

GIFCT Artificial Intelligence Working Group Report: AI Threats and Opportunities

GIFCT Year 5 Working Group

December 2025



GIFCT
Global Internet Forum
to Counter Terrorism

Table of Contents

Introducing GIFCT Year 5 Working Groups	3
Sectoral Breakdown of Working Group Participants	3
GIFCT Year 5 Working Group Topics	4
Artificial Intelligence: Threats and Opportunities	6
I. Main Threats & Challenges	7
II. Mitigation Strategies	10
III. Next Steps & GIFCT's Role	12
GIFCT Year 5 Working Group Participant Affiliations	14

Introducing GIFCT Year 5 Working Groups

In February 2025, GIFCT launched its Year 5 Working Groups to facilitate dialogue, foster understanding, and produce outputs to directly support our mission of preventing terrorists and violent extremists from exploiting digital platforms across a range of sectors, geographies, and disciplines. Started in 2020, GIFCT Working Groups contribute to growing our organizational capacity to deliver guidance and solutions to technology companies and practitioners working to counter terrorism and violent extremism, and offer multi-stakeholder perspectives on critical challenges and opportunities.¹

Overall, this year’s three thematic Working Groups convened 178 participants from 40 countries across 6 continents with 39% drawn from civil society (5% advocacy, 14% academia, and 20% practitioners), 23% representing governments, and 38% in tech.

Sectoral Breakdown of Working Group Participants

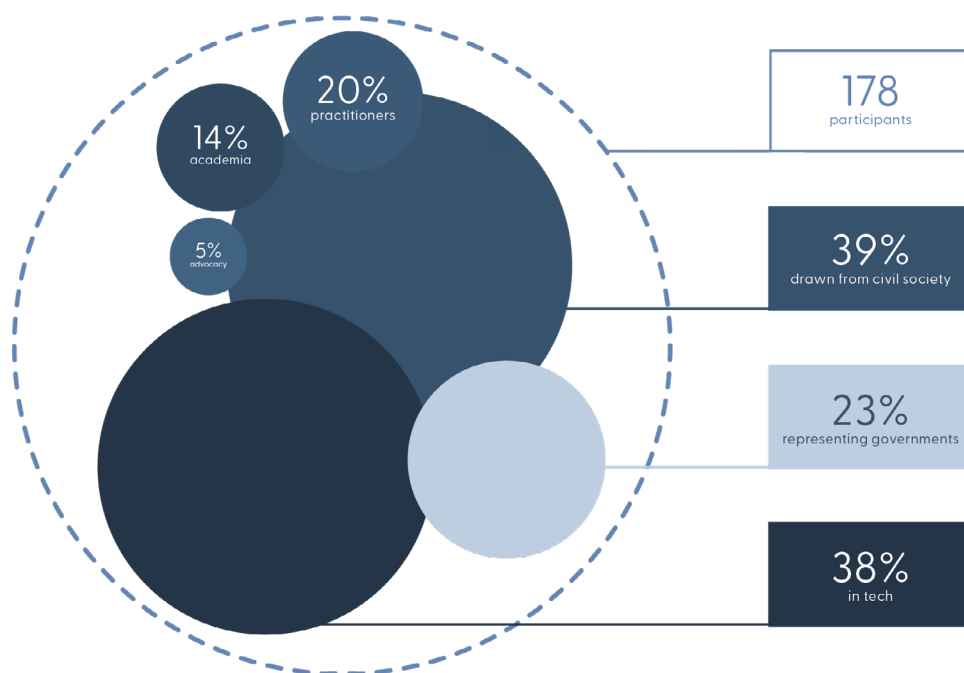


Figure 1: 2025 GIFCT Working Group participants by sector.

1. Working Group outputs are produced by independent experts and do not necessarily represent the views of GIFCT, its members, or the GIFCT Operating Board.

GIFCT Year 5 Working Group Topics

The 2025 GIFCT Working Groups focused on the following three themes:

Investigators Community of Practice

The Investigators Community of Practice (ICOP) brought together a network of investigation, analytic, incident response, and operational trust and safety (T&S) professionals from GIFCT member companies who met monthly throughout 2025. ICOP and its monthly meetings built off-of and iterate-on GIFCT's Working Group structure, and fostered a community of mutual learning between GIFCT and its members. Through ICOP, GIFCT created an ongoing and practical partnership with member T&S practitioner teams. Specifically, ICOP served as a destination for members and GIFCT staff to learn from one another, and brainstorm new GIFCT information sharing solutions. Periodically, ICOP leveraged external experts to provide substantive input for participants in the form of dedicated topic-oriented briefs.

Each ICOP session was focused on a challenge facing T&S operation teams and discussed: (1) threat landscape, (2) best practices, and (3) collective/GIFCT solutions. Sessions included a combination of member-company presentations, structured group discussions, and a GIFCT solutions focus group.

Artificial Intelligence: Threats and Opportunities

This Working Group consolidated and established actionable, cross-sector best practices and standards for AI safety products related to exploitation by terrorist and violent extremist (TVE) actors. Drawing from industry experience, the group mapped TVE threats, identified effective mitigation strategies, and analyzed overlaps in different companies' approaches to develop best practices tailored to specific product types. This effort was conducted in collaboration with government and civil society practitioners to incorporate diverse perspectives and ensure comprehensive, sector-wide impact.

Addressing Youth Radicalization and Mobilization

This Working Group focused on identifying the current trends in youth radicalization and mobilization, and identifying lessons learned from prevention and positive intervention strategies to address these dynamics. This group highlighted best practices while connecting industry, practitioners, and experts to enhance cross-sector efforts. Through a series of structured

multistakeholder dialogues, this group mapped evolutions in both the threats and responses, considering in particular how terrorists and violent extremists have targeted younger audiences online.

The group examined lessons learned from practice and programs, including positive interventions, counter-speech or “counter-narrative” work, and wider PCVE engagements, building from previous GIFCT Working Groups. Key findings and insights gleaned from the group, as well as the identification of relevant tools and resources, have helped equip practitioners in online safety efforts to build resilience in young online users and further positive intervention efforts.

Artificial Intelligence: Threats and Opportunities

Dr. Nagham El Karhili, Charley Gleeson, Skip Gilmour

Introduction

GIFCT's Artificial Intelligence Working Group (the AIWG) mapped terrorist and violent extremist (TVE) threats across Artificial Intelligence (AI) surfaces and consolidated actionable, cross-sector standards and best practices for mitigating the exploitation of AI systems by TVE actors. Acknowledging the persistent limitations and evolving vulnerabilities intrinsic to AI architectures—systems that by design continually change and thus are susceptible to new challenges—the Working Group incorporated these constraints into its recommendations. Drawing on industry expertise and consultations with government and civil society practitioners, the Working Group identified realistic threat scenarios, proposed mitigations, and articulated GIFCT's prospective supportive role across sectors.

This white paper examines the evolving intersection of AI² and terrorist and violent extremist content (TVEC) and activity.³ It synthesizes findings from the AIWG, which both mapped TVE threats across AI surfaces and consolidated mitigation strategies. The paper provides insights into the specific ways AI can be exploited by threat actors and proposes practical interventions. Its goal is to support AI developers, policy makers, researchers, and civil society in understanding and responding to emerging threats while maintaining ethical and human-rights-aligned approaches.

To guide its research, the AIWG posed the following research questions:

1. What are the main TVE threats when it comes to AI products and surfaces?
2. How have these threats been addressed, and which issues remain unresolved?
3. What guardrails (if any) should be built into AI models to ensure safe use?
4. What collaborations have AI developers engaged in with other entities—government, private sector, or civil society—to address TVE threats?
5. How does AI fit into or complicate compliance with existing governance frameworks?



The subsequent sections of this paper address these questions through three approaches:

-  Mapping the main challenges and threats posed by AI to the TVE ecosystem, as currently observed by Working Group participants;

.....

2. Here we use Wang's definition of "Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources." For more information, please see: Pei Wang, "On Defining Artificial Intelligence," *Journal of Artificial General Intelligence* 10, no. 2 (2019): 1-37, doi: [10.2478/jagi-2019-0002](https://doi.org/10.2478/jagi-2019-0002).

3. TVE activity refers to online actions that may not result in or produce content, but must be analyzed via other user signals, such as interactions, financing, direct communications, gameplay, etc.

-  Detailing mitigation strategies at both the model and collaborative levels;
-  Identifying next steps to proactively manage AI-enabled TVE threats.

By grounding the discussion in concrete cases, signals, and practical mitigations, this paper aims to translate abstract concerns about AI misuse into actionable insights for multiple stakeholders.

I. Main Threats & Challenges

TVE Exploitation of AI

The integration of AI into online ecosystems has created new avenues for TVEC and activity that GIFCT has been following closely.⁴ AI's capacity to generate, personalize, and distribute content at scale presents challenges that span technical, operational, and societal dimensions. This section details the realistic threats posed by TVE exploitation of AI technologies and aims to clarify discussions of the dangers, ensuring there is no inflation or downplaying of the evolving threat landscape.

AI-Enhanced Attack Planning and Operational Facilitation (Mobilization)

Threat actors increasingly leverage AI to support attack planning and operational activities. Recent incidents, including attacks in New Orleans,⁵ Las Vegas,⁶ Palm Springs,⁷ and Pirkkala,⁸ reportedly involved AI in planning or ideation processes. Additional foiled plots and court cases, such as a federal case alleging AI-assisted chemical research, illustrate emerging risks.

Priority threat activities include the use of AI to retrieve attack-relevant knowledge on weapons or tactics (1) and assisting attack planning and establishing online infrastructure for cyber attacks (3). While using AI systems to plan for Chemical, Biological, Radiological, and Nuclear (CBRN)-focused attacks is less likely, the impact of such activity is higher than other similar threats (2). AI-enhanced extended reality (XR) tools and simulations are also emerging, enabling attackers to rehearse or ideate novel tactics (4).⁹ Although AI is unlikely to replace all steps in mobilization

.....

4. For more information on our latest report please see: https://gifct.org/wp-content/uploads/2025/04/GIFCT-25WG-0425-AI_Report-Web-1.1.pdf

5. "New Orleans attacker wore Meta smart glasses - what else do we know?" BBC News, 2025, <https://www.bbc.com/news/articles/c205ek63433o>.

6. "Tesla Cybertruck bomber used ChatGPT to plan Las Vegas attack, police say," CBS News, 2025, <https://www.cbsnews.com/news/las-vegas-cybertruck-explosion-fire-chatgpt-plan/>.

7. "FBI says Palm Springs bombing suspects used AI chat program to help plan attack," CNBC, 2025, <https://www.cnbc.com/2025/06/04/fbi-palm-springs-bombing-ai-chat.html>.

8. Luke Baumgartner, "AI at the Centre: Violent Extremist Exploitation in Pirkkala," July 14, 2025, <https://gnet-research.org/2025/07/14/ai-at-the-centre-violent-extremist-exploitation-in-pirkkala/>.

9. For more information, please see: Broderick McDonald et al., "Immersive Technologies," in *Routledge Handbook of Online Violent Extremism*, eds. Suraj Lakhani, Julian Droogan, Stuart Macdonald, and Lydia Khalil (London: Routledge, forthcoming 2025).

pipelines, it amplifies existing low-tech methods by accelerating content creation, grooming, and target research.

Violent Extremist Radicalization

Beyond operational support, AI has the ability to transform the creation and dissemination of TVEC, lowering the barrier to entry and reducing the need for technical skills. Generative models enable the rapid translation,¹⁰ mass production, and refinement of text, images, video, memes, and coded messages (5), often to evade existing moderation systems (6).¹¹ This capacity allows TVE actors to produce both overt and coded content efficiently, increasing volume, quality, and reach. Signals to monitor include surges in AI-modified content, increasing sophistication of TVEC, and cross-platform dissemination that may be aided by AI-facilitated dissemination networks (7).

AI can also facilitate radicalization when exploited by TVE actors, not only through content but also via interactive, anthropomorphic personas and generative conversational interaction (8). Chatbots and AI companions can mimic TVE actors, building emotional attachment, entrenching dependency, and isolating vulnerable individuals. Memory and personalization features amplify harmful viewpoints over time, creating reinforcement loops that normalize violence. Vulnerable populations, including those with lower digital literacy and high exposure to AI tools, are particularly susceptible to these dynamics.

AI-Supported Recruitment

While a less prevalent threat than mobilization and radicalization activities, TVE exploitation of AI technologies also extends to recruitment efforts. These efforts focus on network and audience analyses to pinpoint potentially vulnerable individuals who may be susceptible to TVE messaging (9).¹² This analysis often feed into more consistent radicalization tactics, including propaganda personalization and AI-assisted outreach to identified audiences.

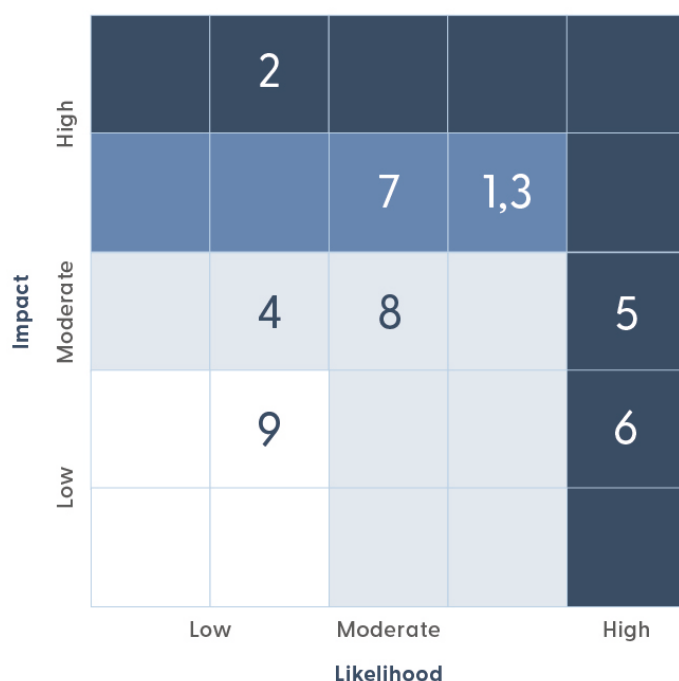
The threat matrix in Figure 2 summarizes the likelihood and impact of the potential threats.

.....

10. For more information, please see: Alessandro Bolpagni and Eleonora Ristuccia, "AI-powered Translation: How AI Tools Could Shape a New Frontier of IS Propaganda Dissemination," July 9, 2025, <https://gnet-research.org/2025/07/09/ai-powered-translation-how-ai-tools-could-shape-a-new-frontier-of-is-propaganda-dissemination/>.

11. For more information, please see: Broderick McDonald, "Extremists are Seeping Back into the Mainstream: Algorithmic Detection and Evasion Tactics on Social Media Platforms," October 31, 2022, <https://gnet-research.org/2022/10/31/extremists-are-seeping-back-into-the-mainstream-algorithmic-detection-and-evasion-tactics-on-social-media-platforms/>.

12. For more information, please see: Fabrizio Minniti, "Automated Recruitment: Artificial Intelligence, ISKP, and Extremist Radicalisation," April 11, 2025, <https://gnet-research.org/2025/04/11/automated-recruitment-artificial-intelligence-iskp-and-extremist-radicalisation/> and Mariam Shah, "The Digital Weaponry of Radicalisation: AI and the Recruitment Nexus," July 4, 2024, <https://gnet-research.org/2024/07/04/the-digital-weaponry-of-radicalisation-ai-and-the-recruitment-nexus/>.



Mobilization

1. Attack planning for attacks e.g. target selection, weapons acquisition
2. Attack planning (CBRN specific weapons acquisition)
3. Attack planning for cyber attacks and activity e.g. creation of jailbreak code
4. Use of XR technologies to rehearse / plan office attacks

Radicalization

5. Propaganda creation e.g. imagery, videos, manifesto
6. Propaganda manipulation for purposes of moderation evasion
7. Propaganda distribution e.g. targeted networks, content uploads
8. Use of TVE extremist personas

Recruitment

9. AI to identify targets for recruitment e.g. propaganda individualization

Figure 2: Threat matrix mapping of TVE exploitation of AI technologies, focusing on mobilization, radicalization, and recruitment.¹³

Factors Increasing the Threat of TVE Exploitation of AI Technologies

Threats are rarely confined to a single platform and frequently span multiple services. Combined with evolving and mixed ideologies that incorporate elements of cybercrime, child exploitation, or nihilistic violence, these risks become increasingly complex. Mapping user journeys across platforms is critical to understanding these dynamics, but privacy considerations, as well as platform limitations, complicate such monitoring. Regulatory asymmetries—where some jurisdictions impose strict AI controls while others remain permissive—further exacerbate risk by providing “safe havens” for malicious TVE actors.

The emergence of agentic AI systems with autonomous capabilities compounds these challenges, introducing new potential threat vectors that require complex governance considerations. Risks also extend to exclusion and bias: safety tooling may fail to account for marginalized or vulnerable populations, amplifying the disproportionate impact of AI misuse. Addressing these vulnerabilities

13. The scoring mechanism was consolidated by GIFCT from Working Group insights and aimed at mapping the relative likelihood and impact of each threat type. The mechanism was based on subject-matter expertise, evidence of use cases from completed and foiled attacks, and Working Group discussions of specific threat types.

in alignment with global human rights frameworks, such as the UN Guiding Principles¹⁴ and EU AI Act¹⁵ risk classifications, ensures that mitigation strategies do not inadvertently stigmatize communities or over-police speech.

A central challenge is that many companies struggle to anticipate or keep pace with rapidly evolving threats, leaving gaps in their defenses for TVE actors to exploit. Misuse often begins with subtle signals—such as jailbreaks¹⁶ or model chaining¹⁷—that appear benign in isolation but escalate into serious vulnerabilities when overlooked. Without systematic monitoring, these signals are easily missed.

This dynamic creates a systemic issue: TVE exploitation often emerges unevenly across platforms, regions, and languages, making detection inconsistent. Smaller AI startups, with limited resources, are particularly vulnerable. Without sustained monitoring, cross-platform intelligence, and mechanisms to identify early TVE-specific warning signs, companies will remain reactive rather than proactive. Malicious TVE actors capitalize on regulatory differences by targeting the most permissive or least-prepared environments for AI misuse.

II. Mitigation Strategies

The evolving landscape of AI presents both significant opportunities and emerging risks in countering TVEC and activity. While a variety of mitigation strategies currently exist, evidence remains limited on whether these measures are sufficiently deployed or over-applied to new AI systems. Developers and platform operators face complex trade-offs: prioritizing privacy for first-party AI use can limit monitoring capabilities, reducing model refusals may increase exposure to determined bad actors, and behaviors that initially appear benign can later prove risky.

Figure 3 broadly outlines current mitigation strategies at both the model (individual AI companies) and collaborative (multi-stakeholder) levels. Because AI systems evolve continuously, effective mitigation requires iterative refinement, with insights from deployment informing product-, model-, and collaboration-level improvements and vice versa. Please note that this figure is not exhaustive, rather, it is meant to give a snapshot of the current state of the sector.

.....

14. More information here: https://unsce.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

15. More information here: <https://artificialintelligenceact.eu/>

16. Defined by the Cambridge Dictionary as “the act of making changes to a piece of technology so you can use it in a way that was not intended by the company or person that produced it, and may not be allowed.” For more information, please see: <https://dictionary.cambridge.org/dictionary/english/jailbreak>.

17. Defined by Moveworks as “a technique in data science where multiple machine learning models are linked in a sequence to deliver a better output.” For more information, please see: Moveworks, “What is Model Chaining?” 2025, <https://www.moveworks.com/us/en/resources/ai-terms-glossary/model-chaining>.

	Product & Model Level			Collaborative Level
	Pre-launch	Hardware / technological	Content moderatio	
Description	<ul style="list-style-type: none"> Internal and external red-teaming¹⁸ Safety-by-design Harm refusal layers Fine-tuning of models based on TVE datasets 	<ul style="list-style-type: none"> Prevention of model modifications Encryption of safeguards TVE-specific detection tools Logic locks 	<ul style="list-style-type: none"> Detection and removal of violative outputs User reporting and remediation Alignment with community standards/terms of service Multilingual moderation Prompt filtering 	<ul style="list-style-type: none"> Multi-stakeholder red-teaming External expert reviews Continuous risk assessments Engagement and partnerships with law enforcement and civil society Trusted flagger reporting mechanisms
Examples	<ul style="list-style-type: none"> MITRE Attack Framework¹⁹ Agent2Agent Security²⁰ 	<ul style="list-style-type: none"> Microsoft Prompt Shield²¹ Microsoft Comms Compliance custom policies²² ML/stats-based behavioral anomaly detection in chatbot interaction logs²³ 	<ul style="list-style-type: none"> EU AI Act alignment²⁴ Temporary AI Code of Conduct²⁵ 	<ul style="list-style-type: none"> GPH²⁶ and the AU Polarization Lab²⁷ external expert reviews
Limitations and gaps	<ul style="list-style-type: none"> Red-teaming quality varies drastically Requires complex and sensitive datasets Can be cost-intensive 	<ul style="list-style-type: none"> Hardware safeguards are typically untested at scale Cannot address all behavioral risks May be undesirable as constraining innovation 	<ul style="list-style-type: none"> Linguistic and regional gaps Risk of over-moderation Misclassification of free speech Small companies are often under-resourced 	<ul style="list-style-type: none"> Resource-intensive Requires sustained engagement Legal and privacy constraints and concerns Limited uptake from small/startup AI firms
	<ul style="list-style-type: none"> System-level risks persist when some models (including open-source models) lack safeguards and adequate mitigations Creation of TVE AI systems can undermine mitigations elsewhere Mitigation efforts can rarely prevent determined misuse 			

Figure 3: Mapping of current and future mitigation strategies for tackling TVE exploitation of AI technologies.

.....

18. For more information, see: Anthropic, "Strengthening our Safeguards Through Collaboration with US CAISI and UK AISI," September 12, 2025, <https://www.anthropic.com/news/strengthening-our-safeguards-through-collaboration-with-us-caisi-and-uk-aisi>; OpenAI, "Working with US CAISI and UK AISI to Build More Secure AI Systems," September 12, 2025, <https://openai.com/index/us-caisi-uk-aisi-ai-update/>; Daniel Fabian, "Google's AI Red Team: The Ethical Hackers Making AI Safer," July 19, 2023, <https://blog.google/technology/safety-security/googles-ai-red-team-the-ethical-hackers-making-ai-safer/>; Meta, "MART: Improving LLM Safety with Multi-round Automatic Red-Teaming," April 5, 2024, <https://ai.meta.com/research/publications/mart-improving-llm-safety-with-multi-round-automatic-red-teaming/>.

19. "ATT&CK," MITRE, 2025, <https://attack.mitre.org/>.

20. "A2AS: Agentic AI Security Framework," A2AS.org, 2025, <https://www.a2as.org/>.

21. "Prompt Shields," Microsoft Ignite, November 17-21, 2025, <https://learn.microsoft.com/en-us/azure/ai-services/content-safety/concepts/jailbreak-detection>.

22. "Create and Manage Communication Compliance Policies," Microsoft Ignite, November 17-21, 2025, <https://learn.microsoft.com/en-us/purview/communication-compliance-policies>.

23. Amardeep Kumar, Danish Ali Khan, and Ruhul Amin, "Ensuring Secure Conversational Commerce: Anomaly Detection in Chatbot Interactions," *Expert Systems with Applications* 295 (2025): 128769, doi: [10.1016/j.eswa.2025.128769](https://doi.org/10.1016/j.eswa.2025.128769).

24. Future of Life Institute, "The AI Act Explorer," EU Artificial Intelligence Act, 2025, <https://artificialintelligenceact.eu/ai-act-explorer/>.

25. "Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems," Government of Canada, September, 2023, <https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems>.

26. "Global Project Against Hate and Extremism," Global Project Against Hate and Extremism, 2025, <https://globalextrémism.org/>.





27. "Polarization & Extremism Research & Innovation Lab," American University, 2025, <https://perilresearch.com/>.

III. Next Steps & GIFCT's Role

As AI continues to evolve, so too do the methods and vectors for TVE exploitation. GIFCT recognizes that addressing these emerging threats requires cross-platform visibility, collaborative mitigation, and proactive engagement with both technical and policy stakeholders. This section outlines GIFCT's strategic next steps and the role it intends to play in supporting the safe and responsible use of AI within the digital ecosystem.

Building a Centralized Knowledge Hub [Near-term Goal]

GIFCT plans to establish a consolidated portal resource for GIFCT member companies to track TVE threat trends and facilitate collaborative information sharing. Key features of this knowledge hub will include:

-  An initial dataset tagged according to the Incident Response Framework (IRF), incorporating Member-specific workflows;
-  Continuous updates on research into the societal and behavioral impacts of AI;
-  Collaborative scenario building and testing around areas of concern to help drive consistent evaluation comparisons and mitigation strategies across model builders;
-  Sharing threat indicators and continued cross-sector intelligence sharing, especially in the context of incident response and logging.

By providing a centralized repository, GIFCT can provide tech companies with actionable intelligence on AI-enabled TVE risks while maintaining user privacy.

Mapping Cross-Platform User Journeys [Medium-term Goal]

As AI-enabled threats are inherently cross-platform, generative AI further exacerbates this dynamic, enabling mass production of localized or demographically targeted content across multiple micro-channels and surfaces.

GIFCT aims to develop a privacy-respecting system for mapping these cross-platform journeys, enabling the tracing of coordinated campaigns without compromising personally identifiable information. By identifying entry points, bridging nodes, and mapping cross-platform transitions, this system will enable partners to detect patterns, measure narrative velocity, and flag signals such as bot-like activity or suspicious fine-tuning events.

Near-term implementation will focus on pilot initiatives with GIFCT Members, with initial emphasis on propaganda and recruitment vectors. Privacy-preserving analytics and dashboard views will provide vetted partners and researchers with actionable insights while maintaining strict safeguards.

Policy Engagement, Research, and Global Capacity Building [Long-term Goal]

GIFCT will continue to serve as a bridge between technical expertise, policy frameworks, and research communities, driving initiatives that enhance global preparedness against AI-enabled threats. Key priorities include:

Policy alignment:

- 🌐 Mapping AI mitigation strategies to existing regulatory frameworks (e.g., European Union's Artificial Intelligence Act, National Institute of Standards and Technology's Risk Management Framework, and Department of Homeland Security's Safety and Security Guidelines for Critical Infrastructure Owners and Operators);
- 🌐 Supporting voluntary standards, tailoring member guidance to TVE-specific concerns to reduce duplication and strengthen compliance pathways.

Research and evidence base:

- 🌐 Commissioning studies on public perception thresholds for harmful content;
- 🌐 Evaluating the societal impacts of AI-enabled TVE narratives;
- 🌐 Promoting interdisciplinary research on human-AI interaction, chatbot-related harms, and online polarization

Global and regional capacity building:

- 🌐 Supporting underrepresented regions through investment in multilingual research, regionally informed red-teaming, safety tooling, and model evaluation, enabling local institutions to contribute meaningfully to AI threat mitigation.

Through these efforts, GIFCT aims to strengthen global collaboration, enhance preparedness for emerging AI-enabled threats, and provide actionable guidance that empowers stakeholders across sectors to anticipate, monitor, and mitigate TVE risks.

GIFCT Year 5 Working Group Participant Affiliations²⁸

Academia	Advocacy	Practitioner & Researcher	Government & Intergovernmental	Tech
Center for Cyber Strategy and Policy, University of Cincinnati	All Tech Is Human	ATCO	African Union Commission	Airbnb
Centre for Human Rights, University of Pretoria	Dignity in Difference	Centinel	African Union Counter Terrorism Centre	Anthropic
Centre for Land Warfare Studies, New Delhi	Global Center on Cooperative Security	Centre for Action and Prevention against Radicalization of Individuals	AI Security Institute, United Kingdom	Bitly
Georgia State University	Life After Hate	Centre for Trustworthy Technology	Birmingham City Council, United Kingdom	Clubhouse
Indonesia International Islamic University	Local Youth Corner Cameroon	Christchurch Call	Civipol, France	Dailymotion
International Policing and Public Protection Research Institute, Anglia Ruskin University	Middle East Institute	Control Risks	Classification Office, New Zealand	Discord
Istanbul University	Next Wave, The International Center for Children and Global Security	Diálogo Americas	Counter Terrorism Department, Pakistan	Dropbox
Leiden University	Parents for Peace	Digital Catpult	Counter-Terrorism Committee, United Nations	Giphy
Northwestern University	Republican Public Association	Extremism and Gaming Research Network	Department of Communities and Justice, Australia	Google
Oxford Internet Institute		FBI National Citizens Academy Alumni Association	Department of Health, Australia	Linkedin
Royal Holloway, University of London		InfraGard	Department of Home Affairs, Australia	Meta
Said Business School		Global Community Engagement and Resilience Fund	Department of Internal Affairs, New Zealand	Microsoft
Sapienza University of Rome		Hedayah	Department of State, United States of America	Nexi Group
Strathmore University		Infobae	eSafety Commissioner, Australia	Pan Africa Gaming Group
Swansea University		Institute for Security Studies Africa	European Commission	Patreon

28. This table highlights participants across all Year 5 Working Groups

Universidad Anahuac		Institute for Security Studies	European External Action Service	Pinterest
University at Albany, State University of New York		Institute for Strategic Dialogue	Federal Ministry of the Interior, Germany	Somtel
University of Auckland		Irregular Warfare Initiative	Global Initiative on Electronic Evidence, United Nations	Streamable
University of Bristol		Jihadoscope	Hellenic Police, Greece	TikTok
University of Cambridge		Moonshot	Home Office, United Kingdom	Twitch
University of Muenster		Mythos Labs	Interministerial Committee for the Prevention of Crime and Radicalisation, France	Vimeo
University of Oxford		Online Safety Exchange	Ministry of Foreign Affairs	X
		Peace Research Institute Frankfurt	National Counter Terrorism Agency, Indonesia	Yahoo
		PVE Works	National Counter Terrorism Fusion Centre, Ministry of National Security, Ghana	Yubo
		Refslund Analytics	National Counterterrorism Center, Office of the Director of National Intelligence, United States of America	
		Tiaki Akoako	National Defence University, Pakistan	
		Uppsala Conflict Data Programme	Ofcom, United Kingdom	
		VoxPol Institute	Organization for Security and Co-operation in Europe	
		International Development Research Center	Public Safety Canada	
			Strong Cities Network	
			United Nations Counter-Terrorism Committee Executive Directorate	
			United Nations Interregional Crime and Justice Research Institute	
			United Nations Office of Counter-Terrorism	
			United Nations Office on Drugs and Crime	

Figure 4: Working Group affiliations highlighting participants across all Year 5 Working Groups

Copyright © Global Internet Forum to Counter Terrorism 2025

Recommended citation: GIFCT Artificial Intelligence Working Group Report: AI Threats and Opportunities. Washington, D.C.: Global Internet Forum to Counter Terrorism, 2025.

GIFCT is a 501(c)(3) non-profit organization and tech-led initiative with over 35 member tech companies offering unique settings for diverse stakeholders to identify and solve the most complex global challenges at the intersection of terrorism and technology. GIFCT's mission is to prevent terrorists and violent extremists (TVE) from exploiting digital platforms. In every aspect of our work, we aim to be transparent, inclusive, and respectful of the fundamental and universal human rights that TVE seek to undermine.



www.gifct.org



outreach@gifct.org