KNOWLEDGE IS POWER

Aaron Rosenmund *highlights the six most dangerous new threats security teams need to know about*

The rise of AI presents both extraordinary opportunities and intimidating challenges in cyber security. While AI can easily identify and exploit vulnerabilities, deploying it without robust security measures introduces significant risks.

As the technology evolves, many organisations prioritise AI innovation at the expense of security, leaving their systems vulnerable. This underscores the need for established security frameworks and ongoing education about the dynamic risks AI presents.

Organisations can effectively mitigate risks and safeguard operations by prioritising AI security and supporting cyber security professionals. Each year, experts at the RSA Conference discussed six of the largest new attacks and threats – and what actionable steps businesses can take to address them.

WEAKNESSES IN CODE

If you give GPT-4 a list of real-world vulnerabilities, it can exploit 87 percent of them autonomously. That means that hackers can use a publicly disclosed list to target companies. While this is the approach human hackers have historically taken – finding gaps in defences – the speed and ease at which an AI can do it changes the game. Applications become highly vulnerable if unpatched, with zero-day (publicly known, unpatched vulnerabilities) and one-day attacks (patched, but not applied) posing significant threats.

To effectively prevent AI-driven attacks, security teams must adopt a proactive approach by leveraging AI for defence – fighting fire with fire. Since human teams can't patch vulnerabilities as fast as malicious AI can detect them, ensuring businesses are fully equipped with AI capabilities is crucial. This requires security teams to be well versed in AI, while automating purple testing (where the role of both the attacker and defender are simulated) can create a continuous feedback loop of simulated attacks and responsive remediation strategies.

Despite their best efforts, humans remain the weakest link in the security stack

Aaron Rosenmund
is Senior Director of Content Security and Curriculum at Pluralsight.

GENAI EXPLOITED AND WEAPONISED

"We need to have GenAI in our company" is a common phrase heard in businesses since the explosion of the technology. While innovations are moving quickly into products as businesses look to capitalise on new tech, they also open a significant entry point for attackers to exploit and weaponise against the business.

A survey by IBM found that 70 percent of C-suite respondents believe that, when it comes to AI, innovation takes precedence over security. And while 82 percent said: "secure and trustworthy AI is essential to the success of their business," only 24 percent of them said they're actually securing their GenAI products.

The risks of an insecure large language model (LLM) are significant and far-reaching. For example, attackers can exploit prompt injection to manipulate AI into revealing sensitive data or performing unauthorised actions, while training data poisoning can corrupt AI during its learning phase, leading to harmful outcomes like backdoor attacks. Looking ahead, there are concerns that attackers may co-opt AI systems to launch coordinated attacks, making it crucial to ensure newly branded AI is not moonlighting as a criminal.

To protect LLMs, start by adopting established frameworks like Google's Secure AI framework (SAIF) and Nist's AI Risk Management Framework. Conduct comprehensive modelling and data validation while enforcing the principle of least privilege.

SOPHISTICATED SPEAR PHISHING

In an era where a voice can be cloned from just a three-second sample of someone talking and verifying their identity, things will get very tough, very fast. We can't rely on companies to AI-generate protective content, as even subtle telltale signs can be easily erased.

This poses a significant challenge for remote identity verification, particularly in distributed workforces. To combat this, focus on strategies to establish and reestablish identities for customers and employees. Leverage AI to detect unusual behaviour, but remember that despite their efforts, humans remain the weakest link in the security stack.

SEXTORTION OF EMPLOYEES

Sextortion is an uncomfortable yet crucial topic to address in the age of GenAI. While the concept isn't new – like those emails threatening to expose what's on someone's hard drive – AI advancements have made it a growing threat, and anyone can be a target. Unfortunately, it usually doesn't end when payment is made. We're seeing attackers using an 'alternate' form of payment, such as giving access to a network, installing malware or any way that compromises a system.

Executives are most at risk, so implementing an executive protection program is vital. Educate the entire company on what sextortion is and what an attack might look like. It's uncomfortable, but these attacks thrive in an environment of ignorance.

THE SPEED AND EASE AT WHICH AN AI CAN FIND GAPS IN DEFENCES CHANGES THE GAME

MFA INTERCEPTION

Push notifications from Multi-Factor Authenticfication are becoming a nuisance, leading many users to click through them without giving them much thought. This makes them vulnerable to "attacker-in-the-middle" attacks, where attackers trick users into logging into a fake site. Once logged in, attackers capture the user's credentials and MFA code to access the real account.

While MFA is still better than nothing, it's not a silver bullet. To enhance security, users are required to enter a code from the login screen, as attackers won't have access to it. Add context to push notifications, such as sign-in location and tighten authentication measures for unusual login times.

LACK OF TRAINED CYBER PROFESSIONALS

The shortage of cyber security professionals is a well-known issue, with 71 percent of organisations having unfilled cyber security positions. This shortage leaves security teams understaffed and burned out, a problem exacerbated by the rise of AI.

Cyber security professionals must be more alert than ever, given the AI tools threat actors now have. However, only 12 percent have significant experience working with AI.

Focussing on upskilling your cyber security team in AI-based defence strategies and leveraging AI to reduce the burden of their job can be beneficial. Tasks like inbound message filtering, summarising incident reports, process automation and filtering bug bounty challenges can all be automated. Supporting employees with resources to stay informed on the way threat actors use AI and upskill on knowledge gaps will make for a more engaged and better-equipped team ready to defend against criminals.

In the face of significant cyber security threats, taking proactive and tangible steps to safeguard employees and the organisation is crucial. Addressing issues like the shortage of cyber security professionals, AI-driven attacks and sextortion requires a deliberate approach – from upskilling your team in AI defence to creating a supportive work environment. By staying vigilant and proactive, businesses can effectively minimise risks and enhance their overall security posture ●