JAMES STEVENSON

TECHNOLOGY EVOLVES THE TACTICS: PREPARING FOR THE RISE OF TERRORIST AI HARMS

Terrorist groups, like the societies they emerge from, adapt to new technologies. As AI capabilities evolve, so too do the tactics of extremist actors. While the full effects may take years to observe, as the technologies continue to develop, we are starting to see them directly alter extremism tradecraft.

From AI-driven denial-of-service attacks and adaptive malware, to drone swarms, and autonomous vehicles - the malicious use and wider adoption of machine learning and artificial intelligence, by terrorist groups, has been widely speculated in recently months. Earlier this year, the UK Government published "The Terrorism Acts in 2023 report of the Independent Reviewer of Terrorism Legislation" a report, which among other areas, considers seven categories of potential terrorism harm that may result from use of Generative AI.

This article builds on these conclusions, providing real world examples of terrorist uses of AI.

PROPAGANDA AND PRODUCTIVITY INNOVATION

Synthetic propaganda is nothing new, deepfakes (audio and visual) have historically been used to support terrorism. With the increasing availability of generative AI tools, we're seeing its weaponisation by terrorist organisations for propaganda generation. ISIS' media division, News Harvest, have adopted generative AI to produce human-like news presenters resembling shows such as CNN and Al Jazeera - generating video, audio, and text content tailored for propaganda in various languages.

CHATBOT RADICALISATION

With the boom of generative AI, people have been flocking to these AI chatbots to converse and share ideas. These chatbots provide a real potential threat of steering users towards terrorist ideals and views - whether in closed-loop private one-to-one chats or in public multi-user settings.

We've recently seen this phenomenon with the xAI chatbot, Grok. On 8 July, 2025, Grok began posting content praising Adolf Hitler, using antisemitic stereotypes, and even referred to itself as "MechaHitler". Similarly, before these concerns on X, Gab's AI chatbot had been observed generating Holocaust denial and other conspiracy theorist content.

ATTACK FACILITATION AND INNOVATION

Generative AI may be used to obtain, or refine practical instructions, training, or support for acts of terror. Such tactics are increasingly common in pro-IS propaganda networks. For example in March 2025, a pro-IS account released an AI-generated video featuring a digital avatar providing instructions for making a bomb using common household items.

Generative AI can also be seen to provide innovative recommendations for new types of attacks. On 6 December 2022, a frequent user of a Rocket.Chat server run by Islamic State published a post claiming to have used the free version of ChatGPT to ask how best to "establish the Caliphate." According to the user, the AI responded with a series of operational steps, which he claimed showed it was "smarter than most activists." He also asserted that the response was original and included the full text of the alleged reply.

MODERATION EVASION

Recently the International Center For Counter Terrorism released findings observed from far right clusters on mainstream chat-based platforms (Reddit, Telegram, X, and 4chan) where users shared tactics for generating harmful yet lawful AI content purposefully designed to not be flagged by content moderation systems. Evasion methods included: platform switching (users temporarily migrated from mainstream to fringe platforms to create harmful content, then returned to mainstream sites to spread it more widely), self-censorship (users deliberately avoid sharing technical detail to reduce the risk of content removal), delegation (users ask others to perform tasks that could get them flagged), banishment sharing (users discuss their past bans to inform others about platform moderation limits), censorship evasion education (banned users share strategies

> **Chatbots provide a real potential threat of steering users towards terrorist ideals and views - whether in closed-loop private one-to-one chats or in public multi-user settings.**

to continue their activities elsewhere, effectively teaching others how to bypass restrictions), and finally tool and site recommendations (individuals suggest alternative platforms or tools with fewer safety or moderation measures).

SOCIAL DEGRADATION

Finally, this category covers AI driven content that spreads distrust, conspiracy narratives or social division without directly urging terrorism.

We see this behaviour in the events less than three hours after the Southport murders. Where on 29 July 2024, the X account Europe Invasion shared an AI-generated image showing knife-wielding men in traditional Muslim dress outside Parliament, next to a crying child in a Union Jack T-shirt. The post amassed around 900,000 views and introduced the slogan "Protect our children". Similarly, an anti-immigration Facebook group created an AI-made image inviting people to a rally in Middlesbrough depicting a large crowd at the cenotaph. Other tools, like Suno, were used to produce xenophobic songs such as "Southport Saga," featuring AI-generated vocals stating things like "hunt them down somehow".

Meanwhile Tech Against Terrorism traced a brand-new TikTok account that began posting only after the Southport attack. Despite having no prior history, its AI-generated protest posters quickly reached 57,000 views, a spread the group links to coordinated bot amplification networks.

CONCLUSION

The adoption of AI by terrorist groups is here, it's not going away, and will only continue to amplify terrorist tradecraft over time. However, our defensive toolkit is evolving also: platforms are now beginning to auto-watermark AI-generated translations and graphics (such as Meta's "Imagined with AI" labels and embedded metadata). AI red-team simulations stress-test counter-terror guardrails against bomb-making or cyber-threat scenarios; shared-hash moderation coalitions, powered by GIFCT and Hasher Matcher Actioner (HMA), now synchronise detection of mutated extremist media across platforms; and in education, pre-bunking curricula like the "Bad News" game paired with browser AI fact-checkers have reported cuts of susceptibility to disinformation. These defences, are already being piloted, tested, and put to practice - and they're only the beginning.

James Stevenson is a security researcher with a decade of experience in the computer security and research industry. He is currently a PhD candidate at the University of Bristol working at the intersection of machine learning, computer science, and social science. His research focuses on detecting and predicting extremist content online, and understanding how extremist groups use AI to enact harms.