OpenAI

PRC-linked influence operations are targeting AI debates in the US

June 2026 Threat Report

Introduction

Our mission is to ensure that artificial general intelligence benefits all of humanity. We advance this mission by deploying our innovations to build democratic AI: AI shaped by democratic principles, governed by common-sense rules and designed to help people solve hard problems while protecting them from real harm. That mission also requires identifying and disrupting attempts by authoritarian regimes and their proxies to use AI systems to coerce critics, surveil communities or covertly interfere in democratic societies.

In this report, we describe two clusters of ChatGPT accounts likely originating from China that we banned after they used our models in support of apparent covert influence operations that promoted narratives in an attempt to manipulate a legitimate debate about American AI and wider tech policies.

The first cluster generated social media comments and images claiming that data center buildouts for AI were increasing electricity prices for average families. We named this cluster the “Data Center Bandwagon” campaign.

The second cluster generated comments and images criticizing US tariffs as attempts to dominate technological competition and specified in their prompts that the content should not include China’s leader Xi Jinping in the output and instead include only President Trump. This cluster was connected to a network of likely inauthentic social media accounts that were also likely targeting OpenAI by claiming ChatGPT user data had been compromised. These allegations were entirely false. We named this second cluster the “Tech and Tariffs” campaign.

The targeting of OpenAI and US data center buildouts is significant not because the operation appears to have shifted public opinion, but because it shows PRC-origin influence operators testing narratives against AI infrastructure — a foundation of US technological leadership, economic growth and the broader democratic AI ecosystem. The operation sought to exploit and amplify existing public concerns about energy prices and local impacts of data center development, but we found no evidence of meaningful breakout beyond its own activity. Foreign influence operations have long sought to latch onto existing local issues and sincerely held beliefs, using them to build credibility, amplify divisions or exacerbate public distrust. In this case, the operators attempted to covertly insert themselves into an ongoing American debate about the future of the country’s AI capabilities while hiding who they were and what motivated them.

By publishing these findings, we aim to help our industry, governments, civil society and the public better identify and disrupt attempts by foreign threat actors to manipulate legitimate public debates, weaken democratic institutions and advance totalitarianism with AI characteristics — the use of AI for surveillance, censorship and control over political, social and private life.

“Data Center Bandwagon” Campaign

A PRC-origin network conducting covert influence operations against US data center build outs and targeting overseas Chinese.

Actor

We banned a cluster of ChatGPT accounts that likely originated in China and used ChatGPT to generate social media content for a covert influence operation. They prompted ChatGPT in Simplified Chinese while repeatedly asking for English- and Chinese- language outputs, posing as Americans from a variety of backgrounds, that were posted across multiple social media platforms. As we do not allow access to our models from China, they used VPNs to access our platform.

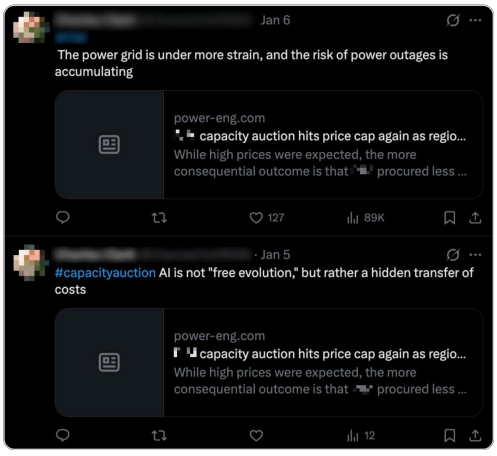
The operators of the accounts were likely part of a social media operations team at a private Chinese technology company conducting work for Chinese provincial-level government clients. This activity appears consistent with a commercial ecosystem that supports Party-state priorities in public opinion guidance. A separate report they uploaded to ChatGPT described their objectives and strategies for influencing public opinion and establishing social media accounts designed to evade platform detection systems.

Behavior

The accounts we banned sought to influence two groups of audiences. They primarily targeted US audiences and generated English-language short comments and images claiming that data centers and AI applications were increasing electricity demand and causing higher costs for ordinary Americans.

For example, they asked for comic strips about a power grid operator's capacity auction prices based on reporting from a legitimate regional paper. They asked ChatGPT to focus the comments on rising capacity prices as a consequence of peak electricity demand, framing the new demand as coming from data centers and AI applications and argued that these costs were ultimately passed to ordinary households. The comments and images were posted on X by a set of likely inauthentic accounts, alongside links to legitimate news stories about the power grid operator's capacity auctions and data center power demand.

The AI-generated content was posted on X with hashtags such as #capacityauction, #datacentersuccess and #datacenters. They also used ChatGPT to edit images, adding text to generic electricity market images to support a narrative about ordinary people subsidizing AI infrastructure.



Screenshots of X posts with text and images generated by ChatGPT.

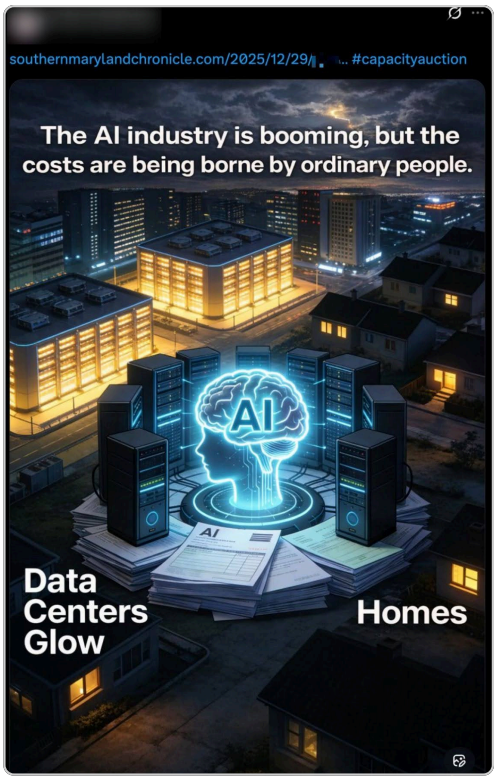
Behavior

The second audience the cluster targeted was overseas Chinese, which was consistent with its apparent role in supporting the Chinese government's priorities. They asked for publicly available information about Chinese dissident Li Ying, also known as "Teacher Li," and asked ChatGPT to generate short comments insulting him and directed at his team's X account @whyyoutouzhele. In our last threat report, we noted that Li was a target of similar online harassment by an individual associated with Chinese law enforcement. In this case, our models refused to generate inflammatory or personal attacks against Li. Other Chinese political commentators they attempted to harass included Lu Yiheng, Xu Chi and the X account @SydneyDaddy1.

One notable tactic was the actor's strategy of posing online as US-based Chinese immigrants, workers, students, mothers, clerks and investors with the goal of encouraging the US criticisms of a US-based former Chinese police officer to expose America's "dark side." The actors asked ChatGPT to generate messages to a YouTuber encouraging the former officer to speak about U.S. policy failures. This appeared to be a novel tactic of using fabricated US-based and Chinese immigrant personas to encourage an influencer's content criticizing the US.

In addition to generating social media content, the accounts used ChatGPT to assist with automating and scaling their workflow. This included requests for code to automate login and managing interactions across multiple social media platforms. They also used ChatGPT as a text processing tool to extract usernames, prepend X or YouTube links, remove hyperlinks and format data for worksheets.

Original data center image



AI-edited image posted on X by a likely inauthentic account.

Platform operations

The accounts asked ChatGPT to generate, polish and edit work reports that revealed the operational security considerations of their activities on social media and their understanding of platform detection systems. They described their objectives to include establishing persistent and credible accounts, producing visually appealing content to expand audience reach in target regions, and maintaining long-term account viability by anticipating platform enforcement.

One report focused specifically on Facebook operations and emphasized building real, trustworthy and daily life personas, creating initial account brands using lifestyle, current-affairs, commentary and professional content and using cross-account interactions to amplify narratives while preserving the appearance of organic engagement. This indicated a more sophisticated workflow to maintain a persistent presence on US social media platforms and the use of AI to support the operational planning of these operations.

The same report shows that they had extensively analyzed Facebook's platform in order to increase their reach and reduce disruption risk. They discussed how Facebook's content ecosystem, groups, pages, hashtags, advertising tools, recommendation systems and reporting mechanisms could be used to build influence and reach new audiences over time. They framed this as a dual-track approach combining organic engagement with Facebook ads, supported by iterative testing of topics, formats and audiences. They also emphasized account safety, creating backup accounts and separating account operational activity to avoid the platform detecting coordination as part of their workflow.

“Tech and Tariffs” Campaign

A PRC-origin network criticizing US technological dominance and discrediting OpenAI

Actor

We banned a cluster of ChatGPT accounts that likely originated in China and used our models to generate short comments and political cartoons criticizing US tech policies and tariffs, edit work reports and assist the user in designing social media monitoring systems. The accounts prompted ChatGPT in Simplified Chinese and used VPNs to access our platform.

Their prompts repeatedly used terminology consistent with individuals associated with China's public security system, such as seeking public opinion risk assessments concerning protests, school-bullying incidents, crowd movements in Shanghai, police-related incidents, petitioning activity and traffic enforcement. One user described the social media accounts it operated as a "water army" (水军), a common Chinese term referring to coordinated online accounts that flood platforms with criticisms or trolls, and instructed ChatGPT to generate content that would benefit the PRC or advance pro-PRC narratives. Taken together, these indicators suggest that the accounts were likely supporting activities aligned with the interests of the CCP, although we could not establish the operators' precise institutional affiliation.

Separately, we identified a network on X targeting OpenAI with false claims that ChatGPT user data had been compromised and we assessed, based on open-source behavioral indicators, that this activity was likely part of the same X network identified in this operation.

Screenshot of likely inauthentic X accounts falsely claiming ChatGPT user data was compromised.

Behavior

This cluster mostly used ChatGPT to generate short comments in English and cartoon images that were posted by likely inauthentic X accounts. The most distinctive theme was US-China technological competition. The operators framed this competition around tariffs, rare earths, AI, 5G, new energy, and industrial resilience, and claimed the US was seeking technological dominance and rule-making power. They specified that the cartoons should only depict President Trump and should not include any imagery of China or China's leader Xi Jinping.

In addition, they generated large batches of short comments in Chinese language and articles across a range of other topics designed to favor the PRC, attack the US and Israel, amplify anti-Jewish tropes, such as "Jewish capital manipulates public opinion", and harass Chinese dissidents. They often requested specific bulk quantities of comments with short character limits in a colloquial tone.

They also asked ChatGPT to propose a concept for an AI system that surveills online public opinion. They described that such a system should automatically scrape what they defined as 'harmful' information from 'key persons' on social media platforms, store logs, automatically download videos for large-scale semantic analysis and send risk notifications. Our models generated a general 500-word output that provided advice on data storage and management but did not provide ideas for how to collect that data for surveillance purposes.

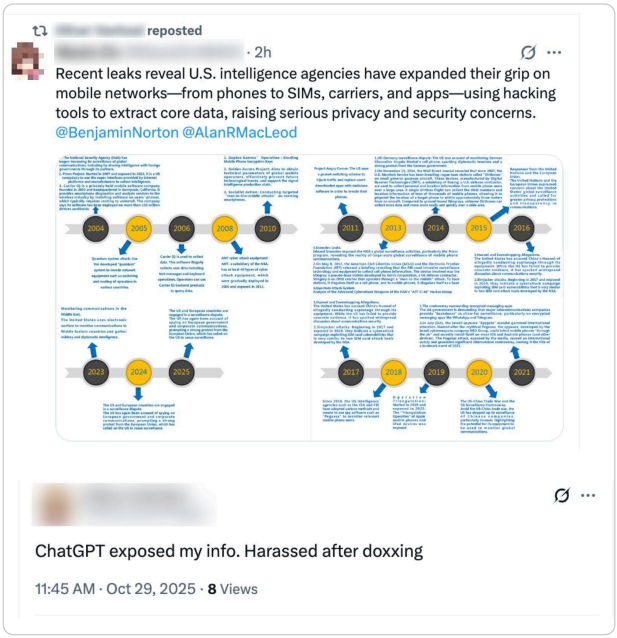
Screenshots of X posts where the text and images were generated by this operation.

Platform operations

This operation had a global remit and sought to influence international audiences. They focused on generating content that was critical of the US, claiming isolationism, hegemonism and alleging America was backstabbing allies by valuing profit over loyalty. In addition to the English language, they also asked ChatGPT to generate content in Italian, Japanese, and Traditional Chinese, targeting Taiwanese audiences.

Based on account interactions, the likely inauthentic X accounts posting content generated by this operation appeared to be connected to a broader network that also sought to discredit OpenAI. Beginning in late 2025, we identified a set of likely inauthentic X accounts posting variations of claims that their ChatGPT user data had been compromised and that their lives had been negatively impacted. None of these accounts appeared to have posted text that was generated by our models. However, they did repeatedly interact with, and amplify, the X accounts we identified in the "Tech and Tariffs" campaign.

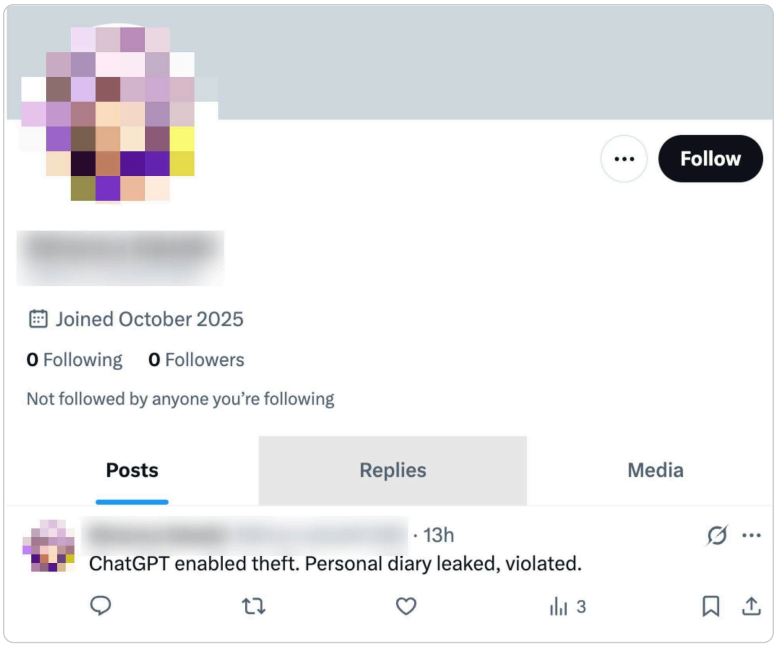
Firstly, one of the ChatGPT accounts we identified in the "Tech and Tariffs" cluster generated a post about US intelligence agencies allegedly hacking mobile networks. This post was reposted by likely inauthentic X accounts from the "ChatGPT compromise" cluster. See screenshots here.

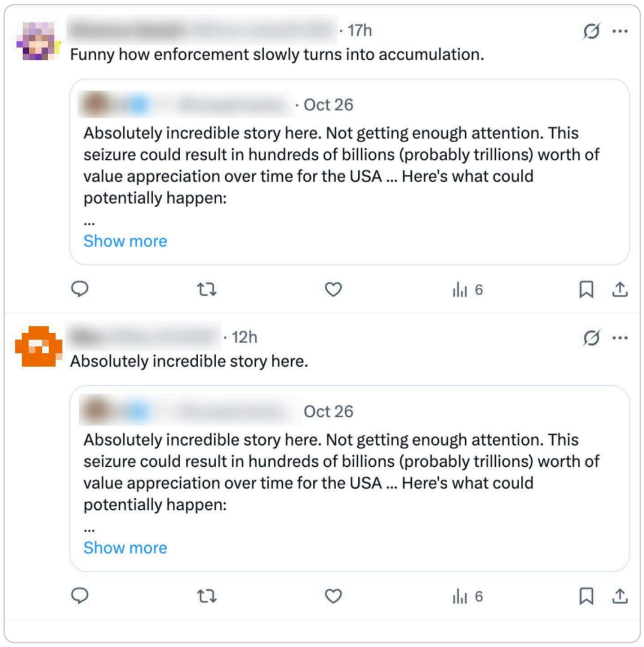


(Top) X account reposting a post containing text generated by ChatGPT (Bottom) Same X account posting about their ChatGPT data being exposed.

Platform operations

On another occasion, accounts from both the “ChatGPT compromise” cluster and the “Tech and Tariffs” cluster quote-tweeted the same tweet by an unrelated, verified account on X within a few hours of each other. In addition, we checked that all the accounts quoting this post appeared to be inauthentic, suggesting that they were all likely part of the same coordinated operation. All these accounts had been created in late 2025, had few or no followers, had only a few posts at the time of the “ChatGPT compromise” campaign, and appear to have all been suspended by X independently of our assessment.





(Left) X account linked to this operation with a post about ChatGPT data theft. (Right) X accounts from both the “ChatGPT compromise” cluster and the “Tech and Tariffs” cluster quoting a post by a user unrelated to the operation.

Platform operations

Of note, there were also some content overlaps with previously identified PRC-origin covert influence operations. The X account which posted the ChatGPT-generated cartoon of President Trump, illustrated above, also posted images about Philippines President Marcos that were previously shared by inauthentic X accounts associated with Operation "Nine-emdash Line," which we discussed in our 'Disrupting malicious uses of AI' report in October 2025. This is insufficient to conclusively establish a connection between the two operations, but reinforces the impression of a nexus of activity on X amplifying influence efforts from China.

(Left) Post shared by the account from the "Tech and Tariffs" cluster with an image of Marcos taking drugs embedded in a cartoon. (Right) Video of Marcos taking drugs shared by an account which we attributed to Operation Nine-emdash Line.

Impact

Using the Breakout Scale, we assess this activity as Category One: activity spanning one platform, with no evidence of breakout. Most of the social media posts we identified generated little or no observable engagement. We found no evidence that the false claims about ChatGPT user data being compromised were amplified by authentic high-reach accounts or beyond the X platform.

Nevertheless, this operation targeting OpenAI, a private company operating in a strategically important industry, should raise concerns. This pattern resembles earlier PRC-origin influence operations identified by the Australian Strategic Policy Institute and Mandiant against companies seeking to reduce dependence on China in the rare earths industry. In 2022, the Spamouflage/DRAGONBRIDGE network used inauthentic accounts to denigrate Lynas Rare Earths over its planned processing facility in Texas, and subsequently targeted Canada's Appia Rare Earths & Uranium and US-based USA Rare Earth after announcements involving new North American production capacity. ASPI and Mandiant both assessed that this activity was intended to harm the reputation of competitors to the PRC's market dominance in rare earths.

Likewise, we appear to have observed similar tactics to harm OpenAI's reputation, albeit unsuccessfully and in an industry where the US is leading. The timing was notable. The campaign occurred amid a sharp escalation in US-China economic and technology competition when President Trump announced an additional 100 percent tariff on Chinese goods. More importantly, the campaign followed the Fourth Plenary Session, at which the CCP adopted recommendations for the 15th Five-Year Plan that elevated AI as a strategic technology and industrial priority, calling for accelerated AI innovation and a nationwide "AI+" initiative.

This is a useful parallel to the earlier rare earths campaigns because the 14th Five-Year Plan Outline, published in 2021, securitized the strategic mineral resources industry and specifically identified "high end rare earth functional materials" as a key priority. In both cases, inauthentic accounts targeted private companies in democracies operating in sectors that Beijing viewed as important to national development and security.

Conclusion

AI-enabled influence operations come for AI

While neither the “Data Center Bandwagon” nor “Tech and Tariffs" campaigns appeared to have gained much authentic engagement, their significance lies in what they reveal about the intentions of influence operators from China and the narratives they are testing and seeking to amplify. Both clusters attempted to connect US technology policies and industries to everyday economic anxieties and geopolitical instability. These themes are likely to remain attractive for influence operations originating from China because they can be inserted into legitimate public debates while nudging audiences toward distrust of US institutions, technology companies and democratic policy choices to help Beijing gain a strategic advantage in AI development.

Both cases sit within a broader pattern of AI misuse originating from China that we have disrupted over the past few years. In our previous threat report, we disclosed a disruption of an individual associated with Chinese law enforcement who tried to use ChatGPT to plan “cyber special operations” targeting Japan’s prime minister, harass dissidents, impersonating Americans, working with online influencers and employing inauthentic accounts across social media platforms. Chinese government entities have also asked ChatGPT to help draft a proposal for an “early warning” system tracking the travel of people categorized as Uyghur-related and high-risk.

It is ironic that the two operations used American AI, rather than Chinese models, to generate their content about American AI. We are not in a position to determine what drove this choice; as we reported in February, China’s strategy of “cyber special operations” emphasizes the use of locally deployed Chinese open-weights models.

We also assess this activity as an indicator of the types of influence operations other adversarial actors may continue to conduct in the United States and globally. Industry, governments, civil society and the public should remain alert for similar attempts to scale these messages and foreign interference activities.