

WILL KNIGHT

BUSINESS JUN 24, 2026 2:45 PM

I Met With China's Top AI Experts. They're Freaking Out, Too

The AI arms race between China and the US has researchers on both sides worried about a “Chernobyl moment.”



PHOTO-ILLUSTRATION: WIRED STAFF; GETTY IMAGES



SAVE THIS STORY



Listen • 6 minutes

JUST OVER A week ago, I attended a major artificial intelligence conference in Zhongguancun, Beijing's bustling high-tech district.

It was packed with fascinating sessions touching on everything from recursive self-improvement—the idea that models can tweak their own code and advance indefinitely—to humanoid robots. And it featured a few legends of computing, including Whitfield Diffie, co-inventor of public-key cryptography, and Andrew Barto, who won the Turing Award with Rich Sutton for his pioneering work on reinforcement learning.

But I left with one takeaway above all else: The US and China should put their fierce AI rivalry to the side.

Frontier AI's cybersecurity and systemic risks are too serious to ignore, and increasingly capable agentic models could soon cause chaos unless the world's AI superpowers can work together. "AI is a global technology with global benefits, global harms, and a consistent tendency for new capabilities to eventually proliferate," Stephen Casper, a computer scientist at MIT who spoke at the conference via video, told me afterward.

Until now, the US has largely viewed China's AI advances as an economic and national security threat. Washington has imposed tight restrictions on chips and chipmaking equipment to stymie the country's development of powerful AI. Most recently, the US government ordered Anthropic to prevent foreign nationals from accessing its most powerful models, Mythos and Fable 5, over national security concerns. In response, Anthropic revoked access for everyone. One company that was of particular concern, WIRED previously revealed, was a South Korean telecom giant with alleged ties to China.

But the conference, organized by the Beijing Academy of Artificial Intelligence, reinforced the idea that both the US and China stand to lose if AI is developed too quickly and recklessly. As AI becomes more powerful, more agentic, and more intertwined with everyday life, the risks that it could be used to conduct cyberattacks or fail in catastrophic ways will only grow. Because the world's two dominant AI powers are responsible for the most advanced models, cooperation between them feels like it

Casper pointed to research showing that the benefits of international collaboration on AI dangers outweigh any national security risks that come from working together. He likened the current situation to how the US and the Soviet Union were forced to work together on nuclear dangers, even as they sought to out-stockpile one another.

“One thing that almost everyone in AI can agree on right now is that AI doesn't need a Chernobyl moment,” Casper said.

One day-long session highlighted the universality of the cyber challenges raised by more advanced AI. This includes new kinds of vulnerabilities in AI-generated code, novel ways of attacking systems enabled by agentic tool use, and automated methods for carrying out social engineering attacks.

After another session, I spoke with Lin Yun, a professor at Shanghai Jiao Tong University who does excellent work on AI and computer security. Yun told me he expects hackers to gain an advantage over the near term, but that new countermeasures, including novel uses of AI, should tip the balance back toward defense over time.

Yun said that even if international cooperation is complicated by competition, it should remain a priority. “If different countries understand the risks in similar ways, it becomes easier to develop shared safety principles and technical standards,” he told me. “The key is to find areas where sharing can reduce systemic risk without exposing sensitive operational details.”

Perhaps the most pressing question for both nations is how to balance openness with risk. Open-weight models have become crucial for research and innovation, with Chinese models proving popular in the US. But as these models advance, it will become more challenging to ensure they don't help hackers identify security vulnerabilities and can't be wielded as cyber weapons.

Over the last few years, Chinese companies have taken the lead in offering highly capable open-weight AI models, like Moonshot's Kimi, Alibaba's Qwen, and Z.ai's GLM. The US has rebooted its own open-weight AI push with models like Nvidia's Nemotron. But we're approaching an inflection point where even less powerful open models could prove dangerous if they're stripped of guardrails. The latest model from China's Z.ai, GLM 5.2, includes frontier agentic and coding capabilities, according to expert analysis. The next generation of open-weight AI models might be just as capable as Fable or Mythos. In fact, just this week, 360 Security Technologies, a top Chinese cybersecurity firm, said it had developed an AI model with hacking capabilities on par with Mythos.

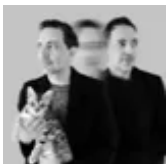
Yun said the industry will need to devise new ways of guaranteeing that open models are

One possible sign of things to come? A source at one of China's leading AI companies, who asked to remain anonymous because they weren't authorized to talk to the press, tells me that security concerns are one reason why some advanced models in China are no longer being released as open source.

*This is an edition of **Will Knight's AI Lab newsletter**. Read previous newsletters **here**.*

You Might Also Like

- **In your inbox:** [Brian Kahn's guide to how the universe works](#)
- Leak exposes members of [Peter Thiel's secretive 'Dialog' society](#)
- **Big Story:** [How Hunter Biden won the internet](#)
- Tesla batteries [keep getting stolen](#) before they even leave the factory
- **Special edition:** [The future of home](#)



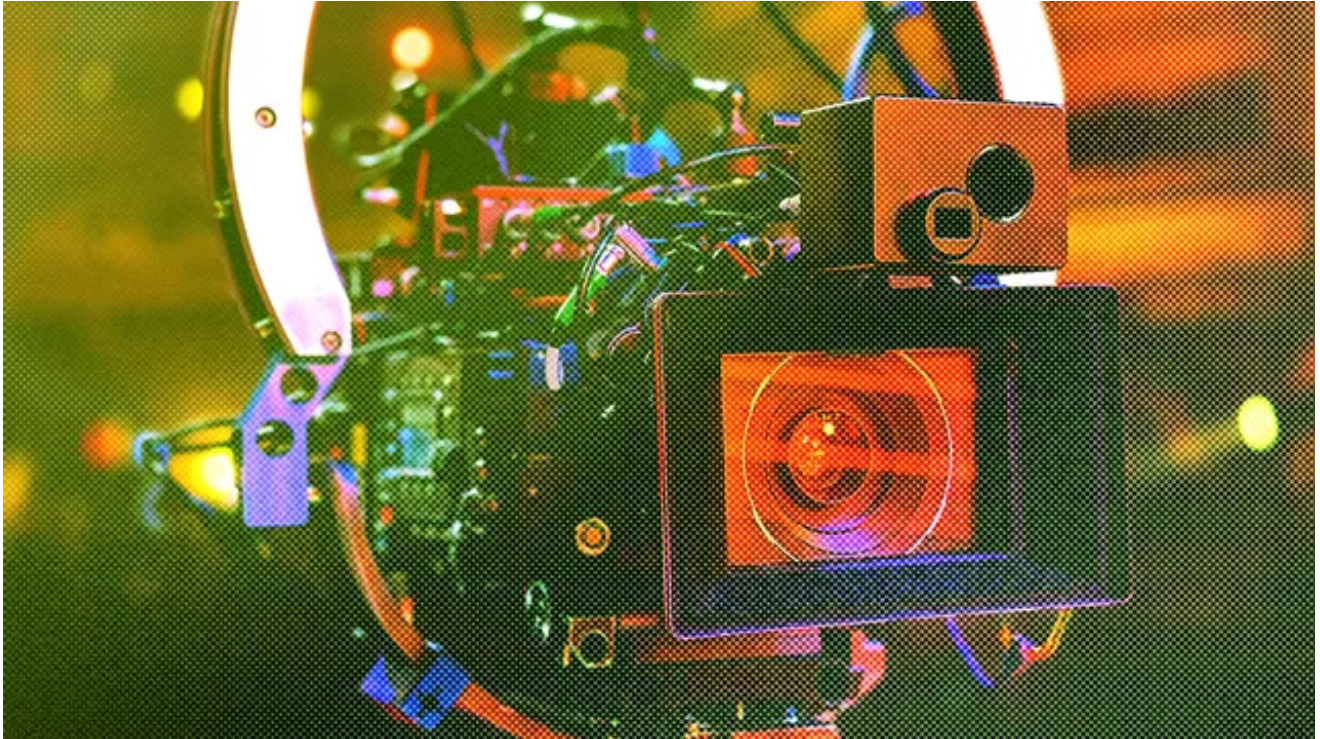
[Will Knight](#) is a senior writer for WIRED, covering artificial intelligence. He writes the [AI Lab newsletter](#), a weekly dispatch from beyond the cutting edge of AI—[sign up here](#). He was previously a senior editor at MIT Technology Review, where he wrote about fundamental advances in AI and China's AI ... [Read More](#)

SENIOR WRITER



TOPICS AI LAB ARTIFICIAL INTELLIGENCE CHINA CYBERSECURITY AGENTIC AI CYBERSECURITY HACKING
ANTHROPIC

READ MORE



Why Amazon Dropped Its OpenAI Movie, Data Center Workers Fight Back, and Meta Leaks Employee Data

The decision by Amazon-owned MGM Studios to drop the OpenAI movie is just part of AI and film industries becoming increasingly intertwined. On *Uncanny Valley*, we look at where this is all headed.

BRIAN BARRETT



Europe is Fed Up and wants its Own AI

It's a stretch to think that the continent can build a top-tier model, but it has an advantage: Donald Trump.

STEVEN LEVY



Truckloads of Tesla Batteries Keep Getting Stolen Before They Even Leave the Factory

Nine major suspected cargo thefts happened at Tesla's Nevada battery factory in January alone, according to sheriff's records obtained by WIRED.

PARESH DAVE



LastPass Users Had Their Data Stolen—Again

Plus: Former national security advisor John Bolton pleads guilty in classified-materials case, Microsoft helps take down major infostealer infrastructure, and more.

LILY HAY NEWMAN



OpenAI Has New AI Models. Here's Why You Can't Use Them

The White House asked OpenAI to delay the rollout of its GPT-5.6 AI models, two weeks after Anthropic had to take its most advanced AI models offline.

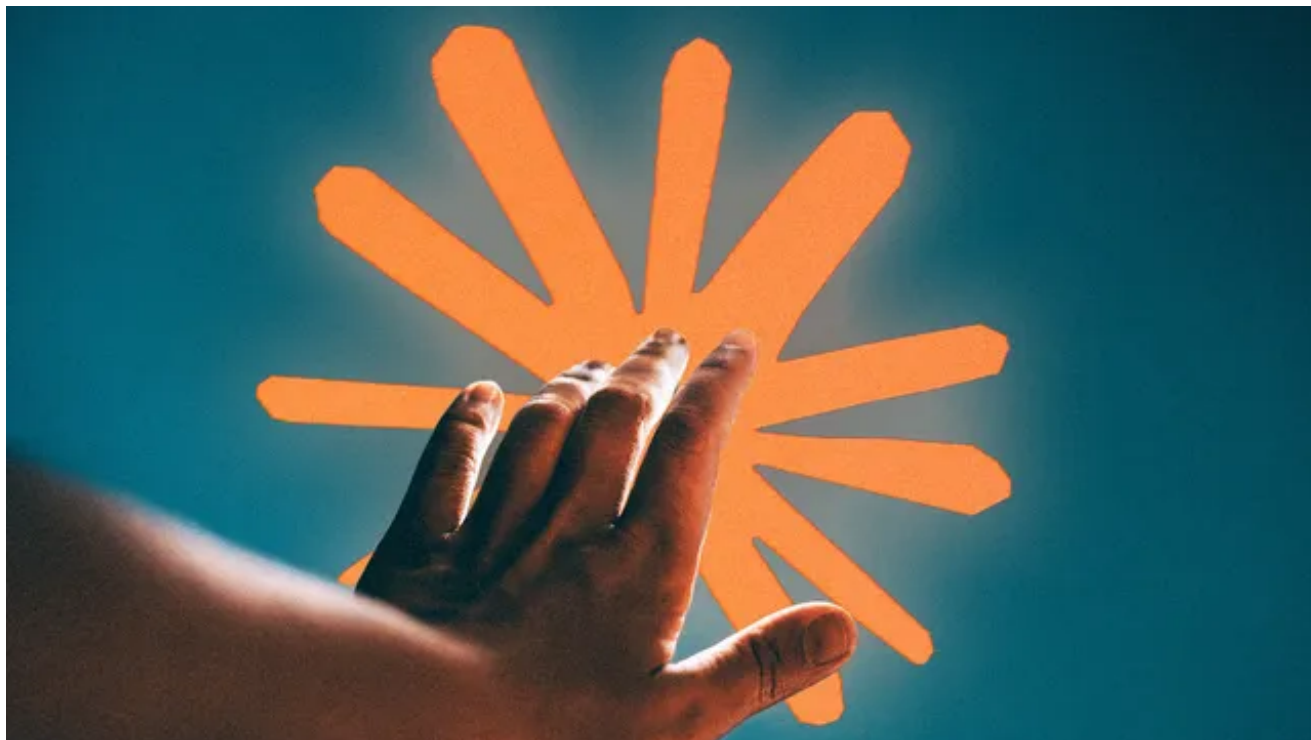
MAXWELL ZEFF



How Qatar Became FIFA's Technology Test Lab

Qatar has become the place where FIFA experiments with the next generation of football technology. The results are already visible across this year's World Cup.

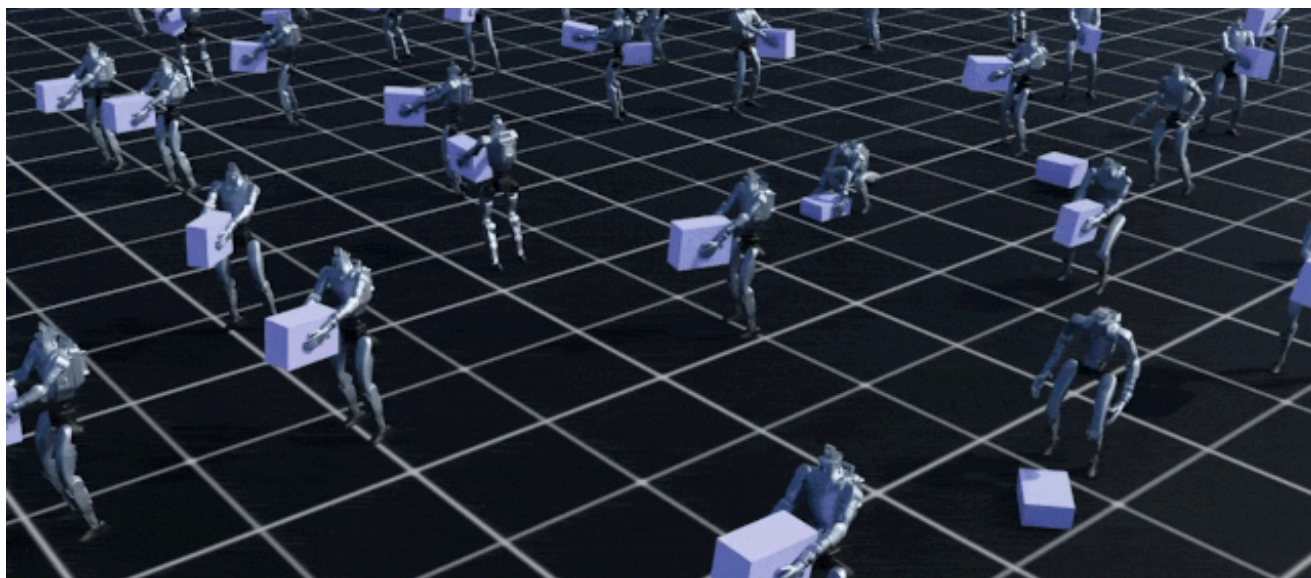
RIM ZREIN



Trump Administration Allows Anthropic to Release Mythos to Select US Organizations

After weeks of negotiations, the White House permitted Anthropic to grant access to its most advanced AI model to a select group of US companies and government agencies.

MAXWELL ZEFF





This Humanoid Robot Is a Terrifyingly Competent Office Intern

Flexion Robotics, a startup founded by ex-Nvidia engineers, has a clever way of training robots to do useful work.

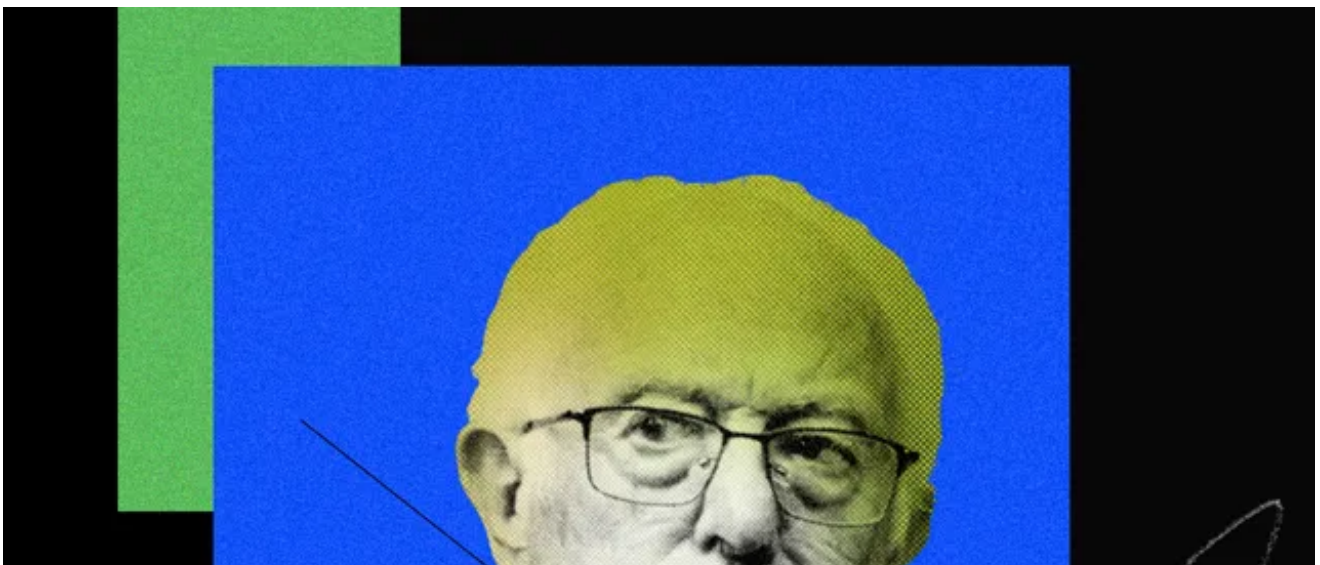
WILL KNIGHT



The 'Almost Homeless' Subreddit Is a Stark Glimpse at Soaring Wealth Inequality

As the billionaire class gets richer, the growing online community is offering tips on how to survive with very little.

MILES KLEE



Bernie Sanders Saw This Coming

For decades, the senator has argued that concentrated wealth threatened American democracy. Now he's betting that frustration with Big Tech, billionaires, and unchecked AI is reaching a tipping point.

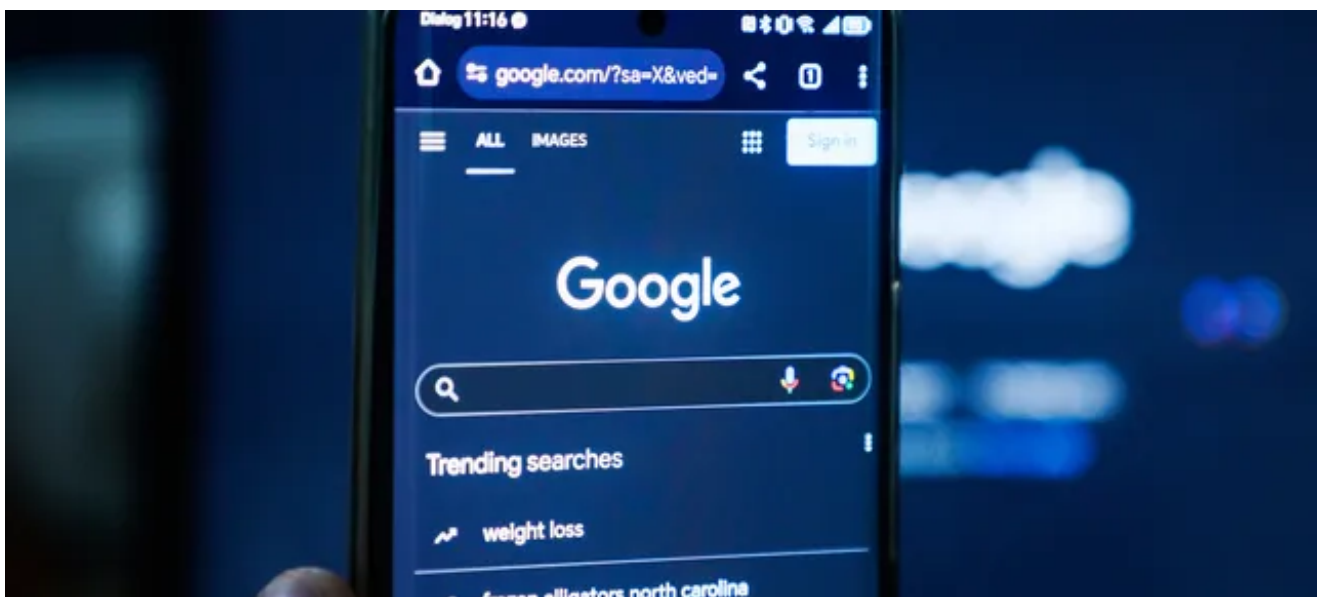
KATIE DRUMMOND

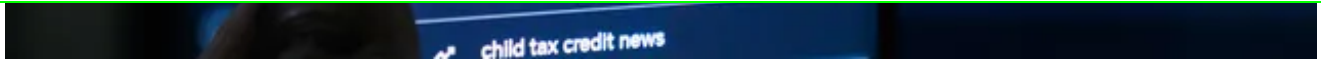


The Busy Bar Is a Gadget to Get People to Leave You Alone

Flipper Devices, a company that built a banned hacking device, now wants to hack your attention span.

BOONE ASHWORTH





Top Google Security Staff Warn Search Data Could Be Hacked if EU Rules Change

Europe's pro-competition proposals could see Google Search and Android systems opened up. The company claims there are serious privacy flaws.

MATT BURGESS



[MANAGE PREFERENCES](#)