

# The Dangers of AI: Far-Right Extremism, Holocaust Distortion, and the Growing Threat to the Jewish Community

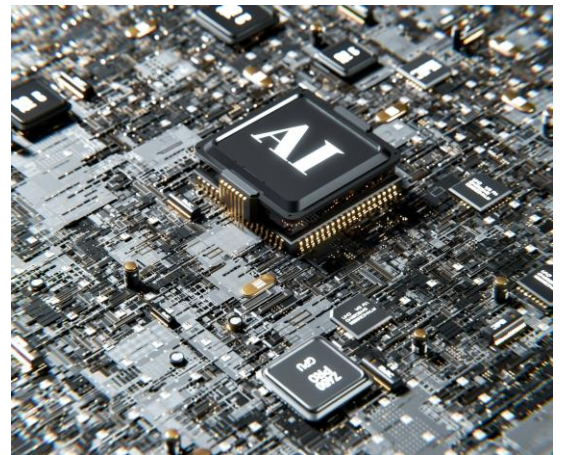
4 Oct 2024  
Rifana Khanum



The rise of AI-generated videos glorifying Nazi ideology and promoting Holocaust distortion has become a troubling trend in the digital space, fuelling a rise in antisemitism. Far-right groups are leveraging artificial intelligence to generate dangerous pro-Nazi content by reanimating Adolf Hitler as a “misunderstood figure”. This hateful and problematic content has spread to tens of millions on social media, fuelling the already growing trend of pro-Hitler material online. This surge allows a flood of misleading narratives to spread, focusing particularly on Holocaust distortion. This wrapped retelling does nothing but stoke the false claims about the historical atrocities committed. Experts propose that this sudden trend could partly be a result of the rising antisemitism following the October 7 Hamas attacks and the conflict that followed. Despite content moderation policies on platforms like X and TikTok, AI-driven hate speech continues to thrive, highlighting the need for stronger measures to control this type of hateful content. UNESCO's latest report warns that without urgent action, AI could be used to distort the Holocaust, fuelling a dangerous rise in antisemitism worldwide.

## The Rise of AI-Generated Pro-Nazi Content

Far-right extremist groups on platforms such as X, TikTok, Instagram and YouTube are using AI to create realistic videos, memes, and other content that glorifies Adolf Hitler, promoting Nazi ideology and distorting historical facts. Many young people frequent these platforms, and exposure to such AI-generated content can radicalise them, contributing to a rise in antisemitism. Furthermore, the accessibility of AI tools makes it easy for anyone to create hateful content; for instance, individuals can readily extract clips from Adolf Hitler's speeches on YouTube and reanimate them. This issue is further compounded by built-in translation features available on many platforms, which broaden the reach of harmful messages to a wider audience.



AI has opened the door to distorted realities and misleading narratives which can skew public perception, undermine historical truth, and foster an environment where Holocaust denial and antisemitism thrive. This is extremely dangerous as they discredit the experiences of Holocaust survivors and victims, erasing the memory of one of history's greatest atrocities and empowering extremist ideologies.

For the Jewish community, the impact is profound. Holocaust distortion and denial invalidate the suffering of survivors, creating feelings of alienation and insecurity. As antisemitism continues to grow both online and offline, Jewish communities face heightened anxiety and vulnerability, increasing their sense of being targets of hate.

## The Failure of Content Moderation Online

While social media platforms like X and TikTok in the UK are slowly becoming more diligent in removing hateful content, including AI-generated hate, they must enhance their efforts in other countries. It is important to acknowledge that tech companies need to enhance their content moderation to effectively remove AI-generated hate speech and misinformation, helping to create a safer online space for everyone.

A report by GTTO partner, the 'NEVER AGAIN' Association highlights the growing concerns regarding X's failure to effectively moderate and remove antisemitic and pro-Nazi content, even after such material has been flagged. The report cites hundreds of shocking examples of extreme hate speech submitted by 'NEVER AGAIN' to X over the past year, many of which have been ignored or left unaddressed. The [findings](#) reveal a lengthy list of 'written posts, images, and videos that incite hatred against various minorities', including the Jewish community. These materials include dehumanising rhetoric, glorification of the Holocaust, and calls for violence. Since August 2023, the 'NEVER AGAIN' Association has [reported](#) 343 cases of extreme hate speech targeting a Polish audience, with only 10 percent removal rate, leaving 90 percent still visible.

This ineffective moderation has serious implications for vulnerable communities, particularly the Jewish community. The spread of antisemitic content online fosters fear and isolation, undermining the ability of Jewish communities to participate in public discourse and feel secure in both digital and physical spaces.

In the UK context, GTTO partner MDI has noted a large surge and increase in antisemitism on X, between September 1, 2023, to March 1, 2024, based on monitoring conducted using the Meltwater platform. The data shows that antisemitic keywords like "I hate Jews" and "Hitler was right" were mentioned 531,000 times, averaging 2.9k daily mentions. Compared to the same period last year, antisemitic mentions rose from 101,000 to 531,000 showing a clear distinction that since October 7th antisemitism has been on the rise.

Similarly, CST (Community Security Trust) a charity that protects British Jews from antisemitism concluded that 630 online antisemitic incidents were recorded, the highest number ever reported for a January-to-June period. This marks a 153% increase compared to the first six months of 2023. They report that the rise in online antisemitism is partly attributed to 'the war in Gaza and the subsequent proliferation of dialogue, debate, information and disinformation on social media platforms, which sometimes slip into anti-Jewish hate'.

In an attempt to respond to these growing concerns, TikTok and X have introduced policies to improve content moderation. Social media platform, X, has introduced a rule that prohibits users from sharing synthetic, manipulated, or out-of-context media that could mislead or harm people. The platform may also label posts with misleading content to help users understand their authenticity and provide additional context. These steps can act as stepping stones towards the elimination and appropriate reaction and response to instances of hate speech particularly those targeting marginalised communities. However, a significant gap persists between the creation of laws and standards and their effective enforcement and implementation, leaving much work to be done in bridging this divide.

Despite the creation of these measures, there is concern for potential loopholes allowing far-right extremist content to get through. For example, posts that use coded language or subtle hints can often avoid detection from moderators. Additionally, what counts as “misleading” can be subjective, allowing harmful content to remain online. By addressing these issues, social media platforms such as X should do better to protect vulnerable communities and combat hate online.

## **The Efficacy of Existing Laws in Combatting the Spread of Illegal Content Online**

When confronted with hate online, we are forced to confront a critical question: the role of the law. More importantly, the implementation and effectiveness of the legal frameworks designed to prevent the spread of illegal content.

In the context of the UK, the Online Safety Act (OSA) seeks to place ‘various responsibilities on online platforms and services that require them to implement systems and processes to ensure the safety of internet users from harms caused by illegal content, such as racially or religiously aggregated public order offences or inciting violence’. In terms of AI and algorithms, the OSA is expected to monitor and regulate AI content alongside greater transparency and accountability. Overall, it is expected to provide a more stringent regulatory oversight of the use of AI in online spaces.

The European counterpart to the Online Safety Act (OSA) is the Digital Services Act (DSA). Adopted in 2022, and still in the process of being implemented, the DSA ‘designates platforms (VLOPs – Very Large Online Platforms) [such as Meta and X] by their size and requires that the platforms implement measures such as risk assessments and takes measures to mitigate these risks’. Nevertheless, despite the efforts to create measurements and laws in place to mitigate the risks of AI and the spread of illegal content, the question of effective implementation remains.

Can these laws and regulatory frameworks effectively tackle the growing risks posed by AI? Can it curtail the growing spread of illegal content? Can it combat the proliferation of hate speech that continues to circulate online?

## **Looking Ahead: Advancing AI and Content Moderation in the Fight Against Online Antisemitism**

Social media platforms like X and TikTok have made some progress in addressing AI-generated hate speech and misinformation. However, a serious concern remains. The alarming rise of antisemitism and glorification of extremist ideologies can lead to devastating real-world consequences, highlighting the need for action beyond mere content removal. The true effectiveness of laws and frameworks designed to combat illegal content and hate directed at marginalised groups remains to be seen in practice.

It is crucial for social media companies to strengthen their moderation efforts. Society must unite in its responsibility to create a safe online space, only through collective action can we ensure this succeeds.

R. Khanum: ‘The Dangers of AI: Far-Right Extremism, Holocaust Distortion, and the Growing Threat to the Jewish Community’. GetTheTrollsOut.org, 4.10.2024.

<https://getthetrollsout.org/articles/the-dangers-of-ai-far-right-extremism-holocaust-distortion-and-the-growing-threat-to-the-jewish-community>