How Iran-Backed Groups Control the Arabic Language Sphere on X

by Samara Azzi (/experts/samara-azzi)

Jan 26, 2024
Also available in
العربية (/ar/policy-analysis/kyf-tsytr-aljmaat-almdwmt-mn-ayran-ly-mjal-allght-alrbyt-ly-mnst-aks)

ABOUT THE AUTHORS

Samara Azzi (/experts/samara-azzi)
Samara Azzi is a venture capitalist based in Geneva.

🕐
Brief Analysis

Until Twitter's moderation policies are strictly enforced and moderation loopholes addressed, coordinated efforts will continue to target independent anti-Hezbollah and anti-Hamas voices on X/Twitter while effectively sharing their own messaging, thus ensuring that in the Middle East, Iran-backed actors maintain their grip on the site.

There is unspeakable violence and cruelty in every war zone. Add to this the fog of war and near instantaneous reportage from conflict zones through social media, and it is daunting for a reader or online activist trying to make sense of it. Misinformation is making this process that much harder.

But attempts to combat misinformation on X (formerly Twitter), or to share information or opinions that are critical of Iranian proxy groups such as Hezbollah, runs up against a pro-Iranian "cyber-army" that is comfortable with gaming the platform's limited moderation controls in English and especially Arabic to suppress criticism and amplify their own narrative. Such efforts have been going on for years, and as a Lebanese online activist attempting to speak out against Hezbollah online during Lebanon's financial crisis in 2019, this gaming of social media content moderation is all too familiar. The success of these coordinated cyber activities likewise highlights the failures of Twitter, and now X, to effectively moderate a platform according to their own content guidelines.

Shadow banning and Self-Censorship

Multiple digital misinformation experts have described the social media content surrounding the war as plagued by an "unprecedented" level of misinformation. Accounts such as @jacksonhinklle, with 2.4 million followers and counting, have widely spread posts despite disinformation specialists repeatedly calling out the account for misattributed videos and false information.

As an online activist, my response when I see an image of video misattributed to Gaza but in fact from the Syrian or

Yemeni war **is to report it ()** . But this is where the system breaks down further. In correctly identifying the source of these images or video, particularly when flagging misinformation on X (Twitter) in Arabic, one can activate a veritable cyber-army of the "axis of resistance" including supporters of Hezbollah, Hamas, and the Hashd Al-Shaabi, which often target such efforts and benefit from the deep pockets of their Iranian backers.

Because of the number of these accounts and the effective nature of the coordination, users expressing a view not in line with the preferred narrative of this cyber-army are likely to find their account reported for false information or another issue that violates X (Twitter's) posting guidelines, although the exact reason is never clear to the user being targeted. What is clear is that these reports often trigger a shadow ban, throttling the algorithm that promotes tweets to a user's followers and, in effect, smothering the reach of the user's message. If enough reports come in from this cyber-army, then the algorithm or monitors can respond by blocking the account and de-platforming the user.

Given my own personal experiences, I can only imagine that these coordinated efforts have a significant influence on the content of those who wish to post information or ideas that are inconvenient or threatening to this cyber-army. I am not a member of any political party, and I post my views on social media independently. Yet I have been on X (Twitter) long enough now to know that to avoid the risk of being shut down by the cyber-armies, I must self-censor. By self-censoring, I avoid writing directly about even things I know to be true but are 'inflammatory' to these cyber-armies, using more diplomatic words instead to get my message across. Even with this self-censorship, I have had many warnings over the years from X (Twitter), and I have been shadow-banned multiple times after posting tweets that criticize Hezbollah.

Identifying a Deeper Problem – X (Twitter's) Moderation Failures

For many independent Lebanese social media users, we've known that content mediation has been broken on Twitter for years—in English but especially in Arabic. Specifically, Lebanese accounts experienced this first hand when trying to post the truth about Lebanon's situation during and after the country's 2019 financial crisis. A number of Lebanese independent activists—myself included—took to social media and blamed the Hezbollah militia for taking over the country, pushing for the default on debt held by Western and Lebanese investors, and ultimately driving the country to ruins.

Access to social media networks can be vital in the Middle East, where independent media is starved for money. Many of these media outlets have collapsed after losing the ability to pay their staff, leaving well-funded media outlets as the default news—owned or funded by those in power in the Middle East and with limited freedom of speech. In this restricted media landscape, the natural inclination is to turn to social media, which in theory promises freedom of speech within established guidelines.

But one by one, the Hezbollah cyber-army shut us down. So aggressive were the efforts to de-platform users that individuals I know had to resort to repeatedly opening up new accounts every few days to stay on the platform. As individuals lacking the resources possessed by these groups, we were unable to effectively "re-platform" and recover our previous following, pushing many of those activists to close down their accounts and give up on the media platform all together. This disparity in coordination—combined with issues regarding Twitter's moderation methods —ensured that our message to the world was killed by a platform once lauded as the democratization of media.

In the face of these repeated attacks, it became clear that there was a third party in this battle between us and the cyber-armies. Twitter's Arabic language moderators were either ignorant of what was repeatedly occurring or complicit. While both possibilities were concerning, the latter was much more dangerous if true.

To find out more, our network decided to see if we could identify the political leanings of those working in Twitter's MENA offices in Dubai. It was not much of an investigation—by visiting the social media accounts of these employees, we quickly identified a number of those who express support for the Free Patriotic movement —a

Christian group allied to Hezbollah in Lebanon—and other pro-moumana'a (Iranian proxy) () views. It was evident to us from these social media accounts that some saw Hassan Nasrallah as a freedom fighter.

Because of a lack of transparency on how Twitter, and later X, moderates its content, it is impossible to know how the employees who expressed these views influenced moderation decisions. Yet it is surprising and frustrating for our group that we could find these accounts associated with employees at one of the world's biggest social media companies. In a region as volatile as the Middle East, the lack of a clear stance on moderating extremist content—especially when those extremists also have positions of power—is deeply concerning.

Nor are these concerns limited to specific political views. In the fight of Israel against Hamas, for example, it is most likely close to impossible to find an Arabic speaker/editor in the Arab world who is openly pro-Israel.

Even Community Notes—directly introduced by Elon Musk to clean up X 's misinformation was gamed by state actors according to a tweet by Musk's friend and business partner David Sachs.

https://x.com/davidsacks/status/1733927451317981512?s=46
(https://x.com/davidsacks/status/1733927451317981512?s=46)

When Elon Musk bought Twitter, I and some of my colleagues thought that there might be more attention payed to the Middle East after he promised to fight Twitter's endemic bot problem and "clean up Twitter." Instead, things have only deteriorated further. There is also little clarity about what future efforts will be undertaken to address this issue. For example, Elon Musk and his friend and business partner David Sacks **suggested (https://x.com/davidsacks/status/1733927451317981512?s=46)** that community notes—themselves introduced to address misinformation—could be manipulated by motivated state actors due to their design, but there is little information for the average user available about how any manipulation of X (Twitter's) moderation tools is being combatted—the last Twitter moderation transparency **update (https://blog.twitter.com/en_us/topics/company/2021/-expanding-access-beyond-information-operations-)** was in 2022. Ultimately, moderation issues have allowed X (Twitter) to become a haven for terrorists groups' own messaging.

Hezbollah: Tweeting with Impunity

On the other hand, what has been posted on X/Twitter demonstrates just how broken the platform is. Lokman Slim, a prominent Lebanese Shia activist who spoke out against Hezbollah and its policies, was killed in the south of Lebanon by four bullets to the head in Feb 2021—almost certainly by Hezbollah, even if Lebanon's legal system will never prosecute his case. Before we woke up to the news of his killing, Hassan Nasrallah's son immediately took to Twitter, appearing to boast about the group's achievement that night. He tweeted "the loss of some is in fact profit and unexpected kindness" followed by the hashtag #no-regrets.

(/sites/default/files/2024-01/tweet%20samara.png)
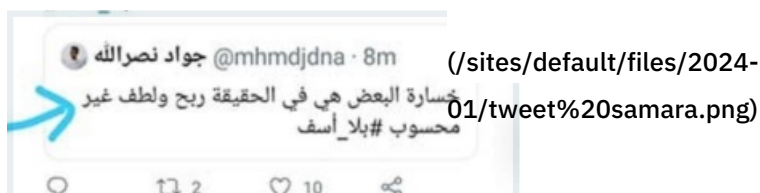
While an independent print media platform would never give space to this person to say such things on the eve of the killing of Lokman Slim, the tweet went viral. The cyber-army joined in a Twitter chorus of glee over Lokman's killing and #no-regrets. Terror was being celebrated, and the apparently unmitigated spread of this celebration in the following days and weeks demonstrated who owns the platform in the Middle East.

X (Twitter) has a policy addressing such tweets, with the most recent version reading

[(https://help.twitter.com/en/rules-and-policies/violent-entities)](https://help.twitter.com/en/rules-and-policies/violent-entities) "you can't affiliate with or promote the activities of violent and hateful entities." Moreover, it states that "we will remove any accounts maintained by individual perpetrators of terrorist violent extremist, or mass violent attacks, and may also remove posts disseminating manifestos or other content produced by perpetrators." But in Lebanon, as in much of the Middle East, those who repeatedly violate this policy are not fringe actors. Instead, Twitter's moderation policy is easily manipulated to target the very people it purportedly seeks to "platform."

Both before and after Musk's purchase of Twitter, the company leadership have expressed concerns about the balance of moderation versus free speech. But when it comes to this concern, it is worth returning to Friedrich Hayek's *The Constitution of Liberty*. In this book, Hayek discusses the ideals of freedom that he believes have and should continue to guide western societal ideals. One key point here is that the tolerance of intolerance ultimately hinders rather than promotes a free society. By showing tolerance to intolerance you are really digging up a grave to your own free society.

There are some **signs of change (https://blog.twitter.com/en_us/topics/company/2023/maintaining-the-safety-of-x-in-times-of-conflict)** on this issue in the wake of October 7 and the massive rise of misinformation—X(Twitter) took down **hundreds (https://www.pbs.org/newshour/amp/politics/x-has-taken-down-hundreds-of-hamas-linked-accounts-ceo-says-an-expert-thinks-its-not-enough)** of Hamas linked accounts in the week after October 7, and efforts have continued, including suspending a prominent **Hezbollah account (https://themessenger.com/tech/elon-musks-x-suspends-hezbollah-linked-account)** identified by a *Jerusalem Post* reporter. Just last week, X **suspended (https://x.com/mehmetcanbekli1/status/1744591748733514113?s=48)** a Qassem brigades account after it quickly reached 150,000 followers after launch.

But numerous studies have emphasized the ability of these coordinated cyber-armies can adapt to attempts like this, especially if efforts are ad-hoc instead of systematic. Until Twitter's moderation policies are strictly enforced and moderation loopholes addressed, these coordinated efforts will continue to target independent anti-Hezbollah and anti-Hamas voices on X/Twitter while effectively sharing their own messaging, thus ensuring that in the Middle East, Iran-backed actors maintain their grip on the site. ❖

RECOMMENDED

BRIEF ANALYSIS

Cairo: Stuck between Securing the Red Sea and Avoiding a Perception of Support for Israel

Jan 26, 2024

◆

Mohamed Maher,
Mohamed Farid

(/policy-analysis/cairo-stuck-between-securing-red-sea-and-avoiding-perception-support-israel)

How Hamas Manipulates Gaza Fatality Numbers:

Examining the Male Undercount and Other Problems

Jan 25, 2024
- ◆

Gabriel Epstein

(/policy-analysis/how-hamas-manipulates-gaza-fatality-numbers-examining-male-undercount-and-other)

Can Iraq's Militias Imitate Houthi Successes at Sea?

Jan 24, 2024
- ◆

Hamdi Malik,
Michael Knights

(/policy-analysis/can-iraqs-militias-imitate-houthi-successes-sea)

REGIONS & COUNTRIES

Iran (/policy-analysis/iran)

Lebanon (/policy-analysis/lebanon)

Middle East (/policy-analysis/middle-east)