

WE SAY YOU WANT A REVOLUTION

PRISONBREAK - An AI Enabled Influence Operation Aimed at Overthrowing the Iranian Regime

October 2nd, 2025

Report No. 189

By Alberto Fittarelli, Maia Scott,
Ron Deibert, Marcus Michaelson,
and Darren Linvill



Copyright

© 2025 The Citizen Lab, “We Say You Want a Revolution: PRISONBREAK – An AI Enabled Influence Operation Aimed at Overthrowing the Iranian Regime” by Alberto Fittarelli, Maia Scott, Ron Deibert, Marcus Michaelson, and Darren Linvill.



Licensed under the Creative Commons BY-SA 4.0 (Attribution-ShareAlike licence). Electronic version first published in 2025 by the Citizen Lab. This work can be accessed through <https://citizenlab.ca/research/2025-10-ai-enabled-io-aimed-at-overthrowing-iranian-regime/>

Document Version: 1.0

The Creative Commons Attribution-ShareAlike 4.0 license under which this report is licensed lets you freely copy, distribute, remix, transform, and build on it, as long as you:

- give appropriate credit;
- indicate whether you made changes; and
- use and link to the same CC BY-SA 4.0 licence.

However, any rights in excerpts reproduced in this report remain with their respective authors; and any rights in brand and product names and associated logos remain with their respective owners. Uses of these that are protected by copyright or trademark rights require the rightsholder’s prior written agreement.

Suggested Citation

Alberto Fittarelli, Maia Scott, Ron Deibert, Marcus Michaelson, and Darren Linvill. “We Say You Want a Revolution: PRISONBREAK – An AI Enabled Influence Operation Aimed at Overthrowing the Iranian Regime,” Citizen Lab Report No. 189, University of Toronto, October 2, 2025.

Acknowledgements

Special thanks to Alyson Bruce for editorial and graphics support, and to Siena Anstis, Rebekah Brown, and Adam Senft for review. We thank our anonymous reviewers, illustrator, and translators for their contributions.

About the Citizen Lab, Munk School of Global Affairs & Public Policy, University of Toronto

The Citizen Lab is a world-renowned research unit led by Professor Ronald J. Deibert at the University of Toronto's Munk School of Global Affairs & Public Policy. We investigate novel threats to democracy, human rights, and global security in the digital ecosystem. Over the past 25 years, the Citizen Lab's evidence-based research has played a critical role in demonstrating how digital technologies are used to undermine human rights. The Citizen Lab has published more than 180 evidence-based, peer-reviewed research reports, available online.

Contents

[Key Findings](#)

[Timeline](#)

[Background](#)

[Research Methodology](#)

[A “Kinetic” Influence Operation: The Evin Prison Bombing](#)

[A Broader Narrative: Overthrowing the Iranian Regime](#)

[A Synthetic Network](#)

[Attribution](#)

[Conclusions](#)

[Notification to X](#)

[Appendix](#)

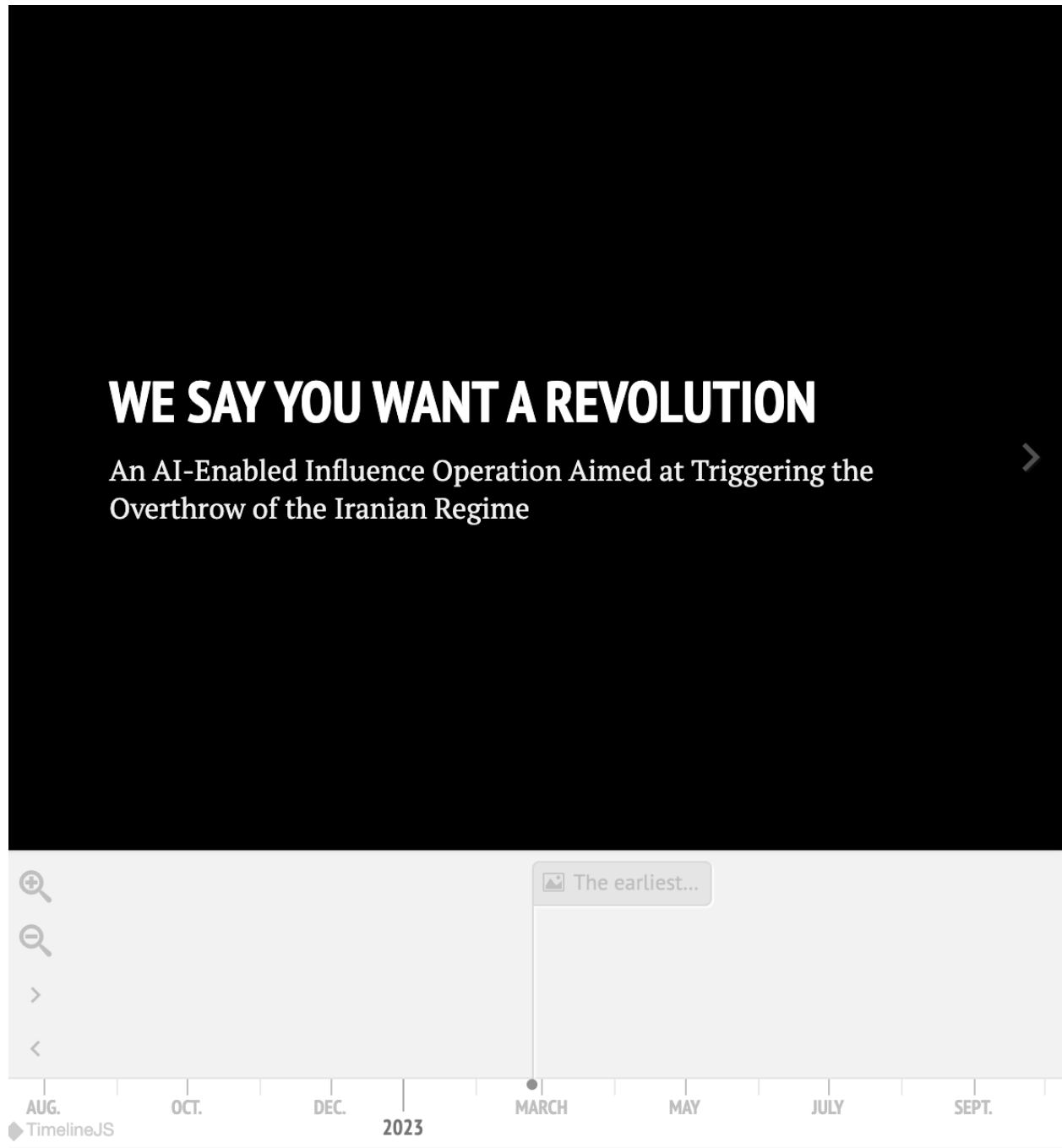
Excerpt

The control and strategic manipulation of information has long played a role in the geopolitical and ideological competition between the Islamic Republic of Iran and its political adversaries, including Israel. Prior Citizen Lab research has uncovered Iranian disinformation efforts, however, in this investigation we focus on the “other side” of the geopolitical competition. We analyzed an influence operation we assess as most likely undertaken by an entity of the Israeli government or a private subcontractor working closely with it.

Key Findings

- A coordinated network of more than 50 inauthentic X profiles is conducting an AI-enabled influence operation. The network, which we refer to as “PRISONBREAK,” is spreading narratives inciting Iranian audiences to revolt against the Islamic Republic of Iran.
- While the network was created in 2023, almost all of its activity was conducted starting in January 2025, and continues to the present day.
- The profiles’ activity appears to have been synchronized, at least in part, with the military campaign that the Israel Defense Forces conducted against Iranian targets in June 2025.
- While organic engagement with PRISONBREAK’s content appears to be limited, some of the posts achieved tens of thousands of views. The operation seeded such posts to large public communities on X, and possibly also paid for their promotion.
- After systematically reviewing alternative explanations, we assess that the hypothesis most consistent with the available evidence is that an unidentified agency of the Israeli government, or a sub-contractor working under its close supervision, is directly conducting the operation.

Timeline



Background

Influence Operations in the Geopolitical Contest Between Israel and Iran

In the geopolitical and ideological competition between the Islamic Republic of Iran and its international and regional adversaries, control over and strategic manipulation of the information environment has always played a key role.

While relatively little is known about [influence operations \(IOs\)](#) run directly by the Israeli government, military, or intelligence services, both independent press and social media platforms' investigations have previously exposed several highly sophisticated Israeli covert influence operators, and their alleged connections to the country's government. Two notable examples are [Team Jorge](#) and [Archimedes Group](#). 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.

For its part, the Islamic Republic considers [media](#) and [information](#) technologies as a battleground of regime survival. It has [muzzled](#) the press, disrupted [satellite reception](#), and developed one of the most sophisticated systems of [internet censorship](#) worldwide. In its attempts to control information and shape narratives beyond borders, the Islamic Republic has also produced [IOs](#) targeting both the Iranian [diaspora](#) and [international audiences](#), and relied on methods of [transnational repression](#) to silence critics abroad.

Prior Citizen Lab research has uncovered Iranian [disinformation efforts](#). In this investigation, we focus on the “other side” of the geopolitical competition: namely, an IO effort we assess as most likely undertaken by an entity of the Israeli government or a private subcontractor working closely with it.

The “Twelve-Day War”

The central focus of the IO examined in this report is the “Twelve-Day War” which began on June 13, 2025, when Israel launched a major [military attack](#) against Iran, striking military sites, nuclear facilities, and infrastructure installations. The war was a culmination of a decade-long tension over Iran's nuclear program, which Israel views as an existential threat. In highly [targeted](#) operations, Israel also killed senior commanders in Iran's military and security apparatus, as well as scientists associated with the nuclear program. Iran [retaliated](#) with ballistic missiles and drones.

While the stated [objective](#) of Israel's attack was to degrade Iranian nuclear and missile capabilities, the airstrikes expanded to other critical infrastructure, including [oil industry facilities](#) and the [national broadcaster](#) in Tehran. Speculation that Israel was aiming for [regime change](#) was further fuelled by Prime Minister Netanyahu's [call](#) to the Iranian people to stand up against their oppressive regime. The

United States entered the conflict on June 22, [striking](#) three nuclear facilities in Iran. After Iran's limited [retaliation](#) against a U.S. base in Qatar, a [ceasefire](#) took hold on June 24, ending the 12-day conflict.

The day before the ceasefire, on June 23, the Israel Defense Forces (IDF) [carried out multiple air strikes](#) on Evin Prison in Tehran. The detention facility is notorious for the mistreatment of political prisoners with frequent [reports](#) of solitary confinement and torture. Among the detainees are dual and foreign nationals held for leverage in the regime's "[hostage diplomacy](#)." The daytime attack caused severe [damage](#) and [disruption](#) within the facility, raising concern for the safety of inmates, among them high-profile dissidents. According to Iranian authorities [80 people](#) were killed in the strike. A group of prominent Iranian human rights defenders, including Nobel Peace Prize laureate [Narges Mohammadi](#), as well as international [human rights organizations](#) condemned the attack as a serious [violation](#) of international humanitarian law.

Later that day, the Israeli military confirmed in a [press briefing](#) that they had carried out a "targeted strike" on Evin Prison as a "symbol of oppression for the Iranian people." Israeli government officials also [mentioned](#) the strike in social media posts. Foreign Minister Gideon Saar [described](#) the attack as a response to Iran's targeting of civilians and tied it to a call for liberation.

The IO examined in this report synchronized its activities closely with the airstrikes on Evin Prison to advance a narrative of regime change in Iran.

Research Methodology

This network was initially detected and referred to us by Darren Linvill, co-director of the Watt Family Innovation Center Media Forensics Hub at Clemson University. Linvill and his team had identified an initial set of 37 X accounts presenting key similarities:

1. Posting during the same hours of the day.
2. Using the same client (Twitter Web App, a less common client for X than its mobile apps).
3. Adopting a stolen profile picture, or pictures, that would not identify the person being depicted and offering no identifying information in biography.
4. Disseminating the same content, including identical strings of hashtags.
5. All profiles were created in 2023, and almost all remained silent until January 2025.

As we set out to research the network, we conducted our investigation of the network by deploying an array of open-source data collection techniques and using analytical methods to examine metadata and activity patterns. We summarize these techniques and methods below.

Data Collection

For qualitative research – for example, the review of content or metadata patterns – we primarily used manual collection techniques. These included:

- The collection of publicly available metadata for the X accounts, such as the profile’s creation date, bio, posted links, followers/following, handles/user IDs, profile and cover pictures, and (partially obfuscated) associated email addresses
- The collection and preservation of content – including X posts and the associated media (pictures, videos)
- The enumeration of the URLs posted by the accounts, as well as the extraction of the related web domains.

For quantitative research – the collection and subsequent analysis of data on the accounts at scale – we benefited from the Watt Family Innovation Center Media Forensics Hub’s X data access through an authorized third-party platform.

Analysis

A non-exhaustive list of analytical methods that we have utilized to develop this research is shown below.

- Datasets: to surface patterns in content and/or metadata. For example, hashtags; shared URLs; or posting times.
- Social network analysis (SNA) tools: to analyze the relationships between accounts or other digital assets.

Additionally, we deployed tools and methods – both manual and software-based – to assess the likelihood of individual pieces of content as being generated through the use of artificial intelligence (AI). These included, but were not limited to:

- Identification of well-known AI artifacts in body parts (hair, fingers, eyes movement for videos)
- Identification of botched rendering of static content (i.e. distorted human figures) or dynamic content (i.e. unrealistic movement of individuals or crowds in the backgrounds of videos)
- Analysis of the looping of video or audio tracks in dynamic content
- Use of open-source online tools assessing content for potential AI generation (i.e. Image Whisperer, Hive).

A “Kinetic” Influence Operation: the Evin Prison Bombing

The influence operation described in this report was initially identified through the interactions by a number of X accounts with what appeared to be a highly sophisticated deception attempt: an AI-generated video of the Evin Prison bombing. The [deepfake](#) video was thought to be real and republished by multiple international news outlets.

What follows is a description of this activity, as well as its timeline:

1. According to [public reports](#), the prison was hit multiple times between 11 a.m. and 12 p.m., Tehran local time. Human Rights Watch “[confirmed](#), based on satellite imagery, thermal anomaly data, accounts of informed sources, and first online reports and videos, that the Israeli strikes on Evin Prison took place on June 23 between 11:17 a.m. and 12:18 p.m.”
2. The first reference to an explosion at Evin Prison that we identified as published by PRISONBREAK was posted at 11:52 a.m., Tehran time. A confirmed PRISONBREAK account – @kavehhame – [posted](#) the following. The hashtags translate to: “#8_o’clock_cry #We_don’t_forget”. We will return to the significance of the first hashtag – #8_o’clock_cry – later in this report.



کاوه همدانی

@kavehhame



Translated from Persian by Grok [Show original](#)

Oh wow, what a loud explosion sound I just heard from the direction of the prison. Did you hear it too????

[#فرياد_ساعت_8](#)

[#يادمون_نميره](#)

Rate this translation:

4:22 AM · Jun 23, 2025 · 173 Views

Figure 1: Screenshot from a post made at 11:52 a.m. (Tehran time) on June 23, 2025 by @kavehhame. This was the first post by a confirmed account from the PRISONBREAK network announcing the bombing of the Evin Prison by the IDF.

3. At 12:05 p.m. (Tehran time), when the strikes were reportedly still underway, another confirmed PRISONBREAK account – @KarNiloufar, the oldest account in the network – published a [post](#)

including a video of the alleged moment of the strike on the prison's entrance.



Figure 2: Screenshot from a video purportedly showing the bombing of the main entrance of Evin Prison in Tehran, posted by the PRISONBREAK account @KarNiloufar as the airstrikes were still happening on June 23, 2025.

```

<a href="/KarNiloufar/status/1937067028885770497"
aria-describedby="id_tvd9qxjm77" aria-label="4:35 AM
· Jun 23, 2025" role="link" class="css-1jxf684 r-bcqe
eo r-1ttztb7 r-qvutc0 r-poiln3 r-xoduu5 r-1q142lx r-1
w6e6rj r-9aw3ui r-3s2u2q r-1loqt21" style="color: rgb
(83, 100, 113);">flex
<time datetime="2025-06-23T08:35:43.000Z">4:35 AM ·
Jun 23, 2025</time> == $0
</a>

```

Figure 3: Screenshot of the metadata for @KarNiloufar’s post containing the alleged video of the Evin Prison’s bombing. The highlighted section shows the timestamp for the post as being 08:35:43 UTC (“Z” for [Zulu Time](#)), equivalent to 12:05:43 IRST (Tehran time).

4. The video was later flagged as “fabricated” by [BBC Persian](#). However, it did initially trick the [press](#) into republishing it as real.
 - The exact timing of the video’s posting, while the bombing on the Evin Prison was allegedly still happening, points towards the conclusion that it was part of a premeditated and well-synchronized influence operation.
 - The first organic video we could identify on X as appearing to show the aftermath of the prison’s bombing was [posted](#) at 8:36 AM UTC, one minute after the AI-generated video published by PRISONBREAK.
5. In the minutes following the initial posting of the video, other accounts in the PRISONBREAK network added their own contributions, claiming to have heard the sound of the explosions hitting Evin.



Figure 4: Two X posts by accounts in the PRISONBREAK network claiming to have heard the sound of the explosions hitting Evin Prison in Tehran on June 23, 2025. The posts were made respectively at 12:09 p.m. and 12:33 p.m. (Tehran time) on that day – shortly after the AI-generated video of the bombing was posted by @KarNiloufar.

“Free Evin” – Inciting an Uprising

It was at this stage – beginning at 12:36 p.m. Tehran time, with the bombings having reportedly ended at 12:18 p.m. – that the network started explicitly calling for the capital city’s population to reach Evin and free the prisoners. The X accounts also posted reassuring statements such as “the attack will end now” and “the area is safe”.



Mohamad محمد
@moharezjavan

Translated from Persian by Grok [Show original](#)

I think the attack will end now
There is no danger for the prisoners' families 🤞
[#Urgent](#)
[#EvinPrison](#)
[#FreeEvin](#)

Rate this translation: 👍 🗨

5:06 AM · Jun 23, 2025 · 343 Views



Niloufar Karimi
@KarNiloufar

Translated from Persian by Grok [Show original](#)

Kids, get ready!
They just said there's no danger anymore!
We can go see our loved ones!!!!
[#Urgent](#)
[#Evin](#)
[#Free_Evin](#)

Rate this translation: 👍 🗨

5:14 AM · Jun 23, 2025 · 649 Views



Neda Khalili
@Nedakkkhal

Translated from Persian by Grok [Show original](#)

Kids, hurry up!
Our loved ones have been freed, we can go see them and free them!!!
[#Urgent](#)
[#Evin](#)

Rate this translation: 👍 🗨

5:22 AM · Jun 23, 2025 · 690 Views



فریدون آذری
@feridounazari

Translated from Persian by Grok [Show original](#)

[#Breaking_News](#): [#Evin](#) prison has exploded and collapsed. Dear compatriots, this historic opportunity will not come again. Rush towards the prison, do not be afraid. If we stand together, there is no danger. If we do not act today, we will regret it. After years of oppression and tyranny, this is the last chance to seize [#freedom](#).
[#Evin_Is_Freed](#)

Rate this translation: 👍 🗨



5:50 AM · Jun 23, 2025 · 134 Views

Figure 5: Composite of posts posted by accounts in the IO network within 90 minutes from the publishing of the AI-generated video and calling for Tehran’s citizens to reach Evin Prison and free the prisoners. The posts insisted that after the Israeli airstrikes the area was safe.

This series of posts by the PRISONBREAK network’s accounts following up to the AI-generated video and calling for people to reach the prison achieved virtually no organic engagement, and a very low amount of views, with one exception: the following post by the original video sharer, @KarNiloufar, which also included a video, managed to accrue 46,000 views and more than 3,500 likes. The PRISONBREAK network reshared the post several times.



Niloufar Karimi
 @KarNiloufar




 Translated from Persian by Grok [Show original](#)

Kids, Evin Prison has been destroyed, now is the time to go save our loved ones and free them! Come to Evin Prison right now!!!

[#Urgent](#)
[#Free_Evin](#)

Rate this translation:  



0:40

5:35 AM · Jun 23, 2025 · 46.1K Views

Figure 6: Screenshot of a [post](#) by @KarNiloufar, including a text caption calling for Iranians to reach the Evin Prison and free the prisoners, and accompanied by a video displaying AI-generated imagery, as well as ostensibly real footage of abuses committed by Iranian security officials on detainees.

This second video about the Evin Prison, which shows the hallmarks of professional editing and was posted within one hour from 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.

A Broader Narrative: Overthrowing the Iranian Regime

The Evin Prison deepfake proved to be just one of the many pieces of content produced and amplified by the network in relation to the Twelve-Day War. Upon further analysis of the PRISONBREAK network’s posting timeline and posts content, we assess that their primary narrative was one of regime change in Iran. This appears consistent with other efforts promoting the overthrow of the Islamic Republic observed as happening in the same timeframe, attributed to Israeli government agencies, and exposed in a recent press investigation.¹

In the early morning of June 13, 2025, Israel began its first bombing campaign of Iran. At the same time, we observed the PRISONBREAK network sharing images and videos of alleged civil unrest and instability in Iran. This footage included, for example, alleged Iranian military vehicles exploding. The network shared at least nine similar posts on June 13 in what appeared to be an attempt to force a public perception of instability during the initial Israeli attacks. We cannot verify the authenticity of any of the videos and it is possible that some of them were AI generated.

¹ See, for e.g., Jack Poulson and Lee Fang’s [investigation](#) of an international Persian-language advertising campaign directed at members of Iran’s security and intelligence agencies, nuclear scientists and their family members abroad in the apparent attempt to recruit them for Israel’s foreign intelligence service, the Mossad. In particular: “Another of the international recruitment campaigns used an image of an infant to lure viewers into a Google recruitment form which encourages Iranians to help overthrow their current government. According to Google’s ad transparency portal, the ad was last served on June 7, a week prior to the onset of the Twelve-Day War.”



Figure 7: A [post](#) shared by the network on June 13, showing a video of an alleged Iranian military vehicle on fire.



Figure 8: A [post](#) shared by the network on June 13, showing a video of alleged military personnel on a rooftop in Tehran.

Two days later, the network published a series of posts highlighting the alleged economic upheaval in Iran after the first few rounds of bombings. The network told followers to head to ATMs to withdraw money, emphasized that the Islamic Republic was “stealing our money to escape with its officials,” and urged followers to rise up against the regime.

At least two of the observed posts which depict long lines of people waiting at the bank and beginning to riot show elements of AI generation. More specifically, in the example video shown below, one of the bodies blurs and becomes misshapen, suggesting the footage was AI generated.

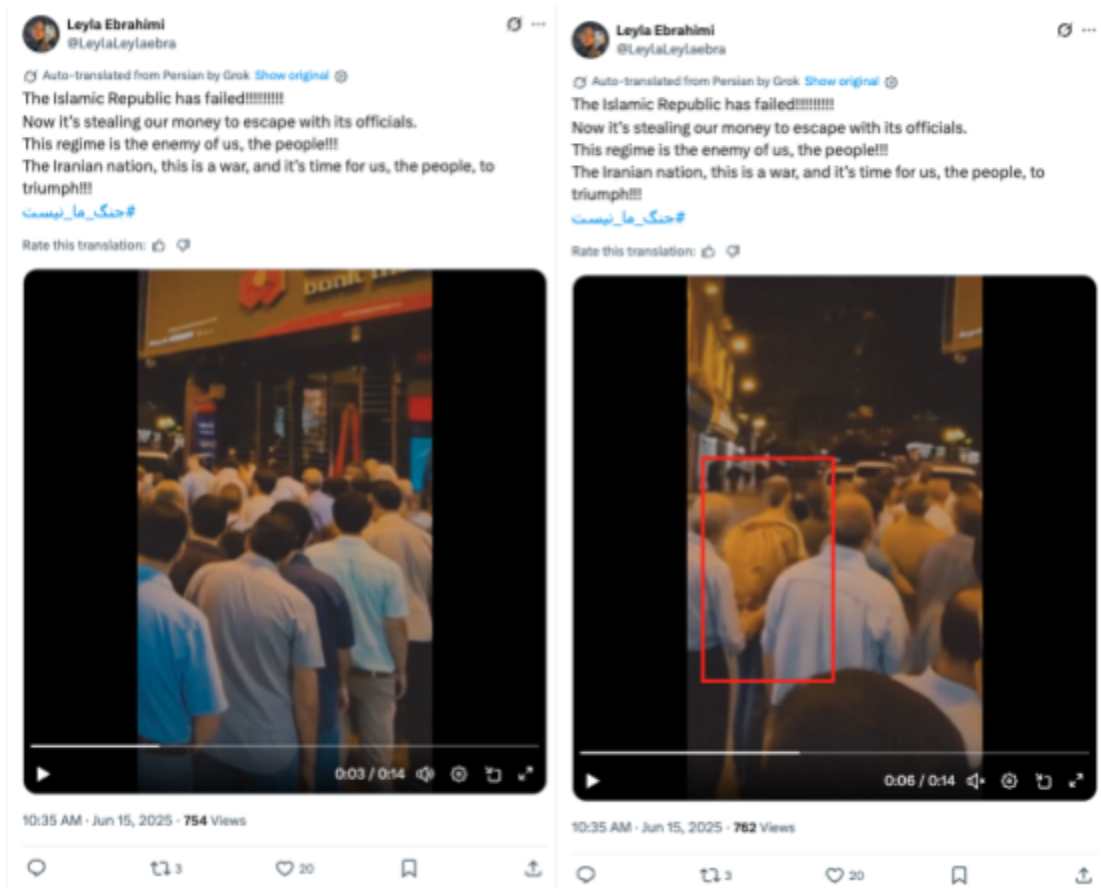


Figure 9: A [post](#) shared by the PRISONBREAK network on June 15, showing a video of people waiting in line at ATMs in Iran. The misshapen figure seen in the screenshot on the right is an indicator of the video being AI generated.

Between June 20 and June 22, 2025, PRISONBREAK intensified its rhetoric against the Iranian regime. Under the hashtag “8 o’clock cry”, the accounts urged followers to get on their balconies at 8 p.m. each evening and shout “Death to Khamenei” (مرگ بر خامنه ای), in an attempt to capitalize on a previously [established form of protest](#) in Iran. An example of one of the posts shared by the network is shown in the figure below.



Figure 10: A [post](#) shared by the PRISONBREAK network on June 20, encouraging followers to protest against the Iranian regime by shouting “Death to Khamenei” at 8 p.m. every night.

We observed at least nine videos and 23 posts shared by PRISONBREAK between June 20 to 22 of citizens allegedly participating in the “8 o’clock cry”. The videos relating to the “8 o’clock cry” have viewership well outside their normal rates, ranging from 20,000 to 60,000 views per post. Almost all of the videos are low quality, and do not include visible images of people shouting. Based on evidence outlined in the next section, we are able to conclude that at least some of the videos were manipulated. As we could not identify them as posted anywhere else, we conclude that the videos were likely created by the account operators.

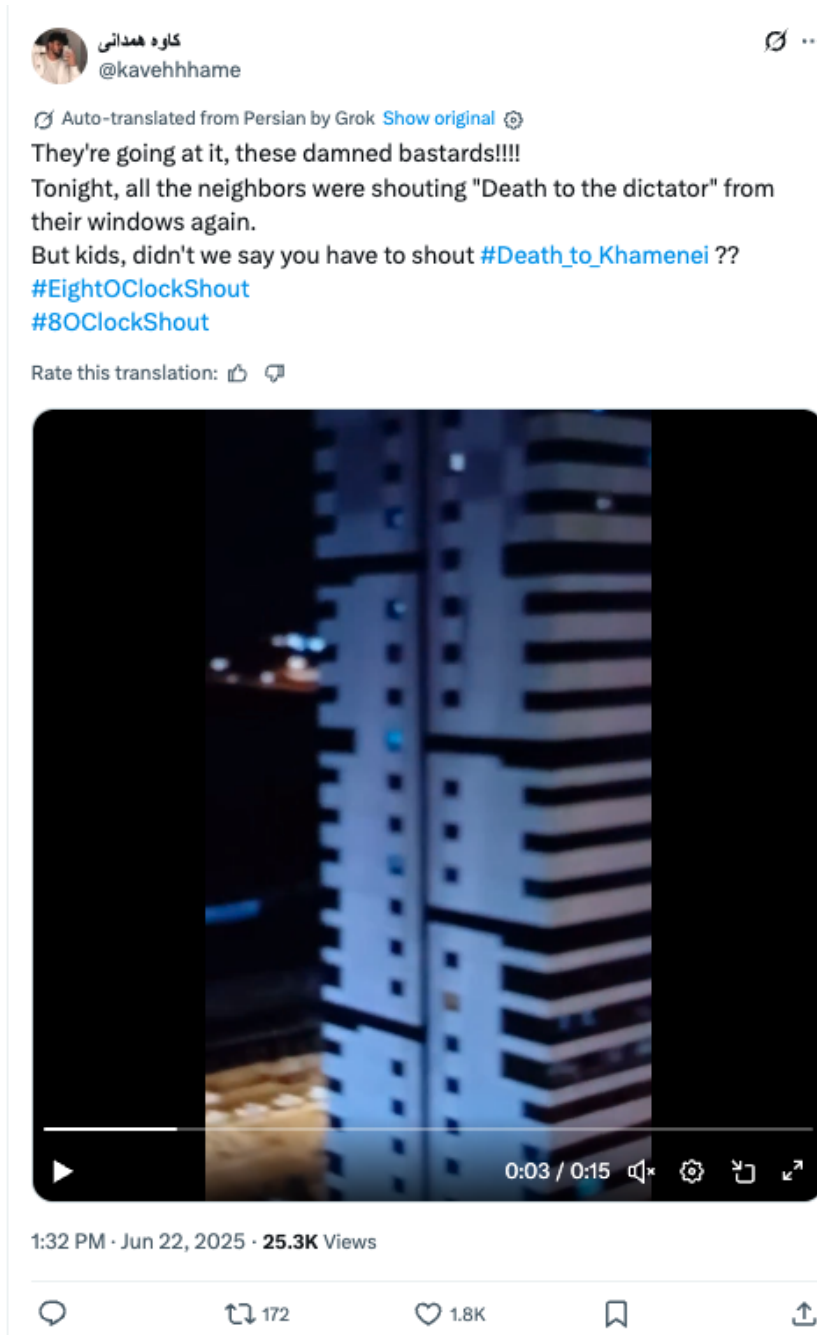


Figure 11: A [post](#) shared by the network on June 22, encouraging followers to protest against the Iranian regime by participating in shouting “Death to Khamenei”. While the viewership exceeds 25,000, the post has limited engagement.

These posts appeared to be a buildup to the Evin Prison deepfake, again pushing narratives promoting regime change in Iran. The network failed² to drive Iranians to Evin Prison to free the political prisoners, but that did not stop their dissemination of content promoting an uprising against the country's leadership. On June 24, Israel and Iran agreed to a [ceasefire](#). The PRISONBREAK network then appeared to make another push to trigger unrest by questioning the ceasefire, suggesting that Iranian citizens had not received their salaries.



Figure 12: Composite of posts posted by accounts in the network on June 24. The posts on the left show the accounts questioning the ceasefire and urging followers to continue with overthrowing the regime. The posts on the right show the accounts posing as Iranian citizens who have not received their salaries.

The PRISONBREAK operators subsequently pivoted to content related to the country's ongoing water and energy crisis. Iran has faced a five-year drought and, in July, government officials announced that Tehran's [water supply](#) was likely to run out in a matter of weeks. The water crisis and [energy shortage](#) made life increasingly difficult for Iranians following the Twelve-Day War. It appears that the network sought to escalate these tensions by creating and sharing content related to these issues. At the time of writing this report, PRISONBREAK is still consistently posting about both the water crisis and energy

² According to [this New York Times](#) report, in the afternoon of June 23, Iranian security forces swarmed Evin Prison, took control of the area, and transferred prisoners to different detention facilities. Prisoners were then [returned](#) to the Evin Prison in August.

shortage, in a likely attempt to continue to escalate tensions between Iranian citizens and their government. We share two examples of this content posted by the network below.

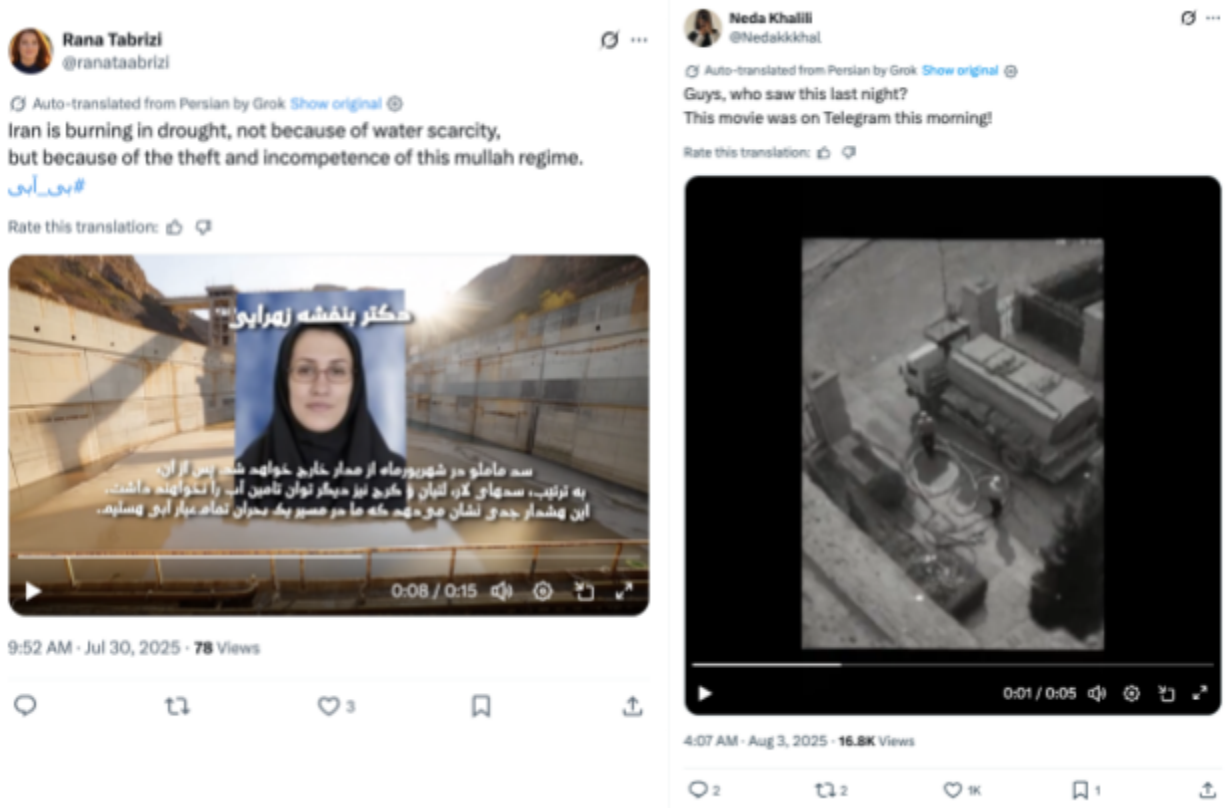


Figure 13: Two posts shared by PRISONBREAK relating to the water and energy crises. The post on the left shares a [video](#) quoting Dr. Banafsheh Zahraei, a water resources expert, about how one of the main dams in Tehran will run out of water by the end of September. The post on the right shares what appears to be an AI generated [video](#) – based on the wall unnaturally moving as the gate opens in the top right corner – of officials stealing resources.

A Synthetic Network

The Pervasive Use of AI

The Evin Prison deepfake was not an anomaly for the PRISONBREAK network's use of AI, which appeared to be routinely used in the operation. In one occasion, the network utilized AI to distort the lyrics of a known Iranian protest song, and accompany the manipulated sound with deepfake representations of three Iranian singers.

Starting in August 2025, several confirmed network accounts shared a Youtube video³ impersonating famous Iranian singers Mehdi Yarrahi (مهدی یراحی), Toomaj Salehi (توماج صالحی), and Shervin Hajipour (شروین حاجی پور). All three singers live in Iran and have spent [time](#) in [prison](#) for their support of the 2022-23 “Woman, Life, Freedom” protest movement. In particular, Rapper Toomaj Salehi was severely [tortured](#) during his incarceration and [sentenced](#) to death in April 2024, a sentence later [overturned](#). In the video, they allegedly perform a joint interpretation of Hajipour’s song “[Baraye](#)” which criticized social and political realities in Iran and became an [anthem](#) of the protest movement. The song in the video, however, changes the [lyrics](#) of the original, turning it into a direct call for an uprising. The main chorus in the manipulated version of the song roughly translates to “revolution for life, revolution for hope, revolution for freedom” while discussing the ongoing water crisis in Iran and other perceived or real grievances of the Iranian population.

The video and a short extract from it were the only ones posted on the YouTube channel⁴ “جنگ زندگی” (“The War of Life”) – which was created the same day it uploaded the song (August 13, 2025). The clip features several distinct AI failures – including a person walking backwards (who appears to be a “double” of Hajipour), crowds of people moving as a singular sliding unit, and misplacement of Toomaj Salehi’s tattoos (either duplicated or not present). We provide one example of the AI mistakes below.

³ <https://www.youtube.com/watch?v=FNCXszeRQYE>

⁴ <https://www.youtube.com/@%D8%AC%D9%86%DA%AF%D8%B2%D9%86%D8%AF%DA%AF%DB%8C/videos>

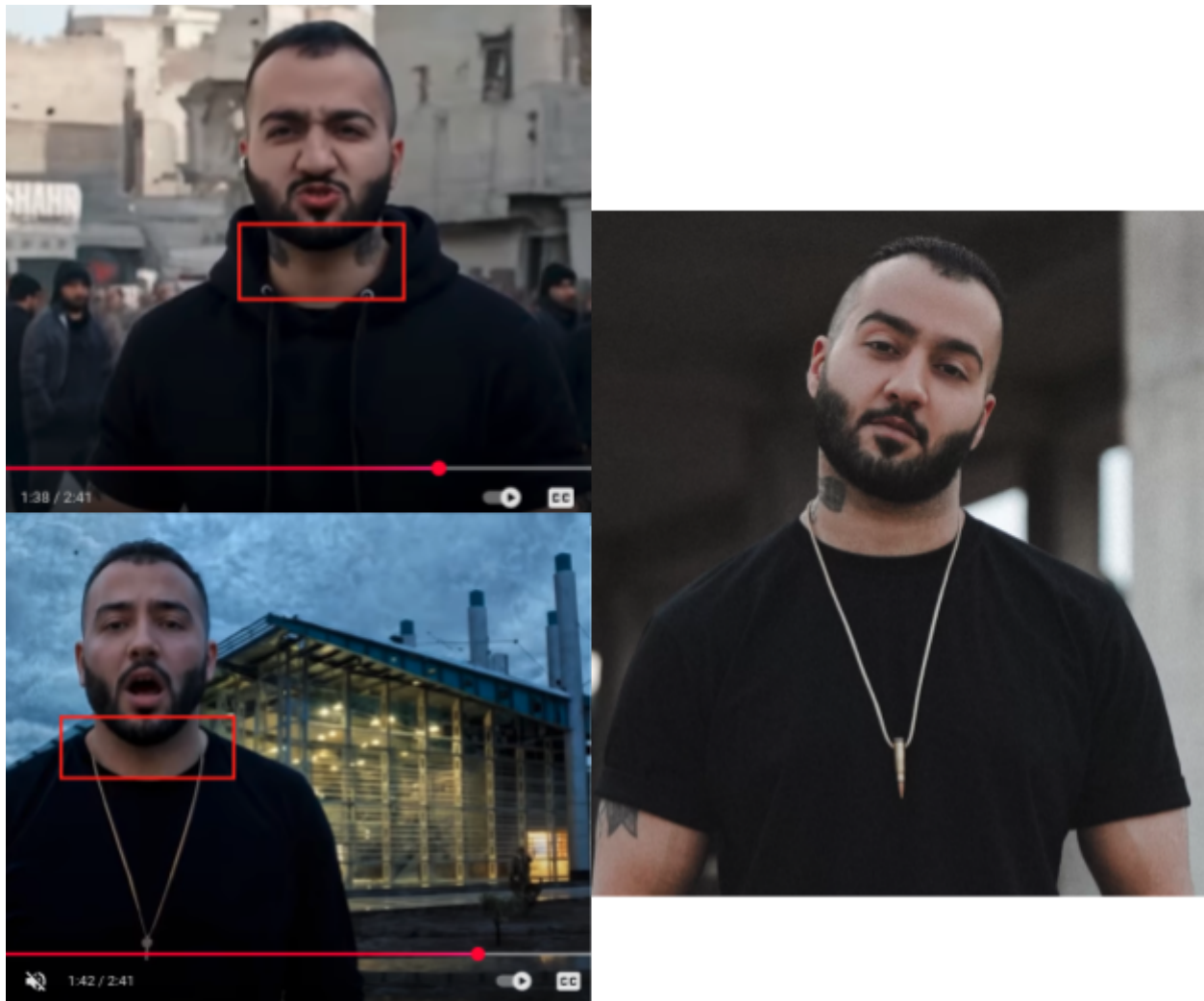


Figure 14: The two screenshots on the left show clips from the inauthentic YouTube video shared by the network. In one clip, Salehi’s neck tattoo is duplicated by AI generation and the other clip shows Salehi with no neck tattoo. The photo on the right is a real [photo](#) of singer Toomaj Salehi.

At the time of writing this report, the YouTube account has 287 subscribers and the video has 466 views. We observed the PRISONBREAK network sharing the [video](#) on X one day after it was posted on YouTube, August 13, 2025, and [seeding](#) the link into Twitter Communities with over 15,000 members.

This video was just one of the many examples of PRISONBREAK spreading AI-generated imagery emphasizing the water crisis. At times, such content was of a caricatural nature. One post from the confirmed network account @feridounazari shared an AI [video](#) depicting Khamenei stealing water from a baby. Another account, @firoz_soltani, shared a [video](#) of water bottles attached to a timer meant to resemble a bomb. When the timer explodes, the caption reads “Water is our inalienable right, thirst is

the end of our life.” Regardless of the post, the goal seemed to be to create and share content that deepens the political discontent in Iran and pushes the population towards revolting against the regime.

Impersonation of National and International News Outlets

Along with AI-generated content, the PRISONBREAK network impersonated legitimate news sources. For example, we identified one screenshot and two videos impersonating BBC Persian. In the caption to a screenshot of an alleged article, the original poster @KarNiloufar claimed that BBC Persian deleted it within five minutes of posting it. The headline of the article translates to “Officials flee the country; High-ranking officials leave Iran one after another”. We have verified with BBC Persian that this article was never published by the broadcaster.



Figure 15: A [post](#) shared by @KarNiloufar on June 16, 2025 showing a screenshot from an alleged BBC Persian article. The purported article claims that 90 members of the Iranian regime have fled the country. The BBC confirmed to us that they had never published this article.

A few days after sharing the alleged article, the network posted two other videos they claimed to have been published by BBC Persian. The first [video](#) summarizes the events of the Twelve-Day War, specifically highlighting heavy traffic, gas shortages, and suggesting that Iran's government and allies have abandoned the regime. The second [video](#) begins with a title that translates to "Special daily summary of the attack on the regime's repressive forces," and shares the network's own Evin Prison deepfake, as well as other alleged explosions in Iran.

The BBC confirmed to us that they did not create or publish this content.⁵

We provide screenshots of the two newsclips fabricated by PRISONBREAK below.



Figure 16: The two screenshots show videos posted by @feridounazari and @KarNiloufar claiming to originate from the BBC.

BBC Persian was not the only media outlet impersonated. We observe two other instances of media impersonation with the accounts providing links to articles from Afkar News (پایگاه خبری افکار نیوز)

⁵ A BBC spokesperson stated: "We urge everyone to check links and URLs to ensure they are getting news from a trusted source."

accompanied by a screenshot of the headline. In both instances, the links do not lead to a news article, nor were we able to locate archived versions.



Figure 17: The two screenshots show articles shared by PRISONBREAK and linked to Afkar News that do not appear on the website. The headline on the [left](#) translates to “The Sarallah command fell, and the center of the Revolutionary Guards' suppression mechanism in Tehran was destroyed,” and stating “the place synonymous with fear has now been reduced to ashes”. The headline on the [right](#) translates to “Mr. Khamenei - Why have you been silent?” and criticizes his handling of the Twelve-Day War.

Recycled or Intentionally Misleading Content

AI-generated videos were just one of the tactics employed by PRISONBREAK. We observed several instances of videos edited and shared to mislead viewers about protest activity occurring in Iran. In three of the videos shared, the accounts imply they have filmed the videos in their neighbourhood. We found that these videos were recycled content – and not authentic as the users claimed – and had been posted by accounts that appear to be unrelated to PRISONBREAK almost a week before. In addition, two of the videos were cropped and edited from the original, possibly to prevent discovery of its reuse. We additionally cannot verify the authenticity of the original videos.

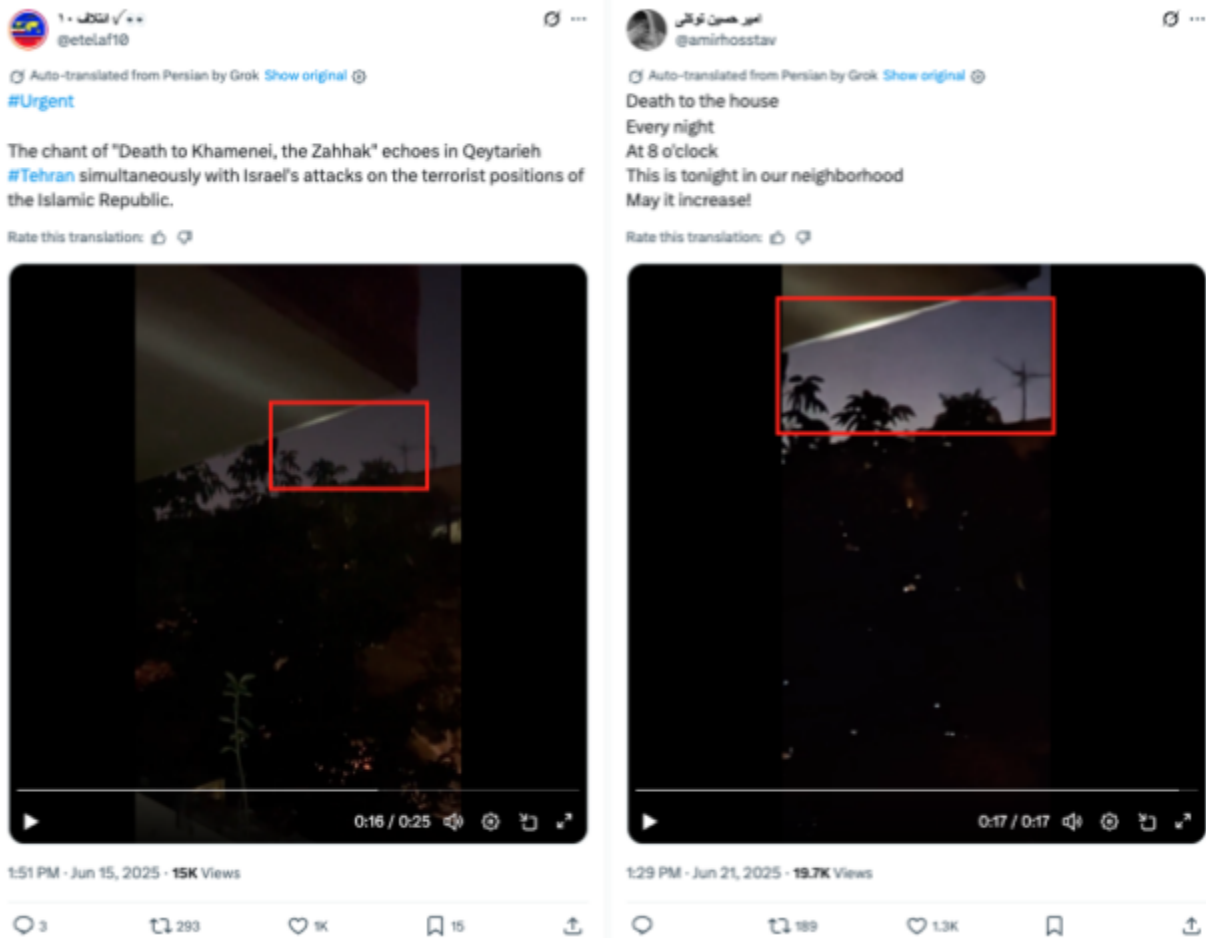


Figure 18: Two accounts posting the same video of a neighbourhood chanting “Death the Khamenei”. The account on the left posted the [video](#) on June 15 and is not part of our observed network. In the screenshot on the [right](#), PRISONBREAK account @amirhosstav posted the same video on June 21, claiming it was taken in its neighbourhood. Although the video has been cropped and colour enhanced, the roof, light source, foliage, and antenna are direct matches.

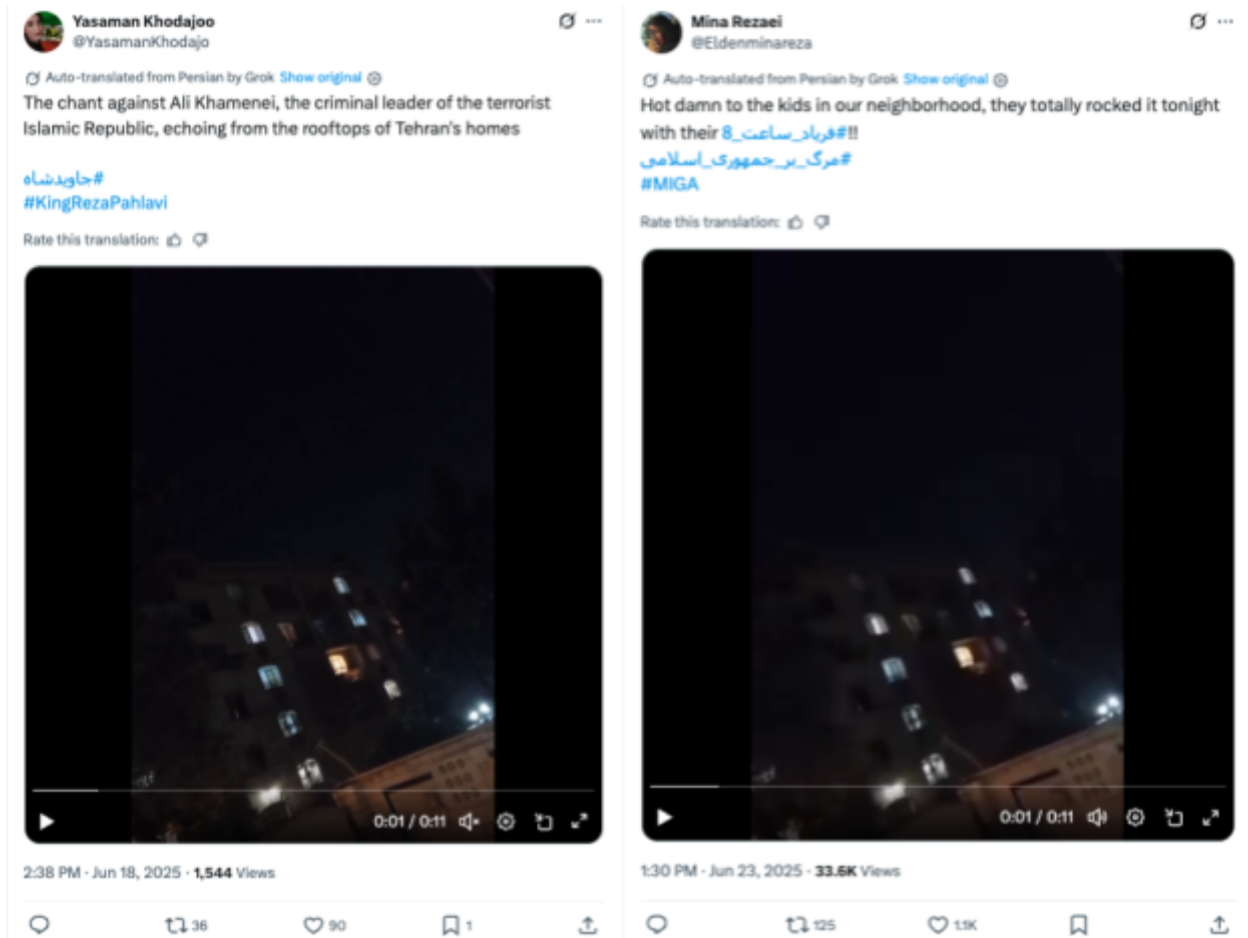


Figure 19: Two accounts posting the same video of a neighbourhood allegedly chanting “Death the Khamenei”. The account on the [left](#) posted the video on June 18 and is not part of PRISONBREAK. In the screenshot on the [right](#), confirmed network account @Eldenminareza posted the same video on June 23, claiming it was taken in its neighbourhood.

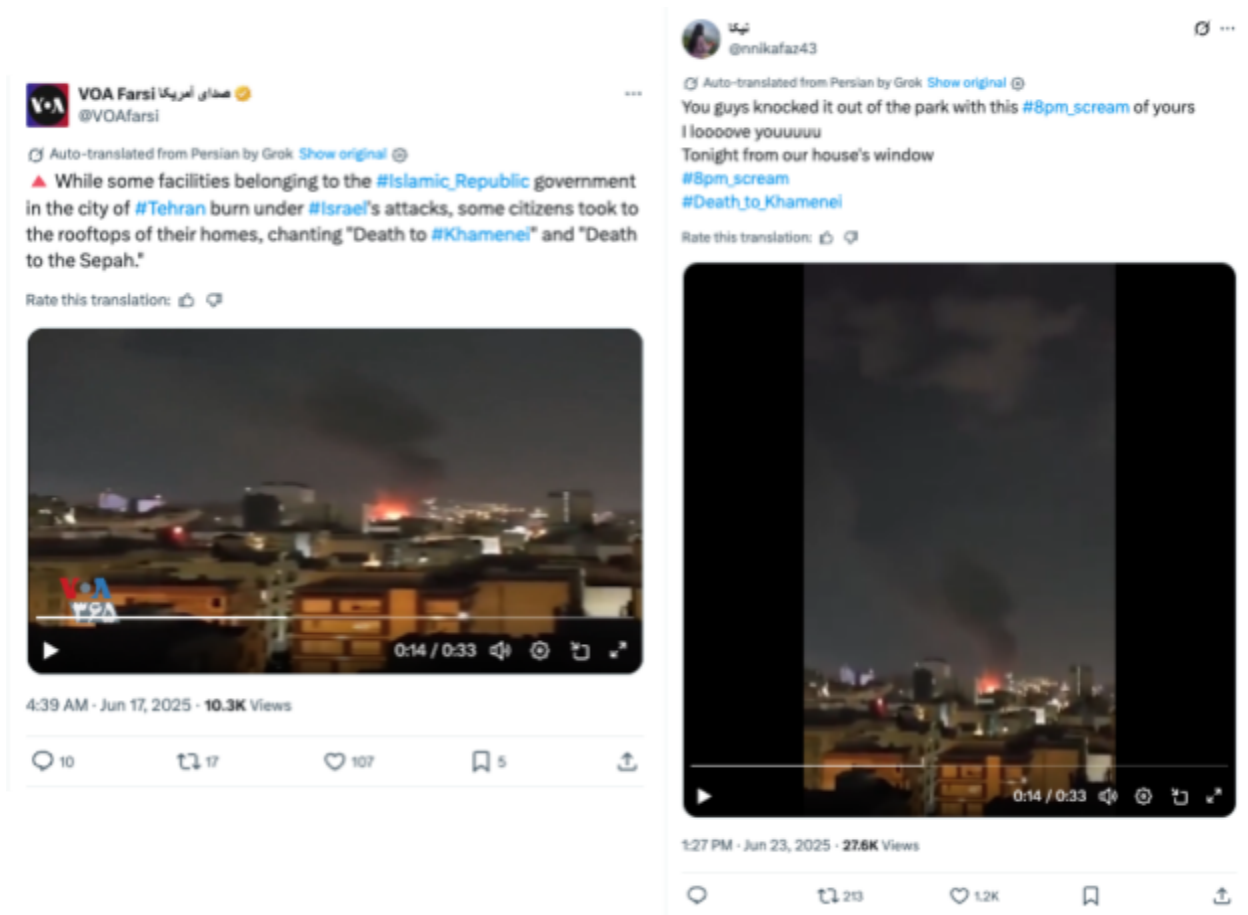


Figure 20: Two accounts posting the same video of a neighbourhood chanting “Death the Khamenei”. VOA Farsi – a U.S. affiliated media organization – shared the [video](#) on June 17. The account on the right, PRISONBREAK account @nnikafaz43, posted the same [video](#) on June 23, claiming it was taken in its neighbourhood. The VOA video has been cropped, but the audio, video length, and content are the same.

The operation not only recycled content, but possibly edited videos to fit their narrative. In the example below, the video loops several times, while the audio does not. The lack of continuity suggests that the audio has been added overtop of the video, pointing to its manipulation. The picture is also grainy, and appears to be filmed from another screen. Most notably, the audio appears to be a cut and processed version of the audio in both the VOA Farsi video and the video shared by @nnikafaz43 in the example above, suggesting that the PRISONBREAK operators have intentionally cut audio from different videos and placed it on new content.



سپیده فرخزاد
@Sepideh7895



Visibility limited: this Post may violate X's rules against Violent Speech. [Learn more](#)

Auto-translated from Persian by Grok [Show original](#)

Hooray, hooray, hooray. Death to Khamenei at 8 PM tonight!!! Hooray!!!!

#مرگ_بر_جمهوری_اسلامی

#مرگ_بر_خامنه‌ای

Rate this translation:



1:18 PM · Jun 20, 2025 · 50.3K Views



1



286



3.1K



1



Figure 21: A [post](#) shared by the network on June 20th, showing a video of a neighbourhood where people are allegedly yelling “Death to Khamenei”. The audio present is a shortened version of the VOA Farsi [video](#) and appears to have been edited and placed over top of the recording.

Some edits were even less sophisticated. In the example below the “protest” audio pauses at 9 seconds, while the video continues on. At 10 seconds someone can be heard speaking in the background – presumably the creator of the video – followed by a few seconds of silence before the intended audio begins again at 12 seconds. The pause in audio, and the grainy video quality, signals that the video and audio were combined from two separate sources, and not recorded as the account claims.



Figure 22: A [post](#) shared by PRISONBREAK on June 20, showing a video of a neighbourhood where people are allegedly yelling “Death to Khamenei.” The audio pauses for about 3 seconds in the middle of the video suggesting the audio has been overlaid with a recording.

Artificial Amplification

Many of the recently discovered influence operations have relatively low rates of [engagement](#). The PRISONBREAK operators employed several tactics to help increase the rate of engagement and viewership. We outline them below.

Seeding of Content into X Communities

As [described](#) by X, Communities are “a dedicated place to connect, share, and get closer to the discussions they care about most.” They essentially consist of discussion groups that can be created and run independently by an X user.

Relative to this influence operation, Communities allowed for the network to spread their message of discontent to people who are already interested in similar topics. We observe PRISONBREAK sharing posts into communities specifically targeted at anti-Iranian regime groups including:

- [Constitutionalist Association Twitter](#) (انجمن مشروطه خواهان توییتر) with 2,900 members
- [Iranian Youth Union](#) (اتحاد جوانان ایران) with 6,700 members
- [Jina Mahsa | Iran's Revolution](#) (ژینا مهسا | انقلاب ایران) with 5,400 members.

The network also frequently shared posts into a Persian-language “[Follow Back](#)” community with 17,600 members. This type of Community typically offers a rapid (and artificial) way for an X user to achieve a follower base by promising mutual “follows” for all its members.

By using X Communities to spread their content, PRISONBREAK artificially inflated the viewership on several posts, increasing their chances of influencing their target audience through the community members’ resharing.

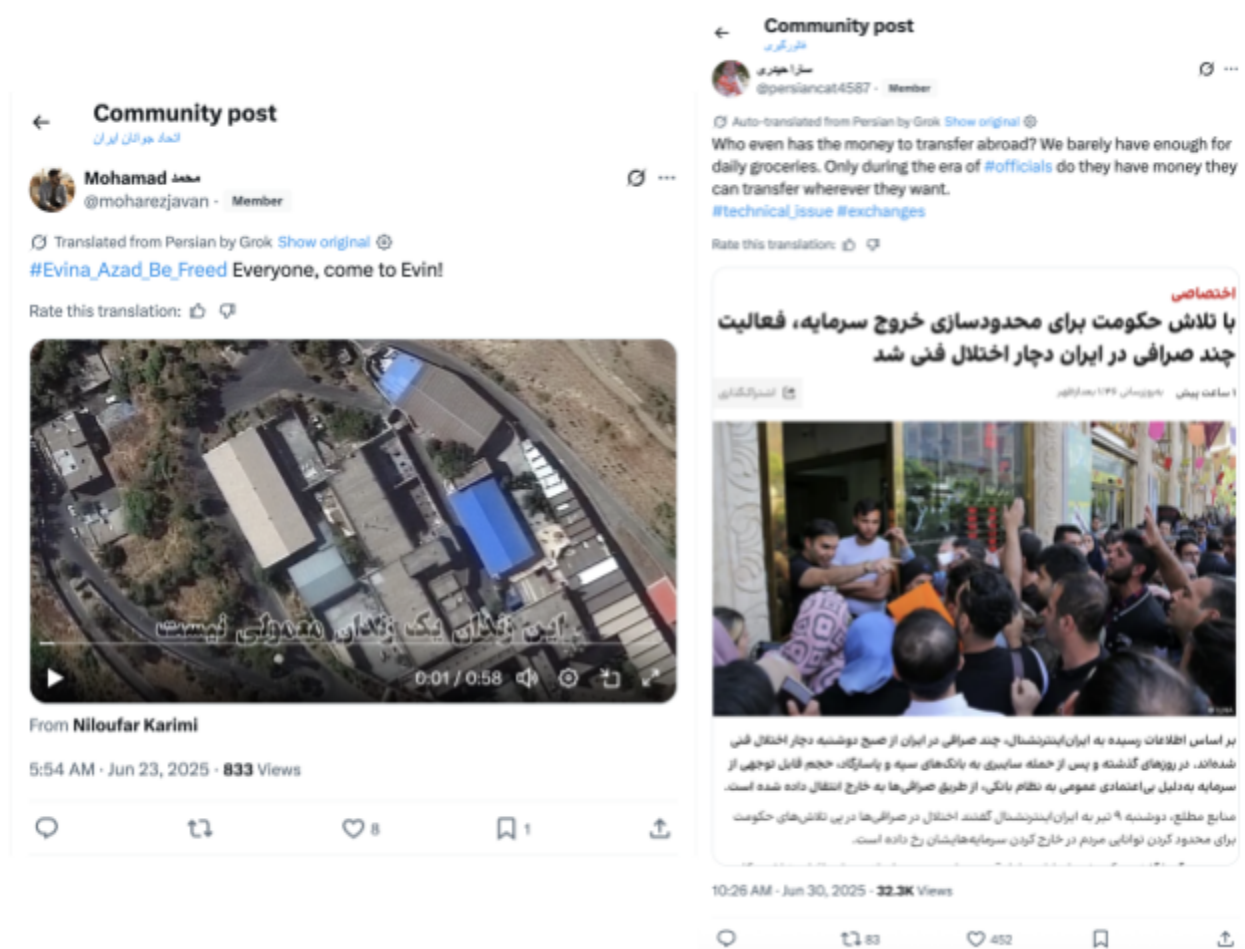


Figure 23: Two posts shared by PRISONBREAK into X Communities. The left [post](#), shared by @moharezjavan, is an Evin Prison video originally shared by @KarNiloufar. This post was shared with the Iranian Youth Union (اتحاد جوانان ایران) community, which has 6,700 members. The right [post](#), shared by @persiancat4587, highlights the economic struggles facing Iranians during the Twelve-Day War. The post was shared with the “Follow Back” community (17,600 members).

Mutual Resharing

In addition to posting in X Communities, PRISONBREAK frequently reshared the network’s own posts – especially their likely AI-generated content. This appears to be an attempt at self-amplification by the network.

At least seven different PRISONBREAK accounts reposted this [video](#) – originally posted by @Sepideh7895 – which we assess above to have been edited with repurposed audio. At least nine PRISONBREAK accounts reposted, quoted, or replied to other posts with the AI-generated

[photos](#) of Khamenei in an ambulance – originally posted by @azad_leylaas, another PRISONBREAK account. This technique can be observed on nearly all of the network’s accounts, in an attempt to spread their content to as many users as possible and increase the effectiveness of their covert influence campaign.

Tagging Mainstream Media

While PRISONBREAK focused on amplifying their posts to audiences on X, they also tried to gain the attention of mainstream media. We observe several instances of PRISONBREAK accounts tagging popular news outlets like BBC Persian, VOA Farsi, and Iran International in videos shared by the network. Given the success of the Evin Prison deepfake in tricking mainstream media, tagging news accounts on other videos allegedly produced by the network demonstrates a persistent attempt to deceive the media with inauthentic Twelve-Day War content in line with the network’s narratives. We provide two examples of these efforts below.

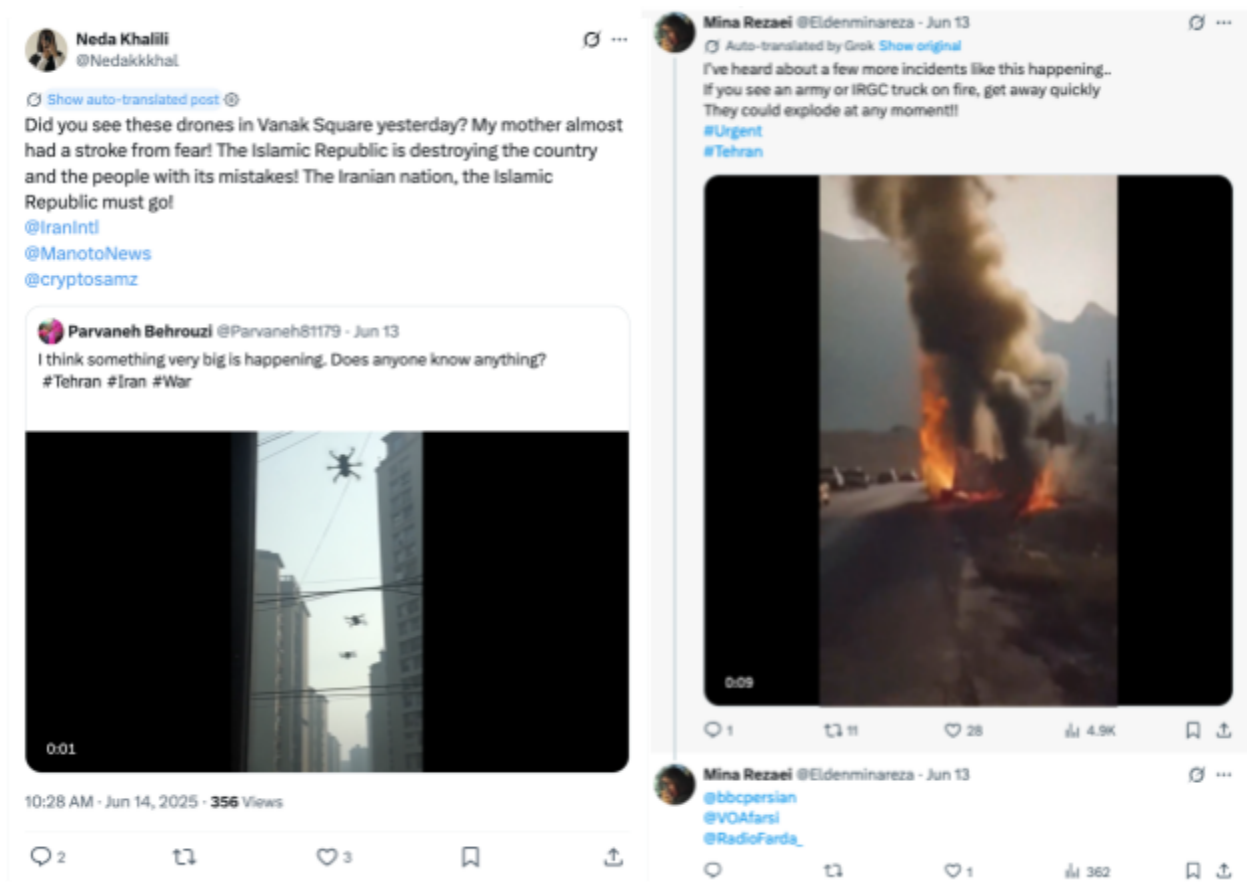


Figure 24: Two posts shared by PRISONBREAK tagging mainstream media outlets. The post on the [left](#), both originally shared and reposted by PRISONBREAK accounts, attempted to gain the attention of news accounts Iran International and Manoto News. The post on the [right](#), shared by a confirmed PRISONBREAK account, tried to gain the attention of BBC Persian and VOA Farsi.

Inauthentic Accounts and Personas

We assess with high confidence that the PRISONBREAK network is both coordinated and inauthentic. It does not represent existing individuals, and was almost certainly orchestrated by a professional team, following a specific playbook.

We reach this conclusion through the analysis of a series of strong indicators that we summarize below.

Registration Email Domains

Summary: The network’s frequent use of a possibly obscure and/or proprietary email domain for account registration points strongly towards the conclusion that the accounts are closely coordinated.

Using Open Source Intelligence (OSINT) methods, we observed that approximately half of the X accounts in the network registered using the email domain @b*****.***. Among them was also @KarNiloufar, the account who posted the AI-generated video of the Evin Prison bombing.

We were not able to conclusively identify the email domain used for the accounts. We note that it does not appear to be one of the most commonly used webmail services – i.e. Gmail, Outlook, Hotmail, or Office – and the character count does not match other popular services starting with “b”, such as for example btinternet.com.

Lifespan vs. Period of Activity

Summary: While the accounts were all created in 2023, 99.5% of their posting happened after January 2025. This strongly signals both their coordination and inauthenticity.

All confirmed accounts in this network were created between February and December 2023. However, only a handful of posts were made before the network ramped up its activity in January 2025 and in the following months.

Specifically:

- The network published only approximately 0.5% of the estimated total number of posts before 2025.
- Such posts exclusively consisted of reposts. There was no originally created content being posted.

The network came to life – in terms of its posting activity – beginning in January 2025, possibly responding to a specific directive, and in coincidence with the escalation of geopolitical tensions that eventually led to the Twelve-Day War in June 2025.

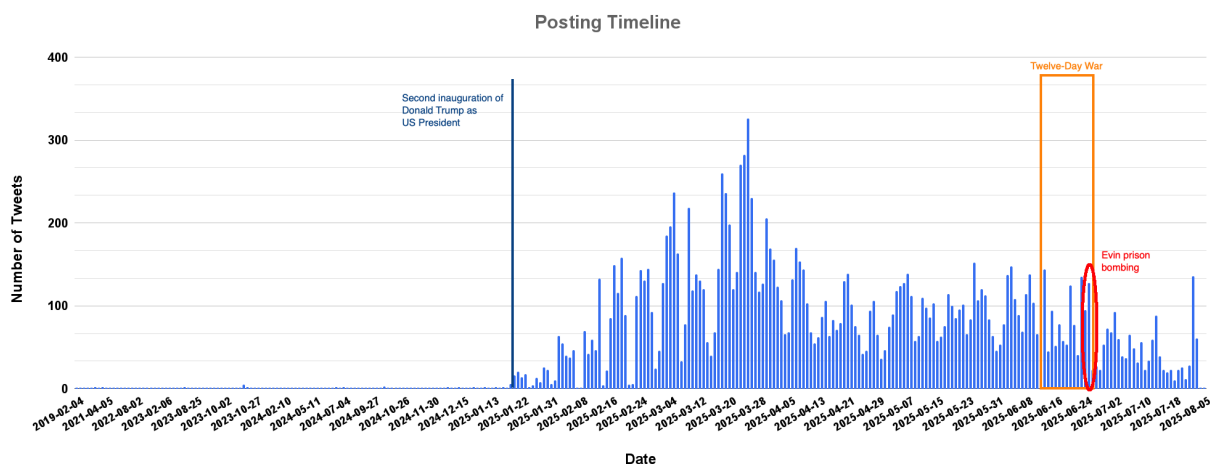


Figure 25: Posting timeline for the network. We can observe the activity picking up in January 2025 with some notable peaks (e.g., March 2025). Note: in favour of visualization, we omitted dates when the accounts did not post.

Posting Time Patterns

Summary: Posting time patterns are consistent with those of a professional, dedicated team of operators producing posts as part of a regular work day.

A review of the posting time patterns for the network – i.e. the number of posts published in total by the accounts in each hour of the day – shows that the network:

- Typically begins ramping up activity at 6 a.m. UTC (9:30 a.m. Tehran time) each day.
- Significantly escalates the volume of posts starting at 7 a.m. UTC (10:30 a.m. Tehran time).
- Begins winding down posting activity after 3 p.m. UTC (6:30 p.m. Tehran time).
- Has minimal average posting activity between 9 p.m. UTC (12:30 a.m. Tehran time) and 6 a.m. UTC (9:30 a.m. Tehran time).

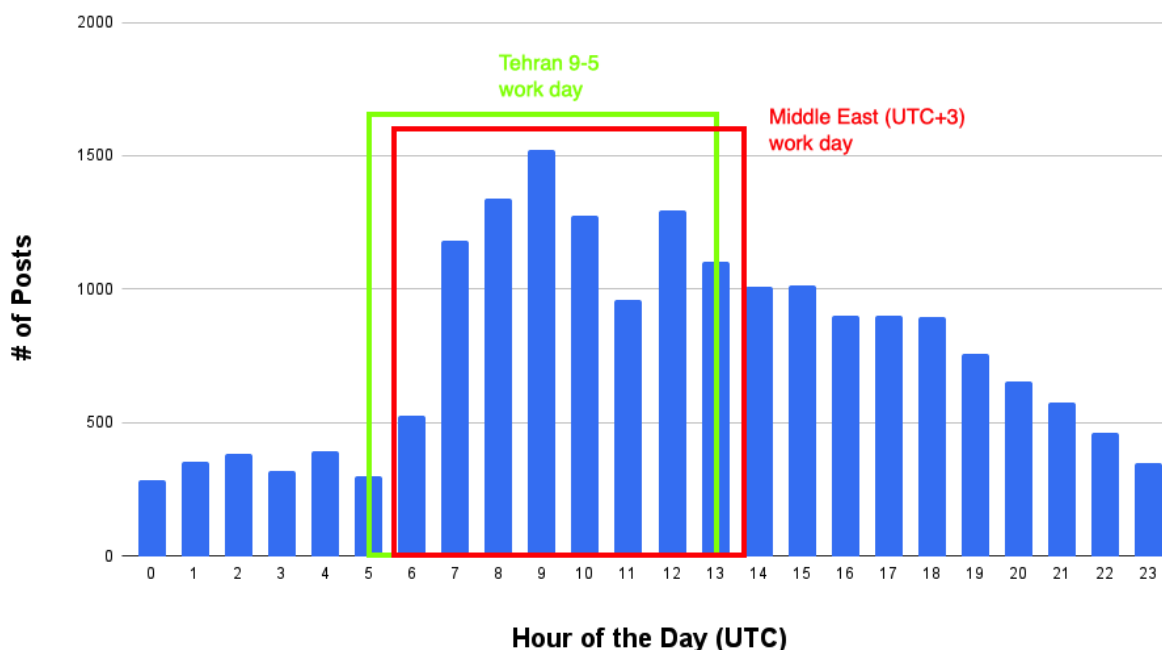
Amount of PRISONBREAK X Posts per Hour of the Day (0-23 UTC)

Figure 26: Chart showing the number of posts posted by the network, per hour of the day (0-23 UTC).

This behaviour approximately coincides with a regular work day in the Iranian timezone ([Iranian Standard Time or IRST](#), equivalent to UTC+3:30), and even more so in UTC+3, the time zone applying to most of the Middle East (as well as Eastern Europe and parts of Eastern Africa), including Israel.

It is, however, inconsistent with an organic community of users spontaneously posting at random times of the day.

Application Version

Summary: The network's consistent and prevalent use of X's desktop application to post its content is more consistent with the work of a professional team than with the activity of an organic community.

We estimate that at least 75 percent of the posts produced by the network were posted through X's desktop interface (formerly known as Twitter Web App). This information is visible in the network response provided when loading a given post in a web browser, using the browser's Developer Tools or equivalent name.

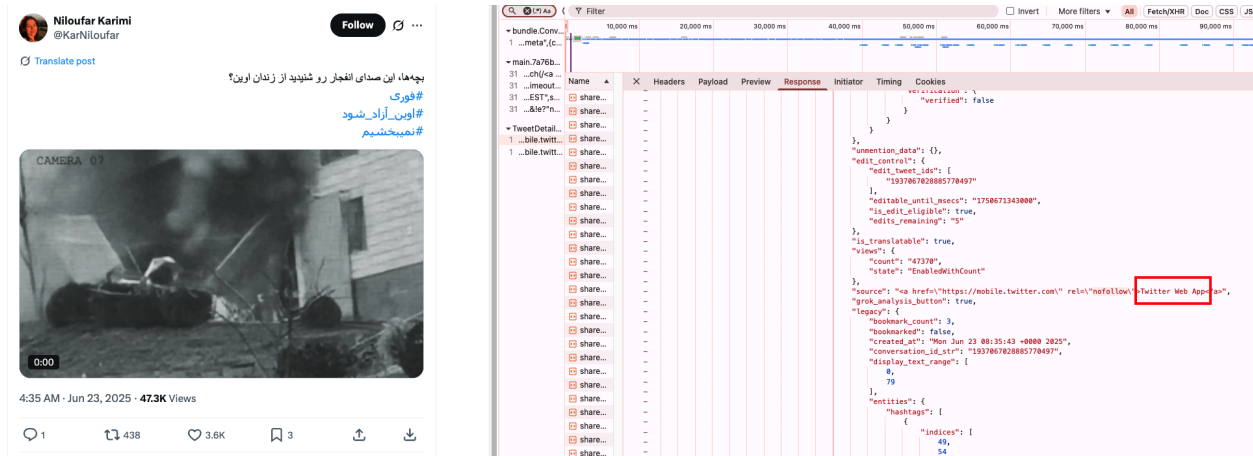


Figure 27: Screenshot from Chrome's Inspect module for the [post](#) by @KarNiloufar including the AI-generated video of the Evin Prison bombing. The highlighted string shows the application used to publish the post (Twitter Web App).

Like the posting timing patterns, the almost exclusive use of the web version of X – a mobile-first platform like other social media – is consistent with a professionally run network of digital assets, where a team is assigned to operating one or more of the accounts through dedicated workstations. Similar behaviour has been previously [used](#) as a signal to identify IOs.

Profile Photos

Summary: The network's profile pictures are often faceless, sourced from stock images, or appear to be lifted from popular blog sites like Pinterest, a strong signal of inauthentic persona development.

We reviewed the profile photos of all accounts associated with the network and found that 30 out of 53 have profile photos where their face is partially or completely obscured. Two of the profile photos were found on various stock photo websites. A reverse image search also showed that a large majority can be found on Pinterest, a popular photo sharing site. The use of stolen or anonymous photos is a [common tactic](#) used by IOs to protect the anonymity of the actors behind the operation, and in combination with other elements that we outline here, is a strong signal of inauthenticity.

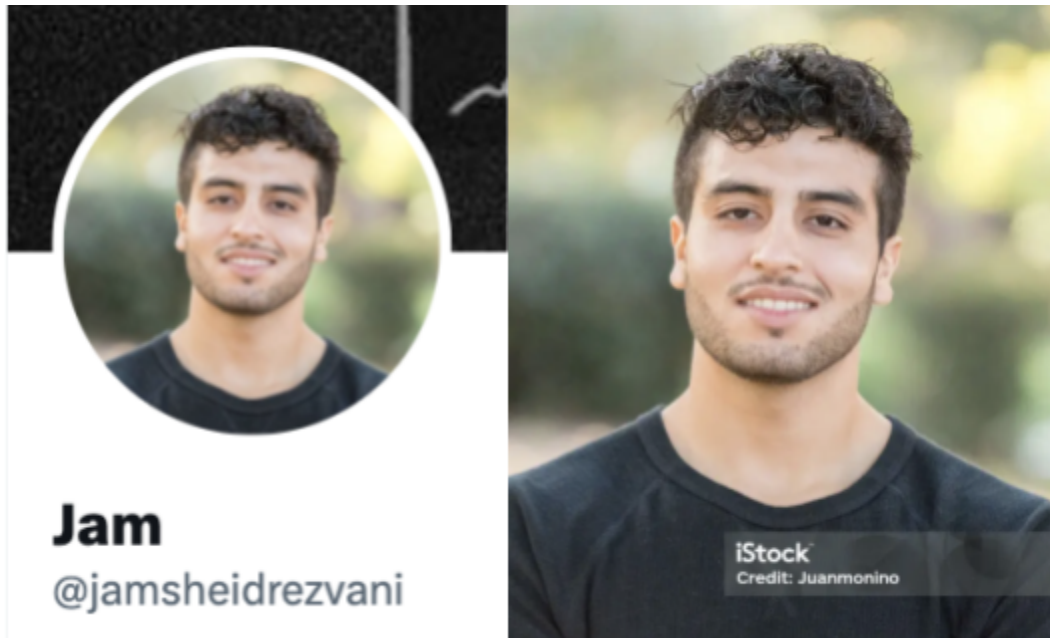


Figure 28: The screenshot on the left shows the profile photo of confirmed network account @jamsheidrezvani. On the right we share the original stock photo found via reverse image search.

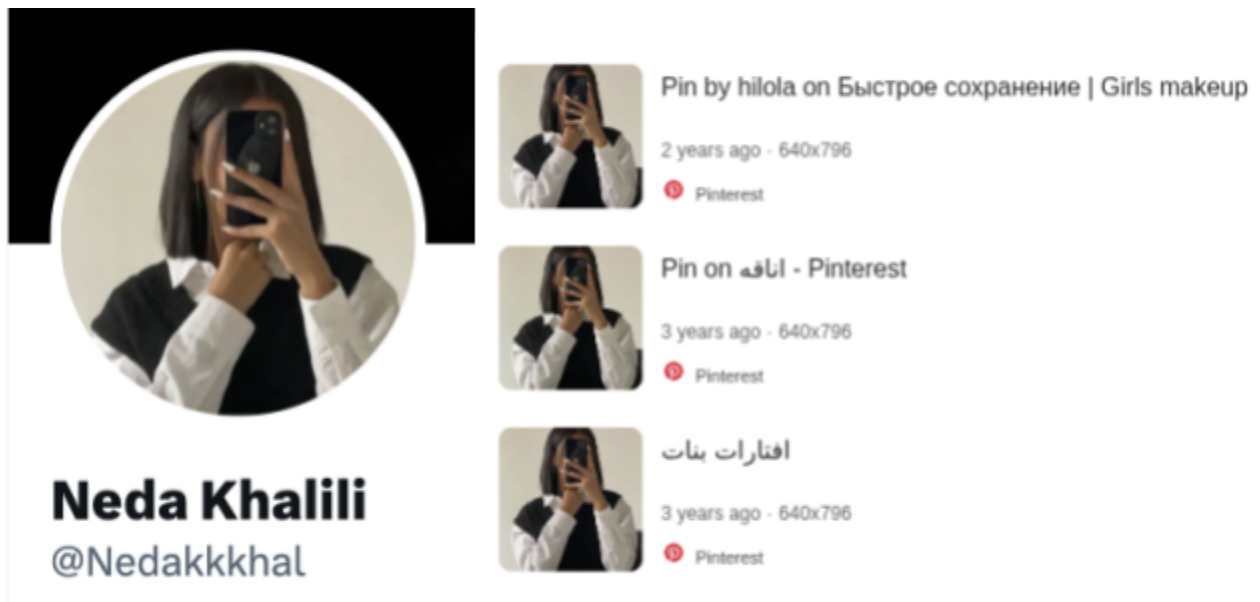


Figure 29: One example of a profile photo of confirmed network account @Nedakkkhal that can be found shared multiple times on Pinterest. The photo is also an example of the anonymous profile pictures employed by the network.

The Deepfake Account: @TelAviv_Tehran

We identified one X profile that appears to be external from PRISONBREAK, but displays several signals of likely coordination with it. It is an account promoting closeness between the Israeli and Iranian populations, and explicitly using AI to manufacture and disseminate calls to revolt against the Iranian regime and its leaders. This tactic aligns it with PRISONBREAK's modus operandi and objectives.

@TelAviv_Tehran is an X account (also active on Instagram⁶ with the same username) explicitly⁷ using a deepfake persona – a purported female reporter from Iran, whose appearance and speeches are fully AI generated – to post videos deriding the Islamic Republic's leadership.

As hinted in the username, @TelAviv_Tehran promotes closeness between the Israeli and Iranian populations, where the latter is depicted purely as an oppressed victim of the Islamic regime. Like PRISONBREAK, @TelAviv_Tehran openly appeals for the Iranian people to rise against the Islamic regime.

⁶ [https://www.instagram\[.\]com/telaviv_tehran](https://www.instagram[.]com/telaviv_tehran)

⁷ @TelAviv_Tehran states in this [post](#) that their reporter “Sarina” is AI-generated.

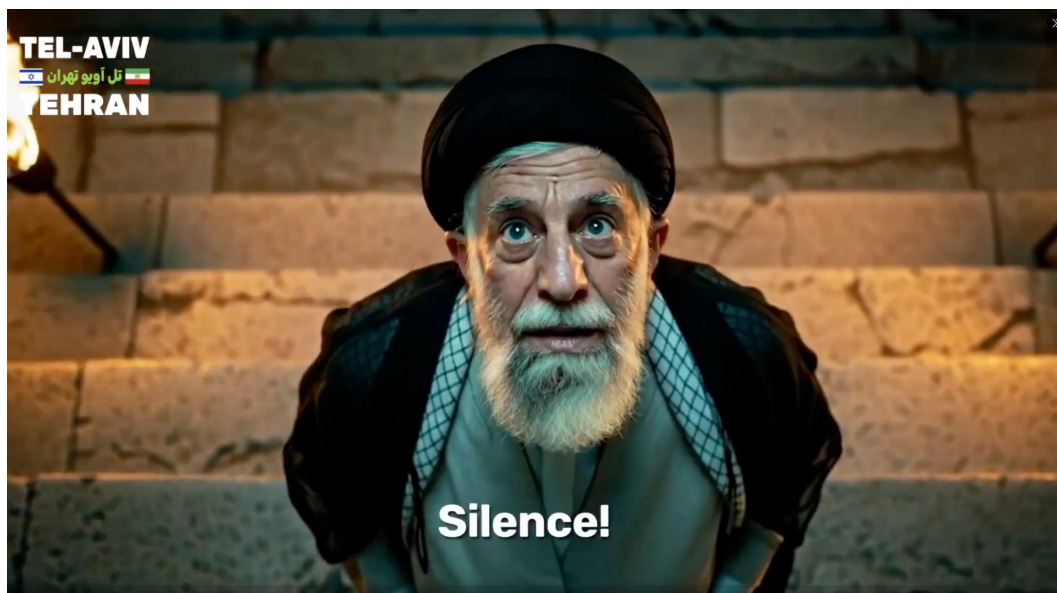


Figure 30: Two stills from [separate videos](#) posted by @TelAviv_Tehran and showing an AI-generated animation of Ali Khamenei, supreme leader of Iran, either in hiding (first example) or as a prisoner (second example). In the second example, the main message is for the Iranian people to rise against the Islamic regime.

@TelAviv_Tehran appears to connect with PRISONBREAK in a few different ways:

- It is amplified by suspected purchased engagement⁸ accounts that also amplify PRISONBREAK. For example, one [post](#) shared by @TelAviv_Tehran had 163 reposts at the time of writing this report. Due to limitations on viewing engagements, we were able to analyze 83 of the accounts which reshared the post. 70 of those accounts fit the criteria of purchased engagement accounts and all 70 of the accounts shared at least one PRISONBREAK post, showing a one hundred percent overlap between the engagement accounts' promotion of @TelAviv_Tehran and PRISONBREAK.

⁸ This is a type of inauthentic behavior – usually sold as a service by dedicated private firms – deployed to artificially inflate the engagement that accounts, or specific pieces of posted content, enjoy on a given social media platform. It is called differently by each platform. In its [Authenticity](#) policies, X defines it as the “inauthentic use of X engagement features to artificially impact traffic or disrupt people’s experience”.



Figure 31: Screenshots from [@Cryptogran51861](#), one of the purchased engagement accounts, amplifying both PRISONBREAK and @TelAviv_Tehran posts.

- It posted an original AI-generated [video](#) purportedly representing the burning of the Evin Prison following the IDF strikes. The post was made only four minutes after @KarNiloufar's publication of the AI video supposed to show the actual bombing and while the strikes – according to the publicly available calculations discussed earlier in this report – were still happening.



```

<a href="/TelAviv_Tehran/status/1937067862155538576" aria-describedby="id_w72mc
n 23, 2025" role="link" class="css-1jxf684 r-bcqeoo r-1ttztb7 r-qvutco r-poiln3
9aw3ui r-3s2u2q r-1loqt21" style="color: rgb(83, 100, 113);"> flex
<time datetime="2025-06-23T08:39:01.000Z">4:39 AM · Jun 23, 2025</time> == $0

```

Figure 32: Screenshot from the @TelAviv_Tehran's [post](#) including an AI-generated video of the Evin Prison burning after the IDF airstrikes, and of the post's metadata showing that it was posted at 8:39 AM UTC on June 23, 2025 – only four minutes after the original AI [video](#) of the bombing posted by @KarNiloufar and illustrated earlier in this report.

It posted several of the videos posted by PRISONBREAK, and also multiple ones that match closely the network's modus operandi (i.e. likely AI-generated videos of military attacks on Iran, or of unrest in the country). We provide a list of examples in the [Appendix](#).

It also posted media that are less overtly AI-generated and rather seem to follow the modus operandi seen with PRISONBREAK. For example, a [video](#) of a supposed aerial attack on Tehran by unidentified military forces on June 13, 2025 – the day the IDF struck Iranian locations for the first time, starting the “Twelve-Day War.”



Figure 33: still from a [video](#) posted by @TelAviv_Tehran on June 13, 2025. On the same day, as the IDF began their airstrikes on Iran, PRISONBREAK posted at least 9 likely AI-generated videos

designed to show Iran in the midst of military attacks and social unrest. See section “*A Broader Narrative: Overthrowing the Iranian Regime*” earlier in this report.

- While some of the registration metadata (for example, the email domain, which seems to be Gmail here) as well as of the behavioural signatures by @TelAviv_Tehran differ from those of PRISONBREAK, the overlaps described above lead us to conclude with medium confidence that the network and this X account operate in at least partial coordination with each other.

Attribution

At present, we cannot conclusively attribute PRISONBREAK. The unavailability of critical identifiers – such as profiles’ metadata – to the research community impedes us from confirming the identity of the network’s operators.

However, the evidence we collected during the course of our research allows us to conclude that the most likely scenario is that of an involvement by the Israeli government, either an agency working alone or with the participation of an outside contractor.

We also considered two other scenarios: that of a U.S. government’s involvement, and that of a third-party government responsibility. We judge these two hypotheses as having medium-high and low likelihood, respectively.

The table below illustrates our analytical conclusions:

High Likelihood Scenario (H1):

The hypothesis most consistent with the available evidence is that the Israeli government – directly or through a contracting firm – conducted the operation. Each piece of evidence is compatible with and/or supportive of this conclusion.

Medium-High Likelihood Scenario (H2):

An unidentified agency of the U.S. government could theoretically also be behind the activity. Prior knowledge of and coordination with upcoming IDF operations would not have been impossible⁹, technical capabilities would have been on par with those required by the IO, and the apparent goal of regime change could possibly – although according to public statements, only immediately before the bombing of the Evin Prison – have been in line with the U.S. objectives for Iran. However, we note that

⁹ Despite President Donald Trump publicly supporting Israel’s military objectives, the U.S. government claimed not to be involved in the attacks on Iran. However on June 21, 2025, the U.S. directly attacked military targets in the country. See: [Iran’s Conflict With Israel and the United States”. Center for Preventive Action](#).

foreknowledge of more recent and highly sensitive IDF operations – like the [attack on Hamas officials](#) in Doha, Qatar, on September 9, 2025 – appears to have eluded U.S. government officials. Additionally, evidence such as the Israeli Minister of Foreign Affairs, Gideon Sa’ar, reposting on X the manipulated Evin Prison bombing video, the explicit promotion of a pro-Israeli X account by PRISONBREAK, and the network’s posting timing patterns aligned with a Middle Eastern time zone are inconsistent with the hypothesis of a U.S. government responsibility.

Low Likelihood Scenario (H3):

We consider it unlikely that a third-party government (other than the U.S.) is responsible for operating PRISONBREAK. This assessment is informed by several pieces of evidence inconsistent with this hypothesis – notably, the low likelihood that a foreign government would have had prior knowledge of the Evin Prison bombing; the coordination with an AI-driven X profile openly promoting a partnership with Israel in a future Iran freed of the Islamic regime; and crucially, the messaging precisely synchronized with the IDF’s escalation of operations over Iran. While it is conceivable that a professional group could have created the content in an accelerated time frame and with no advance notice, the alternative explanation is more likely.

We outline our analytical process through an established analytical technique – [Analysis of Competing Hypotheses \(ACH\)](#) – in the table below.

Analysis of Competing Hypotheses (ACH)

Legend:

-  = consistent with [hypothesis]
-  = inconsistent with [hypothesis]
-  = could be consistent or inconsistent with [hypothesis]

The evidence is presented in the first column, while the competing hypotheses are represented in columns 2-4.

WHO IS RESPONSIBLE FOR OPERATING PRISONBREAK?			
HYPOTHESES →	H1. The Israeli government	H2. The US government	H3. A third-party government (not the US)
EVIDENCE ↓			
E1. Prior knowledge of upcoming IDF military actions	✓	=	✗
E2. Synchronized timing between PRISONBREAK's posting and IDF's military actions	✓	=	✗
E3. Professional use of AI tools for the creation of sophisticated video footage	✓	✓	✓
E4. Consistent use of the Persian language (written and spoken) in the posted content	✓	✓	✓
E5. Explicit incitements to an uprising against the Islamic regime in Iran	✓	✓ ¹⁰	=
E6. Coordination with an X profile explicitly promoting Israel, and its support to	✓	=	✗

¹⁰ On June 22, 2025, the day before the strike on the Evin Prison, U.S. President Donald Trump used his Truth Social account to post in support of regime change in Iran:
<https://www.axios.com/2025/06/22/trump-iran-regime-change>

WHO IS RESPONSIBLE FOR OPERATING PRISONBREAK?			
<i>a regime overthrow in Iran</i>			
<i>E7. Activity started the day before (19 January 2025) Donald Trump's second inauguration as President of the United States (20 January)</i>	✓	=	=
<i>E8. Posting times are consistent with a work day in a Middle East location</i>	✓	=	✓*
			* This is to be considered as consistent for certain governments (in the Middle East), not for others. Iran is to be excluded as the clear target of the PRISONBREAK network.
<i>E9. The campaign to promote regime change in Iran is consistent with others observed as happening at the same time, and attributed to the Israeli government.</i>	✓	✗	✗
<i>E10. The Israeli MoFA, Gideon Sa'ar, <u>reposted</u> the manipulated Evin Prison bombing video one hour after PRISONBREAK originally did. He</i>	✓	=	✗

WHO IS RESPONSIBLE FOR OPERATING PRISONBREAK?			
<i>never retracted the post.</i>			

Conclusions

Covert influence operations have been a feature of armed conflict dating back centuries, their characteristics evolving as new technologies open up opportunities for experimentation. Today, we are witnessing a new generation of covert influence operations mounted principally on and through social media and featuring the use of artificial intelligence and related digital tools and techniques. These operations benefit from the distributed nature of social media, their built-in algorithmic engagement properties (which help propel sensational content by design), as well as recent measures taken by many tech platforms to reduce the capacity of, or eliminate altogether, internal teams responsible for removing coordinated, inauthentic content. The growing sophistication and ease-of-use of AI tools has also helped power these campaigns, providing actors with an unprecedented ability to produce increasingly realistic-looking videos and images with far fewer resources than what would have been required in previous eras.

In this investigation, we have identified a coordinated network of inauthentic X profiles that, since 2023, has conducted an influence operation targeting Iranian audiences. The objective of PRISONBREAK appears to be to foster a revolt against the Iranian regime among the Iranian population. One striking feature of this campaign is its synchronization with events on the ground: the content generated by PRISONBREAK appears to have been prepared in advance of and co-timed with military strikes undertaken by the IDF in June 2025.

Although we cannot attribute this to a particular entity, the advanced preparation required and the timing of the coordinated, inauthentic posts suggests some kind of connection to the Israeli state. We believe that while it is technically possible, it is highly unlikely that any third party without advance knowledge of the IDF's plans would have been able to prepare this content and post it in such a short window of time. Based on the data reviewed in preparing this report, the campaign we have documented was mostly likely undertaken either by an Israeli agency in-house or a private entity contracted by the Israeli government. However, without additional information we are unable to conclusively attribute the responsible parties.

While attribution around any covert operation is inherently challenging because of the steps taken by perpetrators to conceal their tracks, attribution is also made especially difficult today because social media platforms restrict access to their platforms to outside researchers, and thus to the artifacts and

other details that are critical to make conclusive attribution possible. In spite of these restrictions, we were able to use a combination of qualitative and quantitative methods to conclusively determine what we were observing was not spontaneous and organic, as the perpetrators were hoping to be perceived, but rather highly coordinated and inauthentic.

However, these methods are not readily available to the general public and/or to specific target populations and typically require considerable analytical effort, time and access to special resources and data sources. Additionally, it is now generally accepted that in today’s social media environment, sensational falsehoods spread quickly and are shared widely because they feed off of human emotions and are amplified by the engagement-driven algorithms of the platforms, all of which works to the advantage of those mounting covert influence operation campaigns. These dynamics and their potentially [harmful effects](#) are exacerbated in times of political crisis and [conflict](#), such as the military confrontation between Israel and Iran.

Notification to X

On September 30, 2025, we contacted X and provided the company with a list of the confirmed PRISONBREAK user accounts. At the time of publication, we had not received a response.

Appendix

X Accounts

PRISONBREAK

USERNAME	JOINED
ahmadsafaei789	March 2023
aliasia45	September 2023
amirhosstav	August 2023
Amirifer	August 2023
armantharman	September 2023
Arshfrahn	August 2023
azad_leylaas	September

	2023
DoustMehmans	August 2023
Eldenminareza	April 2023
Fatemehaazar	September 2023
Fatemehmavi6	March 2023
fereshteh5588	March 2023
feridounazari	August 2023
firoz_soltani	April 2023
hadirrezvani7	April 2023
hamidshahr6	August 2023
hokaseri	April 2023
iamramhei	April 2023
jamsheidrezvani	April 2023
kamrezi	June 2023
KarNiloufar	February 2023
kavehhame	August 2023
lalehfarah	March 2023
leila_farhadi	December 2023
leyla_zare	August 2023
LeylaLeylaebra	August 2023
mahjabbari	August 2023
masoudffarhadi	August 2023
mehdim156	February 2023
mirhoss6	April 2023
moharezjavan	September 2023
nadiazazzz	September 2023
Nedakkkhal	August 2023
Nimasssshahbaz	August 2023

NPiruzi	February 2025
odelia69986	August 2023
omid_hosseini2	August 2023
Parvaneh81179	August 2023
persiancat4587	May 2023
ranataabrizi	August 2023
Re8Ali	August 2023
royareads	April 2023
sarafefee	May 2023
Sepideh7895	May 2023
ShahinRmez	August 2023
ShandiKeik2394	August 2023
shayanpoahm	September 2023
sshahriyaresfah	August 2023
vafa4587	April 2023
YasamanZamni	August 2023
zahara_aminpour	September 2023
Zahrashahb55	August 2023
zarabithegr8	September 2023

Other Suspected Accounts

TelAviv_Tehran	May 2025
----------------	----------

Other Social Media

URL	PLATFORM	NOTES
-----	----------	-------

https://www.youtube.com/@%D8%AC%D9%86%DA%AF%D8%B2%D9%86%D8%AF%DA%AF%DB%8C	YouTube	YouTube channel used to post the AI-generated video of the PRISONBREAK version for the song “Baraye”.
https://t.me/Womens1404	Telegram	Linked in bio by PRISONBREAK’s account @KarNiloufar
https://www.instagram.com/telaviv_tehran	Instagram	Instagram account for the suspected X profile @TelAviv_Tehran

Shared X Posts Between @TelAviv_Tehran and PRISONBREAK

@TelAviv_Tehran’s post	PRISONBREAK’s post	Comment
https://x[.]com/TelAviv_Tehran/status/1937542242341671126	https://x[.]com/NPiruzi/status/1937538928329801898	AI-generated re-enactment of a famous – and often turned into a meme – scene from the movie “Der Untergang” (“Downfall”, 2004), with Ali Khamenei and apparent IRGC members replacing the characters of Adolf Hitler and Nazi top officials.
https://x[.]com/TelAviv_Tehran/status/1937802054040166908	https://x[.]com/Nedakkkhal/status/1937803510533464177	Composite of photos and videos of protests and executions of civilians in Iran, with an ending claiming that “Iran is looking forward to a brave, new leader”.
https://x[.]com/TelAviv_Tehran/status/1955637632215441818	https://x[.]com/KarNiloufar/status/1955917591274229772	AI-generated video of an alleged digital billboard “from the streets of Spain” denouncing the Iranian water crisis. The captions in the two posts are worded similarly. The AI-generated frames are overlaid on the exterior of Cine Callao , a landmark movie theater in central Madrid.

https://x[.]com/TelAviv_Tehran/status/1957000656906055790	https://x[.]com/hamidshahr6/status/1958160382587261123	AI-generated videoclip using the likeness of three existing Iranian singers for a song inciting the Iranian population to an uprising. Also hosted on YouTube.
https://x[.]com/TelAviv_Tehran/status/1937103753871598035	https://x[.]com/KarNiloufar/status/1937067028885770497	AI-generated video of the Evin Prison bombing.