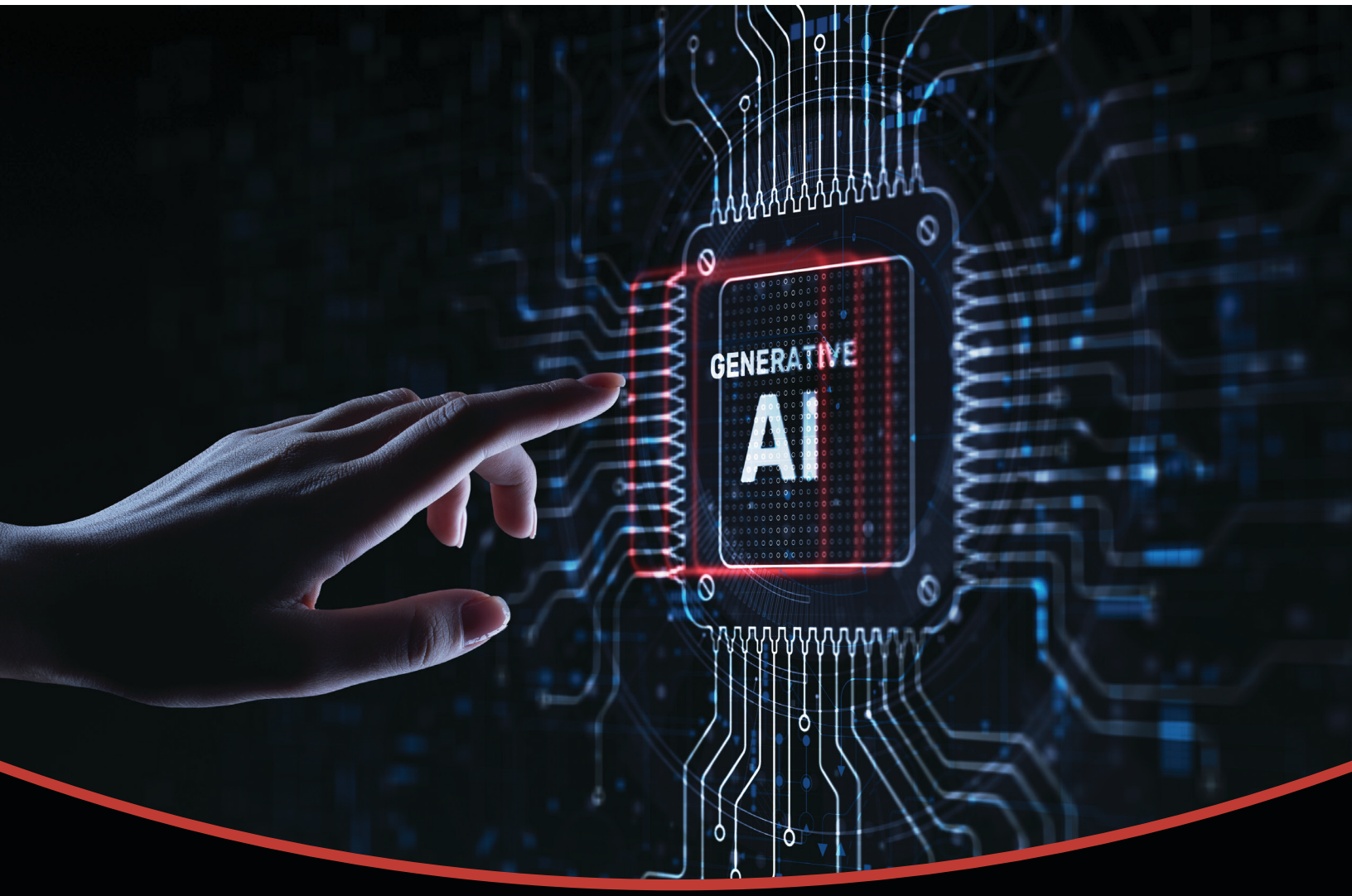
NETWITNESS

WHITEPAPER



The Race for the Generative AI Security Prize

How defenders can position themselves for the win by using 'GenAI'

The emergence of generative artificial intelligence (AI} models will almost certainly produce not only new ways for threat actors to launch attacks, but more threat actors, as well.

Using the large language models (LLMs) underpinning generative AI, almost anyone, now, can launch a fairly sophisticated phishing campaign, for instance.

But as helpful as LLMs may be for planning and launching attacks, they can be equally useful for cyber defenders. In fact, cyber defense may have an edge in at least one respect where the the use of these models is concerned - especially where defender teams have a long history of working with LLMs.

Because to function effectively, "generative AI" and the LLM algorithms underlying it need a critical tool: data. Lots of data. To break into a targeted system or network using GenAI would require very specific data about that system or network - data that an organization could well protect from unauthorized eyes by using GenAI.

Using proprietary, clean and sound data to develop or enhance an existing LLM can make for an effective GenAI cyber-defense tool. At the 2023 RSA conference, several major organizations rolled out GenAI-powered cybersecurity features, including one that alerts developers to coding errors and helps them fix those flaws.[1]

This paper will explore the threats that generative AI could pose to organizations' systems, networks, and data. We will look at who is behind these threats, what motivates them, and how they might use GenAI to attack.

1. Forbes, "Generative AI: Friend and Foe: httP.s:[/www.forbes.com/sites/heatherwishartsmith/2023/06/06/generative-ai-cv.bersecuritY.-friend-and-foe/

And we'll consider why cybersecurity defense teams may be well positioned to out-maneuver attackers using the technology - because of something we have that they may not.

Who's out to get you, and why

Not all cybercrime consists of financially motivated attempts to install ransomware: state-sponsored threat actors may wish to disrupt operations or to spy; hacktivists may aim to make a statement or draw public attention to a cause. All these groups are potential users of generative AI.

Generative AI, which is a form of AI that can generate text, images, and other media based on queries provided to it, emerged for public use at the end of November 2022.

Millions since then have used it to write letters, summarize notes, and generate other forms of text.

Malicious actors, the ultimate early adopters, are no exception: mere months later, cybercriminals were discussing possibilities for its use in dark web forums. One said he was using it to write LLM-powered technology to create encryption scripts - and this person is reportedly not a developer. Others say they are creating file stealers and using the technology for fraudulent schemes.[2]

But AI experts and researchers have had access to the technology for many years, and are equally skilled in its use, if not more so. We should soon be positioned to use the powers of LLMs for good, not evil.

2. Checkpoint Research, "OPWNAI: Cybercriminals Starting to Use ChatGPT," httP.:s://research.checkQQint.com{2023/oP.wnai-cv.bercriminals-starting-to-use-chatgP-!L

Threat? Or Defense?

There is a plethora of ways in which attackers might use LLMs, and just as many generative-AI-assisted defense scenarios to block those attempts. In other words, whatever our adversaries can do, we might also be able to do. Although we don't know precisely the quantity or nature of the data that attackers possess, we defenders certainly have our own, unique data at our behest as well as years of highly technical experience with AI.

Here's another reason for optimism: As GenAI makes it easier to form and carry out attacks, it also reduces the expertise and skills needed to become a hacker. The result may be more adversaries, but will threats increase commensurately? Quantity is not the same as quality: many gen-AI-enabled adversaries will be low-level actors. GenAI could enable defenders to increase our ranks and our skill levels, as well, and could dramatically improve defense overall.

For instance, by explaining complex concepts and tasks in easy-to-understand terms, GenAI defense could enable those with less training and expertise to do certain cybersecurity jobs - and do them well.

It could also help make up for shortages of qualified cyber personnel by automating routine tasks, freeing up experienced cyber teams to work on thornier issues.

Tit for tat

Let's consider some ways attackers might use the technology, and also how it might help defenders to detect, prevent, and respond to these threats:

- **THREAT:** Generating fake videos that inflame viewers against a political leader or cause.
- **COUNTER:** Telling viewers if a video is fake or real.

- **THREAT:** Send phishing emails that lack the spelling and grammatical errors that previously marked them as phony, even mimicking specific organizational styles to send convincing spear-phishing emails.
- **COUNTER:** Warning users when an email is fake and steering them away from the false websites these emails direct them to.

- **THREAT:** Finding vulnerabilities in enterprise applications and software that allow them to slip in undetected.
- **COUNTER:** Alerting developers to coding errors while they are working so they can correct mistakes before launch.

How deep is your data?

There are a number of generative AI, large-language-model applications, some of which are not publicly available. Ethicists and governments including the European Union, now considering an EU AI Act,3 are asking whether certain forms of AI are too risky to release for public use, and what the guardrails should be.

But neither legislators nor even data scientists can foretell which new AI types and uses will emerge. And the cybercrime world is certainly doing its own research.

What's certain is that quality AI, for the most part, requires quality training - which requires massive amounts of quality data. Until recently, much of the data that large organizations possess has sat dormant, unusable by virtue of its sheer volume.

But AI can process and analyze all of it, and even aggregate it to give cyber teams a view of organizational systems and networks that is panoramic and microscopic all at once. Analysts can see the big-picture view, in depth.

In the AI race to the finish line between attackers and defenders, the one with the most, best data wins. This truism gives NetWitness a distinct advantage over bad actors as well as over other security organizations, in these ways:

3. European Parliament News, •AI Act: a step closer to the first rules on Artificial Intelligence," https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-doser-to-the-first-rules-on-artificial-intelligence

- NetWitness has the expertise. Our scientists have worked with machine learning, a form of AI, for more than a decade. We've got a solid grounding in the technology.

- **NetWitness is silo-free.** We speak all three cyber languages: data science, computer science, and threat intelligence. We also deal with data from all an organization's sources: logs, endpoints, and packets, and provide color and context for our clients that defines their users, devices and data origins and destinations.

- **NetWitness empowers our customers.** We're using generative AI to help even less-experienced cybersecurity people detect, protect, and respond to threats as easily and effectively as seasoned pros. Our customers have the highest degree of comprehension and awareness of their own environments as well as of their assets - especially critical in this remote-worker age - and the ability to respond in real time to threats.

- **NetWitness thrives on innovation.** Already our products flag potential threats based on user behaviors. Next up are machine learning-powered features that see everything on our customers' networks as well as potential problem areas - then tell you what you need to do to protect against intrusion.

- **NetWitness has vast repositories of quality data, and we know how to use it.** Many researchers try to start with a solution or desired outcome and work backward to achieve it -but that method doesn't always fit with the datasets they have. We start with the data, because the data tells the story that can lead us to success.

Leveling the cybersecurity field

Alarmists began warning us of danger as soon as, or even before, AI publicly debuted. But the fact is that threat actors continually change tactics, and they will continue to do so. Nothing slows or prohibits well funded, determined, intention-driven threat actors from conducting their nefarious missions. Nothing stops them; nothing ever will.

But generative AI and other LLM-driven technologies aren't necessarily the death knell of any organization or of cybersecurity. These tools are as valuable to defenders, in fact, as they are to attackers.

Questions to ask as you consider which partner to choose to protect your enterprise and assets include: Who has the mathematical expertise? Who has the data? Who has the long history of working with artificial intelligence? In short, whom can you trust with your AI defense program? NetWitness is ready with the answers.

www.netwitness.com