

A.I. Muddies Israel-Hamas War in Unexpected Way

Fakes related to the conflict have been limited and largely unconvincing, but their presence has people doubting real evidence.



By **Tiffany Hsu and Stuart A. Thompson**

Tiffany Hsu and Stuart Thompson monitor the spread of false and misleading information online.

Published Oct. 28, 2023 Updated Oct. 30, 2023

Sign up for the Israel-Hamas War Briefing. The latest news about the conflict. [Get it sent to your inbox.](#)

It was a gruesome image that shot rapidly around the internet: a charred body, described as a deceased child, that was apparently photographed in the opening days of the conflict between Israel and Hamas.

Some observers on social media quickly dismissed it as an “A.I.-generated fake” — created using artificial intelligence tools that can produce photorealistic images with a few clicks.

Several A.I. specialists have since concluded that the technology was probably not involved. By then, however, the doubts about its veracity were already widespread.

Since Hamas’s terror attack on Oct. 7, disinformation watchdogs have feared that fakes created by A.I. tools, including the realistic renderings known as deepfakes, would confuse the public and bolster propaganda efforts.

So far, they have been correct in their prediction that the technology would loom large over the war — but not exactly for the reason they thought.

Disinformation researchers have found relatively few A.I. fakes, and even fewer that are convincing. Yet the mere possibility that A.I. content could be circulating is leading people to dismiss genuine images, video and audio as inauthentic.

On forums and social media platforms like X, Truth Social, Telegram and Reddit, people have accused political figures, media outlets and other users of brazenly trying to manipulate public opinion by creating A.I. content, even when the content is almost certainly genuine.

“Even by the fog of war standards that we are used to, this conflict is particularly messy,” said Hany Farid, a computer science professor at the University of California, Berkeley, and an expert in digital forensics, A.I. and misinformation. “The specter of deepfakes is much, much more significant now — it doesn’t take tens of thousands, it just takes a few, and then you poison the well and everything becomes suspect.”

Artificial intelligence has improved greatly over the past year, allowing nearly anyone to create a persuasive fake by entering text into popular A.I. generators that produce images, video or audio — or by using more sophisticated tools. When a deepfake video of President Volodymyr Zelensky of Ukraine was released in the spring of 2022, it was widely derided as too crude to be real; a similar faked video of President Vladimir V. Putin of Russia was convincing enough for several Russian radio and television networks to air it this June.

Deepfake v. Deepfake

Two deepfakes produced roughly a year apart demonstrate how the technology — now in the hands of everyday users — is advancing.



Note: Video edited to show greater detail.

“What happens when literally everything you see that’s digital could be synthetic?” Bill Marcellino, a senior behavioral scientist and A.I. expert at the RAND Corporation research group, said in a news conference last week. “That certainly sounds like a watershed change in how we trust or don’t trust information.”

Amid highly emotional discussions about Gaza, many happening on social media platforms that have struggled to shield users against graphic and inaccurate content, trust continues to fray. And now, experts say that malicious agents are taking advantage of A.I.’s availability to dismiss authentic content as fake — a concept known as the liar’s dividend.

Their misdirection during the war has been bolstered partly by the presence of some content that was created artificially.

A post on X with 1.8 million views claimed to show soccer fans in a stadium in Madrid holding an enormous Palestinian flag; users noted that the distorted bodies in the image were a telltale sign of A.I. generation. A Hamas-linked account on X shared an image that was meant to show a tent encampment for displaced Israelis but pictured a flag with two blue stars instead of the single star featured on the actual Israeli flag. The post was later removed. Users on Truth Social and a Hamas-linked Telegram channel shared pictures of Prime Minister Benjamin Netanyahu of Israel synthetically rendered to appear covered in blood.

A.I. Fakes Spreading Online

These images, believed to be generated by A.I. tools, have reached large audiences online.

Note: Content obscured by The New York Times.

Far more attention was spent on suspected footage that bore no signs of A.I. tampering, such as video of the director of a bombed hospital in Gaza giving a news conference — called “A.I. generated” by some — which was filmed from different vantage points by multiple sources.

Other examples have been harder to categorize: The Israeli military released a recording of what it described as a wiretapped conversation between two Hamas members, but what some listeners said was spoofed audio (The New York Times, the BBC and CNN reported that they have yet to verify the conversation).

In an attempt to discern truth from A.I., some social media users turned to detection tools, which claim to spot digital manipulation but have proved to be far from reliable. A test by The Times found that image detectors had a spotty track record, sometimes misdiagnosing pictures that were obvious A.I. creations, or labeling real photos as inauthentic.

In the first few days of the war, Mr. Netanyahu shared a series of images on X, claiming they were “horrifying photos of babies murdered and burned” by Hamas. When the conservative commentator Ben Shapiro amplified one of the images on X, he was repeatedly accused of spreading A.I.-generated content.

One post, which garnered more than 21 million views before it was taken down, claimed to have proof that the image of the baby was fake: a screenshot of AI or Not, a detection tool, identifying the image as “generated by AI.” The company later corrected that finding on X, saying that its result was “inconclusive” because the image was compressed and altered to obscure identifying details; the company also said it refined its detector.

“We realized every technology that’s been built has, at one point, been used for evil,” said Anatoly Kvitnitsky, the chief executive of AI or Not, which is based in the San Francisco Bay Area and has six employees. “We came to the conclusion that we are trying to do good, we’re going to keep the service active and do our best to make sure that we are purveyors of the truth. But we did think about that — are we causing more confusion, more chaos?”

AI or Not is working to show users which parts of an image are suspected of being A.I.-generated, Mr. Kvitnitsky said.

Available A.I. detection services could potentially be helpful as part of a larger suite of tools, but are dangerous when treated like the final word on content authenticity, said Henry Ajder, an expert on manipulated and synthetic media.

Deepfake detection tools, he said, “provide a false solution to a much more complex and difficult-to-solve problem.”

Rather than relying on detection services, initiatives like the Coalition for Content Provenance and Authenticity and companies like Google are exploring tactics that would identify the source and history of media files. The solutions are far from perfect — two groups of researchers recently found that existing watermarking technology is easy to remove or evade — but proponents say they could help restore some confidence in the quality of content.

“Proving what’s fake is going to be a pointless endeavor and we’re just going to boil the ocean trying to do it,” said Chester Wisniewski, an executive at the cybersecurity firm Sophos. “It’s never going to work, and we need to just double down on how we can start validating what’s real.”

For now, social media users looking to deceive the public are relying far less on photorealistic A.I. images than on old footage from previous conflicts or disasters, which they falsely portray as the current situation in Gaza, according to Alex Mahadevan, the director of the Poynter media literacy program MediaWise.

“People will believe anything that confirms their beliefs or makes them emotional,” he said. “It doesn’t matter how good it is, or how novel it looks, or anything like that.”

Tiffany Hsu reports on misinformation and disinformation and its origins, movement and consequences. She has been a journalist for more than two decades. [More about Tiffany Hsu](#)

Stuart A. Thompson writes about how false and misleading information spreads online and how it affects people around the world. He focuses on misinformation, disinformation and other misleading content. [More about Stuart A. Thompson](#)

A version of this article appears in print on , Section A, Page 7 of the New York edition with the headline: Artificial Intelligence Muddies Israel-Hamas War, Casting Doubt on Images