Picture credit: Adobe Stock

FIGHTING FIRE WITH FIRE

Brett Raybould *explains the benefits of using AI to limit AI attacks*

Artificial Intelligence is the latest tool to capture the public's imagination, thanks to the generative systems that now write (and sometimes even code) for us. Attackers are already using the technology to threaten legitimate users with super-charged phishing attacks. The best way to stop them, therefore, is to use AI for our own purposes.

For decades, attackers and defenders have been navigating an intricate dance together. They follow a simple cycle; attackers find a weak spot in the defender's systems and begin to exploit it. The defender realises the problem and plugs the hole. So, the attacker moves on to find another attack vector. And they move like this, toe-to-toe, around the battlefield.

Defenders would rather not have to dance at all of course, but as long as the two sides are evenly matched, this is manageable. Attacks will happen, but smart companies can defend against them. While one side might gain the upper hand for a while, there has been a rough equilibrium over time.

However, two trends are now combining to destabilise this delicate equilibrium. The first has evolved over the last decade or two. It's the gradual emergence of the browser as a key conduit in enterprise IT. The browser's popularity as a business tool has risen sharply along with SaaS applications since around 2010. While on-premises apps that aren't accessed via the browser won't go away entirely, they're typically in the minority these days.

Between them, the browser and email account for over 80 percent of actions leading to security breaches

Instead, employees access most of their computing resources through a browser window.

Windows are transparent, browsers are not. What happens in the browser is often invisible. Websites use browsers' JavaScript engines to process data in ways that security teams cannot see. Even legitimate applications often obfuscate their traffic in the browser to avoid performance-hindering security inspections.

As the web browser became the most critical tool for knowledge workers today, a new breed of attacker emerged, one that has learned to weaponise the browser's ubiquity and opacity to target employees and company networks.

The 2023 Verizon Breach Data Report shows web applications as the top action vector for attack, often through the use of stolen credentials. Email – often accessed via a browser – is the second. Between them, the browser and email account for over 80 percent of actions leading to security breaches or incidents.

As a result, a broad array of cyber threats have evolved beyond the cross-site scripting and man-in-the-middle attacks, using the browser to avoid detection. These browser-based threats are getting worse as attackers use more adaptive techniques designed to evade traditional detection systems like firewalls and anti-virus software.

At Menlo Security, we call these highly evasive adaptive threat_techniques. They are leveraged to compromise browsers, gain initial access to the endpoint and ultimately deploy threats like ransomware or malware and are unmatched in their ability to evade detection, making them the perfect vector for attackers.

Here are some of the ways that they subvert traditional security, becoming a risk to your company.

SEO POISONING

SEO poisoning uses black hat SEO to get malicious content to the top of search engines, lulling the user into a false sense of security. Another, MFA bypass, uses reverse proxies to collect multi-factor authentication tokens and hijack victims' online sessions.

Some highly evasive adaptive attacks are multi-staged. For example, an SEO poisoning attack might use a legitimate domain such as Microsoft 365 to host its malicious content. That makes it difficult for link scanners to detect and block the traffic based on domain reputation alone.

PASSWORD-PROTECTING FILES

Once the user is persuaded to click a malicious link on the site, it might send them a password-protected document containing the malware loader. Many traditional content scanners will allow such documents through to avoid disrupting business workflows.

HTML SMUGGLING

Alternatively, they can use another pernicious evasive adaptive technique called HTML smuggling, which uses obfuscated JavaScript to construct the attack malware on the client side, dodging file scanners.

AVOIDING EMAIL SECURITY

One way that attackers prevent detection by email scanners is to avoid using that channel altogether. Instead of trying to sneak malicious links or files to victims in email messages, they'll use social media systems, including business-focused ones, to message users and deliver attacks via the browser.

The second trend that threatens to give attackers the upper hand began almost 70 years ago, at Dartmouth. Artificial intelligence was ahead of its time then, but thanks to cloud computing and software optimised for GPU training, it came into its own around the same time as SaaS did. Today, we stand at the cusp of a new era of generative AI.

Generative AI uses the same underlying neural network technology as machine learning, but thrives on far larger models. This, along with its innovative algorithms, allows it to create new information rather than merely classify existing data.

ISOLATION TECHNOLOGY CAN SEE WHERE SECURE WEB GATEWAYS AND FIREWALLS CANNOT

Just like any technology, generative AI can be used for good or bad purposes. For every hundred people using it to write business emails or harmless limericks about their dog, there will be one using it for nefarious purposes.

With artificial intelligence more readily available, the threat of AI-based cyber attacks is, of course, significantly on the rise. Namely, criminals are using generative AI to scale up phishing attacks and even to generate disruptive strains of malware, officials have said.

According to a survey of IT professionals from Blackberry, more than seven in 10 feel that foreign states are likely to already be using ChatGPT for malicious purposes against other nations.

Black hats have already learnt how to jailbreak legitimate systems like ChatGPT. As some of these large language models are open source, criminal entrepreneurs have already produced 'dark' versions specifically designed to help scammers and malware producers deliver their attacks, lowering the barrier to entry for attackers. WormGPT and FraudGPT are two of the earliest we have seen.

In this sense, ChatGPT could democratise cyber crime in the way that we've seen ransomware-as-a-service take hold. This would lead to a massive spike in the volume of attacks we witness globally.

LLMs such as WormGPT are designed to help scammers and phishers. They don't contain any of the protections that prevent LLMs like ChatGPT or Google's BARD from generating damaging text like phishing emails.

Instead, WormGPT and FraudGPT are specifically trained to produce fraudulent phishing texts. For example, an SMS message that encourages the receiver to click on a malicious short link, or a longer email that could be used in a business email compromise attack.

These tools' clear, grammatically correct text makes phishing emails more convincing. It means cyber criminals don't need to rely on their first language or dialect. They can use generative AI to translate phishing emails effectively across many vocabularies.

It additionally allows phishing scammers to create emails at scale, especially useful for snowshoe spammers. This term applies to the process of bulk registering domains to launch thousands of short-run phishing campaigns.

AI-BASED DEFENCE SYSTEMS CAN 'SEE' IMAGES THAT SCAMMERS INSERT IN EMAILS

This presents a significant issue for organisations. Security awareness training is typically centred around spotting discrepancies such as misspellings and awkward subject line titles, which could indicate that something is suspicious. With these typical indicators disappearing through ever more convincing phishing campaigns, companies are more at risk.

How do we fight the coming onslaught of AI-powered attacks? AI is a powerful tool for cyber criminals, but it can also be a robust form of defence. Attackers and the malicious websites that serve their traffic leave behind a digital exhaust that offers deep insights into their real intentions. Everything from the URL that a malicious site is using through to the elements used on a web page and the images it displays offer powerful clues.

Human operators cannot comb through that amount of data, but AI is the perfect tool to analyse large amounts of data at speed – more than a human operator ever could – and spot patterns that deviate from the norm. It can marry computer vision with web page analysis and URL risk scoring to understand what's happening in the browser and how it might be trying to dupe a user into unwittingly downloading malicious content.

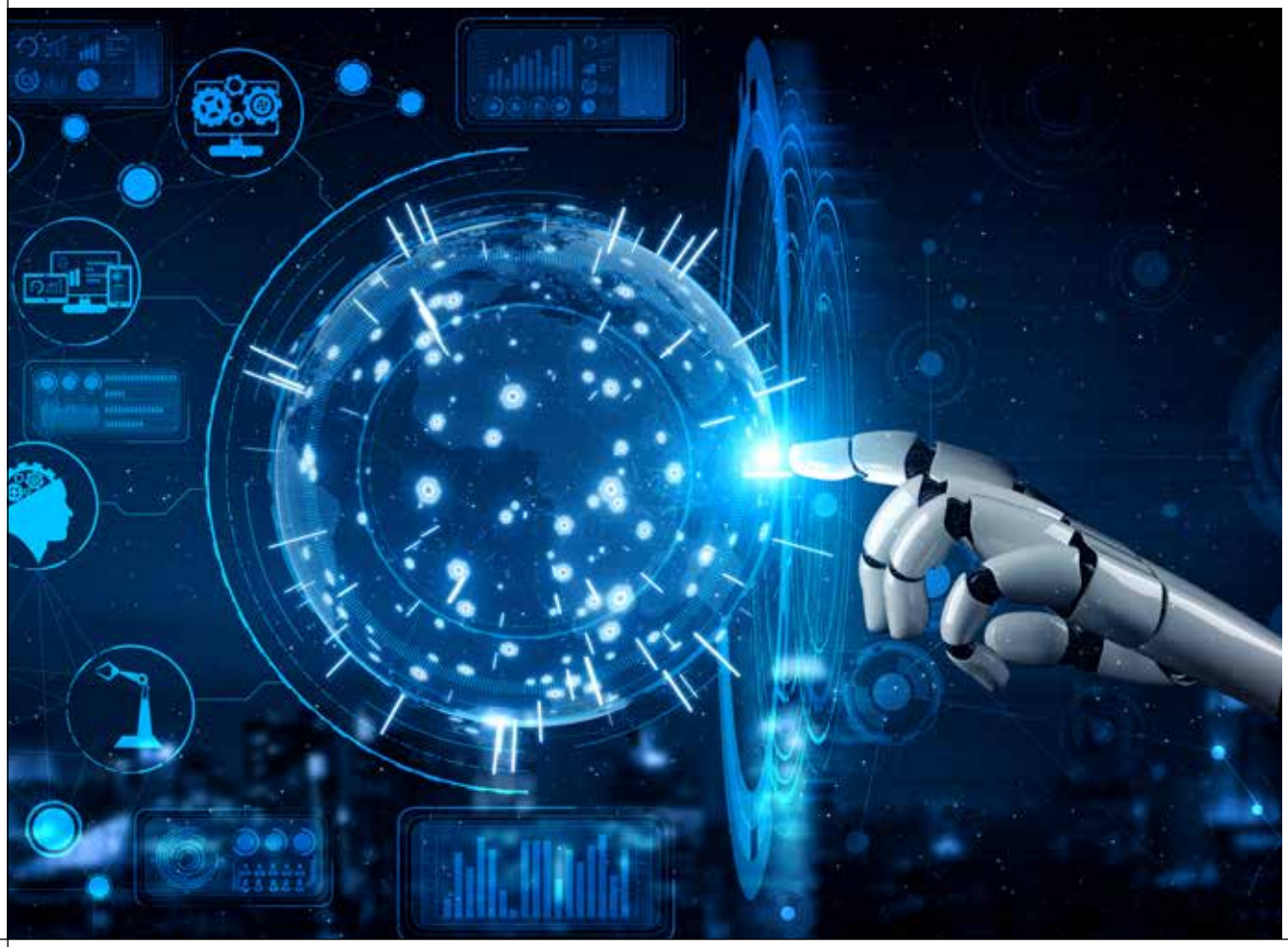
AI-based defence systems can use computer vision to 'see' images that scammers insert in emails or web pages to trick readers. The data that AI analyses comes in the form of network traffic flowing to and from the browser, but also from within the browser itself. They can apply sophisticated URL risk scoring mechanisms, combining them with an analysis of web page elements. When passed through constantly updated machine learning models, this data can determine the intent of a website in real-time.

Isolation technology can see where conventional tools like secure web gateways and firewalls cannot. It can ensure that all active content is executed in a cloud-based browser, rather than on a user's end device, ensuring that malicious payloads never have the opportunity to reach the target endpoint. It can then use machine learning to spot the telltale signs of a highly evasive adaptive threat attacks and raise the alarm.

As attackers begin to use multiple malicious generative AI tools already available on the dark web, the race is on to expand our defences and stabilise the balance of power in cyber security before it's too late. Machine learning capabilities in security products will soon become mandatory to spot sophisticated, automated attacks before they happen ●

Brett Raybould is EMEA Solutions Architect at Menlo Security

AI is the perfect tool to analyse large amounts of data at speed and spot patterns that deviate from the norm



Picture credit: Adobe Stock