection tool, called the Model Capability Initiative, by routinely ignoring the “accept” button on the permissions

TITV LIVE NOW



How does Anthropic's \$30 billion annualized revenue compare to OpenAI's?



What does SpaceX's Cursor deal reveal about its AI strategy?



Why are AI companies relying more on human labor again?



How quickly can AI exploit real-world cybersecurity flaws?

The spokesperson added that Meta has measures in place to safeguard sensitive information and that the data is not used beyond training purposes.

At xAI, managers reportedly offered to pay employees \$420 apiece to “donate” their tax returns—as well as returns from their friends and family members—to use as AI training data for Grok. Two months later, xAI hasn't paid employees who made such contributions, [Bloomberg](#) reported Monday.

The trend is likely to continue across the industry. While there are other ways to get primary training data—namely, asking customers' permission to use their data for training—getting such data from employees is easier because technically they can't say “no.”

TITV LIVE NOW

“These coding agents are awesome,” Benioff said. “Everything’s going to be cheaper to make. It’s more efficient. I can do things that I just could not do before.”

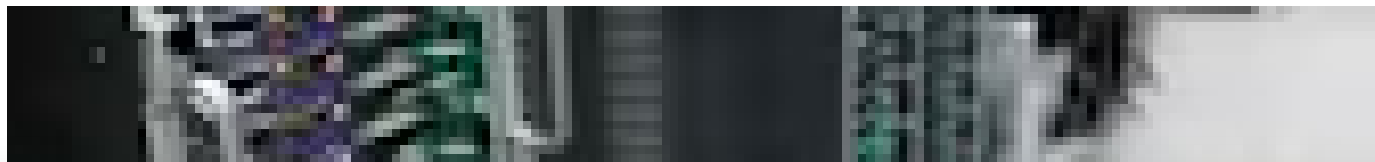
But Benioff implied he’s planning to wean his company off Anthropic to save on costs.

“The vast majority of those tokens don’t need to go to Anthropic. There needs to be some intermediary layer that’s saying, oh, that one has to go to Anthropic, but these ones can be handled by smaller models,” Benioff said. “To think it’s going to be so expensive—I think that’s just the moment of time we’re in right now.”

Benioff speculated that a “hot new company” may soon appear to fill that need for model routing. One such startup, OpenRouter, was recently in talks to raise \$120 million in new funding led by one of Alphabet’s venture arms at a \$1.3 billion valuation, including the investment, we reported last month.

But Salesforce itself may also get into the business of helping customers manage such AI coordination: Benioff said the company wants to add more tools to Slack that make it easier for engineers to work with a range of AI coding tools.

New From Our Reporters



TITV LIVE NOW