Israeli cabinet likely behind AI-generated disinfo campaign in Iran: *Researchers*

brookings.edu

By Derek B. Johnson
Reporter

PERSPECTIVE

A coordinated Israeli-backed network of social media accounts pushed anti-government propaganda — including deepfakes and other AI-generated content — to Iranians as real-world kinetic attacks were happening, with the goal of fomenting revolt among the country's people, according to researchers at Citizen Lab.

In research released this week, the nonprofit — along with Clemson University disinformation researcher Darren Linvill — said the so-called PRISONBREAK campaign was primarily carried out by a network of 50-some accounts on X created in 2023, but was largely dormant until this year.

The group "routinely used" AI-generated imagery and video in their operations to try to stoke unrest among Iran's population, mimic real news outlets to spread false content, and encourage the overthrow of the Iranian government.

Israel's military campaign in Gaza, launched following a coordinated attack by Hamas in October 2023, eventually expanded to include air strikes in Lebanon and Yemen.

In June, the Israel Defense Forces launched an attack against Iranian nuclear facilities while also targeting senior Iranian military leaders and scientists for assassination. Those strikes expanded to other Iranian targets, like oil facilities, national broadcasters, and a strike on Evin Prison in Tehran.

In the early days of the conflict, the networks shared images and videos — of uncertain authenticity — claiming to show Iran in a state of chaos and instability.

One widely circulated video, likely altered with AI, depicted people standing in line at an ATM before breaking into a riot, accompanied by messages like "The Islamic Republic has failed!" and "This regime is the enemy of us, the people!"

But the bulk of Citizen Lab's research focused on the period between June 13–24, 2023, during the "12-Day War" between Israel and Iran, and social media activity during and after a real June 24 Israeli air strike on Evin Prison. The facility is known for housing thousands of political prisoners.

The strike happened between 11:17 a.m. and 12:18 p.m. Iranian local time. By 11:52 a.m., accounts associated with the network began posting about the attack, and at 12:05 p.m., one posted an AI-generated video purporting to show footage of the attack, tricking several news outlets into sharing the content as genuine.

"The exact timing of the video's posting, while the bombing on Evin Prison was allegedly still happening, points towards the conclusion that it was part of a premeditated and well-synchronized influence operation," wrote researchers Alberto Fittarelli, Maia Scott, Ron Deibert, Marcus Michaelsen, and Linvill.

Other accounts from the network began quickly piling on, spreading word of the explosions, and by 12:36 p.m., accounts were explicitly calling for Iranian citizens to march on the prison and free the prisoners.

Most of the posts failed to gain traction with online audiences, except for one. A message calling on "kids" to storm Evin Prison to free their "loved ones" also contained a video with AI-generated imagery. It managed to rack up more than 46,000 views and 3,500 likes.

"This second video about the Evin Prison, which shows the hallmarks of professional editing and was posted within one hour of the end of the bombings, further strongly suggests that the PRISONBREAK network's operators had prior knowledge of the Israeli military action, and were prepared to coordinate with it," researchers wrote.

Those posts and others by PRISONBREAK operators led researchers to believe the campaign — still active as of today — is being carried out by either an Israeli cabinet agency or a subcontractor working on behalf of the Israeli cabinet.

The press office for the Israeli embassy in Washington, D.C., did not immediately respond to a request for comment from CyberScoop.

> This second video about the Evin Prison, which shows the hallmarks of professional editing and was posted within one hour of the end of the bombings, further strongly suggests that the PRISONBREAK network's operators had prior knowledge of the Israeli military action, and were prepared to coordinate with it.

Debris of the main entrance of the Evin prison, which was destroyed in Israeli strikes, is pictured in northern Tehran, Iran, on July 1, 2025.
MORTEZA NIKOUBAZL/NURPHOTO

Despots, democracies fuel disinformation ecosystem

It's not the first time the Israeli cabinet has been tied to an online influence campaign related to the Gaza conflict, nor would it be the first time the entity has reportedly tapped private industry to wage information warfare.

Last year, researchers at Meta, OpenAI, Digital Forensic Research Lab, and independent disinformation researcher Marc Owen Jones all tracked activity from a similar network on Facebook, X, and Instagram that targeted Canadian and US users with posts calling for the release of Israeli captives kidnapped by Hamas, criticism of US campus protests against Israeli military operations, and attacks against the United Nations Relief and Works Agency.

Meta and OpenAI both flagged STOIC, a firm based in Tel Aviv that is believed to be working on behalf of the Israeli cabinet, as behind much of the activity.

Citizen Lab's report identified two other Israeli firms, Team Jorge and Archimedes Group, that sell disinformation-for-hire services to cabinet clients.

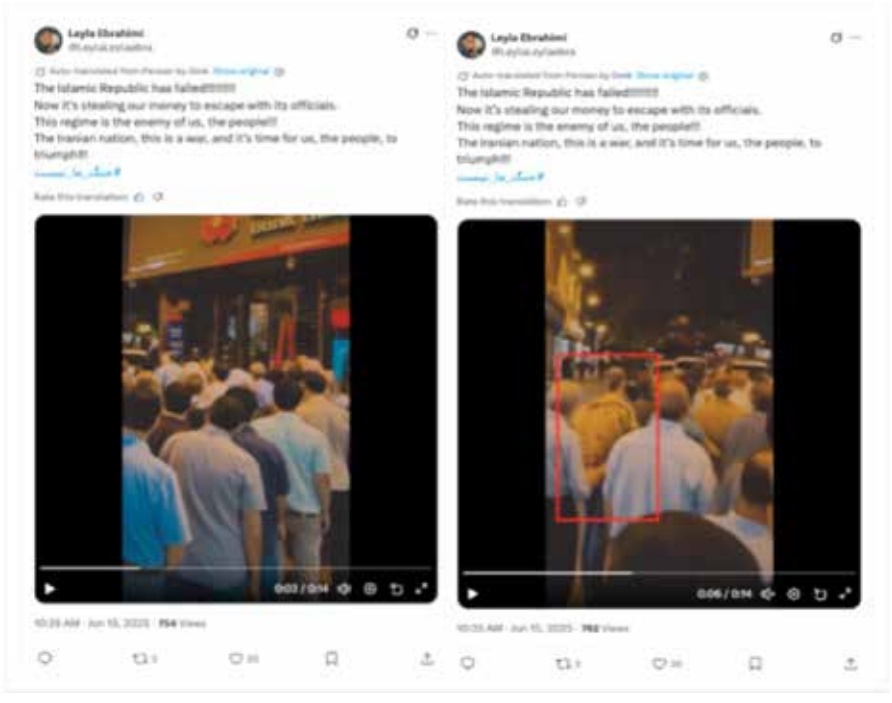
"Both companies offered their services to a wide array of clients globally, used advanced technologies to build and conduct their covert campaigns, and advertised existing or prior connections to the Israeli intelligence community," Citizen Lab researchers wrote.

While Western threat intelligence companies and media outlets can present disinformation campaigns as mostly a tool of autocratic or authoritarian countries, researchers have warned that democratic governments and private industry are increasingly playing key roles in information warfare.

David Agranovich, Meta's senior policy director for threat disruption, told CyberScoop last year that commercial marketing firms provide governments an additional layer of obfuscation when attempting to manipulate public opinion without leaving direct digital fingerprints.

"These services essentially democratize access to sophisticated influence or surveillance capabilities, while hiding the client who's behind them," Agranovich said.

The full article first appeared on CyberScoop.



An X post shared by the PRISONBREAK network on June 15, 2025, shows a video of people waiting in line at ATMs in Iran. The misshapen figure seen in the screenshot on the right is an indicator of the video being AI-generated.
CYBERSCOOP