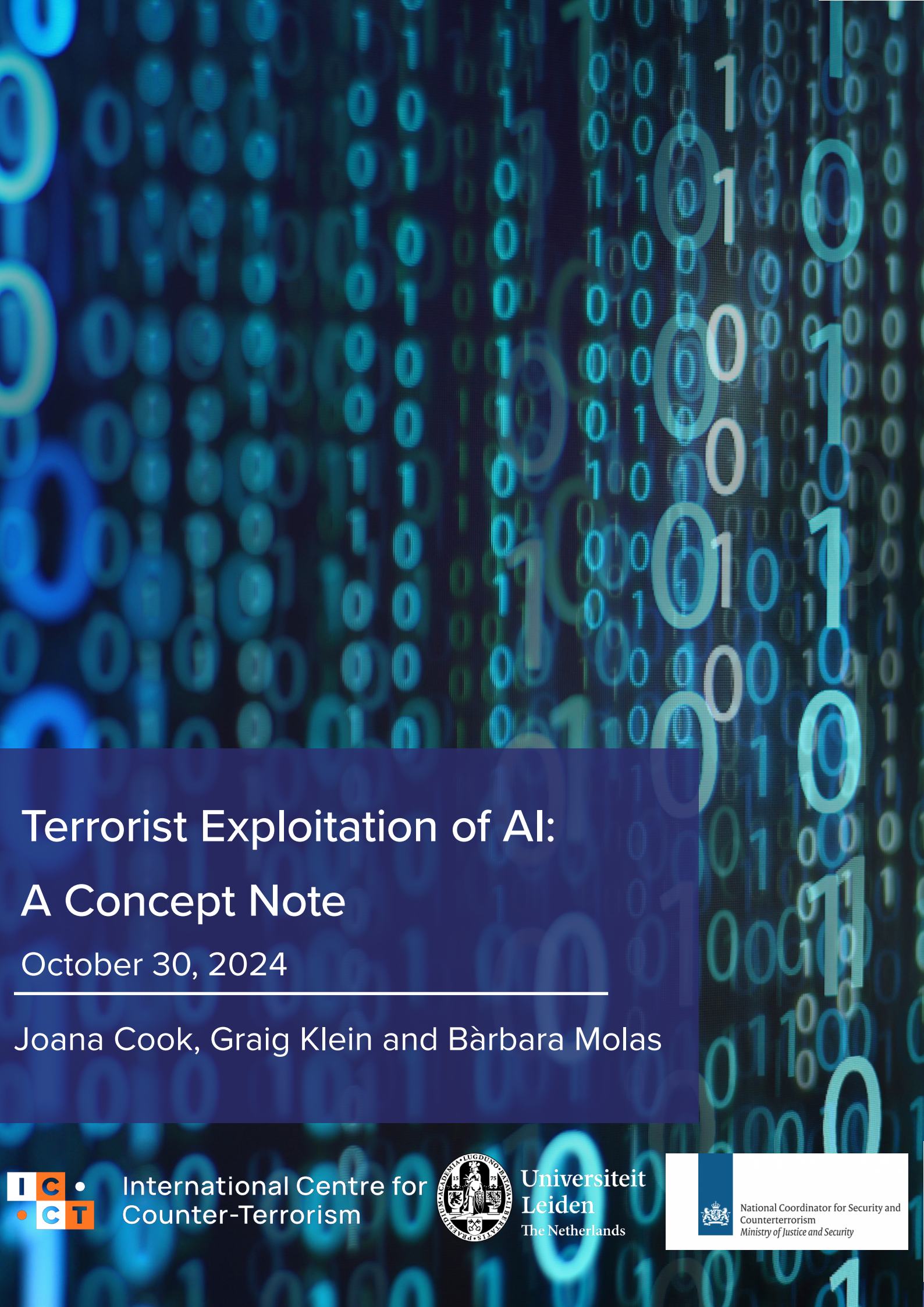

# Terrorist Exploitation of AI: A Concept Note

October 30, 2024

Joana Cook, Graig Klein and Bàrbara Molas

International Centre for Counter-Terrorism

Universiteit Leiden
The Netherlands

National Coordinator for Security and Counterterrorism
Ministry of Justice and Security

# Terrorist Exploitation of AI: A Concept Note

Joana Cook, Graig Klein and Bàrbara Molas

October 30, 2024

International Centre for Counter-Terrorism

Universiteit Leiden
The Netherlands

National Coordinator for Security and Counterterrorism
*Ministry of Justice and Security*

# About ICCT

The International Centre for Counter-Terrorism (ICCT) is an independent think and do tank providing multidisciplinary policy advice and practical, solution-oriented implementation support on prevention and the rule of law, two vital pillars of effective counter-terrorism.

ICCT's work focuses on themes at the intersection of countering violent extremism and criminal justice sector responses, as well as human rights-related aspects of counter-terrorism. The major project areas concern countering violent extremism, rule of law, foreign fighters, country and regional analysis, rehabilitation, civil society engagement and victims' voices. Functioning as a nucleus within the international counter-terrorism network, ICCT connects experts, policymakers, civil society actors and practitioners from different fields by providing a platform for productive collaboration, practical analysis, and exchange of experiences and expertise, with the ultimate aim of identifying innovative and comprehensive approaches to preventing and countering terrorism.

## Licensing and Distribution



## Research Assistance

This concept note was produced with research assistance from Dyon van Velzen and Nina Prillwitz.

# Contents

# Glossary

- **Algorithmic bias:** the systematic errors in a computer system that favour certain groups over others based on their design and the data they are trained on.[1]
- **Autonomous weapons:** machines capable of independently searching, selecting and engaging targets without human intervention.[2]
- **Artificial General Intelligence (AGI):** a type of AI capable of performing any intellectual task a human can do. Unlike current narrow AI systems that specialise in specific tasks, AGI would have the ability to generalise knowledge and skills across various domains.[3]
- **Artificial Intelligence (AI):** technology that enables computers and machines to perform tasks that normally require human intelligence. It can simulate human learning, reasoning, problem-solving, decision-making, and language understanding through algorithms.[4]
- **Artificial Superintelligence (ASI):** a hypothetical level of AI that surpasses human intelligence in every aspect, including creativity, problem-solving, and emotional intelligence.[5]
- **Centralised Artificial Intelligence:** managed by a single entity (company), and data is collected and stored in a central location, making it easier to manage and maintain. However, this also introduces risks, such as data monopolies, and potential misuse of power.[6]
- **Cyberterrorism:** the use of digital technology to launch attacks on e.g. computer systems, networks, and the information stored within them, which causes physical, political, economic, or other damage. Intended to induce fear or coerce entities to act in a way that furthers any political, ideological, or religious objective (e.g. hacking, spreading propaganda, or disrupting critical infrastructure).[7]
- **Decentralised Artificial Intelligence:** distributes the processing, storage and maintenance across multiple nodes,[8] making it more democratic, accessible, and secure, although it poses challenges in terms of regulation.[9]
- **Deep-learning:** a subset of machine learning that has the ability to perform unsupervised learning of unstructured data by using multilayered neural networks. Simulates the complex decision-making power of the human brain.[10]
- **Generative Artificial Intelligence:** a type of AI that can generate new and original content based on the user's prompt or request. This includes but is not limited to text, images, videos, audio, and software code. It uses deep-learning models to understand patterns in data and generate new similar content.[11]

1 ''What is Algorithmic Bias,'' *Data Camp,* July 17, 2023, https://www.datacamp.com/blog/what-is-algorithmic-bias.
2 Neil Davison, ''What you need to know about autonomous weapons,'' *International Committee of the Red Cross*, 26 July 2023, https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons.
3 Will Douglas Heaven, "Google DeepMind wants to define what counts as artificial general intelligence", *MIT Technology Review*, November 2023, https://www.technologyreview.com/2023/11/16/1083498/google-deepmind-what-is-artificial-general-intelligence-agi/.; "What Is Artificial General Intelligence (AGI) and How To Prepare For It" *IMD*, last updated October 2024, https://www.imd.org/blog/digital-transformation/artificial-general-intelligence-agi/.
4 Andreas Kaplan, and Michael Haenlein, "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", *Business Horizons* 62, no. 1, (2019), https://doi.org/10.1016/j.bushor.2018.08.004.; Shiona McCallum, Chris Vallance, Tom Gerken & Jennifer Clarke "What is AI, how does it work and what can it be used for?" *BBC News*, 13 May 2024. https://www.bbc.com/news/technology-65855333.
5 Tim Mucci and Cole Stryker "What is artificial superintelligence?" *IBM,* 18 December 2023, https://www.ibm.com/topics/artificial-superintelligence.
6 "What is the Difference Between Centralized and Decentralized AI?" *Venice*, 26 September 2024. https://venice.ai/blog/what-is-the-difference-between-centralized-and-decentralized-ai
7 Jordan Plotnek and Jill Slay, "Cyber Terrorism: A Homogenized Taxonomy and Definition." *Computers & Security* 102, (2021), https://doi.org/10.1016/j.cose.2020.102145.
8 "Decentralized AI: Pros and Cons" *Zerocap*, 30 May 2024. https://zerocap.com/insights/snippets/decentralized-ai-pros-cons/
9 "Regulatory Landscape for Decentralized AI: Navigating Legal Frameworks", *Medium*, 2 May 2024. https://medium.com/sample-dcentai-blog/regulatory-landscape-for-decentralized-ai-navigating-legal-frameworks-c2880881e832
10 Jim Holdsworth and Mark Scapicchio, "What is deep learning?," *IBM*, 17 June 2024, https://www.ibm.com/topics/deep-learning.
11 Cole Stryker and Mark Scapicchio, "What is generative AI?," *IBM*, 22 March 2024, https://www.ibm.com/topics/generative-ai.

- **Large Language Models (LLM):** a type of AI designed to understand, generate and manipulate human language. Models are trained by vast datasets of text and built using deep-learning techniques. These models can perform tasks like answering questions, translating languages, and summarising content.[12]
- **Radicalisation:** the process by which individuals or groups come to adopt increasingly extreme political, social, or religious ideals, often leading them to support or engage in acts of violence.[13]
- **Terrorism:** the unlawful use of violence or threat of violence, outside of regular warfare and typically against civilians, to achieve political, ideological, religious or other objectives. Terrorist acts aim to create fear and send messages to broader audiences beyond the immediate victims.[14]
- **Violent extremism:** ideologies or actions that justify or promote the use of violence to achieve ideological, political, or social goals. Extremist groups often reject democratic processes and resort to force to impose their beliefs.[15]

12 "What are LLMs?," *IBM*, n.d., https://www.ibm.com/topics/large-language-models.
13 Caroline Logan, Randy Borum and Paul Gill. '*Violent Extremism: A handbook of risk assessment and management*' (UCL Press, 2023).
14 Alex P. Schmidt, "Defining Terrorism," *International Centre for Counter Terrorism*, 2023. https://doi.org/10.19165/2023.3.01.
15 Ibid.

# Introduction

This concept note provides a brief overview of the subject of terrorist exploitation of Artificial Intelligence (AI). It first describes key concepts and areas of focus in current research. It then summarises some of the contemporary research on several key sub-topics of AI and emerging terrorist threats, including: AI for operational purposes; AI content and content flows; AI and large language models; AI and state terrorism; and AI as an independent entity. It finally references some further recommended readings.

## What is Artificial Intelligence?

AI is a general-purpose technology designed to make human activity more efficient and improve the well-being of people.[16] AI systems operate at significant levels of automation and include various iterations, such as algorithmic AI, generative AI, large language models (LLMs), and deep learning machines.

Generative AI and LLMs, in particular, gained widespread attention with the release of ChatGPT in November 2022, marking a turning point for these technologies.[17] Generative AI platforms use machine learning to generate high-quality images, audio, songs, videos, and multifunctional simulations through "prompts" or instructions based on the data they were trained on.[18] As with any other new and emerging technology, AI provides opportunities for terrorist exploitation.[19] However, it can also be considered in terms of opportunities and, for example, the potential use of AI for countering terrorism.[20]

## AI and Emerging (Terrorist) Threats

Despite the implementation of barriers and policies by generative AI platforms aimed at curbing the creation of violent, discriminatory, and harmful content, recent studies suggest that AI, in particular LLMs, have the potential to "enable terrorists to learn, plan, and propagate their activities with greater efficiency, accuracy, and impact than ever before."[21] Such enhancements of terrorist capabilities also include enhancing their cyber capabilities, enabling physical attacks through automation, facilitating the financing of their activities, and generating tailored material as well as amplifying its spread. Indeed, ill-intentioned users continue to adapt, finding ways to "jailbreak" – deliberately circumventing the ethical and operational boundaries designed by AI platforms – to generate harmful content that may enhance their capabilities to radicalise, recruit, and target victims in popular online spaces.[22]

In response to the evolving capabilities of AI, counter-terrorism experts have begun to study the weaponisation of this technology by extremists and its impacts on the field. The body of literature can be divided into two types of research. The first type is theoretical and aims to anticipate the different ways in which users can exploit AI.[23] This approach has hypothesised on the potential

16 "What is artificial intelligence (AI)?," *ISO*, n.d.; "AI Principles Overview," *OECD.AI*, last modified May 2024, https://oecd.ai/en/ai-principles
17 Cole Stryker and Mark Scapicchio, "What is generative AI?," *IBM*, 22 March 2024, https://www.ibm.com/topics/generative-ai.
18 Kinza Yasar, "What is an AI prompt?", *TechTarget*, September 2023, https://www.techtarget.com/searchenterpriseai/definition/AI-prompt#:~:text=AI%20prompts%20provide%20explicit%20instructions,to%20produce%20the%20desired%20outputs.
19 Clarisa Nelu, "Exploitation of Generative AI by Terrorist Groups," *International Centre for Counter-Terrorism*, 10 June 2024, https://www.icct.nl/publication/exploitation-generative-ai-terrorist-groups.
20 Stuart Macdonald, Ashley Mattheis and David Wells. *Using Artificial Intelligence and Machine Learning to Identify Terrorist Content Online* (Tech Against Terrorism Europe, 2024) https://tate.techagainstterrorism.org/news/tcoaireport.
21 Gabriel Weimann et al., "Generating Terror: The Risks of Generative AI Exploitation," *Combating Terrorism Center*, January 2024, https://ctc.westpoint.edu/generating-terror-the-risks-of-generative-ai-exploitation/.
22 Anil et al., "Many-shot jailbreaking," *Anthropic*, 2024, https://www.anthropic.com/research/many-shot-jailbreaking.
23 Darya Bazarkina, "Current and Future Threats of the Malicious Use of Artificial Intelligence by Terrorists: Psychological Aspects," in *The Palgrave Handbook of Malicious Use of AI and Psychological Security,* ed. Evgeny Pashentsev (Palgrave Macmillan, 2023); Miriam Fernandez and Harith Alani, "Artificial Intelligence and Online Extremism: Challenges and Opportunities," in *Predictive Policing and Artificial Intelligence,* eds.

uses of AI based on terrorists' and extremist ideologies, goals, and agendas. The second, and more recent type of research, aims to test AI tools and functionalities in controlled environments in order to similarly anticipate potential harmful uses.[24] In such experiments, experts induce platforms to generate harmful content. The results of this type of research have been particularly helpful in identifying the most vulnerable spaces to jailbreak.[25]

# 1. AI for Terrorist Operational Purposes (e.g. Weapons Systems)

Terrorists already utilise weapons and weapons systems that benefit from AI technologies, and some future potential exploitation of AI scenarios have been identified. Current state-of-the-art general-purpose AI systems are capable of autonomously executing many simple tasks, but some evaluations have shown that they struggle with more complex ones, or ones that involve many steps.[26] That said, autonomous or AI-powered military drones have been fulfilling vital functions, including reconnaissance, data analysis, object detection, and targeting in combat settings, including in Ukraine.[27] There are also "unmanned aerial systems", or drones which are remotely piloted using AI. According to the United Nations Security Council Counter-Terrorism Committee, these are some of the primary terrorist threats in relation to AI-powered weapons.[28] Terrorist groups already deploy drones and drone swarms to perpetrate attacks.[29] By 2023, at least 65 violent non-state actors had the capability to deploy drones.[30] AI communication and linkages between drones enable swarm formation stability under human control. If humans are removed from the loop safety mechanisms, AI can allow the weapons systems to autonomously identify targets and fire ammunition.[31] We have already seen this with Ukraine's Caesar 155mm truck-mounted self-propelled weapon, which, upon receiving a fire mission, is able to move, halt, deploy, load, and fire without an operator.[32] Though still speculative, it is possible that terrorists could control drones in such a fashion or equip these AI-enabled drones with chemical or biological weapons.[33]

---

John McDaniel and Ken Pease (Routledge Frontiers of Criminal Justice, 2021); Miles Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," *Future of Humanity Institute*, 2018, https://doi.org/10.48550/arXiv.1802.07228.

24 "Advanced AI evaluations at AISI: May update," *AI Safety Institute*, 20 May 2024, https://www.aisi.gov.uk/work/advanced-ai-evaluations-may-update; Weimann et al., "Generating Terror"; Miron Lakomy, "Artificial Intelligence as a Terrorism Enabler? Understanding the Potential Impact of Chatbots and Image Generators on Online Terrorist Activities," *Studies in conflict and terrorism,* 15 December 2023, https://doi.org/10.1080/1057610x.2023.2259195.

25 Weimann et al., "Generating Terror"; Miron Lakomy, "Artificial Intelligence as a Terrorism Enabler? Understanding the Potential Impact of Chatbots and Image Generators on Online Terrorist Activities".

26 Bengio et al, "International Scientific Report on the Safety of Advanced AI: Interim Report. International Scientific Report on the Safety of Advanced AI", Report 1 (interim report). May 2024. p. 65.
https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai

27 "Introduction: The Power of Drones and the Introduction of Artificial Intelligence", *Vision Platform,* 27 January 2024, https://visionplatform.ai/artificial-intelligence-drones/.

28 "Preventing Terrorists from Using Emerging Technologies," *Vision of Humanity*, n.d.
https://www.visionofhumanity.org/preventing-terrorists-from-using-emerging-technologies/.

29 Robert J. Bunker, Terrorist and Insurgency Unmanned Aerial Vehicles: Use, Potentials, and Military Applications, (Strategic Studies Institute and US Army War College Press, 2015); Yannick Veilleux-Lepage and Emil Archambault, "A Comparative Study of Violent Non-state Drone Use," International Centre for Counter-Terrorism, 9 December 2022, https://doi.org/10.19165/2022.3.01.

30 Institute for Economics & Peace, "*Global Terrorism Index 2023: Measuring the Impact of Terrorism,*" 2023, http://visionofhumanity.org/resources.

31 "*Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes,*" United Nations Counter-Terrorism Centre and United Nations Interregional Crime and Justice Research Institute, 2021; Birgitta Dresp-Langley, "The Weaponization of Artificial Intelligence: What the Public Needs to be Aware of," *Frontiers in Artificial Intelligence* 6, (8 March 2023), https://doi.org/10.3389/frai.2023.1154184.; Nelu, "Exploitation of Generative AI"; Susan Sims, "Emerging Terrorist Threats: Everything, Everywhere, All at Once?," in *Emerging Technologies and Terrorism: An American Perspective* ed. Susan Sim, Eric Hartunian, and Paul J. Milas (US Army War College Press, 2024).

32 Stephen W. Miller,"Ukraine's Caesar Artillery Incorporates Artificial Intelligence", *Armada International,* 10 January 2024, https://www.armadainternational.com/2024/01/ukraines-caesar-artillery-incorporating-artificial-intelligence/.

33 Roger Brent, T. Greg McKelvey Jr., and Jason Matheny, "The New Bioweapons: How Synthetic Biology Could Destasbilize the World," *Foreign Affairs,* 20 August 2024; Sims, ''Emerging Terrorist Threats''.

The risk of drone-enabled biological attacks is amplified by the potential for terrorists to hack biometric databases and use retina, fingerprint, facial recognition and other biomarkers to access laboratories.[34] Gaining access to biomarkers and high-resolution photographs also creates the potential for 3D printing of facial models to bypass security,[35] and improves passport counterfeiting operations.[36] Similarly, while advancements in facial recognition software have allowed law enforcement, security services, and emergency/rescue services to search for specific individuals in their operations, terrorist groups too may one day integrate such technologies and software into their drones that would improve target identification, surveillance, and enable targeted strikes against specific individuals.[37]

Alongside the increased use of AI in warfare, the use of autonomous vehicles in terrorist attacks is expected to increase with the growth of self-driving cars.[38] Videos have already shown the Islamic State experimenting with self-driving vehicles and placing mannequins and heat signature replicating tools and technology inside the vehicles to counter detection systems.[39] The networked technology that self-driving cars rely on creates opportunities for terrorists to hack into the network to manipulate and disrupt communication, perpetrate human-free "suicide" vehicle bombings,[40] or create traffic jams or roadblocks that prevent emergency response and security services from reaching an attack location.[41] Beyond vehicles, the increasingly networked nature of government and civilian devices and systems in smart cities and smart homes increases the potential scale and impact of a terrorist hack or attack.[42]

Finally, terrorists could leverage the abundance of open-source data and AI-powered data analytics to predict countermeasures and attack responses to increase the impact and effectiveness of terrorist attacks.[43] By evaluating previous security service responses to attacks and civilian movement in different locations before, during, and after an attack, terrorists could use predictive analyses to maximise both the likelihood of successfully penetrating a target and the resulting carnage. While it has yet to be confirmed whether terrorists are using AI, Natural Language Processing (NLP), LLMs, etc. to process and analyse data, there is evidence that terrorists have realised this potential and use augmented and virtual reality tools to collect information for attack planning and practice.[44]

34 Sarah Lohmann, "ChatGPT, Artificial Intelligence, and the Terrorist Toolbox," *in Emerging Technologies and Terrorism: An American Perspective*, ed. Susan Sim, Eric Hartunian, and Paul J. Milas (US Army War College Press, 2024).
35 Giulia Carbonaro, "Can Videos Uploaded on Social Media Allow Hackers to Steal Your Biometric Data?," Euronews (website), 27 October 2022, https://www.euronews.com/next/2022/10 /27/can-videos-uploaded-on-social-media-allow-hackers-to-steal-your-biometric-data.
36 "*Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes*," United Nations Counter-Terrorism Centre and United Nations Interregional Crime and Justice Research Institute, 2021.
37 Ibid.
38 Ibid.
39 Ibid.
40 Nelu, "Exploitation of Generative AI''; Sarah Lohmann, "ChatGPT, Artificial Intelligence, and the Terrorist Toolbox," in *Emerging Technologies and Terrorism: An American Perspective*, ed. Susan Sim, Eric Hartunian, and Paul J. Milas (US Army War College Press, 2024).
41 "*Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes,*" United Nations Counter-Terrorism Centre and United Nations Interregional Crime and Justice Research Institute, 2021..
42 Ibid.
43 Ibid; Branislav Todorovic and Darko Trifunovic, "Prevention of (Ab-)Use of the Internet for Terrorist Plotting and Related Purposes," in *Handbook of Terrorism Prevention and Preparedness,* ed. Alex P. Schmid (International Centre for Counter-Terrorism, 2021); Nikita Vashishta, "Artificial Intelligence-Assisted Terrorism: A New Era of Conflict," *Vivekanada International Foundation,* 29 August 2023, https://www.vifindia.org/article/2023/august/29/Artificial-Intelligence-assisted-Terrorism-A-New-Era-of-Conflict.
44 Susan Sims, "Emerging Terrorist Threats''.

# 2. AI Content and Content Flows (e.g. Propaganda and Tactical Learning)

Generative AI has the potential to change online communication and media, especially regarding misinformation and disinformation, but more generally concerning content and content flows, specifically the production of text, images, animations, music and videos, and their circulation. Using synthetic (fake) images, videos, or audio that align with a terrorist organisation's values could help increase the breadth and amplification of propaganda produced, increase the quality of messages through tailored content, and affect attitudes and emotions more effectively.[45] In the context of conflict, for instance, AI-generated images have been used to undermine the enemy by picturing injured or deceased young people or babies.[46] A recent example of this occurred in October 2023 during the continuing conflict between Israel and Hamas, where AI-generated images were employed by extremist actors to depict exaggerated scenes of destruction and suffering to manipulate public sentiment and rally support for their cause.[47]

Beyond its role in producing propaganda, generative AI can enhance tactical learning within terrorist organisations. AI text-generating tools can be used to develop training capabilities with minimal external guidance, such as providing instructions for circumventing content moderation, generating tutorials on emergency medical responses on the battlefield or analysing the effectiveness of various tactical strategies.[48] The Islamic State, for instance, has gone as far as to publish a guide on how to securely use generative AI, and supporters of the terrorist group have expressed an interest in further using AI to boost the scale and scope of its public content.[49] They have also explored methods to use AI-generated content to bypass detection and censorship, ensuring that their messages reach broader audiences without being flagged or removed.[50]

# 3. AI and Large Language Models (LLMs) (Centralised and Decentralised Platforms)

LLMs have already been used to produce and spread extremist texts on social media,[51] maintain personalised communication with recruits as chatbots, and generate propaganda.[52] The generated content effectively mimics "interactive, informational, and influential content," which can be exploited to radicalise individuals into adopting violent extremist ideologies and behaviours.[53] Other risks associated with LLMs include social scams, the spread of misinformation and disinformation, and easy access to malicious coding.[54] LLMs include both centralised and decentralised platforms. The key distinction between centralised and decentralised AI lies in how control and processing are managed. Decentralised models may be less monitored and open to exploitation.

45 Maggie Engler, "Considerations of the Impacts of Generative AI on Online Terrorism and Extremism" *GIFCT Red Team Working Group*, 20 September 2023. https://gifct.org/wp-content/uploads/2023/09/GIFCT-23WG-0823-GenerativeAI-1.1.pdf
46 Nelu, "Exploitation of Generative AI".
47 Tech Against Terrorism, "*Early Terrorist Experimentation with Generative Artificial Intelligence Services*." Tech Against Terrorism Briefing, November 2023, https://techagainstterrorism.org/news/early-terrorist-adoption-of-generative-ai.
48 Weimann et al., "Generating Terror".
49 The Soufan Centre, "Terrorist Groups Looking to AI to Enhance Propaganda and Recruitment Efforts," 3 October 2024, https://thesoufancenter.org/intelbrief-2024-october-3/.
50 Tech Against Terrorism, "Early Terrorist Experimentation".
51 Maggie Engler, "Considerations of the Impacts of Generative AI on Online Terrorism and Extremism".
52 Stephane Baele, "Artificial Intelligence And Extremism: The Threat Of Language Models For Propaganda Purposes," *CREST Security Review* 16, 2022; Weimann et al. "Generating Terror," 17-24.
53 Kris McGuffie and Alex Newhouse, "The radicalization risks of GPT-3 and advanced neural language models," *arXiv*, 2020, https://doi.org/10.48550/arXiv.2009.06807.
54 Puczyńska et al., "LLMs in jihadist terrorism"; Ben Buchanan, Andrew Lohn, Micah Musser,and Katerina Sedova "Truth, Lies, and Automation: How Language Models Could Change Disinformation." *CSET*, 1 May 2021; Europol, "*ChatGPT - The impact of Large Language Models on Law Enforcement, a Tech Watch Flash Report from the Europol Innovation Lab*", Publications Office of the European Union, 2023.

There are already concrete examples demonstrating the severity of these risks. For instance, in the United Kingdom, a man plotted to kill Queen Elizabeth after being influenced by conversations with a chatbot.[55] In another case, the far-right created its own chatbot and even manipulated a Microsoft chatbot into adopting white supremacist ideologies.[56] On platforms like Character.ai, users can design bots that mimic specific characters, including members of violent extremist groups.[57] Furthermore, Islamic State groups have used bots on Telegram to disseminate their propaganda.[58] One of the challenges extremist groups have traditionally faced is the production and dissemination of content, particularly due to a lack of translators, which limits their ability to reach and radicalise wider global audiences. However, LLMs can easily overcome this barrier.[59] Indeed, terrorist organisations have already begun leveraging LLMs to amplify the spread of extremist content. These models are used to generate articles and instructional materials on how to conduct attacks,[60] as well as to assist in planning and executing operations by analysing large datasets, identifying patterns, and providing strategic insights, which can enhance the effectiveness of terrorist activities.[61] This essentially lowers the financial and expertise barriers to committing acts of terror.[62] Extremist groups have even gone as far as developing their own chatbots, and have used open-source models to foster online radicalisation, promote violence, and imitate victims of violence and stereotypes.[63] A key aspect of the development of such chatbots is their centralised or decentralised nature.

# 4. AI and State-Sponsored Terrorism

In light of the rapid technological developments in domains such as communication, encryption, surveillance and (facial) recognition, AI is also increasingly becoming an important tool in state activities, including state-support of violent non-state groups in some cases. Through the implementation of recognition and prediction algorithms, states are already using AI-driven systems to identify and keep track of potential targets. Most prominently, the Israeli intelligence service has developed a system called *The Gospel* to identify targets in Gaza, enabling them to identify "up to 100 new targets every day", whereas human intelligence officers were only able to identify 50 per year.[64] Due to the large degree of autonomy on which these systems operate, requiring very little human input, terrorist groups could greatly benefit from obtaining access to such technologies. For example, states could use this rapid target identification to guide violent groups more accurately or allow them to more efficiently gather their own intelligence in cases where they own such capabilities.

55 Tom Singleton, Tom Gerken and Liv McMahon, "How a Chatbot Encouraged a Man Who Wanted to Kill the Queen," *BBC News*, October 6, 2023, https://www.bbc.com/news/technology-67012224.
56 Siegel, "RedPilled AI"; Koblenz-Stenzler et al., "Navigating Far-Right Extremism".
57 Chris Vallance and Imran Rahman-Jones. 'Urgent need for terrorism AI laws, warns think tank', *BBC News*, 4 January 2024, https://www.bbc.com/news/technology-67872767.
58 Yannick Veilleux-Lepage, Chelsea Daymon and Emil Archambault, Learning from Foes: How Racially and Ethnically Motivated Violent Extremists Embrace and Mimic Islamic State's Use of Emerging Technologies. *Global Network on Extremism and Technology* (2022).
59 Hanny Sari and Muhamad Syauqillah, "The Role of Translation in ISIS Propaganda: International Online Radicalization Methods and Its Effects on Extremism in Indonesia," *International Journal of Science and Society*, (2022): 319-36. https://doi.org/10.54783/ijsoc.v4i4.578.
60 NCTC, DHS, FBI "*First Responders Toolbox. Violent Extremists Use of Generative AI*," 6 May 2024, https://www.dni.gov/index.php/nctc-how-we-work/joint-ct-assessment-team/first-responder-toolbox.
61 Weimann et al., "Generating Terror".
62 Puczyńska et al., "Large Language Models".
63 Siegel, 'RedPilled AI''; Koblentz-Stenzler, "Navigating Far-Right Extremism"; Stuart A. Thompson, "New Uncensored Chatbots Ignite a Free-Speech Fracas" *New York Times*, 2 July 2023, https://www.nytimes.com/2023/07/02/technology/ai-chatbots-misinformation-free-speech.html; Liram Koblentz-Stenzler and Uri Klempner, "Navigating Far-Right Extremism in the Era of Artificial Intelligence," *Global Network on Extremism and Technology*, 25 January 2024.
64 Geoff Brumfiel, "Israel is using an AI system to find targets in Gaza. Experts say it's just the start,'' *NPR*, 14 December 2023, https://www.npr.org/2023/12/14/1218643254/israel-is-using-an-ai-system-to-find-targets-in-gaza-experts-say-its-just-the-st.

Furthermore, AI enables states to further obfuscate their sponsorship of non-state armed groups. Through encrypted messaging and sophisticated networks of communication, the developing online space provides diverse potential for hiding a state's role.[65] In a more direct way, states have already made use of unmanned air systems to aid non-state groups in combat without sending military personnel.[66] For example, in 2022, Iran directly provided lethal unmanned aerial systems (UAS) to the Kata'ib Hizballah, Harakat al-Nujaba, and Asa'ib Ahl al-Haq groups in Iraq,[67] with the UN reporting at least seven similar cases in 2023.[68] Lastly, with the omnipresence of social media and messaging networks, states have access to a low-cost way of inciting violence, for example through using botnets to exploit social tensions.[69]

In a peaceful context, in both democracies and autocracies, AI has been employed to steer public opinion, "spread disinformation," and automatically censor critical online content.[70] Authoritarian regimes have harnessed AI to produce propaganda that reinforces state narratives. For instance, in China, AI has been used to conduct large-scale surveillance and to develop models that integrate with censorship systems, ensuring that online discussions remain within the boundaries set by the state.[71] Democracies are not immune to these practices; while they do not use AI for overt control, they still employ it in ways that can influence public opinion. For example, models with built-in assumptions can present skewed or incomplete information that subtly alters how people perceive political and social issues. This effect is amplified by the fact that only a few large companies control the main AI models, giving them significant control over public discussions.[72] Finally, state-aligned groups are also using LLMs as a tool for cyberwarfare. At the moment, they are mostly using it for research-oriented purposes and troubleshooting, but they are expanding. This technology could also be distributed to terrorist groups and aid them for research and intelligence purposes when they do not have sufficient trained personnel.[73]

## 5. AI as an Independent Entity (as an Uncontrolled Weapon or 'Terrorist')

Some developers are working to create general-purpose AI systems that can act with increasing autonomy, which means that they would be capable of operating, taking actions, or making decisions without the express intent or oversight of a human.[74] While such companies hope for these developments to solve human health problems or relieve human workers of mundane tasks, less oversight for AI systems would involve a wide range of societal risks.[75] These concerns mainly pertain to artificial general intelligence (AGI), systems that can rival human cognitive skills, and artificial superintelligence (ASI), machines with the capacity to exceed human intelligence.

65 Daniel Byman, "Understanding, and Misunderstanding, State Sponsorship of Terrorism," *Studies in Conflict and Terrorism* 45, no. 12 (2020): 1031–49, https://doi.org/10.1080/1057610x.2020.1738682.
66 United Nations Office of Counter-Terrorism, Global Counter-Terrorism Programme on Autonomous and Remotely Operated Systems (AROS Programme) and Conflict Armament Research, "Global Report on the Acquisition, Weaponization and Deployment of Unmanned Aircraft Systems by Non-State Armed Groups for Terrorism-related Purposes," *United Nations Global Counter-Terrorism Strategy* (2023), https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/unoct_car_global_report_web_en.pdf.
67 "Country Reports on Terrorism 2022," *US Department of State, Bureau of Counterterrorism*, 2022, https://www.state.gov/reports/country-reports-on-terrorism-2022/.
68 United Nations, "Global Report on the Acquisition, Weaponization and Deployment of Unmanned Aircraft Systems".
69 Byman, "Understanding, and Misunderstanding, State Sponsorship of Terrorism."
70 Tate Ryan-Mosley, "How generative AI is boosting the spread of disinformation and propaganda", *MIT Technology Review*, 4 October 2023.
71 Feldstein, Steven. "The Consequences of Generative AI for Democracy, Governance and War." *Survival* (London) 65, no. 5 (2023): 117–42. 123. doi:10.1080/00396338.2023.2261260.
72 Ibid, 120.
73 Nelu, "Exploitation of Generative AI".
74 Bengio et al, "International Scientific Report on the Safety of Advanced AI: Interim Report. International Scientific Report on the Safety of Advanced AI".
75 "When will the first weakly general AI system be devised, tested, and publicly announced?" *Metacalculus*, https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/,

Currently, no such systems exist,[76] but there are different estimates of when such capabilities may be achieved.[77]

There are numerous risk factors involved in increasingly autonomous AI, but three are worth stressing. Above all, the lack of transparency regarding how AI models actually work, including how they are trained or used, makes liability harder to determine, making regulation and enforcement more difficult. Secondly, the fact that developers competing for market share may have limited incentives to mitigate risks can increase such risks.[78] Finally, as the AI race pushes states to develop autonomous systems and weapons as rapidly as possible, states might have little time for reflection on the long-term effects of these technologies upon society.[79]

While it is widely recognised that while AI systems may bring great benefits to humanity, AI systems could also do immense harm to humans.[80] The current rapid progress in the development of AI systems may not end before it presents a broad range of capabilities that exceed our own capacities.[81] While some argue that current AI models are currently not nearly capable of a takeover,[82] others argue that AI has already surpassed human-level performance in some areas, giving rise to fears that AGI and ASI may not be far off.[83] Put bluntly by one academic, the potential development of ASI poses three existential risks including human extinction or extreme harm to humanity.[84] It is difficult to prevent AI models from being misused and nearly impossible to stop a model's capabilities from proliferating, which results in the fact that dangerous AI capabilities can rise quickly and unexpectedly.[85] Moreover, nobody yet knows how to train powerful AI systems to be strictly helpful, honest, and harmless.[86]

Scholars such as Yampolskiy, a pioneer in the field of AI safety, suggest that "development of the technology should be slowed or suspended until AI safety can be assured and controls established."[87] Many scientists agree that AI risks are real and that they need to be mitigated before further development. Industry self-regulation seems to be an important first step, but wider societal discussions and government intervention will be needed to create standards and to ensure compliance with them.[88] There have been calls for the creation of an international AI safety organisation, and in May 2023, CEOs of the world's leading AI companies and experts expressed their concerns in an open letter entitled "Statement on AI Risk", released by the California-based non-profit Center for AI Safety.[89] The letter reads one sentence: "Mitigating the risks of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[90] In short, with the possible development of AGI or ASI, there is also a concern about the potential for AI itself to become a 'weapon' or 'source of terror' (though not necessarily ideologically motivated), which could harm humanity.

76 Zachary Kallenborn, ''Policy makers should plan for superintelligent AI, even if it never happens,'' *Bulletin of the Atomic Scientists*, December 21, 2023, https://thebulletin.org/2023/12/policy-makers-should-plan-for-superintelligent-ai-even-if-it-never-happens/
77 "List of p(doom) values", *PauseAI*, https://pauseai.info/pdoom.
78 Bengio et al, "International Scientific Report on the Safety of Advanced AI: Interim Report. International Scientific Report on the Safety of Advanced AI" p. 66.
79 Pax for Peace, "*State of AI: Artificial intelligence, the military and increasingly autonomous weapons*", 2020, p.4.
80 "List of p(doom) value", n.d. *PauseAI*, https://pauseai.info/pdoom.
81 Core views on AI Safety: When, Why, What and How?". 8 March 2023, *Anthropic*.https://www.anthropic.com/news/core-views-on-ai-safety
82 Baum, Seth D. "Assessing the Risk of Takeover Catastrophe from Large Language Models." *Risk Analysis*, 2024. doi:10.1111/risa.14353.
83 Kevin Roose, "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn". *The New York Times*. 30 May 2023. https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html.
84 Betty Coffman, "Q&A: UofL AI safety expert says artificial superintelligence could harm humanity", *UofL News*, 15 July 2024, https://www.uoflnews.com/section/science-and-tech/qa-uofl-ai-safety-expert-says-artificial-superintelligence-could-harm-humanity/.
85 Anderljung, Markus, Joslyn Barnhart, Anton Korinek, Jade Leung, Cullen O'Keefe, Jess Whittlestone, Shahar Avin, et al. "*Frontier AI Regulation: Managing Emerging Risks to Public Safety,*" 2023. doi:10.48550/arxiv.2307.03718.
86 Core views on AI Safety: When, Why, What and How?". 8 March 2023, *Anthropic*. https://www.anthropic.com/news/core-views-on-ai-safety.
87 Betty Coffman, "Q&A".
88 Anderljung et al., "Frontier AI Regulation".
89 Billy Perrigo, "AI Is as Risky as Pandemics and Nuclear War, Top CEOs Say, Urging Global Cooperation", *TIME*, 30 May 2023, https://time.com/6283386/ai-risk-openai-deepmind-letter/; Kevin Roose, "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn". *The New York Times*. 30 May 2023. https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html.
90 Zachary Kallenborn, ''Policy makers should plan for superintelligent AI, even if it never happens,'' Bulletin of the Atomic Scientists, December 21, 2023, https://thebulletin.org/2023/12/policy-makers-should-plan-for-superintelligent-ai-even-if-it-never-happens/

## Recommended Readings

Baele, S., and Brace, L. (2024). "AI Extremism: Technology, Tactics, Actors," *Vox-Pol*. https://voxpol.eu/new-vox-pol-report-ai-extremism-technologies-tactics-actors/

Molas, B., and Lopes, H. (2024). "Say it's only fictional": How the Far-right is Jailbreaking AI and What Can Be Done About It. *International Centre for Counter-Terrorism*. https://www.icct.nl/sites/default/files/2024-10/Molas%20and%20Lopes.pdf

Wells, D. (2024). "The next paradigm-shattering threat? Right-sizing the potential impacts of generative AI on terrorism." *The Middle East Institute*. https://www.mei.edu/publications/next-paradigm-shattering-threat-right-sizing-potential-impacts-generative-ai-terrorism

## For Further Information

Antinori, A. (2019). Terrorism and deepfake: From hybrid warfare to post-truth warfare in a hybrid world. In *ECIAIR 2019 European Conference on the Impact of Artificial Intelligence and Robotics* (p. 23). Academic Conferences and publishing limited.

Bazarkina, D. (2023). Current and future threats of the malicious use of artificial intelligence by terrorists: Psychological aspects. In *Springer eBooks* (pp. 251–272). https://doi.org/10.1007/978-3-031-22552-9_10

Blanchard, A., and Hall, J. K. C. (2023). Terrorism and Autonomous Weapon Systems: Future Threat or Science Fiction? *CETaS Expert Analysis (June 2023)*.

Esmailzadeh, Y., and Motaghi, E. (2024). International Terrorism and Social Threats of Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4893952

Fernandez, M., and Alani, H. (2021). Artificial intelligence and online extremism. In *Routledge eBooks* (pp. 132–162). https://doi.org/10.4324/9780429265365-7

Haas, M. C., and Fischer, S. C. (2020). The evolution of targeted killing practices: Autonomous weapons, future conflict, and the international order. In *The Transformation of Targeted Killing and International Order* (pp. 107-132). Routledge.

Hussain, G. (2025). Artificial Intelligence (AI) and Radicalization. In *Cyber Security in the Age of Artificial Intelligence and Autonomous Weapons* (pp. 58-69). CRC Press.

Ibrahim, A., and Shuja, S. F. (2024). Artificial Intelligence led Lethal Autonomous Weapon Systems and Terrorism. *CISS Insight Journal*, *12*(1), P24-55.

Lakomy, M. (2023). Artificial intelligence as a terrorism enabler? Understanding the potential impact of chatbots and image generators on online terrorist activities. *Studies in Conflict and Terrorism*, 1–21. https://doi.org/10.1080/1057610x.2023.2259195

Lohmann, S. (2024). *ChatGPT, Artificial Intelligence, and the Terrorist Toolbox. An American Perspective*, 23.

Nelu, C. (2024). Exploitation of generative AI by terrorist groups. *International Centre for Counter-Terrorism.* https://www.icct.nl/publication/exploitation-generative-ai-terrorist-groups

Rickli, J. M., and Liang, C. (2024). "New and Emerging Technologies for Terrorists" in *The Routledge Companion to Terrorism Studies: New Perspectives and Topics*, eds Max Abrahms.

Tech Against Terrorism. (2024). Mapping far-right terrorist propaganda online. https://techagainstterrorism.org/hubfs/TCAP_Report_Mapping_Far-right_Terrorist_Propaganda_Online%20(1).pdf?hsCtaTracking=0020654a-aaca-4ccb-850e-772879bfe6ee%7Ce611b3e7-6423-44dd-859e-088bd4d5c535

# About the Authors

## Joana Cook

Joana Cook is a Senior Project Manager at ICCT and Editor-in-Chief of the ICCT journal. She is also an Assistant Professor of Terrorism and Political Violence in the Faculty of Governance and Global Affairs, Leiden University. Her research more broadly focuses on women and gender in violent extremism, countering violent extremism, and counter-terrorism practices. More recent scholarly interests include non-state actor governance, and factors and pathways to radicalisation.

Joana is also a Research Affiliate at the Department of War Studies, King's College London and an Associate Fellow at the International Centre for the Study of Radicalization; an adjunct lecturer at Johns Hopkins University; a non-resident Fellow at the Program on Extremism at George Washington University; a Research Affiliate with the Canadian Network for Research on Terrorism, Security and Society (TSAS); and a Digital Fellow at the Montreal Institute for Genocide and Human Rights Studies (MIGS), Concordia University. She is a graduate of King's College London where she completed her MA and PhD in the Department of War Studies (BA University of Regina). She has presented her research to senior government and security audiences in a number of countries, and at institutions such as the UN Security Council, NATO, the Parliamentary Assembly of the Council of Europe, and the Counter-Daesh Communications Cell, amongst others. She has also been featured in media such as Time, the Telegraph, the Huffington Post, the Washington Post, the New York Times and on BBC World News, CNN, Sky News, BBC Radio, the National Post and CBC. In May 2019 she did her first TEDx talk on women in security. She holds a BA in Political Science from the University of Regina, an MA in Conflict, Security and Development, and PhD in War Studies (both from King's College London).

## Graig Klein

Dr. Graig R. Klein is an Assistant Professor in the Institute of Security and Global Affairs (ISGA) at Leiden University. His research focuses on the strategic use of political violence by non-state actors and governments. He leads the European Union funded research project Terrorist Group Adaptation & Lessons for Counterterrorism (ERC, TERGAP, 101116436), which offers insights into terrorists' uses of violence, attack target selection, and tactical and strategic decision-making in response to counterterrorism. At the core of his research is investigating and analysing complex socio-political phenomenon and threats of violence using a variety of statistical and big-data analytical tools. Previously, under a grant funded by the US Office of the Director of National Intelligence, he was a co-investigator creating the FOCUSdata Project, a database of over 1.5 million Russian, Iranian, Chinese, and North Korean foreign ministry statements and state-controlled newspapers articles from 2004-2020, to study dis-information practices and reactions to adversaries' policies and behaviors.

He has been an Academic Principal Investigator at the World Bank, collaborated on urgent international security problems at NORAD/US Northern Command, and presented his research at the 2024 NATO Center of Excellence – Defence Against Terrorism (COE-DAT)

Terrorism Experts Conference. His research is published in leading peer-review journals, in international media outlets, and by the International Centre for Counter-Terrorism. He holds a PhD in Political Science (specialising in International Relations) from Binghamton University (SUNY), a MA in International Peace & Conflict Resolution from American University, and a BA in Political Science from Binghamton University.

## Bàrbara Molas

Dr. Bàrbara Molas joined ICCT in August 2022 as a Research Fellow for the Current and Emerging Threats Programme. She received her PhD from York University (2021, Toronto) and is an expert on far-right ideology, online radicalisation, and prevention. She has an international consulting background having worked with intergovernmental organisations, national prosecution services, and Big Tech companies.

Molas is also the co-editor of Responses to the COVID-19 Pandemic by the Radical Right (Columbia University Press, 2020) and The Right and the Radical Right in the Americas (Rowman & Littlefield, 2021) as well as the author of Canadian Multiculturalism and the Far Right (Routledge, 2022). In an effort to educate non-expert audiences on emerging far-right threats, Molas has also collaborated with openDemocracy, Fair Observer, Rantt Media, the Globe and Mail, and the Global Network on Extremism & Technology (GNET). Given her expertise, she has been interviewed by several entities such as the New Statesman, TRT World, Globo News, Grid, de Volkskrant, and the Canadian Defence Association Institute.