Israel-Hamas war poses early disinformation test for Meta's Threads

Publication Date: 2023-10-13

Author: Nick Robins-Early

Section: Technology

Tags: Meta, Threads, Instagram, X, Israel, Hamas, Gaza, features

Article URL: https://www.theguardian.com/technology/2023/oct/13/instagram-threads-misinformation-israel-hamas

When Meta launched Threads, its Instagram-linked Twitter clone, in July, the company promised a kinder and friendlier experience than the divisive content and extremism that often dominate other social networks. Now, as social media users seek out information on the Israel-Hamas war, the young app is facing its first test amid the rampant misinformation emerging from the conflict. On Telegram and X, previously known as Twitter, repurposed videos, doctored photos and manipulated media falsely claiming to document the war have circulated widely. The quantity of these posts, and scale of their reach, have alarmed fact-checkers, disinformation monitors and extremism experts, who have criticized these social networks for allowing misinformation to flourish. Meanwhile, Threads has avoided touting itself as a place for real-time information or a window into the conflict. The result is that while Threads still contains numerous posts about the conflict, it is a visibly different experience from what is being amplified on X. Dina Sadek, a research fellow at the Atlantic Council's Digital Forensic Research Lab (DFRLab), said she had seen far fewer of the falsified videos and images that are going viral on other platforms. Several posts viewed by he Guardian that contained debunked misinformation, such as one that falsely claimed to show kidnapped Israeli children being held in cages, had limited engagement. Unlike Telegram and X, Threads has explicitly deprioritized news content, with the head of Instagram, Adam Mosseri, posting on Tuesday that amplifying news "would be too risky given the maturity of the platform". "We're not anti-news," Mosseri wrote. "But, we're also not going go [sic] to amplify news on the platform." One potential reason for the apparent lack of viral falsehoods on Threads, Sadek said, is that the platform is simply not used in the same way or at the same scale as others such as X or Telegram. "The thing with Threads is it's generally less used, especially when broadcasting developments from a conflict like this one," she said. "People are still building their base followers there." When Threads debuted, some of the earliest adopters and biggest accounts were brands and celebrities that posted anodyne content that the social network amplified on to users' feeds. Following an initial rush of 100 million users joining the app, the number of active users plummeted by about 80% after a month. Whether Threads continues to avoid amplifying content about the war, and how the misinformation landscape on the platform changes, is still an open question. Meta's other properties, Facebook and Instagram, have struggled for years to rein in misinformation and incitements to violence. Facebook has been utilized to stoke ethnic cleansing and civil war in countries such as Ethiopia and Myanmar, while extremist movements of varying stripes have evaded the platforms' moderators to share propaganda and organize. Meta did not respond to questions from the Guardian on Threads' content-moderation policies. The financial motivation for posting on Threads is also different from that of social networks where users can directly make money off their content. X, which has made a huge push under its owner Elon Musk to

attract creators and influencers as it struggles to make up for lost advertising revenue, allows users who pay $8 a month for verification to monetize their content through ads or subscriptions. The monetization has given these blue checkmark users an incentive to drive engagement through any means available, misinformation experts have warned, leading to unscrupulous users posting misleading, false or incendiary content. “So far, misinformation about the Israel-Hamas war appears to be most prevalent on X,” said Jack Brewster, an editor at the media analysis firm NewsGuard. “Many of the false and unsubstantiated claims NewsGuard identified were boosted by verified X accounts. Some of these accounts appeared to have been set up recently to gain virality.” Under Musk’s ownership, X’s misinformation problem intensified after the hollowing out of the company’s content moderation through layoffs, attrition and reliance on automated systems. Several of the large accounts on X that have spread misinformation or branded themselves as chroniclers of the war are not active on Threads. These include two accounts with a history of posting falsified videos – one of which posted antisemitic epithets – that Musk endorsed to his 160 million followers. “We saw a lot of violent content that was not moderated at all, especially in languages outside of English,” said Moustafa Ayad, the executive director for Africa, the Middle East and Asia at the non-profit Institute for Strategic Dialogue. The European Union issued warnings to Musk and the Meta CEO, Mark Zuckerberg, this week, urging them to remove any content that violated EU laws and instructed them to respond within 24 hours. X’s CEO, Linda Yaccarino, announced on Thursday that the platform had removed hundreds of Hamas-linked accounts and labeled tens of thousands of posts. Meta responded Friday in a blog post, announcing that it had created a special operations center for content moderation specifically for posts in Hebrew and Arabic as well as removing a cluster of fake Facebook and Instagram accounts linked to a previous Hamas misinformation operation. In addition, the company said it had removed or labeled 795,000 pieces of content as “disturbing”, which limits a post’s reach, since the start of the conflict on 7 October. The company altered its moderation policies to respond to the war, directing algorithms on Facebook, Instagram and Threads to flag more posts to moderators and expanding the scope of its violence and incitement policy to remove content that identified hostages. It also restricted hashtags related to the war on Instagram and limited the use of Facebook and Instagram Live by users with previous policy violations. In order to prioritize the safety of those kidnapped by Hamas, we are temporarily expanding our Violence and Incitement policy and removing content that clearly identifies hostages.” X has increasingly leaned on its Community Notes system, which allows users to add context or flag content as inaccurate, but already there have been examples of holes in the system. In one case, a video posted by Donald Trump Jr showing Hamas militants killing Israelis was flagged as disinformation, although Wired reported that an open-source intelligence researcher determined it was real. “It can’t just be left up to users,” Ayad said. “You need experts that understand these landscapes and the contexts.” Violent videos and influencer operations from the Israel-Hamas conflict have also moved among social media platforms, creating a pipeline between more closed-off social media spaces like Telegram and Discord and the public view of X. One Iranian Revolutionary Guards-linked account with more than 370,000 subscribers on Telegram, for instance, has been posting a barrage of propaganda that disinformation monitors have then seen reposted on other platforms. It has made the conflict, already a dangerous situation in which it’s difficult to get reliable information, even harder for the average person to parse. “All of these different ecosystems are currently intersecting, along with state-influenced operations that exist on these platforms as well,” Ayad said. “It’s creating a very noxious mix.”