

TECH GIANTS

Meta failed to flag AI video during 2025 Israel-Iran war, Oversight Board says

The board says Meta should have labeled a viral AI-generated video of alleged damage in Haifa during the 2025 Israel-Iran war, urging platforms to do more to help users spot synthetic media in conflicts.

By [ANANYA BHATTACHARYA](#) 

10 MARCH 2026

Meta allowed an artificial intelligence-generated video to go unchecked on Facebook during the 12-day war between Israel and Iran in June 2025, according to a new ruling published by the Oversight Board today.

The fabricated video, posted by a user in the Philippines posing as a news source, showed extensive damage to buildings in Israel's third-largest city, Haifa. Six users had reported it, and a similar video had earlier been debunked by credible news sites on TikTok. But Meta took no action despite the video being flagged, according to the Oversight Board, often described as the "supreme court" for Meta's content moderation decisions. The board overturned Meta's decision to leave the video online without a "High Risk AI" label.

While the content did not warrant removal — there was no imminent threat of physical harm or violence — its inauthenticity should have been clearly flagged, the board said.

"As the quantity and quality of AI-generated content increase, its impact on people and societies will be profound," the board noted in its case decision.



As the quantity and quality of AI-generated content increase, its impact on people and societies will be profound."

The board's finding comes at a time when such AI-generated videos are rampant amid the ongoing U.S.-Israel war against Iran. Both Israel and Iran state actors are generating deepfakes at a more rapid pace than in times of peace, according to NewsGuard, a platform that rates the reliability of news and information online. A

BBC analysis found that smaller social media creators are using AI tools to create and monetize fake war imagery.

Meta relies on metadata to determine which content is AI-generated, the company admitted to the board during the investigation. This largely applies only to static images, however, and even then, users can strip metadata from images before uploading to social media. Mostly, platforms rely on self-disclosure — tools to detect and flag manipulated audio and video are still underdeveloped.

The Oversight Board emphasized Meta has to “do more” to help users identify AI-generated content in armed conflicts, including providing details about the origin of media, investing in stronger detection tools, and developing better methods for labeling — and to do it all in a timely way.

In an interview with *Rest of World* last month, Oversight Board member Sudhir Krishnaswamy said the board’s mandate will possibly “be less individual case-based and more structured” to make broad-based reforms and recommendations as AI proliferates.

Experts warn that platforms will have to deal with unprecedented levels of misinformation in the age of AI. “While other conflicts, including those in Armenia, Ukraine, and Gaza, have seen copious recycled images, fake live streams, and excerpted computer game takes, AI-generated content related to the Iran-Israel conflict has taken disinformation to an industrial level,” Mahsa Alimardani and Sam Gregory, human rights researchers at international nonprofit Witness, wrote in June 2025. 

Ananya Bhattacharya  is a reporter for Rest of World covering South Asia's tech scene. She is based in Mumbai, India.