BROOKINGS

Posting murder on social media platforms

Darrell M. West

October 12, 2023

→ As the ongoing conflict between Israel and Hamas claims the lives of thousands of Palestinian and Israeli civilians, both genuine and fraudulent footage of violent attacks have been posted on social media platforms.

→ The practice of posting videos of murder and violence online is not new, and can serve sometimes-conflicting purposes of informing or radicalizing members of the public.

→ Several social media companies have downgraded their "trust and safety" teams, which has exacerbated the difficulty of both moderating violent content and combatting misinformation.

'X' logo is seen on the top of the headquarters of the messaging platform X, formerly known as Twitter, in downtown San Francisco, California, U.S., July 30, 2023. The European Commission has launched an investigation into Elon Musk's X social media platform to see whether it complies with new EU tech rules on illegal and harmful content following the spread of disinformation on its platform after Hamas' attack on Israel. REUTERS/Carlos Barria/ File Photo

TechTank
Improving Technology Policy

It was shocking to learn from CNN News ↗ on Tuesday that an Israeli grandmother's murder by Hamas militants was photographed by her executioners and posted to her Facebook account. According to her grandson, who was interviewed by anchor Jake Tapper, the shocking video appeared in his News Feed and likely that of her friends and family members because it was uploaded from her own account. Facebook algorithms are used to spotting violent content that appears on sites run by extremists, but not ordinary people with no history of posting violent videos.

This tragedy reveals the depth of the depravity taking place in Israel this week but also the challenges facing social media platforms during times of war. Tech companies are going to have to figure out how to identify violent content and remove it before it reaches thousands or millions of people. If it takes Facebook, Twitter, YouTube, Twitch, Parler, and Telegram hours to take down such material, as sometimes has been the case, terror organizations are virtually guaranteed large numbers of viewers of their barbaric and inhumane actions.

Extremist organizations long have broadcast gruesome videos with hopes of shocking the world and recruiting followers to their cause. For example, ISIS often taped beheadings ↗ and put them on the internet for recruitment purposes ↗. Although nauseating to typical viewers, such actions were fruitful in convincing thousands of people to travel from America, the United Kingdom, France, and other places around the world to join that cause ↗.

Some mass shooters also have live-streamed their killings. They apparently do this to document their rampages and serve as a role model to others who want to murder other people. The Christchurch ↗ shooter in New Zealand did this as did the killer in a Buffalo grocery store shooting ↗. In the latter case, the gunman live-streamed his rampage on Twitch. Although only a handful of people were watching at the time and the stream was removed within minutes, one viewer kept a copy and posted it on other websites ↗ and message boards. Before long, millions of people had viewed the tape.

It therefore comes as little surprise that Hamas terrorists recorded their murders and posted them for many to see. They wanted to document their onslaught and convince others to join their movement. Ambushing people in the early morning hours in their own homes, at music festivals, or on the street is not sufficient for terrorists during a digital era. They also want to achieve broader strategic goals, and thus need to document their actions and push them out to large numbers of viewers. In that way, they can terrorize innocent civilians around the world, while also recruiting followers with similar hatred for Israel.

Social media firms have algorithms that are good at spotting violence but spotty in assessing motives. How can they distinguish graphic scenes being broadcast from historical or other cases designed to inform people about atrocities versus contemporary scenes of decapitations and carnage designed to radicalize people and attract allies? There is legitimate interest in informing people about some kinds of historic and contemporary violence and some of that public education does involve scenes that are horrific or brutal in nature. We witnessed this in the war on Ukraine, for example, where videos raised public awareness about Russia's premeditated brutality. The old adage that people should "never forget" atrocities means that some graphic videos may need to be online and available for people to remember.

Yet algorithms are terrible at measuring motives and distinguishing videos meant legitimately to inform versus those designed to radicalize and inflame public passions. State actors as well as terrorist organizations are posting videos in hopes of demonizing their opposition and influencing public perceptions. On Twitter, Israeli Prime Minister Benjamin Netanyahu has shared compilation videos showing Israeli airstrikes on Gaza, one of which has attracted nearly 85 million ⌐ views.

It is one of the challenges of our AI-driven world that technology is springing ahead and being used for nefarious purposes. Videos are becoming one of the primary sources of information for many people who prefer to learn by watching as opposed to reading. They therefore are being deployed by criminals and terrorists to torment, manipulate, and scare people.

The recent trends with several social media companies of downgrading ⌐ their content moderation efforts and getting rid of their human trust and safety divisions are the opposite of what is needed in a world that can be violent, extreme, and dangerous. We need human guardrails to protect people from violent content, extreme actions, and behavior that is criminal in nature. Leading firms need to rededicate their resources to these important priorities and reinvigorate their content moderation practices. Unless we get a handle on social media content, it will be difficult for countries to protect themselves and their citizens from criminals and terrorists and insulate viewers from the radicalizing power of extreme actions broadcast online.

In addition, content moderation is vital to fighting disinformation. There have already been rampant false narratives ⌐ about the Israel-Hamas war on X (formerly Twitter) and Telegram, among other platforms. Bad actors are using outdated footage, fake videos, and made-up stories to inflame people and win the information war, and even experienced researchers are struggling to identify and corroborate accurate information. Without serious efforts to mitigate inaccurate material, social media content could distort public perceptions about a serious, ongoing conflict.

AUTHORS

Darrell M. West Senior Fellow - Center for Technology Innovation, **Douglas Dillon Chair in Governmental Studies**

Acknowledgements and disclosures