TechXplore

FraudGPT and other malicious AIs are the new frontier of online threats. What can we do?

July 24 2024, by Bayu Anggorojati, Arif Perdana and Derry Wijaya

Credit: CC0 Public Domain

The internet, a vast and indispensable resource for modern society, has a darker side where malicious activities thrive.

From identity theft to sophisticated malware attacks, cyber criminals keep coming up with new scam methods.

Widely available generative artificial intelligence (AI) tools have now added a new layer of complexity to the cyber security landscape. Staying on top of your online security is more important than ever.

The rise of dark LLMs

One of the most sinister adaptations of current AI is the creation of "dark LLMs" (large language models).

These uncensored versions of everyday AI systems like ChatGPT are re-engineered for criminal activities. They operate without ethical constraints and with alarming precision and speed.

Cyber criminals deploy dark LLMs to automate and enhance phishing campaigns, create sophisticated malware and generate scam content.

To achieve this, they engage in LLM "jailbreaking"—using prompts to get the model to bypass its built-in safeguards and filters.

For instance, FraudGPT writes malicious code, creates phishing pages and generates undetectable malware. It offers tools for orchestrating diverse cybercrimes, from credit card fraud to digital impersonation.

FraudGPT is advertised on the dark web and the encrypted messaging app Telegram. Its creator openly markets its capabilities, emphasizing the model's criminal focus.

Another version, WormGPT, produces persuasive phishing emails that can trick even vigilant users. Based on the GPT-J model, WormGPT is also used for creating malware and launching "business email compromise" attacks—targeted phishing of specific organizations.

What can we do to protect ourselves?

Despite the looming threats, there is a silver lining. As the challenges have advanced, so have the ways we can defend against them.

AI-based threat detection tools can monitor malware and respond to cyber attacks more effectively. However, humans need to stay in the mix to keep an eye on how these tools respond, what actions they take, and whether there are vulnerabilities to fix.

You may have heard keeping your software up to date is crucial for security. It might feel like a chore, but it really is a critical defense strategy. Updates patch up the vulnerabilities that cyber criminals try to exploit.

Are your files and data regularly backed up? It's not just about preserving files in case of a system failure. Regular backups are a fundamental protection strategy. You can reclaim your digital life without caving to extortion if you are targeted by a ransomware attack—when criminals lock up your data and demand a ransom payment before they release it.

Cyber criminals who send phishing messages can leave clues like poor grammar, generic greetings, suspicious email addresses, overly urgent requests or suspicious links. Developing an eye for these signs is as essential as locking your door at night.

If you don't already use strong, unique passwords and multi-factor authentication, it's time to do so. This combination multiplies your security, making it dramatically more difficult for criminals to access your accounts.

What can we expect in the future?

Our online existence will continue to intertwine with emerging technologies like AI. We can expect more sophisticated cyber crime tools to emerge, too.

Malicious AI will enhance phishing, create sophisticated malware and improve data mining for targeted attacks. AI-driven hacking tools will become widely available and customizable.

In response, cyber security will have to adapt, too. We can expect automated threat hunting, quantum-resistant encryption, AI tools that help to preserve privacy, stricter regulations and international cooperation.

The role of government regulations

Stricter government regulations on AI are one way to counter these advanced threats. This would involve mandating the ethical development and deployment of AI technologies, ensuring they are equipped with robust security features and adhere to stringent standards.

In addition to tighter regulations, we also need to improve how organizations respond to cyber incidents and what mechanisms there are for mandatory reporting and public disclosure.

By requiring companies to promptly report cyber incidents, authorities can act swiftly. They can mobilize resources to address breaches before they escalate into major crises.

This proactive approach can significantly mitigate the impact of cyber attacks, preserving both public trust and corporate integrity.

Furthermore, cyber crime knows no borders. In the era of AI-powered cyber crime, international collaboration is essential. Effective global cooperation can streamline how authorities track and prosecute cyber criminals, creating a unified front against cyber threats.

As AI-powered malware proliferates, we're at a critical junction in the global tech journey: we need to balance innovation (new AI tools, new features, more data) with security and privacy.

Overall, it's best to be proactive about your own online security. That way you can stay one step ahead in the ever-evolving cyber battleground.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Provided by The Conversation

Citation: FraudGPT and other malicious AIs are the new frontier of online threats. What can we do? (2024, July 24) retrieved 24 July 2024 from https://techxplore.com/news/2024-07-fraudgpt-malicious-ais-frontier-online.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.