Visualizing War under Information Disorder: Generative AI in the Iran–Israel Escalation*

Tamar Mitts†

September 14, 2026

Abstract

Contemporary wars unfold in fragmented information environments in which states compete over both facts and interpretation of events. Generative AI offers new ways to participate in this competition by enabling actors to produce content at scale, make false information appear real, and create more vivid depictions of symbolic ideas through synthetic images and videos. I examine how this technology was used during the June 2025 Iran-Israel escalation using rich data from six social media platforms. I find that AI-generated media increased during the conflict, but not in a way that looks like flooding: it remained far less prevalent than authentic media. In a matched analysis of 400 images, I find that synthetic images were much less likely to depict concrete events and more likely to portray symbolic messages related to resolve, victory, or enemy humiliation. AI-generated content also received significantly more engagement, suggesting its potential to attract attention.

Word count: 9,640

Keywords: propaganda; political communication; generative AI; social media; misinformation; armed conflict

*I thank Matthew Francis, Kevin Greene, Martin Innes, Hannah Kazis-Taylor, Erica Lonergan, Theo Milonopoulos, David Rogers, Jacob Shapiro, and Laura Smith for their terrific comments. I also thank Eden Cohen, Sara Elbaum, Eden Kaduri, Ian Kearns, Christine McNeill, Brian Andres Palacios Paz, Ethan Sheinker, and Elizabeth Yee for their research assistance. This research was supported in part by the Department of Defense Minerva BARI grant, "Influence, Manipulation, and Information Threats as Adversarial Techniques: Events, Evolution, and Effects." I used OpenAI's GPT-5.6 Sol for copyediting. I closely reviewed and revised all suggested language edits and am responsible for the final text.

†Associate Professor, School of International and Public Affairs, Columbia University.

Introduction

In the first days of the military escalation between Iran and Israel in June 2025, images of destruction and ruins became part of the conflict almost as quickly as the strikes themselves. After Iranian missiles hit Tel Aviv, for example, a viral video shared on social media by an Iranian-affiliated news outlet showed extensive devastation in realistic-looking scenes that were later found to be generated by artificial intelligence (Tehran Times, 2025; Cornwell, Hafezi, and Holland, 2025; Factnameh, 2025). At the same time, Israeli residents were closely following updates on their phones and expressed distress and concern over what might come next, with many afraid to leave their homes (Lidman, 2025). Days later, after Israel struck Evin prison in Tehran, the Israeli government shared a video that appeared to show the prison gates destroyed in a precise strike, calling on Iranians to rise up against the regime. This depiction was later found to likely be AI-generated, omitting information on the casualties that the strike caused on the ground (Sa'ar, 2025; Alimardani and Gregory, 2025; Amnesty International, 2025). Many Iranians encountered these competing claims after days of severe restrictions on internet access, with limited ability to determine what had happened or whose account to trust (Alimardani and Gregory, 2025; Burgess, 2025).

These examples illustrate the contemporary crisis around information in international politics, particularly during violent conflict or war. In these contexts, uncertainty about rapidly shifting events can create opportunities for influence and deception that strategic actors can exploit. While wartime propaganda is not new, digital communication technologies, such as social media and generative AI tools, have changed the speed and scale at which it can be created and shared. As a result, various conflict-related accounts, such as claims about military success and blame for civilian harm, or even more abstract or symbolic statements about the conflict's trajectory, power, resolve, or vulnerability, can now reach large audiences almost instantaneously, while verification of such claims can lag behind.

Morse and Pratt refer to this crisis as *information disorder* – a context in which easy and cheap content production, widespread circulation of misleading information, and algorithmic amplification of sensational narratives create a fragmented information environment where even

reaching agreement on facts can be hard (Morse and Pratt, 2025). Farrell and Newman suggest that part of the origin of this problem lies in the institutions created to facilitate the liberal international information order, which were designed to enable open, global information flows but, along the way, also created opportunities for governments and other actors to exploit those flows strategically (Farrell and Newman, 2021). Related work suggests that this crisis is not limited to false claims or sensational narratives; political actors can also withhold relevant information or even undermine the procedures and institutions through which some accounts become accepted as authoritative (Adler and Drieschova, 2021; Lesch, 2023; Carnegie, Clark, and Zucker, 2024; Morse, 2025). The crisis over information, therefore, involves both uncertainty over who can establish a credible account of international events and who can shape the narratives through which their significance is understood.

For states involved in war, this creates a particularly important strategic problem. Conflict actors benefit from offering an explanation – their narrative – for military action that can legitimate it to observing audiences. Governments seek to present their own conduct, whether past or planned, as justifiable and effective, while at the same time portraying their adversaries in a negative light (Goddard and Krebs, 2015; Krebs, 2015; Snyder, 2015; Miskimmon, O'Loughlin, and Roselle, 2013; Morse and Pratt, 2022). Because the same message can reach domestic audiences, adversary publics, and international observers, these communications may pursue several objectives at once (Chubb and Wang, 2023). But under conditions of information disorder, producing a favorable account is insufficient. Since the global information environment, particularly during crises, is saturated with conflicting accounts, states need to compete with other voices to attract attention to their preferred narrative.

Generative artificial intelligence provides new tools to participate in this competition by allowing the production of content – whether text, audio, images, or video – quickly, at relatively low cost, and across multiple languages. In this paper, I focus on visual media, as it seems to play an increasingly important role in wartime communication. Images and videos can be used to document battlefield events, expose civilian harm, and present the results of an operation, but also to communicate symbolic or aspirational ideas that cannot be documented with a camera, such as future victory or

the enemy's humiliation or weakness. Generative tools expand both possibilities: they can fabricate images that mimic real photos of concrete events or create scenes that are clearly imagined. This makes visual content particularly useful for examining how actors involved in conflict use generative AI for strategic purposes.

I distinguish among three patterns of use. The first is *production at scale*. Generative AI allows actors to circulate large amounts of content rapidly, respond quickly to new developments, and adapt messages to different audiences in multiple online spaces. Nyhan et al. (2026) suggest that one of the most likely effects of the growing availability of generative AI tools is the ease with which actors can flood information ecosystems with their outputs, even when AI-generated messages are not especially persuasive. Previous work on how states use flooding suggests that repeated and pervasive content can redirect audience attention, occupy platform space, and crowd out other messages, making the desired narratives difficult to avoid (King, Pan, and Roberts, 2017; Roberts, 2018; Lu and Pan, 2022; Lu et al., 2025). If this use is prominent, then we would expect posts containing AI-generated media to increase sharply during conflict and constitute a meaningful share of the broader wartime information stream.

A second use is *fabricated evidence*, a form of deception in which generative AI is used to create visual material that appear to document events that never actually happened (Chesney and Citron, 2019; Vaccari and Chadwick, 2020; Twomey et al., 2023; Birrer and Just, 2025). Such content can shape how audiences understand the conflict by supplying what seems to be evidence on events – whether attacks, statements by political leaders, or evidence of destruction or civilian harm – that did not occur or by presenting real events in misleading ways (Gartzke and Lindsay, 2015; Lanoszka, 2019; Miller and Robinson, 2019). If this use is common, then we would expect AI-generated visuals to frequently depict concrete events in a way that resembles real photographs or videos.

A third use is *symbolic storytelling*. Generative AI can produce imagined scenes that turn political interpretations or specific narratives into visual stories. These stories can organize impressions of the conflict as a whole, present one's side as righteous and powerful and the enemy as evil

or vulnerable, and make a particular imagined future outcome seem likely to happen. Synthetic images, for example, can create vivid depictions of the enemy leader as humiliated, with their regime collapsing and their population rising against them. They can also portray powerful images of resilience and resolve, or present one's side's victory as already achieved. These images may look like real photos, but they may also be stylized as cinematic, fantastical, or even cartoonish. Their goal is to offer a symbolic interpretation of the conflict or its desired outcome in a way that supports a strategic narrative. If this use is central, then we would expect AI-generated visuals to be less likely to document concrete events and more likely to convey symbolic and hypothetical messages than non-synthetic wartime images. Of course, these three patterns of use are not mutually exclusive. Conflict actors can use them simultaneously within the same context.

I study these predictions in the June 2025 military escalation between Iran and Israel, a 12-day period of direct military confrontation in which the two countries exchanged strikes against one another. The conflict was commonly referred to as ushering in a new era of AI-generated wartime propaganda, as government institutions, military and political figures, state media, and networks aligned with both sides engaged in an information contest to shape the narrative around the war (Myers, Odenheimer, and Solomon, 2025; Alimardani and Gregory, 2025). I draw on a cross-platform collection of 212,763 posts published on Facebook, Instagram, Telegram, TikTok, X, and YouTube between May and July 2025 by 196 official and unofficial accounts linked to Iran and Israel. Of these, 129,736 contained images or videos, which I downloaded for closer analysis. My goal was to capture the public-facing communication of both sides during the military escalation, including official messaging and the wider networks through which their narratives circulated.

To examine how these actors used generative AI during the conflict, I developed a multi-part empirical strategy. First, I used an automated classifier, validated through multiple tests,[1] to identify posts containing AI-generated media, measure their prevalence, and assess whether they had, in fact, flooded the information space during this period. Second, I zoomed in on comparable images posted by the same accounts, on the same platforms, in the same languages, and on the same calendar dates.

[1] See supplementary information for more detail.

I sampled 400 images from the broader corpus, half classified as AI-generated and half as non-AI, forming 200 matched pairs. Human annotators familiar with the context then closely examined the images and coded their relationship to actual events, the messages they communicated, and their visual form.

Third, I examined what happened after these images were posted on social media. I analyzed platform-native engagement metrics to compare the attention received by AI-generated and non-AI media. I also collected a secondary dataset of replies to images posted on X – the platform with the highest prevalence of AI-generated media in the dataset – to examine whether users engaging with AI images showed signs of recognizing them as synthetic. For this purpose, I created measures of whether replies to AI-generated and non-AI images explicitly referred to artificial intelligence or to particular AI tools used to generate synthetic media. I then examined how broadly such references appeared across discussions surrounding the two types of posts.

I find several notable patterns. First, the results provide limited support for production at scale. Although both sides increased their use of AI-generated media during the June escalation, the data do not show signs of flooding: both sides used synthetic media relatively sparingly – only 1.7% of the images and videos in the data appear to be AI-generated. These patterns hold even when I relax the threshold for identifying AI-generated media – under the more permissive classification threshold, no more than 3% of the images and videos shared by both sides are classified as AI-generated. This finding is somewhat at odds with expectations about how AI systems would be used in conflict settings (Goldstein et al., 2023; Wack et al., 2025), as well as with accounts of this particular conflict, which was often described as a watershed moment for AI-generated propaganda (Myers, Odenheimer, and Solomon, 2025).

Second, I do not find strong evidence that generative AI was used primarily to fabricate evidence. Such cases were clearly present. For example, Iran-linked accounts shared AI-generated visuals of Israeli F-35 planes being shot down, missile strikes on Tel Aviv, and widespread devastation at Ben Gurion Airport that had not occurred (Alimardani and Gregory, 2025; Myers, Odenheimer, and Solomon, 2025). Similarly, a network assessed as likely linked to the Israeli government circulated

synthetic and manipulated videos portraying unrest and growing opposition to the regime inside Iran, as part of a broader effort to encourage opposition to the Islamic Republic (Fittarelli et al., 2025). Yet across the matched comparison, AI-generated images were far less likely than non-AI images to depict a concrete event: only 30 percent did so, compared with 64 percent of matched non-AI images. They were also much less likely to have a photographic or documentary appearance.

The reply analysis points in a similar direction. If the main use of AI-generated images was to fabricate evidence, then we would expect to see viewers accepting them as authentic, or at minimum, not reacting in a way that shows those users recognize those images as synthetic. However, when analyzing replies to AI-generated images and comparing them to replies to posts without them, I find that the former were much more likely to mention artificial intelligence or a specific AI tool. References to AI or AI tools also appeared across a much larger share of conversations surrounding posts sharing synthetic media. Of course, this does not necessarily mean that most users recognized the images as artificial, but it does suggest that their synthetic character was often visible in public discussion. This makes it less likely that the main use of generative AI in this context was to deceive.

These findings therefore shift attention from flooding and deception to a third, more prominent pattern: *symbolic storytelling*. Interestingly, when examining the themes shared by synthetic and non-synthetic images in my matched sample, I find that the former were significantly more likely to convey general symbolic messages or depict future and hypothetical events. AI-generated images were often used to signal resolve and resilience or to mock and humiliate the enemy through scenes that were cinematic, hypothetical, or even fantastical. For example, images generated with AI were nearly twice as likely to convey general symbolic messages (48 versus 25 percent) and more likely to depict future or hypothetical events (9 versus 3 percent). These patterns suggest that generative AI was primarily used in this context to visualize scenes that had not yet occurred – and, in some cases, could not exist outside the imagination. This allowed actors on both sides to create vivid visual depictions of stories or narratives that could advance their strategic goals.

Perhaps most surprising, given their low prevalence, is how much attention these images attracted on the platforms where they appeared. My analysis shows that posts containing AI-generated media

received significantly more views, likes, and comments than posts containing non-synthetic media, even after accounting for differences across accounts, platforms, posting times, and audience size. While this association does not mean that synthetic images were more persuasive or influential, it suggests that there was something attractive about them (either to humans or to social media algorithms) that made them more visible. It may be that the very features that distinguished AI-generated media from non-synthetic ones – their striking visual forms and "impossible" imagined scenes – also helped them attract attention.

My findings therefore point to a dimension of wartime information contestation that is less studied in existing work: its role in giving symbolic form to political claims. An important strand of scholarship on information in global politics examines how actors manage uncertainty by disclosing, withholding, or strategically manipulating information (Carnegie and Carson, 2018; Carnegie and Carson, 2020). The findings of this study suggest another strategic use in the emerging AI era – giving visual form to claims about future victory or defeat, strength or vulnerability, or placing anticipated and desired outcomes into the visual field of an ongoing crisis. The matched comparison shows that this was not confined to a handful of prominent examples, but systematically distinguished AI-generated images from other wartime visuals. The attention that such images attracted online is consistent with evidence from other domains showing that AI-generated images can be preferred to human-made works and can perform well in aesthetic evaluations and click-through rates (Hees et al., 2025; Hartmann, Exner, and Domdey, 2025). Thus, perhaps the strategic value of AI-enabled propaganda lies not only in its capacity to increase the volume of content that actors can produce (Goldstein et al., 2023; Wack et al., 2025), but also in its ability to attract attention in crowded and contested information environments.

Case, Data, and Research Design

The June 2025 escalation and the social media dataset

The June 2025 escalation between Iran and Israel began with a surprise Israeli attack on Iran on June 13 and was followed by a barrage of Iranian missile strikes on Israel. The two states continued to strike each other in the following days, causing damage to military sites, government and civilian infrastructure, as well as deaths and injuries. While events were unfolding on the ground, a parallel information contest took place in the digital sphere. There, social media accounts of government institutions, military organizations, aligned media, and informal networks on both sides sought to shape the narrative around the war by sharing information – real and fabricated – about what was going on. This conflict offers a useful opportunity to study the use of AI-generated propaganda, as it took place after high-quality image- and video-generation systems became widely accessible. This allows me to examine whether Iran and Israel, and the networks that supported them, used AI-generated media and, if so, in what ways.

I collected public social media content posted by official and unofficial accounts linked to Iran and Israel from six platforms: Facebook, Instagram, Telegram, TikTok, X, and YouTube. The data come from a broader collection on information campaigns by state and non-state actors involved in the Middle East conflict since October 7, 2023. Here, I focus specifically on activity linked to Iran and Israel.[2] I collected the data in two stages. First, I identified official accounts linked to both sides by searching each platform for accounts belonging to governments, state institutions, military bodies, political leaders, and other formal entities. I confirmed the official nature of the accounts by consulting institutional websites and links across platforms.

I then supplemented the official account list with unofficial accounts that regularly circulated or amplified one side's messaging by conducting platform searches and drawing on reports from organizations that published open-source platform investigations. To ensure that I do not attribute unofficial account activity to the official messaging of each country, I classified accounts as official

[2]See SI for more details on the broader dataset.

only when institutional ownership or control was clear. Unofficial accounts, then, were those who consistently advanced one side's messaging, but for which there was no clear evidence that they were owned by government actors. These included aligned media outlets, cyber units, semi-official channels, and other accounts that consistently amplified the actor's messaging. The Supplementary Information provides additional detail on the account-identification process and the full list of accounts included in the dataset.

Since my focus is on the Iran-Israel confrontation of June 2025, I focus on content shared by accounts linked to the two countries from May 1 to July 31, 2025 – a window that allows me to bracket the escalation itself as well as the weeks immediately before and after it. During this period, I prospectively collected public posts from these accounts using a third-party API,[3] retaining all available metadata, including account and post identifiers, post text, publication date, engagement measures, URLs, and attached media. The prospective collection allowed me to capture content posted in real time and archive it, even if it was later removed by the platforms. The data includes 212,763 posts produced by 196 official and unofficial accounts on the six platforms mentioned above. Of these accounts, 95 were linked to Israel and 101 to Iran. During the May–July 2025 window, these accounts shared 129,736 posts containing images or videos, which I downloaded to examine their synthetic nature.

Identifying AI-generated visual media

I identified synthetic visual media in my dataset using Sightengine, a commercial system designed to detect AI-generated content. The system was trained on a large corpus of human-created and artificially generated images covering a wide range of content types – including photography, drawings, artwork, and memes – and is designed to detect images produced by major generative models currently in use, such as Stable Diffusion, Midjourney, Adobe Firefly, Flux, and related systems.The model analyzes the pixel-level properties of images and video frames and does not rely

[3]Posts were collected through the Junkipedia API. See `https://docs.junkipedia.org`.

on metadata such as Exchangeable Image File Format (EXIF) information.[4] Since videos contain more than one frame, the system analyzed one frame of videos in my data. In what follows, I use the term "image" to refer to both still images and video frames. After analyzing each image, the system produced a continuous score between 0 and 1, where higher values indicate greater confidence that the content was generated by an AI model. I classified content as AI-generated if the Sightengine score was 0.9 or higher.

Since the model could potentially miss hybrid or heavily edited images, or even confuse human created digital media with AI-generated content, I took several additional steps to assess the detector's performance. First, to examine the model's ability to identify synthetic images generated with tools available in June 2025, I constructed a benchmark sample of 1,000 synthetic images using Flux.1 Schnell, a text-to-image model that was available during the study period,[5] and used it to examine Sightengine's ability to detect them as synthetic. Using the same 0.9 threshold described above, I found that the model was able to identify 968 images (or 96.8%) as AI-generated. While this test only allows me to assess the model's false negative rate, it is useful because it suggests that, at least for known AI images generated with a model available in June 2025 (the conflict period), it detected the vast majority of them. It does not, however, provide a complete estimate of the detector's performance when applied to the heterogeneous visual material found in the wartime corpus.

I therefore conducted a second validation exercise using images drawn from the Iran-Israel wartime image corpus. For this test, I selected a blinded and stratified sample of 300 images that covered different ranges of Sightengine scores (0.5, 0.7, 0.8, 0.9, and 0.95), and coded each of them for whether they appeared – to the human eye – to be AI-generated or not. I found that 79.5% of the images that had a score of 0.9 or higher were coded by humans as appearing AI-generated, 15.9% as not AI-generated, and 4.6% as unclear. The agreement between the human coding and the detector was higher when the classification threshold was 0.95: for this cutoff, 85% were judged by humans to be generated by AI. To make sure that my results are robust to various classification

[4]For more information, see: `https://sightengine.com/docs/ai-generated-image-detection`.

[5]See Table SI-5 for details on the model and the image generation process.

thresholds, I estimated my results using the stricter 0.95 threshold, as well as the lower thresholds. The Supplementary Information shows that the estimated prevalence of AI-generated media in the corpus remains similar across classification thresholds ranging from 0.5 to 0.95, and that my findings on engagement with AI-generated media, as well as the results from the hand-coded matched sample and the reply analysis, are substantively similar under alternative specifications. (Table SI-4; Figures SI-1–SI-2). Figure SI-3 in the Supplementary Information further illustrates how continuous classifier scores map onto corpus visuals around the 0.9 cutoff. Figure 1 below provides additional examples of visuals classified as AI-generated from Iran- and Israel-linked accounts.

Matched hand coding of visual form and communicative function

My next step was to examine the themes that were communicated through the images and videos shared by both sides during the conflict and, in particular, whether the nature of the visual themes differed between AI-generated and non-AI-generated images. For this purpose, I constructed a matched sample of 400 posts shared during the June 13-24 escalation – 200 posts containing AI-generated imagery and 200 that did not – which I used for hand coding. To ensure that the images were comparable, I matched each AI post with a non-AI post shared by the same account, on the same platform, in the same language, and on the same date. I conducted the matching without replacement and restricted each account to contribute no more than 20 pairs. Because actors could choose which messages to express through synthetic imagery, these comparisons are descriptive, i.e., they do not purport to identify a causal role of AI.

Two human coders annotated all 400 images independently. When annotating the images, the coders were shown the image, the text of the post sharing the image, an English translation of the text (where needed), and the affiliation of the posting account – Iran or Israel. The coders did not know whether an image had been classified as AI-generated, although the synthetic character of some images may have been visually apparent.

I developed a codebook to evaluate various aspects of the images. First, I examined the images'

Panel A. Iran-linked accounts



Panel B. Israel-linked accounts



Figure (1) **Illustrative AI-classified visuals circulated by Iran- and Israel-linked accounts.**

relationship to real-world events, coding whether they documented a concrete event, dramatized or symbolically represented an event, portrayed a hypothetical or threatened event, or communicated a general, symbolic message that did not relate to an event. Then, I looked closely at the message that each image communicated. I coded how images portrayed one's own side – e.g., by glorifying the country, leader, or government, or celebrating military power – and how they depicted the

enemy, for example, by showing the enemy as humiliated, weak, or vulnerable. Categories were not mutually exclusive: one image could simultaneously portray one's own side and the enemy's side.

Finally, I examined the images' visual form by coding whether the image looked like a photograph captured with a camera or whether it used different visual styles, such as a cinematic or theatrical composition, a meme-like style, or fantastical depictions that could not be real. When examining the visual form, I also coded whether the image showed national or religious symbols and whether it visually displayed military hardware, as well as destruction, injury, or suffering.[6]

In my analysis, I examined the share of AI and non-AI images assigned to each category. I report the difference between them and its statistical significance using Benjamini-Hochberg adjustments within each family of comparisons. The Supplementary Information provides the full matching procedure, and tabular results for the analyses reported below.

Engagement and public recognition

Finally, I studied the reactions that synthetic images received on the platforms on which they were shared. To evaluate whether audience engagement differed between posts containing AI and non-AI images, I compared platform-native measures of views, likes, comments, and shares. I first compared raw differences between the two groups and then estimated separate models for each outcome that adjusted for follower count and posting activity, including fixed effects for account and date.[7] Since actors may choose to use AI-generated media for topics that are already expected to attract attention, and platforms may promote certain accounts or posts in a way that allows them to gain more visibility, this analysis is correlational.

In addition, I explored whether audiences exposed to AI images showed signs of recognizing

[6] After the coders completed the independent coding, I calculated intercoder agreement using the original annotations (see Table SI-7). I reviewed the cases in which the coders disagreed and reconciled the disagreements before analysis. In the analysis reported below, I use the reconciled coding.

[7] For each engagement outcome $k \in$ views, likes, comments, shares, I estimate

$$\log(1+Y_{iat}^{(k)}) = \beta_k AI_{iat} + \gamma_{1k} Followers_{iat} + \gamma_{2k} PostCount_{iat} + \lambda_k^{\top} Platform_{iat} + \alpha_a + \delta_t + \varepsilon_{iat}^{(k)},$$

where i indexes posts, a indexes focal accounts, and t indexes calendar dates. AI_{iat} equals one when a focal post contains visual media classified as AI-generated; $Platform_{iat}$ denotes a vector of platform indicators; and α_a and δ_t denote account and date fixed effects. Standard errors are clustered at the account level.

them as synthetic by collecting reply data from X, the platform on which AI-generated images were most prevalent and first-level replies could be collected systematically. In this analysis, I used a separate set of posts drawn from the X subset of the full cross-platform corpus used in the engagement analysis. To construct this sample, I first identified the X posts containing images that were classified as AI-generated and collected their replies.[8] I then collected a comparison set of X posts not classified as containing AI images that was matched in size to the AI-post set, and retrieved their replies using the same tool. The final reply sample contains 61,500 replies to 1,545 posts on X: 39,169 replies to 682 posts containing AI images and 22,331 replies to 863 posts without them. This sample consists of the X posts in the original actor-based dataset for which I could find at least one eligible reply after removing duplicates and restricting the data to first-level replies in English, Arabic, Persian, or Hebrew.

I identified replies that referred to AI generally or named a particular AI tool or model – particularly those used for image or video generation – using language-specific dictionaries. The general dictionary captured references to AI-generated content, generative AI, deepfakes, and standalone variants of the abbreviation "AI" in English, Arabic, Persian, and Hebrew. The tool-specific dictionary included references to named generative-AI tools and models, including language-specific transliterations. Replies that included at least one keyword from either dictionary were coded as referring to AI, and zero otherwise.[9]

I then used this measure to study whether replies that mention AI or AI tools are more likely to be directed at posts containing AI images. I examined how frequently recognition appears among replies by regressing the AI recognition measure at the individual reply level on an indicator for whether the focal post – to which the reply was directed – contained an AI-generated image.[10] In

[8] I used the Apify "Twitter Tweet Reply Scraper," which allows extraction of first-level replies to a given tweet ID via the X API endpoints exposed through Apify's scraping infrastructure. For more information, see: `https://www.apify.com`.

[9] Full dictionary lists are reported in Tables SI-10–SI-12. The full dictionary and implementation rules appear in Section in the Supplementary Information.

[10] For reply r to focal post p, shared by account a on calendar date t, I estimate $R_{rpat} = \beta AI_p + \alpha_a + \delta_t + \varepsilon_{rpat}$, where R_{rpat} equals one when the reply refers to AI generally or names an AI tool or model, and AI_p equals one when the focal post contains an image classified as AI-generated. α_a and δ_t denote focal-account and calendar-date fixed effects. Standard errors are clustered by focal post.

addition, I measured how broadly replies referring to AI were distributed across posts containing synthetic images by calculating the share of posts whose replies contained at least one reference to AI or AI tools. Both measures allow me to assess public recognition of AI among users who chose to reply. What they do not allow me to do is estimate the share of all viewers who detected an image as synthetic, or whether the replies that mentioned AI or AI tools were approving or critical.

Results

AI-generated media increased but did not dominate the information environment

Figure 2 shows time trends in posting volume by Iran- and Israel-affiliated accounts, as well as audience attention and engagement with the content they shared. I find that the volume of posts from both sides increased over time, with a clear jump after October 7, 2023, and a much more significant increase during the days of the June escalation, when daily posting reached nearly 6,000 posts at its peak. Audience attention shows similar patterns: daily views of content posted by Iran- and Israel-affiliated accounts approached one billion, and daily interactions exceeded 15 million during the June 13-24, 2025 escalation. This illustrates how much both sides sought to reach audiences during the war and how much attention their efforts seemed to have received. It also offers important context for studying the prevalence of synthetic media: this was a period in which both sides shared an unusually high volume of content.

If generative AI enabled these actors to increase content production to such high levels, then we would expect the prevalence of synthetic content to rise significantly during the conflict period. Yet Figure 3 shows a different pattern: posts containing AI media, while somewhat more prevalent during the days of the escalation, did not dominate the information stream. Panel A shows that, of a total of 212,763 posts shared by both sides from May to July 2025, 129,736 contained visual media, and of these, only 2,236 contained material classified as AI-generated. In other words, only about 1.1% of all posts and 1.7% of posts with media contained AI-generated images and

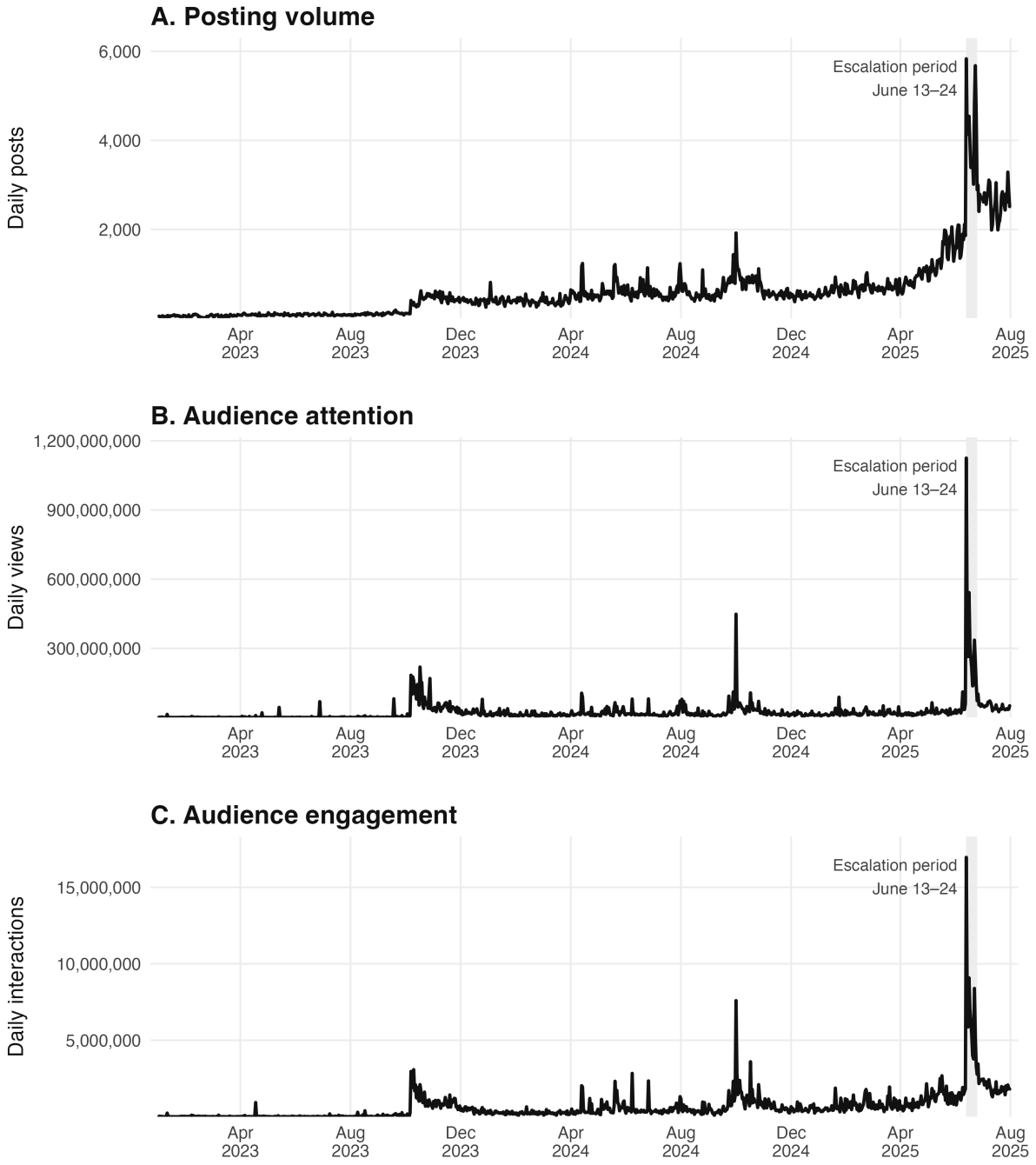


Figure (2) **Posting activity and audience response over time.** Panel A shows the daily number of posts published by accounts linked to Iran and Israel. Panel B shows daily views. Panel C shows daily interactions, defined as the sum of likes, shares, and comments. The shaded region marks the June 13-24, 2025 escalation.

videos. One might wonder whether this result is an artifact of using a restrictive classification rule for AI-generated images. However, when reducing the Sightengine threshold from 0.9 to a highly permissive 0.5, I find that only 2.8% of posts with media were classified as AI-generated

(see Table SI-4 and Figure SI-1).

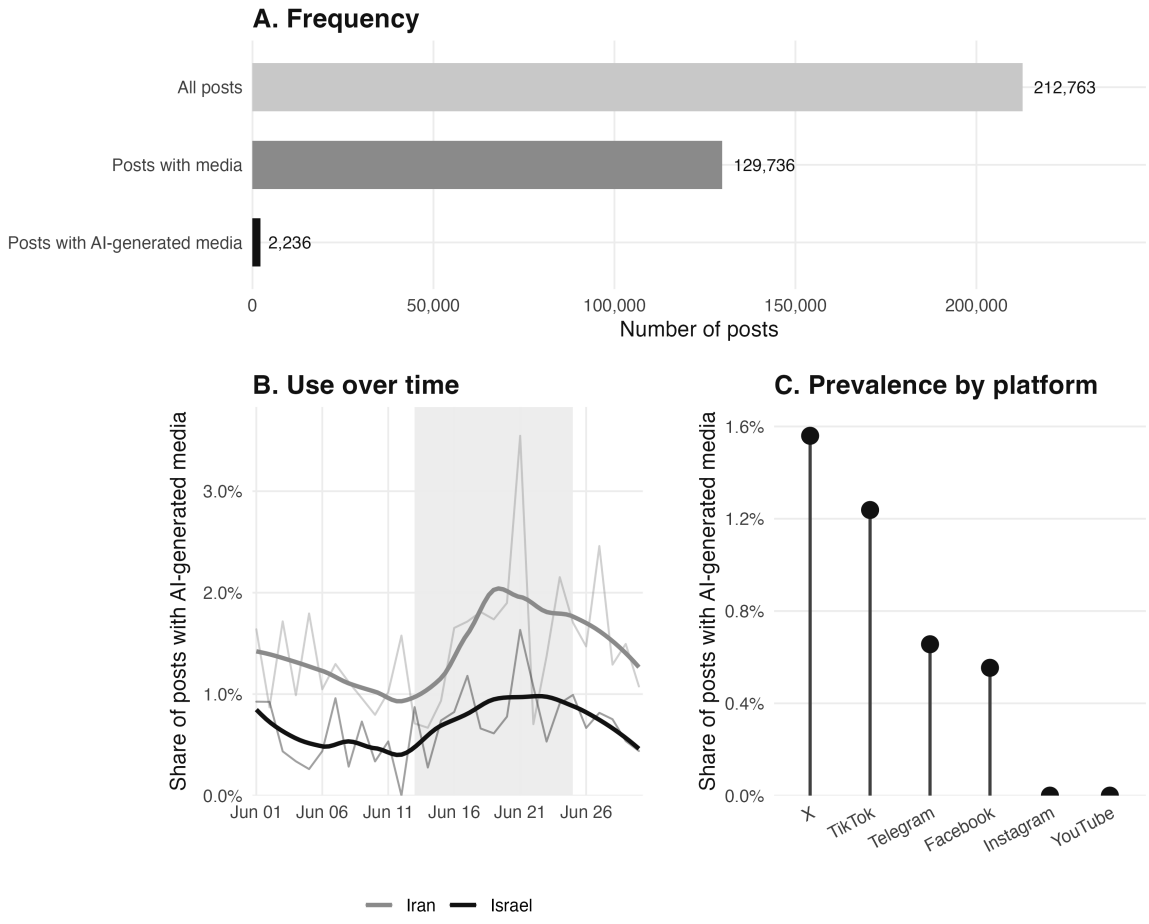


Figure (3) **Prevalence of AI-generated media.** Panel A compares all posts, posts containing any visual media, and posts containing AI-generated media in the May–July 2025 window. Panel B shows the daily share of posts containing AI-generated media for Iran- and Israel-linked accounts during June 2025; thin lines show daily prevalence and thicker lines show smoothed trends, while the shaded region marks the escalation. Panel C reports the share of posts containing AI-generated media by platform.

Panel B shows this pattern over time. I find that AI-generated content became more prevalent for both sides during the escalation, with Iran-linked accounts showing a larger increase, but the prevalence of synthetic visuals remained modest even during the peak of the conflict. Panel C shows that the prevalence of AI-generated media also varied across platforms, with X and TikTok showing the highest share of such content.

These patterns are not consistent with the flooding hypothesis. Even though generative AI tools can significantly decrease production costs and enable disseminating high volumes of content, I do

not find that this was the primary way that Iran- and Israel-affiliated accounts used this technology in this conflict. Interestingly, this finding runs against experts' expectations about the way in which AI systems would be used in influence operations (Goldstein et al., 2023; Wack et al., 2025), and adds context to descriptions of this particular conflict, which was often described as a turning point for AI-generated propaganda (Myers, Odenheimer, and Solomon, 2025). Even though there were thousands of posts sharing AI-generated content, they were far from dominating the information environment.

Synthetic images were focused more on symbolic messages and less on documenting events

To examine the other two patterns of use – fabricating evidence and symbolic storytelling – I next study the themes that were communicated through AI and non-AI images using the matched hand-coded sample. Panel A of Figure 4 shows the extent to which images related to real-world events, such as an attack, a military operation, a public statement, or visible evidence of damage, or did something else, like dramatizing an event, portraying a hypothetical event, or communicating a general, symbolic message. I find systematic differences: while 64% of the matched non-AI images depicted a concrete event, only 30% of synthetic images did. Similarly, while 48% of AI-generated images conveyed a general symbolic message, only 25% of non-synthetic images did. I also find that 9% of AI images represented a future or hypothetical event, compared with 3% of non-AI images. Dramatized versions of real events were also somewhat more common among synthetic images, though the difference was smaller.

While AI-generated media was less likely to document real-world events, that does not mean that it lacked references to the conflict. A general symbolic image could be attached to a specific military development, and a hypothetical scene could refer to a real event. There are plenty of examples of these cases in my data. For example, one synthetic image shared by Israel's Ministry of Foreign Affairs on X portrayed Ali Khamenei, Iran's supreme leader, visibly afraid and surrounded by skeletons representing senior military commanders who were killed in the Israeli strikes that

started the escalation. Similarly, an AI-generated image shared by Khamenei's X account depicted terrified Israeli soldiers fleeing beneath an enormous ring of fire in the sky, with missiles descending around them.[11] What generative AI tools seem to have allowed these actors to do was reduce their reliance on available photographs or footage for messaging campaigns. The technology made it easier to depict broad claims and possible futures, symbolic interpretations of real events or actors, or even the emotional states of leaders that had no direct photographic equivalent.

[11] See `https://x.com/IsraelPersian/status/1935040122350711195` and `https://x.com/Khamenei_fa/status/1936042178867859702` for the original posts. Figure 1 shows additional examples.

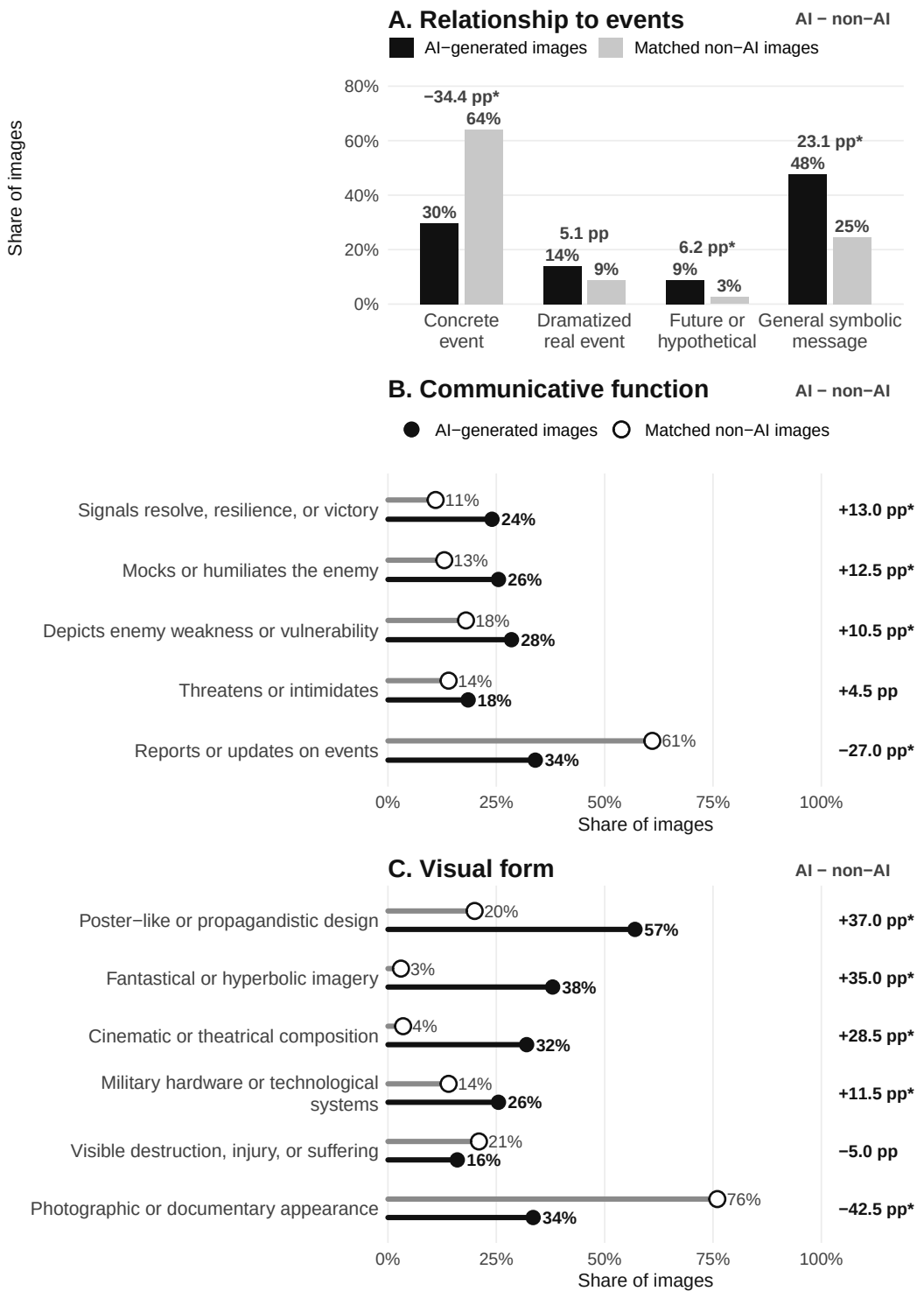


Figure (4) **AI-generated visuals were less documentary and more symbolic and stylized.** The figure compares 200 AI-generated images with 200 matched non-AI images. Panel A shows images' relationship to events, Panel B shows the images' communicative function, and Panel C reports the images' visual form. The "AI − non-AI" on the right of Panels B and C show percentage-point differences . Asterisks indicate Benjamini–Hochberg adjusted $p < .05$ within panel. Full matched-pair estimates and 95% confidence intervals are reported in Table SI-6 in the Supplementary Information.

Panel B shows variation in the messages that were communicated through the images. I find that AI-generated images were 12.5 percentage points more likely to mock or humiliate the enemy – for example, by depicting the leader or government of the rival in a humiliating position or in a way that presented them as ridiculous. Synthetic media were also 13 percentage points more likely to signal the resolve and resilience of one's own side, sometimes through national or religious symbols or by presenting weapons or other military hardware in an epic way. Israel-affiliated accounts, for example, used AI systems to create messages that sought to portray the country as stronger and its military as superior.[12] AI-generated images were also somewhat more likely to communicate threatening or intimidating messages. Iran-affiliated accounts, for instance, often shared dramatic scenes of destruction in Israel alongside terms like "welcome to hell" or "you will be miserable" in Hebrew.[13] These messages contrasted with the tendency of non-AI images to focus on reporting updates or offering details about events.

In Panel C, I examine the visual form of the images – that is, how the images looked to the human eye. I studied various dimensions, including whether the images looked like real photographs or something more cinematic or poster-like. I also examined whether the images featured military hardware and whether they showed visible scenes of destruction, injury, or suffering. I find quite large differences between AI and non-AI images. For example, 38% of AI-generated images had a fantastical or hyperbolic form, compared with only 3% of non-AI images. Similarly, 57% used a poster-like or propagandistic design, compared with 20% of matched non-AI images. I also find large differences in cinematic or theatrical composition, with 32% of the synthetic images having this visual style, compared with only 4% of non-synthetic images. Again, this contrasted with the visual style of non-AI images, 76% of which looked like photos taken with a camera.

These are large differences in the aesthetic form of the visual content shared by both sides during the conflict. Since the images were closely matched by actor, platform, account, and time, the differences suggest a clear, possibly strategic, shift in the communicative goal of this new type

[12] See `https://x.com/IDFFarsi/status/1937142264075100593` and `https://www.tiktok.com/@israel/video/7516926464726420743` for examples.

[13] See `https://x.com/Baran22667157/status/1936544224788590825` and `https://t.me/sepah_pasdaran/164340` for examples.

of media. The finding that AI-generated images were significantly less likely to have a photographic or documentary appearance suggests that the main goal of AI-generated media was perhaps not to fabricate evidence. Of course, some AI images did look like photographs, and the Sightengine classifier may have missed the most convincing synthetic material.[14] That being said, the patterns that emerge from the analysis of the matched images show that AI-generated images were much more likely to use visual styles that departed from realistic photographic forms.

While the visual style of AI and non-AI images differed along the above categories, the differences were less pronounced for the display of military hardware and visible destruction. Synthetic images were somewhat more likely to show weapons or technological systems – often placing them in idealized scenes of technological power or combining them with national symbols. At the same time, they were not more likely than non-synthetic images to show destruction, injury, or suffering. This suggests that AI-generated media were not used to add more graphic violence to the information environment, but instead to disseminate visual representations of symbolic or imagined depictions that supported each side's narrative. These patterns are not driven by images closer to the classifier threshold: the differences in visual form and relationship to events remain substantively unchanged when I restrict the analysis to matched pairs whose AI member received a score of at least 0.95 (Figure SI-2).

The matched comparison thus shows that synthetic images were especially useful for visualizing claims that cameras could not easily capture. Their distinctive visual form seemed to have allowed Iran- and Israel-linked accounts to share vivid interpretations of their preferred narratives about the war's character and trajectory. In the next section, I examine whether this form of wartime messaging also attracted greater audience response.

[14]To examine whether the Sightengine classifier was less likely to identify images with a photographic appearance, I hand-coded the 1,000 known AI-generated images in the Flux sample for whether they had a photographic or plausibly real appearance – that is, whether the image, as it appears, could plausibly have been taken with a camera. The model identified 97.2% of the 720 photographic images and 95.7% of the 280 non-photographic images at the 0.90 threshold; the difference was not statistically significant (Fisher's exact test, $p = .23$).

AI-generated images attracted disproportionate attention

Panel A of Figure 5 reports raw average engagement. I find that posts with AI-generated media received significantly more attention than posts without such media: while posts with AI images received an average of 60,390 views, 1,191 likes, 92 comments, and 358 shares, those without AI media received an average of 33,182 views, 759 likes, 65 comments, and 190 shares. Of course, these descriptive differences may be confounded by other factors: synthetic media, for example, may be concentrated among larger accounts, on particular platforms, or posted during high-attention days.

Thus, in Panel B I present the adjusted associations, which control for the size of the account (measured by the number of followers), the level of the account's activity on the platform at the time of posting, and the platform on which the post appeared, while including fixed effects for account and date. Even after accounting for these confounders, I find that posts containing AI images were associated with 64% more views, 46% more likes, and 24% more comments than posts without AI images. Interestingly, I do not find a difference in shares, for which the adjusted association was very close to zero and statistically insignificant. I examined the robustness of these findings to different AI-image classification thresholds ranging from 0.5 to 0.95 and found that the estimates remain almost identical (see Panel B of Figure SI-1).

Because the analysis is observational, these estimates do not mean that AI-generated media caused greater engagement. There are various alternative explanations for these patterns. For example, actors may reserve synthetic images for communication on emotionally charged topics or use them to make provocative claims that are more likely to grab attention. The results also leave open how viewers interpreted the message: comments on AI images may be hostile, so a higher count does not mean 'positive' engagement, and likes may reflect partisan selection. What the patterns show, however, is that AI-generated media received disproportionate attention even after accounting for major differences in source and timing.

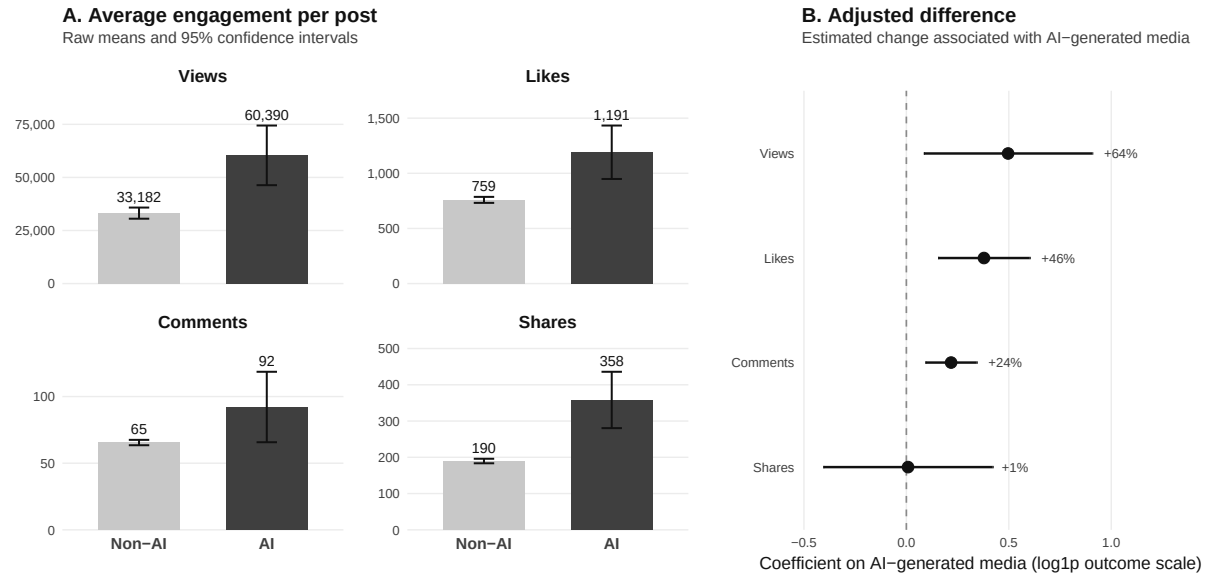


Figure (5) **Engagement with posts containing AI-generated media.** Panel A reports raw means and 95 percent confidence intervals for views, likes, comments, and shares. Panel B reports the coefficient on AI-generated media from separate $\log(1+y)$ models with source, platform, date, follower-count, and posting-volume adjustments. Percentage labels equal $100 \times [\exp(\hat{\beta}) - 1]$. Horizontal lines are 95 percent confidence intervals.

Synthetic images were more likely to receive replies that identified their aritficiality

Finally, I examine whether audiences exposed to synthetic images showed signs of recognizing them as artificial. If the main purpose of AI-generated media was to make it look real, then we would expect those interacting with it to accept it as authentic. I evaluate this possibility in Figure 6, where I examine whether user replies on X that mentioned AI or AI tools varied between the two image types. Panel A shows that synthetic images were significantly more likely to receive replies that mentioned AI or AI tools – while 2.5% of replies to AI-image posts mentioned AI only 0.5% of replies to non-AI posts did. Panel B shows that this sort of public recognition was distributed across a broad set of conversations in replies to synthetic images: 20.2% of posts that shared AI media received at least one AI-mention reply, compared with 3.8% of posts that did not share it.[15]

The reply analysis, however, should be interpreted with caution. First, those who write replies

[15] I find that the estimated recognition gap is similar under the conservative 0.95 classification (Figure SI-2).

are a selected subset of social media users; my analysis cannot reveal how those who did not reply reacted to the images. In addition, replies that mention AI or AI tools may have various meanings: they can call out the artificiality of the image, accuse the account sharing it of deception, admire the production quality of AI tools, or simply laugh at what the image shows. Thus, while the results do not mean that all of those who saw AI images recognized them as synthetic, their artificial nature stood out in user replies to posts sharing them. Interestingly, despite their synthetic nature being publicly discussed, they still attracted substantial attention.

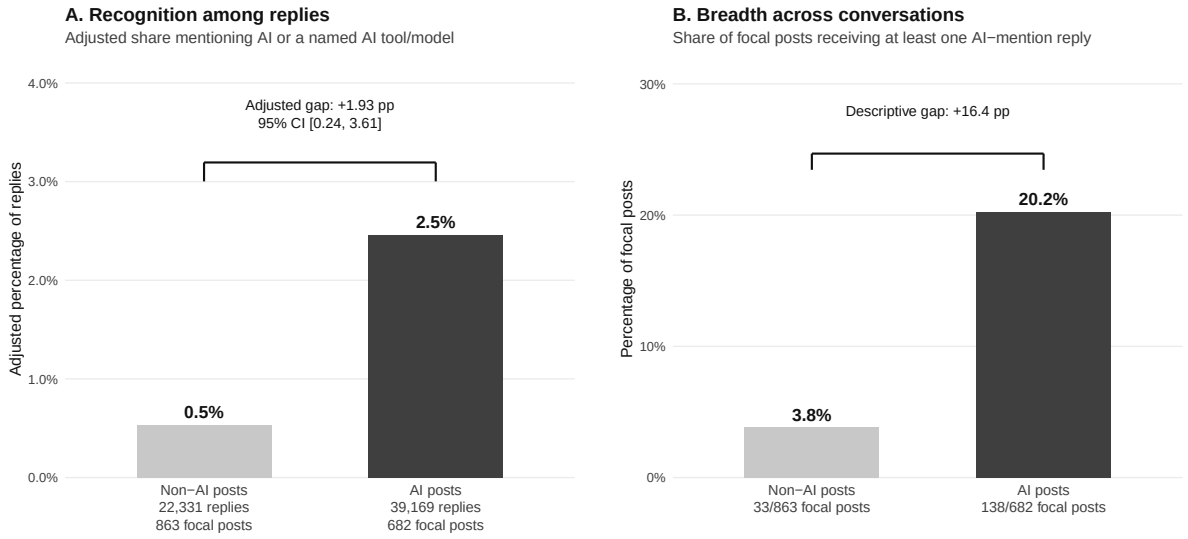


Figure (6) **Public recognition of AI-generated images in X replies.** Replies to posts containing AI-generated images were more likely to mention AI or a named AI tool or model. Panel A reports average predicted probabilities from a reply-level linear probability model with focal-account and calendar-date fixed effects; standard errors are clustered by focal post. The bracket reports the adjusted difference and its 95% confidence interval. Panel B provides a descriptive comparison of the share of focal posts receiving at least one reply that mentions AI or a named AI tool or model.

Taken together, the analysis presented above points to an arena of competition in the global information disorder that goes beyond false information or debates over facts and has not been closely examined in the literature. The patterns I found suggest that the contest over narratives, or interpretations, of the conflict – which seems to take on a new form with the spread of generative AI tools – may become an important part of states' strategic messaging during war. This technology has expanded the visual content that actors can use for this competition, allowing the creation of vivid

scenes that communicate symbolic messages or imagined narratives that cannot be easily captured in photos. While my analysis cannot determine whether this visual style was the reason for the greater engagement with AI images, it suggests that using generative AI for symbolic storytelling may be an important tool for grabbing attention in a crowded information environment.

Discussion

In this paper, I examined what the crisis over information in international politics looked like during the June 2025 military escalation between Iran and Israel – a period in which generative AI tools had become widely available. My analysis examined three expectations regarding AI use: the production of synthetic content at scale, the fabrication of evidence, and the creation of symbolic or imagined representations. Drawing on rich multi-platform data, including an analysis of the visual content of more than 100,000 images, I found the strongest support for the third expectation. While AI-generated media increased during the escalation, it remained a small share of the broader content disseminated online during the war. I also find that synthetic images were not predominantly used for depicting concrete events or mimicking real photos. Instead, they were more likely to communicate symbolic or hypothetical messages, mock the adversary or signal resolve, and much more likely to use poster-like, cinematic, or fantastical forms. Perhaps because of their distinct visual style, AI-generated images attracted more attention on the platforms where they were shared.

These patterns suggest that generative AI has expanded the ways in which actors can visually represent claims about war. While state actors have long had at their disposal various ways to issue threats, defend military action, celebrate success, and criticize their enemies, generative tools enabled them to turn such claims into visual scenes. A notable pattern in the synthetic images shared by Iran and Israel during the June 2025 escalation was the vivid 'story' they told: some depicted the enemy leader as humiliated, afraid, or pathetic; others showcased national symbols as ascendant; still others gave visual form to desired military outcomes that had not yet occurred.

These findings make several contributions to existing research. First, they shift attention from

the aggregate condition of fragmented information in the international system to the behavior of political actors operating within it. Recent work has shown how open information flows, truth-subversion practices, information suppression, and competing sources of expertise can weaken shared factual baselines and complicate global governance (Farrell and Newman, 2021; Adler and Drieschova, 2021; Carnegie, Clark, and Zucker, 2024; Morse and Pratt, 2025; Morse, 2025). In this study, I offered evidence on one understudied way in which state actors have started to adapt to this competitive environment – by using generative AI to make their preferred narratives more visible amid competing voices.

Second, my results suggest that information disorder involves competition not only over facts, but also over their political interpretation. While synthetic media can fabricate events, and the most convincing forms may be difficult to identify with current detection tools, the dominant patterns of AI use that I observed in my data – beyond its use for general symbolic messages – was for visual representations that painted conflict events and developments in one's preferred narrative. These often depicted who was strong, who was weak, what punishment was deserved, and how the conflict might end. Work on legitimation and reputation in international politics has long documented how leaders use these rhetorical forms (Goddard and Krebs, 2015; Krebs, 2015; Snyder, 2015). Generative AI seems to have allowed actors to give these interpretations visual form, turning claims that would otherwise be expressed in words into vivid scenes that could communicate them more immediately.

More generally, the study provides systematic empirical evidence on how state and aligned actors used generative AI for propaganda during a contemporary war. The data cover a wide range of official and unofficial accounts linked to Iran and Israel across six platforms and a large volume of visual content shared during the June 2025 escalation. This allows me to move beyond prominent examples of synthetic wartime media and examine broader patterns in when AI-generated content was used, as well as what it communicated, and how audiences responded to it. These questions have received growing attention, but have yet to be systematically studied in the context of an international military crisis.

There are several limitations that I would like to highlight. First, the classifier that I used for identifying AI-generated images cannot perfectly detect all synthetic media. While the false-negative test using the Flux benchmark showed that the model did a good job detecting known synthetic images, the blinded validation of 300 images from the Iran-Israel dataset showed some noise in the model's ability to detect synthetic images 'in the wild.' Second, the hand-coded sample of 400 images was designed to estimate differences between matched AI and non-AI content, not the full distribution of every visual category across all platforms. There are possibly many other visual dimensions that could be compared in this context that further research could examine. Third, the engagement analysis, as well as the analysis of audience replies on X, is observational. While I find differences between reactions to synthetic and non-synthetic media, there are various factors that can explain these differences. Finally, in this study I focused on a short military escalation between two countries in the Middle East. It is possible that longer wars, conflicts in other parts of the world, or conflicts involving other types of actors, such as nonstate actors, might exhibit different uses of generative AI.

Future research can build on this work to examine whether the patterns identified here travel to other conflicts or crises. It can also compare settings in which actors favor production at scale or fabricating evidence over symbolic storytelling when using generative AI, or examine which organizations produce synthetic media and how that production relates to diplomatic or military strategy. In addition, future work could examine the causal effects of these messaging efforts. For example, it would be fruitful to test whether symbolic synthetic depictions of the conflict affect audience perceptions of resolve, legitimacy, or enemy vulnerability. The research presented here allows us to understand how state actors sought to shape perceptions using content generated by AI, but it would also be important to study whether these efforts are effective.

As international crises take place in a fragmented and increasingly visually mediated information environment, the political effects of synthetic media may depend as much on what they allow actors to represent as on whether they allow actors to deceive.

References

Adler, Emanuel and Alena Drieschova (2021). "The Epistemological Challenge of Truth Subversion to the Liberal International Order". *International Organization* 75.2, pp. 359–386. DOI: 10.1017/S0020818320000533.

Alimardani, Mahsa and Sam Gregory (July 2025). "Iran-Israel AI War Propaganda Is a Warning to the World". *Carnegie Endowment for International Peace*. URL: https://carnegieendowment.org/research/2025/07/iran-israel-ai-war-propaganda-is-a-warning-to-the-world?lang=en.

Amnesty International (July 2025). *Iran: Deliberate Israeli Attack on Tehran's Evin Prison Must Be Investigated as a War Crime*. Published July 22, 2025. URL: https://www.amnesty.org/en/latest/news/2025/07/iran-deliberate-israeli-attack-on-tehrans-evin-prison-must-be-investigated-as-a-war-crime/.

Birrer, Alena and Natascha Just (2025). "What We Know and Don't Know About Deepfakes: An Investigation into the State of the Research and Regulatory Landscape". *New Media & Society* 27.12, pp. 6819–6838. DOI: 10.1177/14614448241253138.

Burgess, Matt (June 2025). *Iran's Internet Blackout Adds New Dangers for Civilians amid Israeli Bombings*. Published June 18, 2025. URL: https://www.wired.com/story/iran-internet-shutdown-israel/.

Carnegie, Allison and Austin Carson (2018). "The Spotlight's Harsh Glare: Rethinking Publicity and International Order". *International Organization* 72.3, pp. 627–657. DOI: 10.1017/S0020818318000176.

— (2020). *Secrets in Global Governance: Disclosure Dilemmas and the Challenge of International Cooperation in World Politics*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108778114.

Carnegie, Allison, Richard Clark, and Noah Zucker (2024). "Global Governance under Populism: The Challenge of Information Suppression". *World Politics* 76.4, pp. 639–666. DOI: 10.1353/wp.2024.a939189.

Chesney, Robert and Danielle Keats Citron (2019). "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security". *California Law Review* 107.6, pp. 1753–1819. URL: https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security.

Chubb, Andrew and Frances Yaping Wang (2023). "Authoritarian Propaganda Campaigns on Foreign Affairs: Four Birds, One Stone, and the South China Sea Arbitration". *International Studies Quarterly* 67.3, sqad047. DOI: 10.1093/isq/sqad047.

Cornwell, Alexander, Parisa Hafezi, and Steve Holland (June 13, 2025). "Iran strikes back at Israel with missiles over Jerusalem, Tel Aviv". *Reuters*. URL: https://www.reuters.com/world/middle-east/israel-says-it-strikes-iran-amid-nuclear-tensions-2025-06-13/.

Factnameh (July 14, 2025). *The First AI War: How the Iran-Israel Conflict Became a Battlefield for Generative Misinformation*. European Digital Media Observatory. URL: https://edmo.eu/publications/the-first-ai-war-how-the-iran-israel-conflict-became-a-battlefield-for-generative-misinformation/.

Farrell, Henry and Abraham L. Newman (2021). "The Janus Face of the Liberal International Information Order: When Global Institutions Are Self-Undermining". *International Organization* 75.2, pp. 333–358. DOI: 10.1017/S0020818320000302.

Fittarelli, Alberto, Maia Scott, Ron Deibert, Marcus Michaelsen, and Darren Linvill (Oct. 2025). *We Say You Want a Revolution: PRISONBREAK—An AI-Enabled Influence Operation Aimed at Overthrowing the Iranian Regime*. Tech. rep. 189. The Citizen Lab, Munk School of Global Affairs and Public Policy, University of Toronto. URL: https://citizenlab.ca/research/2025-10-ai-enabled-io-aimed-at-overthrowing-iranian-regime/.

Gartzke, Erik and Jon R. Lindsay (2015). "Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace". *Security Studies* 24.2, pp. 316–348. DOI: 10.1080/09636412.2015.1038188.

Goddard, Stacie E. and Ronald R. Krebs (2015). "Rhetoric, Legitimation, and Grand Strategy". *Security Studies* 24.1, pp. 5–36. DOI: 10.1080/09636412.2014.1001198.

Goldstein, Josh A., Girish Sastry, Micah Musser, Renée DiResta, Matthew Gentzel, and Katerina Sedova (Jan. 2023). *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations*. Tech. rep. Center for Security, Emerging Technology, OpenAI, and Stanford Internet Observatory.

Hartmann, Jochen, Yannick Exner, and Samuel Domdey (2025). "The Power of Generative Marketing: Can Generative AI Create Superhuman Visual Marketing Content?" *International Journal of Research in Marketing* 42.1, pp. 13–31. DOI: 10.1016/j.ijresmar.2024.09.002.

Hees, Jules van, Tijl Grootswagers, Genevieve L. Quek, and Manuel Varlet (2025). "Human Perception of Art in the Age of Artificial Intelligence". *Frontiers in Psychology* 15, p. 1497469. DOI: 10.3389/fpsyg.2024.1497469.

King, Gary, Jennifer Pan, and Margaret E. Roberts (2017). "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument". *American Political Science Review* 111.3, pp. 484–501. DOI: 10.1017/S0003055417000144.

Krebs, Ronald R. (2015). "Tell Me a Story: FDR, Narrative, and the Making of the Second World War". *Security Studies* 24.1, pp. 131–170. DOI: 10.1080/09636412.2015.1001215.

Lanoszka, Alexander (2019). "Disinformation in International Politics". *European Journal of International Security* 4.2, pp. 227–248. DOI: 10.1017/eis.2019.6.

Lesch, Max (2023). "The Politics and Practice of International Fact-Finding Missions". *International Studies Review* 25.3, viad034. DOI: 10.1093/isr/viad034.

Lidman, Melanie (June 14, 2025). "Israelis are uneasy as they prepare for more Iranian missiles". *Associated Press*. URL: https://apnews.com/article/israel-iran-missiles-attack-tel-aviv-0d75e7309887a0ef0fd55554e3ba9415.

Lu, Yingdan and Jennifer Pan (2022). "The Pervasive Presence of Chinese Government Content on Douyin Trending Videos". *Computational Communication Research* 4.1, pp. 68–98. DOI: 10.5117/CCR2022.2.002.LU.

Lu, Yingdan, Jennifer Pan, Xu Xu, and Yiqing Xu (2025). "Decentralized Propaganda in the Era of Digital Media: The Massive Presence of the Chinese State on Douyin". *American Journal of Political Science*. Online first, pp. 1–17. DOI: 10.1111/ajps.12990.

Miller, David and Piers Robinson (2019). "Propaganda, Politics and Deception". *The Palgrave Handbook of Deceptive Communication*. Ed. by Tony Docan-Morgan. Cham: Palgrave Macmillan, pp. 969–988. DOI: 10.1007/978-3-319-96334-1.

Miskimmon, Alister, Ben O'Loughlin, and Laura Roselle (2013). *Strategic Narratives: Communication Power and the New World Order*. New York: Routledge. ISBN: 9780415717601. DOI: 10.4324/9781315871264.

Morse, Julia C. (2025). "Information Fragmentation and Global Governance in Hard Times". *Ethics & International Affairs* 39.2, pp. 159–172. DOI: 10.1017/S0892679425100130.

Morse, Julia C. and Tyler Pratt (2022). "Strategies of Contestation: International Law, Domestic Audiences, and Image Management". *The Journal of Politics* 84.4, pp. 2080–2093. DOI: 10.1086/719418.

— (2025). "Information Disorder and Global Politics". *International Organization* 79.S1, S26–S43. DOI: 10.1017/S0020818325101069.

Myers, Steven Lee, Natan Odenheimer, and Erika Solomon (July 2025). "Israel and Iran Usher In New Era of Psychological Warfare". *The New York Times*. URL: https://www.nytimes.com/2025/07/15/technology/israel-iran-psychological-warfare.html.

Nyhan, Brendan, Jennifer Pan, Alexandra Siegel, and Yamil Velez (2026). "Easy to Produce, Hard to Persuade: The Asymmetric Effects of AI on the Online Information Ecosystem". APSA Preprints, version 2, May 7, 2026. DOI: 10.33774/apsa-2026-mn97c-v2.

Roberts, Margaret (2018). *Censored: distraction and diversion inside China's Great Firewall*. Princeton University Press.

Sa'ar, Gideon (June 23, 2025). *We Warned Iran Time and Again: Stop Targeting Civilians! They Continued, Including This Morning. Our Response: Viva la Libertad, Carajo!* Post on X. X. URL: https://x.com/gidonsaar/status/1937083300373541216 (visited on 07/29/2026).

Snyder, Jack (2015). "Dueling Security Stories: Wilson and Lodge Talk Strategy". *Security Studies* 24.1, pp. 171–197. DOI: 10.1080/09636412.2015.1003723.

Tehran Times (June 14, 2025). *#BREAKING Doomsday in Tel Aviv*. X. URL: https://x.com/TehranTimes79/status/1933850991599767749.

Twomey, Jack, Dominic Ching, Matthew P. Aylett, Michael Quayle, Conor Linehan, and Gillian Murphy (2023). "Do Deepfake Videos Undermine Our Epistemic Trust? A Thematic Analysis of Tweets That Discuss Deepfakes in the Russian Invasion of Ukraine". *PLOS ONE* 18.10, e0291668. DOI: 10.1371/journal.pone.0291668.

Vaccari, Cristian and Andrew Chadwick (2020). "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News". *Social Media + Society* 6.1, pp. 1–13. DOI: 10.1177/2056305120903408.

Wack, Morgan, Carl Ehrett, Darren Linvill, and Patrick Warren (2025). "Generative Propaganda: Evidence of AI's Impact from a State-Backed Disinformation Campaign". *PNAS Nexus* 4.4, pgaf083. DOI: 10.1093/pnasnexus/pgaf083.

Supplementary Information

Social media data collection

The broader data from which the June 2025 Iran-Israel dataset is drawn consist of public social media posts shared by official and unofficial accounts linked to Israel, Hamas, Iran, Hezbollah, and the Houthis from January 1, 2023, to July 31, 2025. The data include content shared on Facebook, Instagram, YouTube, X, Telegram, and TikTok.

I collected the data by constructing two lists. First, I identified official accounts belonging to governments, organizations, or formal actor-linked entities on each platform. Second, I expanded this list outward to identify unofficial affiliated accounts by conducting platform searches and drawing on reports from organizations conducting open-source platform investigations. This produced a broader set of accounts that included aligned media outlets, cyber units, semi-official channels, and other accounts that consistently amplified the actor's messaging.

I classified official accounts conservatively: I only considered accounts official when there was direct institutional ownership, and accounts that supported an actor's goals but did not clearly belong to the actor institutionally, I coded as unofficial. While in theory, any social media user could be considered part of an actor's information network, especially if they share and amplify the messaging of the side they support, I distinguish between organized networks and organic posting from sympathetic users. The full list of accounts is shown in Table SI-1.

I collected public posts from these accounts prospectively using a third-party API.[16] For each post, I kept all available metadata, including account ID, post ID, post text, publication date, engagement measures, and attached media. The dataset includes original posts as well as reposts, retweets, and forwards. If posts were deleted by the time I collected the data, they were not included, but because I collected data prospectively since January 27, 2025, the collection likely captured a substantial amount of content from the Iran-Israel escalation of June 2025, even if it was later removed. After collection, I combined the posts from all six platforms into one dataset. Table SI-2 shows the number of accounts and posts by actor in the full dataset. Table SI-3 shows summary statistics for the part of the dataset on which I focus in this paper – content posted by official and unofficial accounts linked to Iran and Israel between May 1 and July 31, 2025.

[16] See `https://docs.junkipedia.org`.

Table (SI-1) List of Accounts by Actor and Platform

Actor	Platform	Account Names
Iran official	Facebook	Foreign Ministry, Islamic Republic of Iran
Iran official	Instagram	Dr. Masoud Pezeshkian; Foreign Ministry of the I.R.IRAN; IRhebrew; Pars Today; پایگاه اطلاع رسانی دولت ;سید ابراهیم رئیسی شهید حسین امیرعبداللهیان
Iran official	Telegram	Islamic Republic of Iran; سپاه سایبری پاسداران IRGC 🏴; سپاه سایبری پاسداران IRGC🚩; کانال اخبار سپاه ;سید ابراهیم رئیسی کانال سپاه پاسداران 🏴; پاسداران 🏴
Iran official	TikTok	Khamenei Quotes
Iran official	Twitter	🇵🇸🇮🇷naghizadeh; Ajatollah Khamenei; Government of the Islamic Republic of Iran; Government of the Islamic Republic of Iran🇮🇷; H.Amirabdollahian امیرعبداللهیان; Iran Defense (unofficial); Iran Defense commentary (unofficial); Iran Defense| نیروهای مسلح جمهوری اسلامی ایران; Iran Military; IRAN Military; Iran Military (Commentary); Islamic Iran Military; Khamenei Media; Khamenei Site Web; Khamenei TV; Khamenei.ir; KHAMENEI.IR | فارسی 🇮🇷; Khamenei.ir Hebrew; Masoud Pezeshkian; Pars Today English; ParsToday English; Saeed Khatibzadeh | سعید خطیب‌زاده ;سید ابراهیم رئیسی ;علی شمخانی وزارت دفاع و پشتیبانی نیروهای مسلح ;محمدباقر قالیباف
Iran unofficial	Facebook	Press TV Highlights; SimayAzadi; חדשות סלבריטאים ;נויה לוי ;נירה לוי ;שירה מעין
Iran unofficial	Instagram	Press TV; Tasnim News Hebrew; עקירת ישראלים 💔; خبرگزاری ایرنا | IRNA News ;خبرگزاری تسنیم ;خبرگزاری فارس ;رسانه نگار ;کار آنلاین
Iran unofficial	Telegram	ישראלים פטריוטים🇮🇱; 🇮🇱ללא צנזורה🇮🇱; Attack Alarm|אזעקת תקיפה; Bring Them Home Now ! מחזירים אותם הביתה ;אנרוף - נוש קטיף-חוזרים הביתה ;עכשיו CAR ONLINE; PARS TODAY פארס טודיי; Press TV; PressTV; seculars of israel; דמעות המלחמה ;זכותנו ;ישראל השנייה ;מחאת הדגלים השחורים ;צבא העם 🎗; خبرگزاری ایرنا ;خبرگزاری تسنیم ;אין קול ;אדוק کانال گنجشک درنده ;خبرگزاری فارس
Iran unofficial	TikTok	🇮🇷IRANTEHRAN🇵🇸🇵🇰; Iran Military; Israel.news; jewishfist
Iran unofficial	Twitter	🇮🇷دلداده ی سید علی🇮🇷; 🇮🇷علیرضا آخونپور🇮🇷; 🚩حاج دوگین; ALIgenab.; Attack Alarm|אזעקת תקיפה; CARONLINE; Cyber Court; Cyber Flood; CyberAv3ngers; Daily Iran Military; Daily Iran News; Daily Iran News Commentary; Fars News Agency; Global Affairs; Gonjeshke Darande; Iran Defense Forces; Iran International English; Iran Military; IRAN Military; Iran Military Monitor; Iran Military Watch; Iran News; Iran News Daily; Iran News Daily Commentary; Iran News@ ; Iran Updates; IRAN WATCH; IranRadio אירן רדיו; IRNA News Agency; KhameneiNews; MonitorX; MonitorX; Moses Staff; Noon; Saeed Al-Samahiji; Sahar Emami; SilencedSirs; SilencedSirs ⬛; Simay Azadi; Tasnim News Agency; TasnimNews Hebrew; אירן מאיר ;ישראל השנייה ;מינצ'יקית (ימינצ'ה) ;מינצ'יקית ;יואל משה ;חדשות טסנים ;הדר קורן ;בחירות עכשיו ;מיליטרי ניוז ;קול דוד ;מרים קוזק ;בחשכה. يلمع في الظلام ;ابوالفضل ;اخبار الايراني ;حسین طاهری 🇮🇷; خبرگزاری ایرنا 🇮🇷; خبرگزاری تسنیم کانال سپاه ;علی اکبر رائفی پور ;سید احسان خاندوزی ;روزنامه جام جم ;رسانه نگار ;خبرگزاری فارس AI NEWS; 🇮🇷 خبرگزاری تسنیم | پاسداران ;IRGC ;محمدرضا گل‌محمدی
Israel official	Facebook	חיל האוויר
Israel official	Instagram	Benjamin Netanyahu - בנימין נתניהו; Israel; Israel Defense Forces; Israel Ministry of Foreign Affairs; Israelpersian; Yariv Levin - יריב לוין; Yoav Gallant יואב גלנט; ישראל כ"ץ | Israel Katz; إسرائيل بالعربية ;ارتش دفاعی اسرائیل به فارسی
Israel official	Telegram	اسرائیل به فارسی ;בנימין נתניהו
Israel official	TikTok	@Israel; Benjamin Netanyahu - נתניהו; IDF; יריב לוין - Yariv Levin; ارتش دفاعی اسرائیل
Israel official	Twitter	Benjamin Netanyahu - בנימין נתניהו; COGAT; Embassy of Israel to the USA; Embassy of Israel to the USA | #IsraelUSA75; Gideon Sa'ar | גדעון סער; Government Press Office 🇮🇱; Israel; Israel Defense Forces; Israel Foreign Ministry; ISRAEL STRONGER; Israel ישראל; Israel ישראל 🇮🇱; Israeli Air Force; Mossad Farsi; Prime Minister of Israel; Ron Dermer - רון דרמר; איתמר בן גביר ;'בצלאל סמוטריץ ;יואב גלנט - Yoav Gallant; יצחק הרצוג Isaac Herzog; ישראל כ"ץ Israel Katz; משרד הביטחון; צבא ההגנה לישראל ;ارتش دفاعی اسرائیل | IDF Farsi ;افيخاي ادرعي ;اسرائیل به فارسی ;کابتن إيلا עמיחי שיקלי - Amichai Chikli; نخست وزیر کشور إسرائيل ;Captain Ella
Israel official	YouTube	Israel Defense Forces; ارتش دفاعی اسرائیل
Israel unofficial	Facebook	Noa Tishby; StandWithUs; The Times of Israel; Times of Israel Persian; Yair Netanyahu; ynet; החדשות - N12
Israel unofficial	Instagram	C14; Shirion Collective; The Times of Israel; Yair Netanyahu - יאיר נתניהו; ynet; ynetnews; החדשות - N12; כאן | תאגיד השידור הישראלי
Israel unofficial	Telegram	تایمز اسرائیل به فارسی ;Ynetnews - Latest Stories; חדשות ynet ;יאיר נתניהו Yair Netanyahu ;צ'אט הכתבים N12
Israel unofficial	TikTok	Hananya Naftali; ILTV news; N12; Ynet News; כאן
Israel unofficial	Twitter	(((noa tishby))); Aaliyah Martinez Paul; Ambassador Gilad Erdan גלעד ארדן; C14; David Saranga; Eylon Levy; Hananya Naftali; ILTV Israel News; ISRAEL MOSSAD; Mossad Commentary; Noa Tishby; Shirion Collective; The Jerusalem Post; The Moral Alliance; The Times of Israel; Times of Israel AR; Yair Netanyahu; Yair Netanyahu🇮🇱; ynet עדכוני; Ynetnews; yolan cohen יולן כהן; Yollan cohen יולן כהן ;החדשות - N12 ;ידיעות אחרונות ;ינון מגל ;כאן ;רשת ב | כאן ;רשת ב | כאן חדשות ;עמית סגל ;עמית סגל Amit Segal; حنانيا نفتالي
Israel unofficial	YouTube	mako IL
Hamas official	Instagram	المكتب الإعلامي الحكومي
Hamas official	Telegram	حركة الجهاد الإسلامي في فلسطين - المكتب الإعلامي ;حركة 🔻 الاعلام العسكرى 🔻 ;أبو عبيدة "الناطق العسكري باسم كتائب القسام" ;كتائب الشهيد عز الدين القسام ;قناة الأقصى ;سرايا القدس الإعلام الحربي (الرسمية) ;سرايا القدس الإعلام الحربي ;حماس وكالة شهاب ;وزارة الصحة الفلسطينية/ غزة ;"الدين القسام" الاحتياطية
Hamas official	TikTok	وزارة الصحة الفلسطينية
Hamas official	Twitter	Palestine Info Center; Quds News Network

Table (SI-1) List of Accounts by Actor and Platform (continued)

Actor	Platform	Account Names
Hamas unofficial	Facebook	وكالة شهاب; دائرة الاوقاف الاسلامية في القدس
Hamas unofficial	Instagram	وكالة شهاب الإخبارية; Quds network	شبكة قدس
Hamas unofficial	Telegram	GAZA NOW IN ENGLISH; Global Resistance News; ISNAD - Palestine Media; Quds News Network; Resistance News Network; Resistance Trench; ResistanceTrench; أخبار الأسرى في سجون الاحتلال; إعْلَامُ الْعَسْكَرِي	حَالَاتُ قَسَّامِيَّة	كَتَائِبَ الْقَسَّام; قدس ميديا; غزة الآن - Gaza Now; شبكة قدس الإخبارية; ترجمات عبرية - شبكة قدس
Hamas unofficial	TikTok	4d4mk; Palestine🇵🇸; Vikkibonbikki
Hamas unofficial	Twitter	🔻TG Wilson🔻; Motasem A Dalloul; Suppressed News; Suppressed News.; Suppressed Nws; Suppressed Nws.; Suppressed Voice; Suppressed Voice.; T.G. Wilson; T.G. Wilson 🔻; TG Wilson; TG Wilson 🔻; The Palestine Chronicle; The Saviour; Warfare Analysis; شبكة قدس الإخبارية; نحو الحرية
Hezbollah unofficial	Facebook	Al-Akhbar English; Ramia al Ibrahim - راميا الابراهيم
Hezbollah unofficial	Instagram	Al-Akhbar English; Ramia Al Ibrahim
Hezbollah unofficial	Telegram	Al-Akhbar English; Global Politics; Hezbollah Martyrs; Hezbollah.ru	Хезболла	; Lebanese News and Updates; ManarTV-EN; War Monitor; قناة المنار🇱🇧 - Al Manar TV; علي شعيب - أخبار الحدود; حزب الله الإلكتروني; إذاعة النور-أخبار; إذاعة النور البث المباشر; ملتقى الإعلام المقاوم - سيميا
Hezbollah unofficial	Twitter	Al Mayadeen English; Al-Akhbar English; Ali Hashem علي هاشم; Hadi; Hadi Nasrallah; Leb Now; ramia al ibrahim - راميا الابراهيم; Sitio Al Manar; War Monitor; إذاعة النور; جريدة الأخبار - Al-Akhbar; قناة المنار; نور فاطمة
Hezbollah unofficial	YouTube	Al Mayadeen English; Lebanon On; شخصيات
Houthis official	Twitter	Yahya Sare'e; حزام الأسد; محمد علي الحوثي; نصر الدين عامر; نصر الدين عامر	Nasruddin Amer
Houthis unofficial	Facebook	Ofoq Foundation مؤسسة أفق
Houthis unofficial	Instagram	Ofoq Foundation مؤسسة أفق
Houthis unofficial	Telegram	Almasirah English; Yemen Military; Yemeni Army; Yemeni Army	Houthis; أعمال تلفزيونية - يحيى الشامي
Houthis unofficial	Twitter	مؤسسة فكرة Ofoq Foundation مؤسسة أفق; قناة المسيرة; ضيف الله الشامي; شاهد المسيرة

Note: Channel names are reported as they appeared in the source data. Some channels used more than one display name over time, including minor variations in punctuation or spelling; these variants are retained to reflect how the accounts presented themselves at collection.

Table (SI-2) Number of Accounts and Posts by Actor

Actor	Official accounts	Unofficial accounts	Total accounts	Posts
Iran	32	88	120	406,750
Israel	46	51	97	182,448
Hamas	14	28	42	1,024,229
Hezbollah	0	32	32	149,369
Houthis	4	10	14	18,443

Table (SI-3) Summary Statistics for the Iran–Israel Sample, May 1–July 31, 2025

Panel A. Sample composition

Measure	Value
Unique accounts	196
Iran-linked accounts	101 (51.5%)
Israel-linked accounts	95 (48.5%)
Official accounts	72 (36.7%)
Unofficial accounts	124 (63.3%)
Platforms	6
Posts	212,763
Iran-linked posts	120,812 (56.8%)
Israel-linked posts	91,951 (43.2%)
Posts with visual media	129,736 (61.0%)
Posts with AI-generated media	2,236 (1.1%)
AI-generated posts among visual posts	1.7%
Posts during June 13–24 escalation	51,307 (24.1%)

Panel B. Post-level engagement and audience measures

Variable	N	Mean	SD	Median	P25	P75
Views per post	212,763	33,467.6	611,389.1	2,150.0	38.0	13,700.0
Likes per post	212,743	763.6	6,367.7	0.0	0.0	55.0
Comments per post	212,763	65.8	474.2	0.0	0.0	5.0
Shares per post	212,763	191.5	1,468.1	0.0	0.0	20.0

Corpus-based validation of AI-image classifications

In the paper, I present results from an analysis that evaluates whether the Sightengine detector was able to recognize images known to have been produced by a generative AI model. However, since all images in this benchmark analysis are synthetic, this analysis cannot show how the classifier performs on the heterogeneous mixture of photographs, posters, memes, illustrations, and video frames contained in the wartime dataset used in this paper. Thus, I conducted an additional validation exercise drawing on images from my dataset.

I created a sample of images whose Sightengine score ranged from 0.5 to 1. To ensure that I have enough images across these scores, I divided the data into five ranges and drew a stratified random sample of 300: 80 images with scores from 0.95 to 1, 80 with scores from 0.9 to 0.949, 80 with scores from 0.8 to 0.899, 30 with scores from 0.5 to 0.799, and 30 with scores below 0.5. I sampled images from Iran-linked and Israel-linked accounts approximately in proportion to their representation in that stratum. I sampled more images in ranges near the 0.9 threshold in order to examine the performance of the detector around the cutoff more closely.

One human coder annotated these images while being blind to all the information surrounding them, such as the Sightengine score, the account that posted the image, the text of the post, and other metadata. The coder annotated each image for whether it appeared to have been generated by generative AI or not.

I calculated the estimated proportions of the human coder judgments in each category using the inverse probability of selection within each stratum. I did not use unweighted proportions in the annotated sample, as these would not represent the distribution of the full corpus because I sampled images differently across score ranges and actor affiliations. Among the 2,238 unique images with Sightengine scores of 0.9 or higher, an estimated 79.5 percent were coded by the human annotator as appearing AI-generated (95 percent CI: 72.7–86.3), 15.9 percent were judged not AI-generated (95 percent CI: 10.1–21.7), and 4.6 percent were coded as unclear (95 percent CI: 0.4–8.7). It is important to note that these comparisons do not reflect the detector's true false-positive or false-negative rates. It is possible for human coders evaluating AI images to miss some synthetic media that looks very similar to real photographs. Thus, we can think of this test as a measure of agreement between the Sightengine classification and the human coding.

Table SI-4 shows how this agreement varies across classification thresholds. In the main analysis, I use a threshold of 0.9. At this threshold, 2,236 images are classified as AI-generated – 1.7% of posts containing visual media. When raising the threshold to 0.95, this prevalence decreases to 1.5%, and when lowering it to a very permissive 0.5, the prevalence increases to 2.8%, which is still low compared with the broader set of visual media shared by both sides during the conflict.

When comparing the two tests across these different thresholds, I find that the Sightengine detector identified almost all of the known Flux-generated images at every cutoff: its performance increased from 95.7% at the 0.95 threshold to 98.2% at the 0.5 threshold. However, its agreement with the human assessment decreased as I relaxed the thresholds. While an estimated 85% of images in my corpus that received a score of 0.95 or higher from Sightengine were coded by the human annotator as AI-generated, this estimated proportion decreased to 79.5% at the 0.9 threshold, 71.8% at the 0.8 threshold, 67.8% at the 0.7 threshold, and 57.4% at the 0.5 threshold. Thus, the 0.9 cutoff allowed me to maintain high sensitivity to known synthetic images while avoiding coding images with lower Sightengine scores as AI-generated, many of which did not appear synthetic to the human eye.

Robustness to alternative classifier thresholds

In Figure SI-1, I examine whether the prevalence and engagement results depend on the threshold that I used to classify AI-generated media. Panel A shows the share of posts with images that were classified as AI-generated across thresholds, both overall and separately for Iran- and Israel-linked accounts. The figure shows that the prevalence of posts sharing AI-generated media does not change much across thresholds and remains low even when using the very relaxed threshold of 0.5. This offers further support for the finding in

Table (SI-4) Sensitivity of AI-media prevalence and validation results to the classifier threshold

Threshold	Interpretation	AI-classified posts	Share of visual posts	Flux benchmark detected	Audited positives judged to appear AI
0.95	Conservative specification	1,966	1.5%	95.7% (957/1,000)	85.0%
0.90	**Main specification**	**2,236**	**1.7%**	**96.8% (968/1,000)**	**79.5%**
0.80	Permissive specification	2,638	2.0%	97.4% (974/1,000)	71.8%
0.70	More permissive specification	2,939	2.3%	97.7% (977/1,000)	67.8%
0.50	Highly permissive stress test	3,653	2.8%	98.2% (982/1,000)	57.4%

the main paper that AI-generated media constituted a small share of the wartime content shared by accounts on both sides.

Panel B replicates the engagement analysis across thresholds, using the same model reported in the main paper. I find that the results remain almost identical across thresholds, both in their magnitude as well as their statistical significance. The engagement results therefore do not appear to be driven by posts located close to the main classification cutoff.

Next, I replicated the analysis of the hand-coded sample to examine how the patterns in the visual form and communicative function of the images looked when using the conservative threshold of 0.95. Because I originally created this sample using the 0.9 threshold, I kept the existing matched pairs in which the AI member received a Sightengine score of at least 0.95, while preserving its original matched non-AI comparison. This left me with 346 images, or 173 of the original 200 pairs.

Panel A of Figure SI-2 shows that using the more conservative threshold did not change the results. AI-generated images remain substantially less likely to document concrete events or to have a photographic or documentary appearance, and substantially more likely to convey general symbolic messages or use cinematic, poster-like, and fantastical visual forms. In fact, I find that the estimated differences are similar to, and in several cases slightly larger than, those obtained at the main threshold.

In Figure SI-2B, I do the same with the reply analysis. I keep replies to posts with AI images that scored 0.95 or higher on Sightengine and drop those that received a score between 0.9 and 0.949. The replies to non-AI image posts remained the comparison here as well. I find that restricting the AI group to posts scoring at least 0.95 produced a very similar estimated difference to the one found when using the 0.9 cutoff: 1.73 vs. 1.93 percentage points. However, the estimate for the 0.95 threshold is less precise, perhaps because this analysis reduced the number of AI focal posts to which replies were directed from 682 to 586. Nonetheless, the similarity of the point estimates indicates that the observed recognition of AI-generated media in my dataset is not concentrated among posts that were immediately above the 0.90 cutoff.

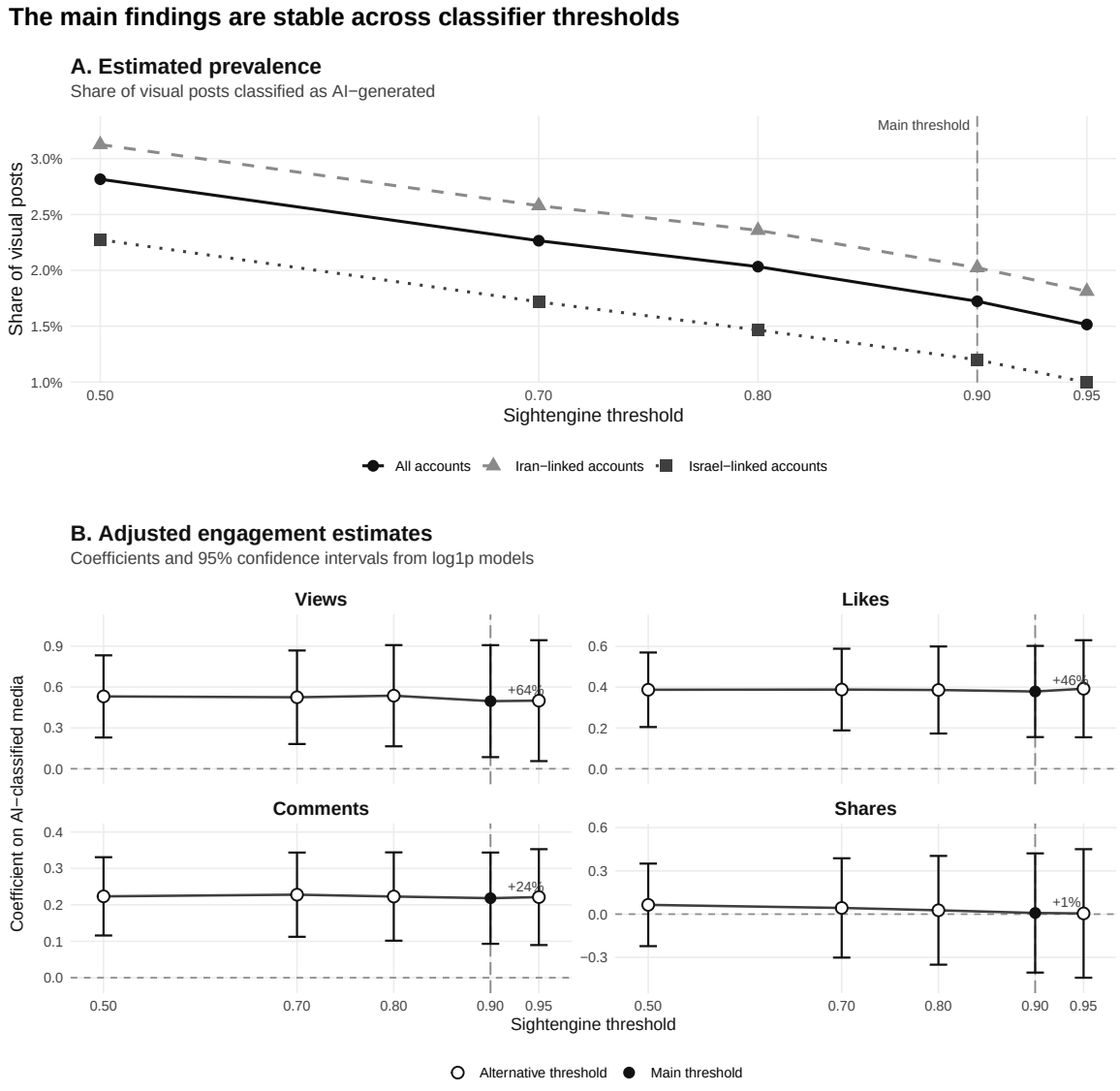


Figure (SI-1) **Sensitivity of the prevalence and engagement results to alternative Sightengine classification thresholds**.

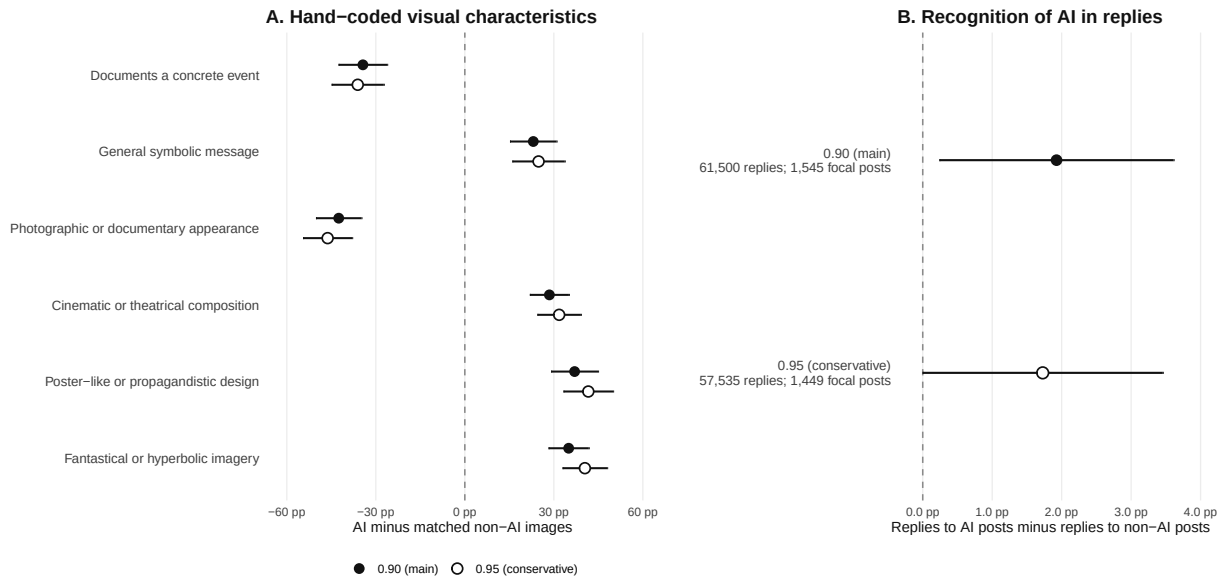


Figure (SI-2) **Robustness of the hand-coded and reply analyses to a conservative Sightengine threshold of 0.95**.

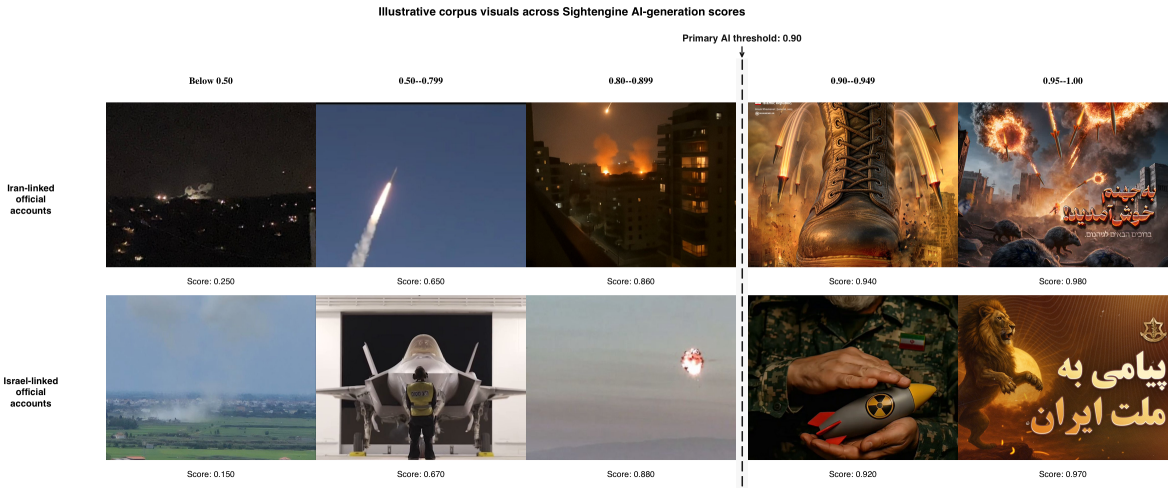


Figure (SI-3) **Illustrative corpus visuals across Sightengine AI-generation scores.**

Table (SI-5) Settings Used to Generate the Flux Validation Benchmark

Feature	Setting
Model	Flux.1 Schnell
Interface	`mflux`
Number of images	1,000
Resolution	1024×1024
Inference steps	4
Quantization	None
Base seed	42
Prompt design	Deterministic prompt generator
Prompt categories	12 weighted content categories
Logged metadata	Prompt, category, seed, model settings, dimensions, generation time
Detection threshold	0.9
Share detected as AI-generated	96.8%

Table (SI-6) Hand-coded comparison of AI-generated and matched non-AI images.

Category	Feature	Non-AI image (%)	AI image (%)	95% CI	BH-adjusted *p*-val
A. Relationship to events	Concrete event	64.1	29.7	[-42.6, -26.2]	<0.001
	Dramatized real event	8.7	13.8	[-1.2, 11.5]	0.115
	Future or hypothetical	2.6	8.7	[2.0, 10.3]	0.006
	General symbolic message	24.6	47.7	[15.2, 30.9]	<0.001
B. Communicative function	Mocks or humiliates the enemy	13.0	25.5	[6.9, 18.1]	<0.001
	Signals resolve, resilience, or victory	11.0	24.0	[6.1, 19.9]	<0.001
	Threatens or intimidates	14.0	18.5	[-1.6, 10.6]	0.151
	Depicts enemy weakness or vulnerability	18.0	28.5	[2.7, 18.3]	0.011
	Reports or updates on events	61.0	34.0	[-34.2, -19.8]	<0.001
C. Visual form	Fantastical or hyperbolic imagery	3.0	38.0	[28.2, 41.8]	<0.001
	Poster-like or propagandistic design	20.0	57.0	[29.1, 44.9]	<0.001
	Cinematic or theatrical composition	3.5	32.0	[21.9, 35.1]	<0.001
	Military hardware or technological systems	14.0	25.5	[4.0, 19.0]	0.004
	Visible destruction, injury, or suffering	21.0	16.0	[-12.5, 2.5]	0.190
	Photographic or documentary appearance	76.0	33.5	[-50.2, -34.8]	<0.001

Note: Entries report the percentage of matched non-AI and AI-generated images assigned to each category. The 95% confidence interval is for the AI-minus-non-AI percentage-point difference from the matched-pair linear probability model. Reported *p*-values use Benjamini–Hochberg adjustments within each family of comparisons. The analyses use the reconciled coding following independent coding by two human annotators.

Table (SI-8) Association between AI-generated media and post engagement

Outcome	Coefficient	Std. error	95% CI, lower	95% CI, upper	Approx. percent difference	p-value	Observations	Accounts
Comments	0.218	0.064	0.093	0.344	24.400	<0.001	212,762.000	196.000
Likes	0.379	0.114	0.155	0.602	46.055	0.001	212,742.000	196.000
Shares	0.008	0.211	-0.405	0.421	0.815	0.969	212,762.000	196.000
Views	0.496	0.210	0.085	0.908	64.290	0.019	212,762.000	196.000

Outcomes are transformed using log(1+x). Models control for follower count and post count and include focal-account and calendar-date fixed effects. Standard errors are clustered by focal account.

Reply dictionary construction and implementation

To identify references to AI or AI tools in the replies to posts sharing AI images and those that dd not, I created two dictionaries. The first dictionary captured general references to AI or AI-generated content in in English, Arabic, Persian, and Hebrew. The second identified mentions of known AI tools and models, such as ChatGPT, Midjourney, DALL-E, Stable Diffusion, Gemini, Grok, Sora, and others, as well as language-specific transliterations. The full dictionaries are reported in Tables SI-10–SI-12.

Before applying the dictionaries, I pre-processed the reply corpus by removing URLs and mentions of account handles, and normalizing white space.I matched the dictionary terms to the replies using case-insensitive matching. Since the replies in my data included various languages, I applied the language-specific AI dictionaries according to the language of each reply, which was available as a metadata variable associated with each reply. I counted standalone variants of “AI” only when they appeared as separate tokens to avoid over-matching on words that simply include the "ai" character sequence. For tool names that are also ordinary words, such as Sora, Flux, Gemini, Runway, and Pika, I required an additional AI, company, or model qualifier to reduce false positives. I coded each reply as whether it was referring to AI or AI tools if it matched either the general AI dictionary or the named tool and model dictionary.

Table (SI-7) Intercoder Agreement for Hand-Coded Visual Characteristics

Category	Agreement (%)
A. Relationship to events	
Concrete event	79.9
Dramatized real event	83.8
Future or hypothetical	87.4
General symbolic message	81.2
B. Communicative function	
Mocks or humiliates the enemy	90.3
Signals resolve, resilience, or victory	67.0
Threatens or intimidates	84.1
Depicts enemy weakness or vulnerability	74.4
Reports or updates on events	83.5
C. Visual form	
Fantastical or hyperbolic imagery	86.1
Poster-like or propagandistic design	78.6
Cinematic or theatrical composition	79.3
Military hardware or technological systems	89.0
Visible destruction, injury, or suffering	92.2
Photographic or documentary appearance	89.3

Note: Percent agreement is calculated from the independent annotations produced by the two coders prior to reconciliation. The substantive analyses reported in the paper use the reconciled coding.

Table (SI-9) AI mentions in replies to posts with and without AI-generated images

Outcome	Difference (percentage points)	Std. error	95% CI, lower	95% CI, upper	p-value	Replies	Focal posts	Focal accounts	Dates
Mentions AI or a named AI tool/model	1.926	0.860	0.240	3.612	0.025	61,494	1,545	80	91

Reply-level linear probability model with focal-account and calendar-date fixed effects. Standard errors are clustered by focal post.

Table (SI-10) Multilingual dictionary for mentions of artificial intelligence

Language	Dictionary entry	Terms and variants matched
English	AI-generated attribution	AI-generated; generated by AI; created by AI; made by, with, or using AI
English	Generative AI	generative AI; gen-AI; gen AI
English	Artificial intelligence	artificial intelligence
English	Deepfake	deepfake; deep fake
English	Standalone AI	AI; A.I.; A I (not embedded within a longer word)
Arabic	AI-generated attribution	مولد/مولدة، منشأ/منشأة، أو مصنوع/مصنوعة بالذكاء الاصطناعي؛ تم إنشاؤه/إنشاؤها بالذكاء الاصطناعي
Arabic	Generative AI	ذكاء اصطناعي توليدي؛ الذكاء الاصطناعي التوليدي
Arabic	Artificial intelligence	ذكاء اصطناعي؛ الذكاء الاصطناعي
Arabic	Arabic AI transliteration	أي آي؛ إي آي؛ إيه آي؛ أيه آي
Arabic	Deepfake	ديب فيك
Arabic	Standalone AI	Latin-script AI variants listed above
Persian	AI-generated attribution	تولید شده با هوش مصنوعی؛ ساخته شده با هوش مصنوعی
Persian	Generative AI	هوش مصنوعی مولد؛ هوش مولد
Persian	Artificial intelligence	هوش مصنوعی
Persian	Persian AI transliteration	ای آی
Persian	Deepfake	دیپ فیک
Persian	Standalone AI	Latin-script AI variants listed above
Hebrew	AI-generated attribution	נוצר/נוצרה על ידי בינה מלאכותית; נוצר/נוצרה ב-AI; מיוצר/מיוצרת ב-AI; תמונה שנוצרה ב-AI
Hebrew	Generative AI	בינה מלאכותית יוצרת; בינה גנרטיבית
Hebrew	Artificial intelligence	בינה מלאכותית
Hebrew	Hebrew AI transliteration	איי איי
Hebrew	Deepfake	דיפ פייק
Hebrew	Standalone AI	Latin-script AI variants listed above

Table (SI-11) Global dictionary of AI tools and models

Tool or model	Terms and variants matched
ChatGPT	ChatGPT; Chat-GPT; optional version suffix 3, 3.5, 4, 4o, or 5
GPT model	GPT-3; GPT-3.5; GPT-4; GPT-4o; GPT-5; GPT Image 1
Midjourney	Midjourney; Mid-Journey
DALL-E	DALL-E; DALL·E; DALL_E; DALL.E; optional version 2 or 3
Stable Diffusion	Stable Diffusion
Adobe Firefly	Adobe Firefly; Firefly AI
FLUX	FLUX.1; FLUX AI; Black Forest Labs
Google Imagen	Google Imagen; Imagen 2; Imagen 3; Imagen 4
Google Gemini	Google Gemini; Gemini AI
Grok	xAI Grok; Grok AI
OpenAI Sora	OpenAI Sora; Sora AI
Runway	RunwayML; Runway AI; Runway Gen-2, Gen-3, or Gen-4
Ideogram	Ideogram.ai; Ideogram AI
Leonardo AI	Leonardo AI
Microsoft image tools	Bing Image Creator; Microsoft Designer; Microsoft Copilot; Copilot AI
Meta AI	Meta AI
Canva AI	Canva AI; Canva Magic Media
Luma AI	Luma AI; Luma Dream Machine
Pika	Pika Labs; Pika AI
Kling AI	Kling AI
Krea AI	Krea AI
Recraft AI	Recraft AI
DreamStudio	DreamStudio
Craiyon	Craiyon
HeyGen	HeyGen
Synthesia	Synthesia
Google Veo	Google Veo; Veo 2; Veo 3

Table (SI-12) Local-language variants of AI tool and model names

Language	Tool or model	Localized name(s)	Required qualifier
Arabic	ChatGPT	شات جي بي تي	-
Arabic	Midjourney	ميدجورني؛ ميدجِرني	-
Arabic	DALL-E	دال إي	-
Arabic	Stable Diffusion	ستيبل ديفيوجن؛ ستابل ديفيوجن	-
Arabic	Google Gemini	جيميني	ذكاء اصطناعي;إي آي;AI
Arabic	Grok	غروك أو جروك	ذكاء اصطناعي;إي آي;AI
Arabic	OpenAI Sora	سورا	ذكاء اصطناعي;إي آي;AI
Persian	ChatGPT	چت جی پی تی	-
Persian	Midjourney	میدجرنی؛ میدجورنی	-
Persian	DALL-E	دال ای	-
Persian	Stable Diffusion	استیبل دیفیوژن	-
Persian	Google Gemini	جمینای	هوش مصنوعی;ای آی;AI
Persian	Grok	گروک	هوش مصنوعی;ای آی;AI
Persian	OpenAI Sora	سورا	هوش مصنوعی;ای آی;AI
Hebrew	ChatGPT	צ'אט ג'יפיטי; צאט גיפיטי	-
Hebrew	Midjourney	מידג'רני	-
Hebrew	DALL-E	דאל אי	-
Hebrew	Stable Diffusion	סטייבל דיפיוז'ן	-
Hebrew	Google Gemini	ג'מיני	בינה מלאכותית;AI
Hebrew	Grok	גרוק	בינה מלאכותית;AI
Hebrew	OpenAI Sora	סורה	בינה מלאכותית;AI