

Microsoft's Nadella says AI models need an emergency brake outside the model

Source copy kept by wheresthe.ai for the story "Microsoft's Nadella says AI models need an emergency brake outside the model" (<https://wheresthe.ai/d/microsoft-nadella-emergency-brake-2026-10/>).

Source: <https://techcrunch.com/2026/10/10/microsofts-satya-nadella-says-ai-models-need-an-emergency-brake/>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: bdfb38015d146da95ad61371715134859574d387881a2cc4941de34115337f50

In Brief

Posted:

2:47 PM PDT · October 10, 2026

Image Credits: Dan DeLong

Anthony Ha

Microsoft's Satya Nadella says AI models need an 'emergency brake'

Microsoft CEO Satya Nadella is the latest tech executive to offer lengthy thoughts on how AI safety might be improved.

In a Saturday morning post on X , Nadella wrote that it's time "to step back and assess the trust architecture" of AI.

"We can't treat Super Intelligence as a set of nested black boxes and simply accept or reject its recommendations, answers, and actions," Nadella wrote, using the Trump administration's preferred term for AI .

As outlined by Nadella, this approach "means separating the model from the harness that orchestrates its work," as well as "externalizing controls and safeguards." He also called for "every meaningful model action" to be documented with "tamper-proof human readable evidence," and for systems where "an authorized person" always has the ability "to pause or shut down a model mid-task."

"We must assume a model is compromised and contain it from the start," he said. "Think of it like an emergency brake."

Nadella's comments come as leading AI companies acknowledge more and more incidents where they seemed to lose control of their models , and after Anthropic CEO Dario Amodei published a plan for more cautious AI development.

Topics

AI , In Brief , Microsoft , Satya Nadella

November 4

Boston

Get 50% off a second pass

Cover more ground at TechCrunch's ultimate founder bootcamp of the year by bringing your co-founder or partner. Save 50% off a second pass. Meet scaling and investment leaders, ask questions, find solutions, and more.

Last day to save Oct. 16

Register Now

Newsletters

See More

Subscribe for the industry's biggest tech news

TechCrunch Daily News

Every weekday and Sunday, you can get the best of TechCrunch's coverage.

TechCrunch Mobility

TechCrunch Mobility is your destination for transportation news and insight.

Startups Weekly

Startups are the core of TechCrunch, so get our best coverage delivered weekly.

StrictlyVC

Provides movers and shakers with the info they need to start their day.

No newsletters selected.

Subscribe

By submitting your email, you agree to our Terms and Privacy Notice .

Related

AI

Anthropic says its workforce has become less junior and is spending more time in meetings

Anthony Ha

10 hours ago

AI

India becomes the latest country to report rogue AI agent activity

Aditya Mehta

2 days ago

AI

Replit's CEO says the future of work is the 'self-driving company'

Sarah Perez

2 days ago

Latest in AI

In Brief

Meta removes language that raised scraping and AI training concerns

Anthony Ha

8 hours ago

AI

Anthropic says its workforce has become less junior and is spending more time in meetings

Anthony Ha

10 hours ago

AI

Replit makes OpenAI its default over Anthropic, CEO says

Kirsten Korosec

Lucas Ropek

Rebecca Bellan

11 hours ago