HC3: Analyst Note

January 17, 2023 TLP:CLEAR Report: 202301171500

Artificial Intelligence and Its Current Potential to Aid in Malware Development

Executive Summary

Artificial intelligence (AI) has now evolved to a point where it can be effectively used by threat actors to develop malware and phishing lures. While the use of AI is still very limited and requires a sophisticated user to make it effective, once this technology becomes more user-friendly, there will be a major paradigm shift in the development of malware. One of the key factors making AI particularly dangerous for the healthcare sector is the ability of a threat actor to use AI to easily and quickly customize attacks against the healthcare sector.

Report

Artificial Intelligence (AI) has most notably been applied to the defensive side of cybersecurity. It has been used to detect threats, vulnerabilities and active attacks, and to automate security tasks. However, because of its known defensive use and because threat actors in cyberspace are known to be highly creative and well-resourced, concerns have been raised in the cybersecurity community about the potential for artificial intelligence to be used for the development of malware.

One example of an AI-powered attack tool produced to study the malicious potential of such technologies is DeepLocker, a set of ultra-targeted and evasive attack tools driven by artificial intelligence. DeepLocker was developed to better understand how artificial intelligence models could be combined with existing malware techniques to create more potent attacks. In the case of DeepLocker, it analyzes the payload distribution lifecycle based on a deep neural network (DNN) AI model to look for appropriate "trigger conditions" in order to reach the intended target.

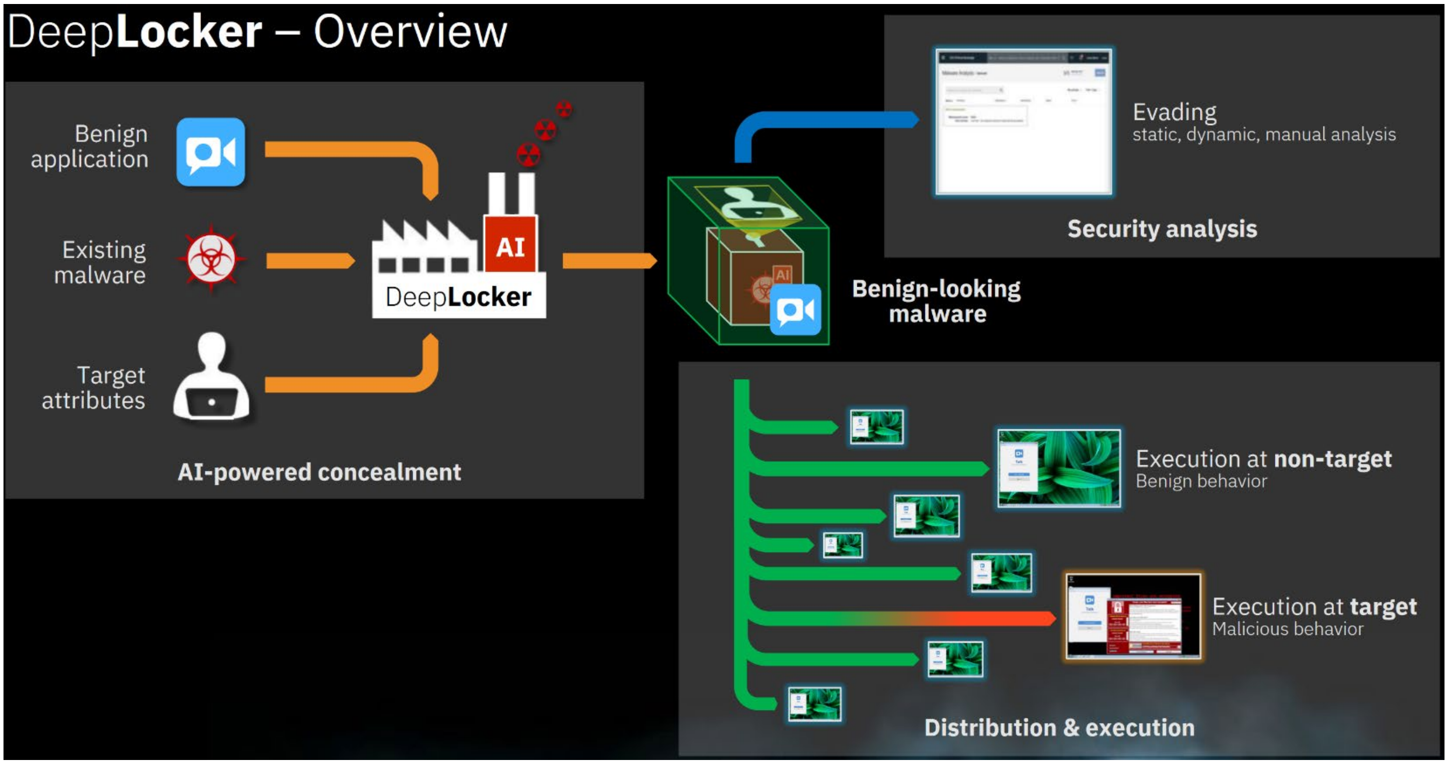

Figure 1: Overview model of IBM's DeepLocker AI malware capability. *(Image source: IBM)*

In November, the artificial intelligence research laboratory OpenAI publicly released a chatbot known as ChatGPT, which is based on its GPT-3.5 language model which was trained on Microsoft Azure. As a chatbot, it's designed to interact with humans and respond to their conversation and requests. Among other things, it can be used to answer questions, write essays, poetry or muisic, and compose e-mails and computer code. Its significant capabilities have raised serious concerns among the top IT companies as being potentially disruptive of existing markets. The platform enjoyed an immediate spike in popularity, garnering a million users in the first six days of its launch. The accessibility and capabilities of the tool as well as its popularity has prompted some prominent members of the cybersecurity community to further investigate how artificial intelligence might be used to develop malware. After testing it, it has been noted for being capable of crafting credible phishing e-mails. Attempts to leverage ChatGPT for malware development purposes have already been identified in the less than two months since it was released:

> "Threat actors with very low technical knowledge – up to zero tech knowledge -- could be able to create malicious tools [with ChatGPT]. It could also make the day-to-day operations of sophisticated cybercriminals much more efficient and easier – like creating different parts of the infection chain."
>
> -- Sergey Shykevich, Threat Intelligence Group Manager at Check Point

- On December 21, 2022 a threat actor posted a Python-based multi-layer encryption/decryption script on an underground hacking forum which he noted he created with the assistance of openAI code. The script had the potential to serve as ransomware and this user in particular has a history of illicit activities including selling access to compromised organizations and stolen databases.
- On December 29, 2022, an individual on an underground hacking forum noted that he was experimenting with ChatGPT to recreate various malware variants. He posted code for a Python-based information stealer that can identify, copy, compress and exfiltrate common file types.
- On December 31, 2022, a threat actor posted to a cybercriminal hacking forum under the discussion title, "Abusing ChatGPT to create Dark Web Marketplaces scripts."
- Researchers from the Universities of California and Virginia, as well as Microsoft, developed a poisoning attack that is capable if tricking code-suggesting engines such as OpenAI's ChatGPT and GitHub's Copilot into recommending malicious code as a response to innocuous requests.

Current artificial intelligence technologies are widely believed to only be at the very beginning of what will likely be a whole array of capabilities that will cut across industries and enter into people's private lives. The cybersecurity community is far from developing mitigations and defenses for such malicious code, and it remains unclear if there will ever be ways to specifically prevent AI-generated malware from being successfully used in attacks. There are already debates and discussions on the ethical use of AI systems and the proper governing models that should be deployed to ensure they are confined appropriately. Some of the resources that are exploring ethics as it applies to AI and appropriate governing models are as follows:

- The Brookings Institution has guidance for developing policy for securing AI decision-making systems. This includes preventing malicious influence from affecting the behavior and output of AI systems.
- The Harvard University Berkman Klein Center has a site dedicated to their research on Ethics and Governance of AI.
- The ETSI Industry Specification Group on Securing Artificial Intelligence is actively developing standards to address the security of AI technologies.

- The Massachusetts Institute of Technology has a site dedicated to Ethics and Governance of Artificial Intelligence.

References

ChatGPT shows promise of using AI to write malware
https://www.cyberscoop.com/chatgpt-ai-malware/

DeepLocker: How AI Can Power a Stealthy New Breed of Malware
https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/

Unleashing DeepLocker – AI Locksmithing
https://i.blackhat.com/us-18/Thu-August-9/us-18-Kirat-DeepLocker-Concealing-Targeted-Attacks-with-AI-Locksmithing.pdf

OPWNAI : Cybercriminals Starting to Use ChatGPT
https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/

Large language model expands natural language understanding, moves beyond English
https://venturebeat.com/ai/large-language-model-expands-natural-language-understanding-moves-beyond-english/

Evaluating Large Language Models Trained on Code
https://arxiv.org/pdf/2107.03374.pdf

OpenAI: ChatGPT
https://openai.com/blog/chatgpt/

Armed With ChatGPT, Cybercriminals Build Malware And Plot Fake Girl Bots
https://www.forbes.com/sites/thomasbrewster/2023/01/06/chatgpt-cybercriminal-malware-female-chatbots/

Artificial Intelligence, a new chapter for Cybersecurity?
https://www.tripwire.com/state-of-security/artificial-intelligence-new-chapter-cybersecurity

OpenAI's new ChatGPT bot: 10 dangerous things it's capable of
https://www.bleepingcomputer.com/news/technology/openais-new-chatgpt-bot-10-dangerous-things-its-capable-of/

OPWNAI : Cybercriminals Starting to Use ChatGPT
https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/

OpwnAI: AI That Can Save the Day or HACK it Away
https://research.checkpoint.com/2022/opwnai-ai-that-can-save-the-day-or-hack-it-away/

Armed With ChatGPT, Cybercriminals Build Malware And Plot Fake Girl Bots
https://www.forbes.com/sites/thomasbrewster/2023/01/06/chatgpt-cybercriminal-malware-female-

chatbots/

People are already trying to get ChatGPT to write malware
https://www.zdnet.com/article/people-are-already-trying-to-get-chatgpt-to-write-malware/

Contact Information

If you have any additional questions, we encourage you to contact us at HC3@hhs.gov.

We want to know how satisfied you are with the resources HC3 provides. Your answers will be anonymous, and we will use the responses to improve all future updates, features, and distributions. Share Your Feedback