Artificial Intelligence and Counterintelligence Concerns Short

Student Guide

May 2025

Center for Development of Security Excellence

Contents

Introduction

Sam receives an urgent text from her colleague. Sam is on vacation and the colleague is covering some of Sam's tasks. The text asks Sam to call them as soon as possible because they need to access something.

Rupert receives a request to connect on social media, from a woman he finds stunning. They start talking. She's sweet, kind, and really seems interested in him and his work.

Haley posts a DOD job online. She receives over 100 applications within a few hours. They *all* seem *so qualified!* It's like they all know exactly which phrases and keywords she's looking for.

In all of these cases, what looks like a normal interaction is actually Artificial Intelligence, or AI, posing a counterintelligence threat.

In this Short, you will identify the threat posed by AI, along with appropriate AI mitigation strategies.

Emerging Threat Landscape

We are at the dawn of the age of AI warfare. Over the past 3 decades, humanity has poured its collective knowledge onto the internet. While individual humans cannot possibly absorb and utilize all that information, AI can.

AI has become commonplace and is available via smart devices and operating systems, while the emergence of quantum computing will advance AI capabilities exponentially. This increases the number of foreign intelligence entities (FIEs) and adversaries using AI attempting to gain access to personnel, information, and facilities.

In this Short, we'll define AI and related terms and explore the threats AI poses. We'll also take a look at some ways it's being used now, including deepfakes, honeypot, and resume swarming. Finally, we'll give you ways to prevent and deter these types of threats.

Deepfake

AI-generated, highly realistic synthetic media that can be abused to threaten an organization's brand, impersonate leaders and financial officers, and enable access to networks, communications, and sensitive information.

Honeypot

A social engineering technique in which AI bots pose as attractive people (usually women), enticing the target to provide something of value, such as information, access, or money.

Resume Swarming

Use of AI to generate large numbers of resumes that are perfectly tailored for the job posting to deceive automated resume screening tools, enabling threat actors to infiltrate an organization by the use of deception.

AI and Counterintelligence Definitions

To better understand the threat posed by the use of AI, we need to first understand some basic terms.

AI is a machine-based system with human input that can make predictions, recommendations, or decisions. It can also influence real or virtual environments.

Counterintelligence, or CI, is the collection of information and activities that aim to protect against espionage, other intelligence activities, sabotage, and assassinations.

Those individuals who commit malicious acts are known as threat actors. This category includes Foreign Intelligence Entities (FIE), adversaries, and criminals.

Distinguishing AI Terminology

There are different applications of AI and machine and/or computer technology. Let's distinguish some of them now.

Recall that AI in general is a machine-based system with human input that can make influential predictions, recommendations, or decisions.

CGI stands for computer generated imagery (images or video) that requires a lot of manual human input. Adobe After Effects is an example of software that uses CGI.

Generative AI takes data and emulates it to "generate" synthetic content. That is, it extrapolates on the human input to create brand new content such as images, videos, audio, text, and other digital content. ChatGPT and Claude are examples of generative AI tools.

Finally, consider machine learning (ML). Machines or systems use AI to "learn" and *improve* from experience without explicit programming. IBM Watson and Microsoft Azure are examples of ML tools.

Cyber vs. Human Uses of AI

Cybersecurity can use AI for *defensive* advantage.

Defensive uses of AI include data and predictive analysis. AI in cybersecurity can enhance threat detection and response, adapt to become predictive, learn intrusion

methods and identify attacks before they're successful, and quickly spot malicious code and anomalous activity.

There are advantages for using AI offensively. Using AI offensively makes it difficult for defensive AI to employ its defenses. AI also learns from unsuccessful attempts; it is less likely to make the same mistakes twice. AI has learning and scraping capabilities that make social engineering far more effective. Finally, AI makes phishing approaches more difficult to detect by avoiding common red flags, such as using generic greetings, incorrect email addresses, or spelling and grammar errors.

As you have learned, the use of AI has changed the CI landscape, both for defensive and offensive purposes.

Offensive Uses of AI

Threat actors use offensive AI to gain identity access. Use of offensive AI also enhances tradecraft, or espionage techniques used for intelligence gathering and assessment, targeting individuals.

Threat actors target high-value individuals at organizations because they have access to trade secrets, financial systems, key strategies, and other sensitive and proprietary intellectual property.

Using AI, they can focus their attacks on highly specific targets to extract only the most useful information. Additionally, adversaries use AI to create and spread misinformation, disinformation, or deception.

Misinformation is false information that is spread, regardless of whether there is intent to mislead.

Disinformation is deliberately misleading or biased information; manipulated narrative or facts; and propaganda.

Finally, they may use offensive AI to enhance non-autonomous weapons. That is, weapons which require human input to function.

AI Tradecraft

Some examples of tradecraft related to AI include spear phishing, deepfakes, resume swarming, and honeypots. We'll take a look at some of these now.

Deepfakes

Deepfakes**,** an increasingly common method of attack, are highly realistic, AI-generated impersonations that pose a risk to both individuals and organizations. Using real-time voice and/or video impersonation, deepfake attacks may threaten an organization's

brand, impersonate leaders, financial officers, or even IT personnel with a goal of enabling access to networks, communications, sensitive information, or even to divert payments or funds.

Deepfake techniques include creating fake online and social media accounts. These fake accounts can then contact high value targets using social engineering.

Social engineering attempts by threat actors may include fraudulent text or voice messages or faked videos to avoid technical defenses. These can then be used to spread disinformation or otherwise pose a counterintelligence threat.

Personas of colleagues may be impersonated using video teleconferencing or phone calls to deploy real-time deepfake contact with voice and video synthesis.

Deepfake Scenario

As mentioned previously, Sam received an urgent text from her colleague. Sam is on vacation and the colleague is covering some of Sam's tasks. The text asks Sam to call them as soon as possible because they need to access something….

In this example, an AI bot searched the internet and built a deepfake profile based on one of Sam's contacts. The AI-enabled voice targeted Sam with active and passive pointers to deepfake the profile.

Next, an AI-enabled actor established telephonic contact to confirm the deepfake's individual identity and credentials. With this contact established, the AI-enabled threat actor was able to solicit or elicit specialized or secure information.

Honeypots

Honeypot is a social engineering technique in which AI bots pose as attractive people, usually women, enticing the target to provide something of value, such as information, access, or money.

Honeypot Scenario

As mentioned previously, Rupert received a request to connect on social media, from a woman he finds stunning. They start talking, and she provides details about hobbies and interests that she has in common with Rupert. She's sweet, kind, and really seems interested in him and his work….

To execute this honeypot scheme, an AI bot was programmed to search LinkedIn for key words, such as the name of the company or an asset, or key roles, such as engineer or developer.

The bot scraped available sources of personally identifiable information, or PII, and demographics, along with associated personal social media data. Using this information, the bot developed a profile for the "whole person" targeting package.

Then, it used social engineering tactics, techniques, and procedures, known as TTP, to create the ideal persona and its online presence. This online presence contacted the target and engaged in conversation, asking to move communications from LinkedIn to WhatsApp.

Finally, the bot used flirtatious and conversationally solicitous techniques until it reached a failure loop, at which point the conversation was handed off to a human to elicit information. In other scenarios, the end goal might be access or money. Note that most targeting that asks to meet in person is fraud or criminal, whereas most information theft targeting seeks to avoid in-person contact.

Resume Swarming

Resume swarming occurs when threat actors use AI to search keywords and qualifications from job postings and develop the "perfect" candidates.

AI can then generate hundreds or thousands of variations of highly qualified, but imaginary, candidates' resumes to apply for open positions at unsuspecting companies.

Alternately, this technique may generate and submit thousands of unqualified resumes along with "perfect" ones.

Resume Swarming Scenario

As mentioned previously, Haley posted a DOD job online. She received over 100 applications within a few hours. They *all* seemed *so qualified!* It's like they all know exactly which phrases and keywords she was looking for….

In this example, an AI bot conducted a keyword analysis on the critical technology job requisition listing by Haley's government agency. The bot generated and submitted dozens of "perfect," but fake, resumes, each tied to a real person, threat actors. Based on those "perfect" resumes, several fake candidates were selected for an interview.

But the AI threat didn't end there.

AI prepared a profile of each interviewer, pulled from the internet and social media, as well as interview preparation guidance. The interviewer profiles were provided to the threat-actor candidates so they could easily build rapport during the interview and answer questions successfully.

Backed by AI, the threat-actors were far more likely to receive an interview, do well during the interview, and to be hired by the government agency.

Once a threat actor became an agency employee, he had access to the agency, government information, information technology, networks, and more.

Mitigation Strategies

Now that we've discussed the strategies employed by threat actors, let's discuss the defense strategies we employ.

One strategy is to identify Critical Program Information (CPI) by asking, "What is the most sensitive and proprietary information exclusive to your organization?" In other words, are there any "trade secrets"?

To ensure you have identified all critical information, be sure to ask, "Where is the CPI contained?" For example, CPI may be held in facilities and devices; systems and networks; or the CPI may be Subject Matter Expert (SME) knowledge.

Your organization can also protect against the AI threat by conducting threat vulnerability assessments (TVA). These assessments consider business and risk prioritization; threat intelligence; impact analysis; existing solutions; and reducing the attack vector.

Additionally, your organization should conduct training and awareness campaigns early and often, with frequent opportunities for feedback and regular updates to address evolving technology.

Ensure personnel understand reporting procedures.

Finally, you can help mitigate the threat by reporting criminal and suspicious activity to the appropriate law enforcement organization in your region as well as to your security manager or facility security officer (FSO).

Review Activity 1

Danny, a data programmer for the DOD, receives a request to add Anna, an attractive woman, on LinkedIn. After adding Anna, she strikes up a conversation and asks to move the conversation to text message. She flirts with Danny and asks many questions about his work. What type of AI tradecraft is being employed?

Select the best response. Check your answers in the Answer Key at the end of this Student Guide.

- ○ Deepfake
- ○ Honeypot

- ○ Resume swarming

Review Activity 2

What are some mitigation strategies for this case?

Select all that apply. Check your answers in the Answer Key at the end of this Student Guide.

- ☐ Provide training and awareness to all personnel
- ☐ Only discuss personal information via a secure messaging app
- ☐ Report criminal and suspicious activity to the appropriate authority
- ☐ Ask for proof of identity before engaging in discussion with an unknown individual

Conclusion

Congratulations. You have completed the *Artificial Intelligence and Counterintelligence Concerns* Short. In this Short, you learned the definition of AI and related terms and explored the threats AI poses. You looked at some of the ways AI is used. And you learned ways to prevent and deter these types of threats.

You should now be able to identify the threat posed by AI, along with appropriate AI mitigation strategies.

Appendix A: Answer Key

Review Activity 1

Danny, a data programmer for the DOD, receives a request to add Anna, an attractive woman, on LinkedIn. After adding Anna, she strikes up a conversation and asks to move the conversation to text message. She flirts with Danny and asks many questions about his work. What type of AI tradecraft is being employed?

- ○ Deepfake
- ⦿ Honeypot (correct response)
- ○ Resume swarming

***Feedback**: In this instance, the adversary is employing the honeypot technique, in which AI bots pose as attractive people (often women) to entice the target to provide something of value.*

Review Activity 2

What are some mitigation strategies for this case?

- ☑ Provide training and awareness to all personnel (correct response)
- ☐ Only discuss personal information via a secure messaging app
- ☑ Report criminal and suspicious activity to the appropriate authority (correct response)
- ☐ Ask for proof of identity before engaging in discussion with an unknown individual

***Feedback**: An effective mitigation strategy for the organization is to provide training and awareness to all personnel. The individual who is targeted should report the suspicious activity to the appropriate authority.*