Threats of Artificial Intelligence

12 MAY 2026

Sienna Fox White



Threats of Artifical Intelligence

1. Introduction

Emerging technology in public and foreign policy has consistently remained a prudent topic for governments worldwide. In recent years, the evolution and rise in usage of Artificial Intelligence (AI) has prompted a more pressing debate, particularly regarding the future of domestic and global threat management. Nobel Peace Prize Winner Henry Kissinger even previously argued that the Artificial Intelligence revolution will be just as significant as the Industrial Revolution itself (Kissinger, 2018), highlighting its potential to signify a new transformative technological era. Additional uncertainties have emerged regarding AI's integration to many facets of

critical infrastructure, such as hospitals, and into the UK's counter-terrorism strategy itself. . The UK government has frequently acknowledged the significance of AI in British 'critical infrastructure dependencies' (Gov.uk, 2022), highlighting the urgency of assessing appropriate ongoing action for the safeguarding of civilians.

Its myriad potential is reshaping not only how governments and the public operate on a daily basis, but how terrorist actors are able to recruit, communicate, and plan attacks. This is of particular concern in Europe, where the number of terrorist attacks doubled to 67 in 2025 (Global Terrorism Index, 2025). No longer a member of the European Union, the UK's security strategy has become less collaborative and increasingly isolated. In recent years, the counterterrorism landscape has undergone a transition from national and international counterinsurgency operations to stronger focus on online threats, particularly from social media (Davis, 2021). As a result, optimised utilisation of AI and collective international understanding of Artificial Intelligence must form a crucial tenet of British anti-terrorism and security strategy in the coming years, both from a domestic and foreign policy standpoint. British counterterrorism policy has been previously based on human-led processes and a post-9/11 understanding of the nature of terrorism. This is a significant concern given that Artificial Intelligence is an increasingly urgent dilemma for national security. (Hall, 2025).

A successful policy response on this matter will be crucial to not only the country's safety from evolving terrorist threats, but also to maintaining and improving the UK's credible positioning as a security actor in the global arena.

2. Assessing the scope of Artificial Intelligence as a threat to security

Artificial Intelligence proves a unique concern for security policy largely due to its exacerbation and streamlining of existent threats born from global internet usage, such as the potential for radicalisation by terrorist groups and propaganda material. The Global Terrorism Index highlights the increasing danger of algorithmic radicalisation for lone actor terrorists (2025) When powered by AI, such algorithms have an increased ability to recommend exponentially more radical content. Artificial intelligence has the dangerous potential to amplify threats born of social media specifically to a critical level; Jonathan Hall's annual 2023 report on Terrorism noted that there were multiple terrorism-related risks associated with generative AI including 'attack planning, propaganda, evading online moderation, deepfake impersonation and identity guessing' (2023).

The uniqueness of the risks can be largely identified by examining its sheer scale, convenience, and accessibility. Of particular concern is the 'democratisation of AI technologies' in which these advanced language models are available on a mass scale with limited restrictions (Fishman, 2). Chatbots such as ChatGPT, Microsoft Copilot, and Claude are available to anyone with an internet connection and email address, marking a shift in the wide availability of such powerful generative technology. From a

broader perspective the increasing possession of internet-enabled devices and literacy levels in the Middle East and Africa in recent years already represents a globally expanding pool of potential recruits for terrorist organisations looking to exploit emerging technologies (Davis, 2021). Using these tools and their implementation into various social media platforms, AI-driven algorithms can 'analyse user behaviour to create personalised content that is more likely to engage and further radicalise individuals' (Sayed et al, 2023).

The real-life consequences of radicalization from online environments in recent decades have been well documented in bloodshed. White supremacist terrorists such as Brenton Tarrant directly credited online communities for his radicalization journey, even livestreaming his attack on social media (Wilson, Dziwulski, 2025). . However, the unparalleled risk that AI contributes can be extrapolated here by reference to the several crucial 'pillars' to radicalisation, identified by Kruglanski (2021), those being: Needs, Narratives and Networks. Within this framework, the 'Needs' element refers to human psychological needs – most notably, a sense of purpose and feelings of belonging. Kruglanski argues that individuals experiencing a lack of purpose or belonging may be more susceptible to extremist ideologies and radicalisation tactics. The 'Narratives' pillar describes the legitimisation of violent action by extremist groups, based on an account that frames the action as key to a sense of belonging and purpose for the individual being radicalised. 'Networks' refer to the personal connections crucial to the radicalisation process, that serve to reinforce and endorse the extremist message portrayed.

This framework proves illuminative for the functions of Artificial Intelligence in increasing the effectiveness of terrorists' recruitment processes, as various AI tools have the power to heighten the effectiveness of these tools on vulnerable individuals. In many cases, once a user views and engages with extremist content, AI-driven algorithms will promote further content of the same variety, as a means of continuing engagement on the same platform. As such, direct recruitment drives by terrorist groups using AI may be a secondary danger to users accessing and generating such content themselves.

An additional threat is the recent and dangerous rise of lone-actor terrorism, in which an individual is not affiliated with any major religious or political groups . Actors of this category nature prove difficult to track by nature as they have limited association with other individuals or groups known to national security agencies (Kenyon, Baker-Beall, Binder, 2021). These threats can be dangerously exacerbated by weaknesses in Artificial Intelligence safeguards, such as users 'jailbreaking AI chatbots using specific prompts to receive detailed instructions for dangerous or violent acts. One user recorded his experience asking a Discord chatbot to tell him a bedtime story in the style of his grandmother depicting a day in her life working in a napalm factory, to which the chatbot returned detailed

instructions for the creation of Napalm.(Franchesci-Bicciari, 2023). Once the loophole became popularised, owners of the site enacted changes to prevent this from happening. However,users exploited the weakness again by simply requesting the chatbot roleplay as an alternative member of the family. Such weaknesses demonstrate the fallibility of interactive AI language models, particularly in comparison with regular internet searches where certain phrases and instructions are wholly blocked, and especially the difficulty of policymaking around system updates – making it a fluid danger. This is made particularly relevant by the fact that pernicious conversations of this context may not be flagged if new and existing loopholes are exploited. As a result, aggrieved individuals may be able to access dangerous information outside of even traditional online knowledge-sharing methods.

As is the case for many such threats based online, the reach of Artificial Intelligence extends far beyond any siloed domestic policy; terrorist actors have the potential for cross-continent strategy and acting using AI chatbots and regular features of the internet to communicate and coordinate terrorist attacks. Radical terrorist groups such as Islamic State already utilise over 100 social media platforms for global recruitment and planning (Davis, 2021).

These risks are also not traditional notions of terrorist actors, but exist in a broader global security spectrum, with increasingly significant consequences. States including Russia and China have been known to utilise Artificial Intelligence in spreading disinformation and propaganda. A 2024 report investigating Russian influence in the 2024 US Presidential election by Microsoft observed that Russia, Iran and China have 'leveraged...generative AI to create content since last summer': China has made use of AI-generated news anchors as a vehicle for state propaganda, Russia is using AI bot-posting on social media, while AI natural language models appear more genuine. The future implications of these uses for AI cannot be known precisely, but malicious usage of AI by hostile states has the potential to act as a significant disruptor for foreign relations, demonstrating the significance of specific policy and international regulation.

3. **The dual-use Dilemma: challenges for domestic policy and international relations**

The predicament of AI integration in counter-terrorism and security strategy presents its own extended debate regarding moral and ethical implications. The emerging digital landscape has been described as a 'chessboard' in which AI 'plays both terrorist and counter-terrorist' (Fishman, p.2). However, the fact remains that many of its features do contribute extensive potential for intelligence gathering and predictive operations . This has been recognised globally and implemented in security policy by many major global actors, including several allies of the United Kingdom.

The newer threat of lone-actor terrorism makes the full scope of counter-terrorism without effective integration of Artificial Intelligence

more difficult: offenders not affiliated with major terrorist groups are less likely to engage on public social media platforms and planning of such an attack does not require easily flaggable activities (such as phone calls and meetings) (Ganor, 2021). To counter threats of this nature and more generally, countries such as the US have utilised AI and machine learning as an effective means of processing huge amounts of data, of which some details may be overlooked by traditional algorithms. These techniques have also been optimised to observe terrorist movement in partnership with drones. This results in an ability to not only know where they are, but complex algorithms process this information to predict where they will be in the future. (Ganor, 2018). Predictive AI has also been utilised by China for policing and surveillance purposes, although the European Parliament has labelled it 'AI-based algorithmic authoritarianism' due to resulting human rights abuses (Kaskina, Cvetkovska, 2).

Worth acknowledging is that reactive policy-making of this nature is aided by the political structure of the respective countries, which includes executive orders in the US, and in more severe cases, strongly centralised power in Russia and China.

The UK has not wholly neglected to recognise and react to dangers posed by AI technologies on counter-terrorism and national security. In 2018, the government funded an AI-based startup to create an algorithm with the ability to detect extremist videos uploaded to social media (Vincent, 2017) . GCHQ, the UK's cyber-security agency, has publicly acknowledged that it is utilising AI to assist in countering a wide variety of cyber threats (GCHQ, 2021). £100million of investment into an AI Safety Institute was announced in 2023. Despite these steps, the most significant legislation passed in recent years, the Online Safety Act, focuses largely on risks emerging from social media platforms rather than AI (ISD, 2024).

Policy stagnation of this nature and the regulatory isolation it may cause is highlighted also by the UK's 2020 departure from the European Union, who have taken steps to implement formal policy surrounding Artificial Intelligence. In 2024, the EU implemented the AI act, which specifically lays down obligations for AI platform developers as a means of expediting removal of potential terrorist content (EU, 2024). While the act highlights that these regulations apply regardless of whether the companies supplying the technology are based in the EU or a third country, the majority of regulations apply only to activities in EU countries. As a result, UK policy must tighten AI regulations specifically with a focus on international cooperation and capitalization on remaining memberships with organisations such as NATO for the interest of physical and cybersecurity.

Additional strengths will include formal agreements with key allies such as the United States and the EU for information-sharing on Artificial Intelligence threats. The Five Country Ministerial Communiqué which includes the United Kingdom, Australia, Canada, the United States and New Zealand, has publicly declared a commitment to information sharing on methods

of establishing frameworks to cope with terrorist related AI threats (Home Office, 2024) and the UK's AI institute has spearheaded an international coalition with Canada and companies such as Amazon for a research project to assess 'AI behaviour and Control' (Department for Science, Information and Technology, 2025) However, formal and tangible security alliances will be needed for both information sharing on specific Artificial Intelligence threats identified combined with the tools to counter them. Another crucial tenet of this roadmap will include alignment and similar levels of strength with existing AI policy implemented by foreign governments to bolster the strength of the UK's standing as a significant global security actor, particularly in the wake of the UK's departure from the EU. Loss of policymaking influence within European institutions as a result of Brexit necessitates novel strategies for the UK to retain its place as both a technological and geopolitical innovator and collaborator in the West. Both spearheading and collaborating with innovative AI policies would serve as both a domestic security step and a success of reinforcing the UK's standing more globally.

While there are evidently agreements and regulations in place for regulation of Artificial Intelligence in general, the United Kingdom must simultaneously hold pace and drive innovation in implementing specific counter-terrorism related AI regulation. These formalised policies and laws, such as adopting a centralised regulation of AI, will be crucial to reinforcing the strength of the UK as not only a strong security actor, but a leading member of allied organisations such as NATO. Counter-terrorism strategies, such as CONTEST and PREVENT, will need to be aligned and synchronised with foreign governments, or risk terrorist organisations exploiting holes identifying and targeting lax regulatory areas. Politically, Britain's standing in global security groups such as the UN and NATO will become increasingly reliant on demonstrating leadership in navigating evolving threats such as Artificial Intelligence rather than merely following a curve. The nature of allyship involves a desire for collaboration, but a lack of strong regulation on data practises, particularly involving AI , may leave global allies reluctant to share information for fear of compromise. Exemplifying proactivity of this nature will require British-led solutions, such as advocating for 'UN-wide' standardisation of Artificial Intelligence guidelines and regulations such as watermarks on Deepfakes and sanctions on technology developers who fail to effectively police their technology according to these guidelines.

These solutions will not only strengthen relations between member states through cooperation but will serve to heighten the international standard for Artificial Intelligence regulation and use in counter-terrorism. Crucial here will be a finding the balance between embedding artificial intelligence techniques into counter-terrorism and security policy, whilst implementing regulations to prevent misuse by malicious actors.

Conclusion:

Artificial Intelligence is undoubtedly reshaping the landscape of global security in addition to the knowledge-gathering concept in its entirety. Key dangers for the UK include the struggle of government policy to adapt fast enough to the fluid nature of intelligent language models, particularly in the wake of Britain's departure from the EU which has demonstrated strong policy response already. However, the potential dangers of AI and its effects upon radicalisation, recruitment and knowledge-gathering are inherently transnational. As a result, national security will increasingly rely upon policy alignment with international allies to increase global resilience against novel threats created by AI, along with consistent awareness of AI implementation by hostile states. Success in this arena will be crucial to international standing; technological competence in the era of Artificial Intelligence and big technology are an increasingly key indicator of a nation's strength. The dilemma of Artificial Intelligence's power and danger existing often in parallel is an unprecedented challenge that transcends borders, thus the ongoing response needed by the UK government will be consistency in co-operation and regulation with foreign governments to tackle this ever-evolving issue, for reasons of both national and international security, along with the welfare of citizens globally.

References

Cabinet Office, National Security and Intelligence, Cabinet Office (2022) Security Policy

Framework. Available at: https://www.gov.uk/government/publications/security-policy-framework/hmg-security-policy-framework

(Accessed: 12th July 2025)

Davis, Aaron L (2021). 'Artificial Intelligence and the Fight Against International

Terrorism.' American Intelligence Journal, vol. 38, no. 2, 2021, pp. 63–73. JSTOR, Available

at: https://www.jstor.org/stable/27168700

(Accessed 23 Sept. 2025)

EU AI Act (2024) Official Journal of the European Union. Available

at: https://artificialintelligenceact.eu/ai-act-explorer/

(Accessed: 30th September 2025)

Fishman, Bar (2024). 'AI's Dual Role in Driving Online Terrorist Content and Counter

Strategies: Is NATO Prepared for AI-Enhanced Extremism?' Available

at: https://www.researchgate.net/publication/384925486_AI%27s_Dual_Role_in_Driving_O

nline_Terrorist_Content_and_Counter_Strategies_Is_NATO_Prepared_for_AI-Enhanced_Extremism

(Accessed: 17th August 2025)

Francheschi-Bicciari, L. (2023) 'Jailbreak tricks Discord's new chatbot into sharing napalm

and meth instructions' TechCrunch. Published April 20th. Available at:

napalm-and-meth-instructions/

(Accessed: 22nd September 2025)
Ganor, B. (2019). 'Artificial or Human: A New Era of Counterterrorism Intelligence?' Studies
in Conflict & Terrorism, p44(7), 605–624. Available
at: https://doi.org/10.1080/1057610X.2019.1568815
(Accessed:15th May 2025)
GCHQ (2021). 'Pioneering a New National Security: the Ethics of Artificial Intelligence'
Available at: https://www.gchq.gov.uk/files/GCHQAIPaper.pdf
(Accessed: 23rd June 2025)
Hall, J. (2023) 'Report of the Independent Reviewer of Terrorism Legislation on the operation of the Terrorism Acts of 2000 and 2006, and the Terrorism Prevention and Investigation Measures Act 2011'. Available at: : https://www.gov.uk/government/publications/the-terrorism-acts-in-2023/the- terrorism-acts-in-2023-report-of-the-independent-reviewer-of-terrorism-legislation- accessible
(Accessed: 17th May 2025)
Hall, J. (2023) 'The Terrorism Acts in 2023: report of the Independent Reviewer of Terrorism
Legislation': https://www.gov.uk/government/publications/the-terrorism-acts-in-2023/the-
terrorism-acts-in-2023-report-of-the-independent-reviewer-of-terrorism-legislation-accessible#:~:text=More%20traditional%20terrorism%20came%20to,a%20political%20and
%20religious%20cause
Home Office (2024) Five Country Ministerial Communiqué 2024 (Accessible). Available
at: https://www.gov.uk/government/publications/five-country-ministerial-communique-
2024/five-country-ministerial-communique-2024-accessible
(Accessed: 27th September 2025)
Institute for Economics and Peace (2025) 'Global Terrorism Index 2025: Measuring the
Impact of Terrorism' Sydney, March. Available at:
(Accessed: 30th September 2025)
Kasinka, R. and Cvetkovska, A. (2024) 'Artificial intelligence (AI) and human rights: Using AI
as a weapon of repression and its impact on human rights', *European Parliament*. Available
at: https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA%282024%29754450%28SUM01%29_EN.pdf
(Accessed: 6th September 2025)
Kissinger, H. (2018) 'How the Englightenment Ends'. The Atlantic, (August 2018) Available
At https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/
(Accessed: 29th September 2025)
Kruglanski, A. W., Bélanger, J. and Gunaratna, R. (2019). The three pillars of radicalization:

Needs, narratives, and networks. Oxford University Press. Available at:
(Accessed: 20th May 2025)
Watts, C. (2024) 'Nation States influence operations ahead of US
Election', *Microsoft Threat Analysis Center*. Available at: https://blogs.microsoft.com/on-the-issues/2024/10/23/as-the-u-s-election-nears-russia-iran-and-china-step-up-influence-efforts/
(Accessed: 13th September 2025)
Sayed, W. S., Noeman, A. M., Abdellatif, A., Abdelrazek, M., Badawy, M. G., Hamed, A., and
Tantawy, S. (2023). 'AI-based adaptive personalized content presentation and exercises
navigation for an effective and engaging E-learning platform'. *Multimedia Tools and Applications*, vol. 82(3), p. 3303-3333 Available at: https://link.springer.com/article/10.1007/s11042-022-13076-8#citeas
(Accessed: 22nd September 2025)
Vincent, J. (2018). 'UK creates machine learning algorithm for small video sites to detect ISIS
propaganda' The Verge. Available at: https://www.theverge.com/2018/2/13/17007136/uk-government-machine-learning-algorithm-isis-propaganda
(Accessed: 12th August 2025)
Wall, C. (2025) 'The Ghost in the Machine: Counterterrorism in the Age of Artificial
Intelligence', Studies in Conflict and Terrorism, pp. 1–27. doi:
10.1080/1057610X.2025.2475850
(Accessed: 23rd September 2025)
Weimann, G., T.Pack, A., Sulciner, R., Scheinin, J., Rapaport, G., Diaz, D. (2024). 'Generating
Terror: the Risks of Generative AI Exploitation' CTC Sentinel. Available at:
(Accessed: 20th August 2025)
Wilson, C.,and Dziwulski, M. (2025). 'Countering the propaganda of terrorists: The deception of Brenton Tarrant' *Journal of Threat Assessment and Management*. Available
at: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ftam0000252
(Accessed: 20th September 2025)
Zitha, L. Z., Pinheiro, M. L., Gonçalves, R. A., & Caridade, S. (2024). 'Recruitment, Affiliation, and Disengagement Among Men in Terrorist Organizations: A Systematic Review'. Social Sciences, vol.13(11), p.609. Available at: https://doi.org/10.3390/socsci13110609
(Accessed: 10th September 2025)

DOI 10.69790/epis10.44933