

Deep Fake in Swords of Iron: A Battle for Public Opinion

Yael Ram and Liran Antebi | No. 1779 | November 5, 2023

Since the outbreak of the war, the amount of false information and fake content on the internet has reached new heights. Much of this content is created using deep fake applications based on artificial intelligence. The fact that any internet user is now able to create synthetic content has made the issue of public diplomacy especially complex. It is vital, therefore, that there be a deep understanding both of the phenomenon and ways to confront the challenges it poses.

In recent years, artificial intelligence applications and tools for producing deep fake content have multiplied and have become cheap and relatively easy to use. At the same time, the credibility of their output has improved, and they now enable the creation of images, sound files, and videos that are difficult – if not impossible – to distinguish from authentic material. [A document](#) published in September 2023 by the US National Security Agency claimed that the extent and ease with which users can create synthetic media pose a significant challenge to national security and could encourage the dissemination of political, social, military, and economic disinformation. This was just the latest in a long string of warnings issued by security establishments.

One of the most common uses of deep fake technology is an attack on individuals, for example, by creating fake pornographic content or [replicating a person's voice](#) for the purpose of theft or extortion. At the same time, deep fake technology has the potential to influence political processes, such as democratic elections. A [deep fake video](#) of Florida Governor Ron DeSantis announcing that he was dropping out of the 2024 presidential race was circulated recently. Another example of how American society and its economy were impacted by deep fake material occurred in May 2023, when [synthetic images of the Pentagon in flames](#) were spread on the internet, leading to a sharp drop in the stock market. The influence of fakes is not solely a result of their quality or the ease with which they can be created but is based mainly on the ability to disseminate them widely and capitalize on their viral potential, thanks to social media platforms.

Until now, the war between Russia and Ukraine was the most prominent test case for the use of deep fake technology to influence public opinion against the backdrop of an international conflict. Both Russian President [Vladimir Putin](#) and Ukrainian President

[Volodymyr Zelenskyy](#) were the subjects of deep fake videos of their making announcements or addressing their respective nations. These videos, however, did not significantly impact the course of the war.

One of the main problems created by the deep fake phenomenon is that people tend to automatically believe content that echoes their position without checking the credibility of the information. This leads to increased extremism, with each side becoming more deeply entrenched in its position. Therefore, when details emerged of the massacre perpetrated by Hamas in southern Israel, many across the world did not believe the reports. In response, Israel's public diplomacy campaign deviated from its traditional reluctance to share graphic images and made a conscious decision [to show pictures and footage](#) of the Hamas atrocities. Yet even after the factual information was presented, many still clung to fake content.

The inability to determine whether a certain content is genuine or fake leads to the [common argument](#) that we are in the throes of a crisis in which the value of truth as the shared foundation of modern society is eroded. Deep fake technology and social media platforms are just another layer in the age of post-truth and fake news, in which more space is given to feelings, faith, and personal opinion than to objective facts. The first to be hurt by the trend are democratic societies that sanctify freedom of speech and pluralism.

Deep Fake in the Anti-Israel Incitement during the Swords of Iran War

According to [some researchers](#), the Israel-Hamas war is unprecedented in terms of the dissemination of deep fake content. As soon as the war broke out, the internet was flooded with various kinds of fake content, opening a new front for Israeli public diplomacy. In one prominent case among many, pro-Palestinian users posted [an image of a baby](#) in the rubble of destroyed buildings in Gaza. The photograph, which was shared widely, appeared on placards at anti-Israel demonstrations across the world and even on the cover of the French newspaper *Liberation*. It quickly became evident, however, that the image was created by artificial intelligence technology and was already in use in February 2023. [According to Wendy McMahon](#), CEO of CBS News, only 10 percent of the more than 1000 videos sent to the bureau's news desk relating to the Israel-Hamas war were genuine. According to McMahon, news organizations across the world are forced to deal with a massive amount of false and fake information, and accordingly, must create standards that ensure content credibility. In addition, [according to TIME magazine](#), at least 14 false claims about the war were viewed around 22 million times on social media platforms X (formerly Twitter), TikTok, and Instagram in the first three days of the war alone.

In addition to efforts to spread propaganda and false information, terrorist organizations are using deep fake technology to sow panic among the public. One Israeli woman, for example, [reported](#) that her WhatsApp account had been hacked, and that in one of her groups, the hackers sent a voice message in which she could be heard yelling "Kidnappers," which generated much panic in the group. Israeli company Clarity examined the message and determined that it was, in all likelihood, based on deep fake technology.

In an effort to deal with the phenomenon, several companies have begun to develop tools capable of identifying deep fake content, for example, “AI or Not,” which can determine whether a certain photograph is authentic or a product of artificial intelligence. While these tools cannot authenticate content with total certainty, their verdict could have far-reaching implications for Israel’s public diplomacy efforts. For example, when the official X account of Prime Minister Benjamin Netanyahu posted a photograph of an infant burned to death by Hamas, it made huge waves in the international media. When the photograph was checked, however, the software claimed that [it had been created by AI](#). As a result, there was a deluge of claims that Israel was spreading disinformation. Although AI or Not subsequently issued a clarification, stating that the image was genuine, and X removed the charge that its authenticity was disputed, for many people the case harmed Israel’s public diplomacy efforts and provided more evidence for people who deny that the massacre even took place. This case highlights the speed with which the battle for public opinion on the internet is won or lost – and the damage that can be caused by presenting false information – even if it is removed or updated at a later stage. In an effort to address this issue, some ten days after the outbreak of the war, Israel’s National Cyber Directorate [issued a call](#) to set up working groups to find ways for the country to counter [fake news](#) online.

The current war has also revealed the massive [harm caused by the changes](#) introduced by Elon Musk, the CEO of X, which have turned the popular platform into a fertile breeding ground for false information. The company did away with verification marks and instead promoted posts by users who pay a monthly subscription fee – irrespective of the content they share. The majority of the employees who were recently fired by the company were responsible for removing disinformation from the platform. After the platform was found to have grossly violated the European Digital Services Act, which bars spreading illegal content and disinformation, the European Commission [launched an investigation](#) into the company.

Technology’s influence, however, is not entirely negative, and some countries – including Israel – have managed to use it in the service of public diplomacy. For example, many [videos and testimonies from the Hamas massacre](#) have been translated into dozens of languages using deep fake technology in the hope of reaching as many users across the world as possible. In addition, virtual reality applications that are also based on deep fake technology have been found [to help with mental health issues](#). The use of these tools, therefore, could help treat people suffering from PTSD as a result of their experiences during the war. Nonetheless, in general terms, the current state of the technology, coupled with the unrestricted dissemination of synthetic content on social media platforms, is a significant challenge.

What Can Be Done?

Despite the need to continue developing tools to locate and identify synthetic online content, the existing tools are limited, do not provide clear-cut answers, and are developed at a pace that cannot keep up with the growing sophistication of the technology used to create fakes. In addition, the social media platforms offer only a

very limited response to people reporting fake information, since the outcome of the report depends largely on the number of people reporting the same content rather than the content itself. Since these platforms are inundated with bots and the pro-Israel camp is vastly outnumbered by the pro-Palestinians, these mechanisms are not especially effective from an Israeli perspective.

The government should take measures similar to those taken against X in light of the quantity of false information about the war on the platform. An [in-depth study](#) on the potential influence of deep fake on national security stressed the importance of cooperation or, alternatively, legal restrictions on social media platforms, which are the main disseminators of false information.

Other solutions are related to social and personal responsibility. Democratic countries should promote programs encouraging digital literacy, which will help users distinguish between fake and genuine content, encourage them to cross-check information and identify reliable sources, and engender awareness of the responsibilities involved in sharing and consuming content on the internet. And yet, it is vitally important that official Israeli entities, led by the IDF Spokesperson's Office, continue to publish evidence and factual information as much as possible and ensure that it reaches target audiences across the world so that the truth can overcome lies and fakes.

Editors of the series: Anat Kurtz, Eldad Shavit and Judith Rosen