Violent Extremists' Use of Generative Artificial Intelligence

Generative artificial intelligence (GenAI) is a rapidly evolving technology that can produce content to include text, audio, video, and images when prompted by a user. Violent extremists and other illicit threat actors—including but not limited to insiders who pose threats and malicious cyber actors—seek inventive ways to exploit existing and emerging technologies, including GenAI, to support a range of violent extremist tactics and other criminal activities. Violent extremists may use GenAI to recruit and radicalize others to violence, disseminate and amplify violent extremist messaging through translation services, text-to-speech, and voice cloning, evade detection of banned content, and plan or train for operations through code generation, instructional chatbots, and cyberattacks.

Foreign terrorist organizations and their supporters have demonstrated interest in GenAI capabilities and explored its potential uses, especially as the technology has become increasingly accessible. As of April 2024, the primary malicious adoption has been to amplify and spread content creation. In support of these efforts, terrorist groups and supporters have released "How To" guides on using large language models, a type of AI that can recognize and generate text. In February 2024, an al-Qai'da-supported group launched a workshop designed to help develop skills using AI and other associated software. In August 2023, an ISIS-aligned tech support group shared an Arabic-language guide with tips to protect data and maintain privacy when using an AI content generator.

This image was generated using AI and is for awareness purposes only.

Notice indistinguishable text in the background

Look for inconsistencies in context, e.g., a missing driver in a vehicle, a car driving on a sidewalk

SCOPE: This product provides insight for public safety officials about possible violent extremists' malicious use of generative artificial intelligence (GenAI) and is intended to highlight potential planning, policy, training, and investigative considerations. This product is not a response to a specific threat against the United States. It provides information about, considerations for, and additional resources related to international terrorism threats and/or threats resulting from violent extremist tactics, techniques, and procedures. In this product, NCTC uses the term "violent extremists" to refer to foreign violent extremists and those US-based violent extremists who are directed, enabled, inspired by, or who otherwise affiliate or collaborate with foreign violent extremists.

GenAI Content Violent Extremists Could Exploit

IMAGES: May be used to spread false or misleading images to alter public perception of facts, disseminate violent extremist media or messaging, and support false narratives.

Check the title, description, comments, and tags. Check for watermark indicating the image was generated using AI through an online search platform. Search for the image using a reverse image search or AI image detector to determine its source. Examine the image for visual distortions that may show inconsistencies like stray pixels, misplaced shapes, and/or random artifacts. Specific examples may include: Change of skin tone near the edge of the face; Box-like shapes and cropped effects around the mouth, eyes, and neck; Look for items that do not match or asymmetry in areas that should have a uniform appearance i.e., spacing of text is inconsistent, too many limbs or teeth.

AUDIO AND VOICE CLONING: May be used to impersonate humans and gain unauthorized access to sensitive information, spread disinformation[a], and/or convince victim(s) to take specific actions based on false narratives.

Consider inconsistent inflection in tone of speech and/or uneven or broken sentences. Assess phrasing to determine if the speaker logically would express themselves the same way. Use contextual clues like background sounds to determine if they are consistent with the speaker's presumed location. Use the context of the message to highlight if it is relevant to recent discussions.

TEXT: May be used to enhance, amplify, and try to legitimize violent extremist messaging with grammatically correct language in multiple languages for dissemination to a global audience. Assess text for misspellings, inconsistent grammar, and/or lack of flow or varying tone in sentences that may indicate it was written by a human; conversely, GenAI tools may use the same phrases or patterns and generic content.

Consider if the source is from a known or verified number, email address, or social media account. Use the context of the message to determine if it sounds like something the sender would send and determine if it is related to a recent discussion.

VIDEO: May be used to create customizable AI-generated features that recite text-to-voice as well as text-to-video features. AI-enabled video alteration technology includes functions that enable face swapping or overlaying facial features on to those in already existing videos, possibly causing distortion and inconsistencies seen in AI-generated image content.

Review for lower-quality sections visible throughout the same video. Assess for changes or out-of-place shadows in the background and/or shadows that are inconsistent with scene lighting. Consider inconsistencies like lack of or irregular blinking or facial microexpressions.

NOTICE: This is a Joint Counterterrorism Assessment Team (JCAT) product. JCAT is a collaboration by NCTC, DHS, the FBI, and state, local, tribal, and territorial government personnel to improve information sharing and enhance public safety. The product promotes coordination among intergovernmental authorities and the private sector in identifying, preventing, and responding to terrorist activities. Consider the enclosed information within the context of existing laws, regulations, authorities, agreements, policies or procedures. For additional information contact us at JCAT@ODNI.GOV. **This document is best printed in 11 X 17.**

[a] Disinformation is the deliberate creation of false reports to mislead, harm, or manipulate a person, social group, organization, or country.

6 MAY 2024
AUTHORED BY NCTC, DHS, FBI

Violent Extremists' Use of Generative Artificial Intelligence *(continued)*

CONSIDERATIONS

As government and private sector organizations increasingly encounter GenAI used for violent extremist messaging, recruitment, planning, or other illicit criminal activities, the following considerations may provide guidance on policy, investigative, and privacy, civil rights, and civil liberty concerns.

Identify: Familiarity with violent extremists' tactics, techniques, and procedures can help first responders identify and mitigate contradictory information that was previously debunked through verified sources.

Document: Similar to evidence found at traditional crime scenes, evidence from online sources should be properly collected, documented, and maintained through chain-of-custody. Gathering possible evidence can assist law enforcement in identifying at-risk individuals, facilitate information sharing between public safety and private sector entities, and be used in court.

Assess: Determine the source of the information, its reliability and credibility, and the intent of the post or message. Establish the best path to respond to or mitigate any potential threat. Violent extremists may create repurposed content using GenAI tools to create illicit content designed to evade detection or system moderation**.**

Report: Reporting suspicious activity through appropriate mechanisms such as state and local fusion centers and the FBI's eGuardian system helps law enforcement officials triage emerging threat information.

POLICY AND PLANNING

- Adopt departmental policies and general guidelines that consider potential CT implications when conducting research in support of the investigative process, including on privacy, civil rights and civil liberties, redress, access to data, and accountability issues. Review and update policies on a recurring basis to account for evolving technologies.
- Train and develop a digitally literate workforce capable of identifying artificially created content through manual detection or automated processes, according to departmental policies. Seek to enhance critical thinking, assess source credibility to identify synthetic media, and employ technical strategies to protect digital ecosystems while protecting privacy, civil rights, and civil liberties.
- Remain aware that any information entered into GenAI tools may be used for future training on the system. Consideration should be given when using third-party or open-source tools for investigative or research purposes, particularly if sensitive or proprietary data is involved.

INVESTIGATIVE

- If trying to determine if content is GenAI, corroborate evidence and ensure a clear and transparent chain-of-custody for digital evidence to demonstrate that it is authentic and has not been altered.
- Use open-source research tools, including reverse image search, to authenticate potential artificially generated content through a third-party search platform or other open-source forensic tools that allow for media verification.
- Remain aware of the legal implications of and policy developments related to emerging technologies including GenAI at the federal and state levels.

INFORMATION SHARING AND COLLABORATION

- Build and maintain cross-sector relationships, including tech sector, academia, and civil society organizations to coordinate and share information and security practices. Adopt a multiagency approach to help identify, respond, and mitigate potential illicit use of GenAI technology.
- Emphasize awareness of the current threat environment to increase understanding of how potential GenAI violent extremist content can be identified and mitigated through training, analysis, and intelligence briefings.
- Consider outreach to the public on the potential use of GenAI by violent extremists to increase transparency, digital literacy, and resilience and counter violent extremist narratives.
- Establish relationships with local media and community leaders and build a team of trusted voices to amplify accurate information.

CYBERCRIME

- Because GenAI has the potential to enable larger-scale, faster, more efficient, and more evasive cyber attacks than traditional cybersecurity tools and capabilities are able to counter, encourage security features like multifactor authentication and Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) to provide an additional security feature for account access.
- Consider hardening personal and organizational social media accounts by applying the strongest security and privacy controls possible, deactivating or deleting profiles that are no longer in use, and removing any personally identifying information from social media profiles. Limit access to personal social media accounts so that violent extremists have less access to images and voice data to exploit and create malicious content.

RESOURCES

Presidential Executive Order 14110—Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 30 October 2023
https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf

US Department of Commerce, National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework, January 2023
https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

The White House, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, October 2022
https://www.whitehouse.gov/ostp/ai-bill-of-rights/

National Security Commission on Artificial Intelligence Final Report, February 2021
https://www.nscai.gov/

DHS Tools for Analysts and Investigators—Suspicious Activity Reporting Indicators and Behaviors
https://www.dhs.gov/publication/suspicious-activity-reporting-indicators-and-behaviors

DEFINITIONS

Deep Fakes: a type of AI-generated, highly realistic synthetic media

Dataset: a collection of data typically designed based on a specific topic source

Generative AI: a model that uses neural networks that teaches computers to identify the patterns and structures within existing data to generate new and original content

Large language model (LLM): a type of AI that uses massively large datasets to learn, predict, and generate new content

Machine Learning (ML): a subset of AI in which systems receive inputs in the form of training data and generate rules that produce outputs

Training: the method to learn parameter values until a model is effective at converting inputs to outputs

NOTICE: This is a Joint Counterterrorism Assessment Team (JCAT) product. JCAT is a collaboration by NCTC, DHS, the FBI, and state, local, tribal, and territorial government personnel to improve information sharing and enhance public safety. The product promotes coordination among intergovernmental authorities and the private sector in identifying, preventing, and responding to terrorist activities. Consider the enclosed information within the context of existing laws, regulations, authorities, agreements, policies or procedures. For additional information contact us at JCAT@ODNI.GOV. **This document is best printed in 11 X 17.**

6 MAY 2024
AUTHORED BY NCTC, DHS, FBI

JCAT

JOINT COUNTERTERRORISM ASSESSMENT TEAM

PRODUCT FEEDBACK

Please use the link below to complete a short survey. Your feedback will help JCAT develop counterterrorism products that support the public safety and private sector community.

https://www.JCAT-url.com

For further information, please email JCAT
jcat@odni.gov

(U) The Joint Counterterrorism Assessment Team (JCAT) is a collaboration by NCTC, DHS, FBI, state, local, tribal, and territorial government personnel to improve information sharing and enhance public safety. The First Responder's Toolbox is an ad hoc, unclassified reference aid intended to promote counterterrorism coordination among federal, state, local, tribal, and territorial government authorities and partnerships with private sector officials in deterring, preventing, disrupting, and responding to terrorist attacks.