Picture credit: AdobeStock

A NEW CHAPTER

Danny Kadyshevitch *explains how Generative AI has democratised fraud and cyber crime*

AI and its transformative potential are on everyone's mind. The generative AI services that have been released in recent years have sparked a seemingly transformative wave of conversation about what this could mean for the very future of humankind. The release of ChatGPT in November 2022 took the world by storm, and even OpenAI – the creator of ChatGPT – is reportedly still amazed by the new uses that the public is finding for its flagship generative AI service. Amid the din of hype and activity, businesses are scrambling to find new ways to use AI to enhance their operations.

There's an important underlying philosophical point here: technology is a morally indiscriminate asset, and it can be used for good or ill. This is exactly what we've seen in the last year: cyber criminals are using AI for their own nefarious ends while cyber security innovators are now using it to stop them. This article covers some of the most exciting breakthroughs in security and fraud prevention, but first, it's critical to understand the dark side.

One researcher with no experience writing malware took just four hours to get ChatGPT to write them a working piece of malware

Many generative AI tools have principally lowered the bar to entry for fraudsters and cyber criminals. Anyone – even those with no technical background or coding experience – can craft sophisticated attacks. Generative AI can now teach how to plot account takeover attacks and automate the creation of professional-looking phishing websites and emails, malware with keylogging tools, and other sophisticated techniques designed to steal credentials and take over customer accounts – without ever writing a line of code.

These tactics can even come from perfectly legitimate chatbots such as OpenAI's ChatGPT – which despite its restrictions – can be gamed into instructing users on these malicious activities. Many of these requests come in the form of hypothetical requests and 'what-if' scenarios." An apparently innocent request for a convincing email template can drive a phishing attack or misinformation campaign that can circumvent traditional anti-fraud and anti-phishing measures, like those designed to spot mistakes, for example.

Generative AI can also be asked to automatically generate malicious code or customise advanced banking trojans or remote access trojans (RATs) to look like legitimate applications. Yes, trojans have been around for decades, but new AI capabilities have made it easier for amateur attackers to draw from advanced reconnaissance information and sophisticated technical abilities that would previously have been inaccessible.

Some note that the race to innovate with generative AI is actually enabling this kind of malicious use. A 2023 paper from researchers at the University of Mississippi notes that the success of services like ChatGPT has driven rapid improvement in the large language models (LLMs) and AI. As a result of competitive pressure, major search engines and service providers have integrated AI chatbots like Bing Chat. The paper notes this rush to market: "leads to gaps in safeguarding mechanisms due to reduced timelines for verification and validation procedures. Astute threat actors can use carefully designed prompts to circumnavigate safeguards on LLMs."

Similarly, the cyber crime and fraud ecosystems are creating their own generative AI tools, such as Fraud GPT and Evil GPT. Fraud GPT was discovered in July 2023, advertised on the Dark Web. This particular generative AI – which looks remarkably similar to ChatGPT – advertises that it can create malicious code, build undetectable malware, identify vulnerabilities and more. The aggregate effect of this innovation is that LLMs make it much easier for anyone to carry out fraud cyber crime.

Generative AI tools have also allowed experienced, professional attackers and crime rings to supercharge their capabilities. They can automate, scale and obfuscate their attacks using these tools. Automation allows organisations to do repeated, menial attacks quicker and generative AI tools are allowing threat actors to do the same.

The work of reconnaissance is made drastically easier too by simply automating the retrieval of open-source information about targets so that threat actors can spot weak points and identify social engineering opportunities. The same is true of code generation as many generative AIs can be used to generate the building blocks of malware, which might otherwise sap time and resources from malicious actors' efforts.

By automating much of the previously manual work of planning, reconnaissance and even malicious code generation, attackers are free to focus on the strategic elements of their campaigns, enabling them to scale their attacks and deploy them more quickly.

Generative AI is also enabling more advanced ways to circumvent security. For example, they can enable polymorphic tactics by generating synthetic code that can effectively obfuscate the signatures of their malware, preventing detection by static analysis. They're also enabling techniques, like steganography, which enables hackers to hide data and malicious code within otherwise innocuous content such as images.

SMART, CONTEXT-AWARE CYBER SECURITY SHOULD BE ABLE TO DETECT WELL-DISGUISED TROJANS

Earlier this year, security researchers proved exactly this when they tricked ChatGPT into helping write a malware tool. One of the researchers – who admitted they had no experience writing malware – said it took them four hours to get ChatGPT to write them a working piece of malware that could take files and insert them into images before sending them out. This tool, the researcher added, successfully evaded security controls and registered no detection on VirusTotal, a popular anti-virus software.

There are yet more innovations that the generative AI revolution is giving fraudsters through services like DeepSwap or Wombo. Deep fakes, voice authentication scams and other identity fraud techniques are already being employed by criminals to deceive more criminals and advance their scams.

Using generative AI services, fraudsters can effectively create synthetic faces and voices to defraud identity recognition systems, create fake identity documents such as ID cards and passports and otherwise impersonate their victims to perpetrate identity fraud. The use of AI in these cases raises the difficulties that many fraud detection systems urgently need to address.

While cyber criminals and fraudsters now find it easier to plan, scale and perpetrate advanced attacks, analysts find it harder and take longer to detect, prioritise and investigate cases as well as orchestrate and update protections.

Static detection and legacy fraud detection systems cannot keep up with the new wave of unknown threats. Traditionally, fraud engines are very good at spotting and stopping known fraud MOs and individual suspicious transactions in, say, an online banking account, but they have a harder time identifying more complex zero-day threats as well as larger patterns of fraud activity. Something needs to change.

Ultimately, organisations will also need to leverage AI to combat this new wave of threats. For example, authentication can be bolstered with

AI-driven fraud detection that includes passive forms of authentication, including privacy-age device fingerprinting and behavioural biometrics, which compare the customer's known devices and typical behaviours to hundreds of signals during the active session. The instant anomalies are detected, identity orchestration can automatically adapt the user journey, challenging the user with phishing-resistant MFA or identity verification, for instance. Conversely, if there's a high level of assurance that the user is indeed the customer, identity orchestration can remove step-up challenges or extend the session length – to improve the user experience (UX) and increase their engagement. With the right solution, pre-made and customisable decisioning rules can meet the company's security and UX requirements while eliminating the cost and complexity of building and maintaining decision logic.

CYBER CRIME AND FRAUD ECOSYSTEMS ARE CREATING THEIR OWN GENERATIVE AI TOOLS

Smart, context-aware cyber security is also able to detect well-disguised trojans and other malware designed to evade standard detection. More robust AI and ML detection models can now analyse event clusters and respond faster to new variants and zero-day malware. The best solutions evaluate signals within the full context of the application flows and up-to-date threat intelligence to identify infected app behaviour, including login overlays and other injection operations that insert malicious code, alter data or hijack app functions.

Similarly, AI and automation can be leveraged for easier, faster and more accurate identity verification that's able to spot fake IDs and deep fakes. In parallel, Transmit Security has developed generative AI models to analyse security events, detect evasive threats and improve analytics. Much like ChatGPT, this conversational AI tool delivers instant answers – in text or graphs – for key insights that will help fraud teams improve security and UX.

Generative AI marks a new chapter in the never ending escalation between attackers and defenders. Yet, too many organisations are still relying on their static detection models to catch AI-powered threats and fraud, which will push their defence and response capabilities to their limits. Defenders need to take advantage of these new developments and use a holistic AI-based approach that corresponds to a new landscape. Companies can now prevent today's artificial intelligence-powered threats and ATO fraud while simultaneously giving customers simple, low-friction experiences. Early adopters stand to win both the customers and revenue ●

Danny Kadyshevitch is a Director of Product Management for Fraud and Security products at Transmit Security. He brings over 15 years of experience in cyber security to his role, having previously built and led product management for the company's Passwordless and MFA Services. Prior to joining Transmit Security, Danny honed his expertise in the 8200 intelligence unit of the IDF and spent seven years in Microsoft's Cloud Security division.

Steganography enables hackers to hide data and malicious code within otherwise innocuous content such as images

Picture credit: Victor Soares/Shutterstock