protiviti®
Face the Future with Confidence

Artificial Intelligence — Security Friend or Foe?

The annual cost of cybercrime is estimated to rise to $6 trillion by 2021.[1] Artificial intelligence (AI), frequently mentioned for its potential to accelerate innovation, boost performance and improve decision-making, is already being applied to defend against cybercrime. Because AI works well with functions that use massive amounts of data and require analysis and judgement, integrating AI-based cybersecurity technology with other defences is a natural choice for cybersecurity professionals.

Today, AI is used more extensively in cybersecurity than in any other function, with 75% of companies using AI technology to detect and ward off cyberthreats, according to the results of a recent global executive survey on AI conducted by Protiviti. Cybersecurity usage of AI is expected to grow nearly 20% by 2021.[2]

AI's significant and compelling benefits come with new risks that need to be managed. As the use of AI increases for both cybersecurity and business applications, hackers will attempt to exploit vulnerabilities.

AI: An Ally in Cybersecurity

Security functions such as evaluating user access logs and managing user access are good examples of the suitability of AI for cybersecurity applications. Logging users' interactions with enterprise systems yields many terabytes of data. These logs record which users are accessing sensitive information, the duration and frequency of their access, and whether they are downloading, emailing or uploading sensitive information to the internet.

Any user might access data inappropriately, even without malicious intent. A few AI use cases in cybersecurity are as follows:

[1] *The Cybersecurity Imperative*, ESI ThoughtLab, 2019: https://www.protiviti.com/sites/default/files/united_states/insights/cybersecurity_imperative_2018.pdf.

[2] *Competing in the Cognitive Age*, Protiviti, 2018: www.protiviti.com/sites/default/files/united_states/insights/ai-ml-global-study-protiviti.pdf.

Flag for Fraud

AI systems can identify patterns from each user's typical activity, determine if those patterns are atypical or concerning, and evaluate any current action against those patterns. AI can swiftly review the massive amount of data contained in user access logs and flag actions that warrant human scrutiny, such as when a user who legitimately and routinely views sensitive data starts downloading that data instead.

Reduce Hacker Dwell Time

Hackers do their best to appear as legitimate users, but AI-enabled scrutiny of user access logs makes it harder for hackers to hide. The term "dwell time" — also known as Mean Time to Identify (MTTI) in cybersecurity lingo — denotes the time from breach to detection. This metric, along with Mean Time to Contain (MTTC), measures the effectiveness of a company's incident response and containment processes. Costs are lowered when breaches are identified and contained more quickly. In 2018, average dwell time was 197 days.[3] AI carries the promise of reducing dwell time to a matter of minutes by monitoring for anomalies in user behaviour in real time.

User Provisioning

AI can be equally useful in user provisioning — creating, updating and deleting users' access privileges. User provisioning started as (and often remains) a manual task, with levels of access to specific systems and data selected for each employee. Some organisations began selecting access rights by role: access rights associated with an accounts payable or a human resources role, for example. But even with role-based security, the process to grant, modify and revoke rights for all personnel — and to keep role definitions in sync as the organisation changes — has remained both labour intensive and error prone.

User provisioning, whether manual or role-based, calls for skilled workers to perform numerous minor tasks with a high potential for error. For instance, a new employee is granted network access and privileges related to approving invoices for payment. Privileges are updated again when the employee assumes responsibility for higher-value invoices. Later, the employee moves to a job in human resources, so security revokes all invoice payment privileges, but grants new privileges to view confidential employee information. When the employee leaves the company, all privileges — including all network, system and data accesses — are revoked.

AI, combined with robotic process automation, can start to automate many of those functions that currently require human intervention. AI systems can monitor employee status, learn from the actions performed when a change takes place, and then take the appropriate actions when it detects an employee status change. The market for security professionals is notoriously tight, and applying an AI solution to user provisioning can free those resources to perform higher-value activities that can't be automated.

Downsides Related to AI and Cybersecurity

While AI's power and efficiency can help accelerate processes that benefit business, that speed and efficiency can also have unintended negative outcomes. The advantages organisations will realise with AI also introduce new risks, and it's critical to manage those risks properly.

It's important, therefore, to evaluate an AI system's potential vulnerabilities from inception, and to design appropriate controls while the system is still under development. Vulnerabilities will exist not only in the system itself, but also in the business processes that arise with the deployment of any new system.

But legitimate enterprises are not the only ones implementing AI to further their goals. Hackers use AI tools to defeat AI-based defences, in what's becoming an arms race in cybersecurity. Around the world, executives anticipate that attacks by hackers who have network access will increase by 247% over the next two years, according to the results of a survey on cybersecurity conducted by Protiviti, ESI ThoughtLab, WSJ Pro Cybersecurity, and other prominent organisations.[4]

[3] *2017 Cost of Data Breach Study*, Ponemon Institute, 2017: https://www.ibm.com/downloads/cas/ZYKLN2E3.

[4] *The Cybersecurity Imperative*, ESI ThoughtLab, 2018: https://www.protiviti.com/sites/default/files/united_states/insights/cybersecurity_imperative_2018.pdf.

To penetrate networks, hackers exploit AI tools in several ways:

- **By overwhelming defensive capabilities** by generating so much traffic against a site that a malicious actor's actions are hidden in the abundance of logs. Such tactics will most certainly overwhelm a human defender but may be foiled by an equally matched AI system.
- **By identifying and analysing an organisation's defences.** An enterprise may have set up a defence that analyses access requests to a certain network area or application to discover possible attack patterns. The attacker can use AI to make requests — and then continuously and intelligently modify them — so that an AI-based cybersecurity application can no longer discern a pattern.
- **By tricking an organisation's AI system into providing access to the network.** Historically, a hacker might attempt a "social engineering" attack — manipulating a person into revealing confidential information that could be used to perpetrate a crime. Now we see hackers attempting to manipulate an organisation's AI defences by repeatedly attempting access until finally providing the right information. This is a particular weakness of AI. It may be easier to deceive an AI "brain" than a well-trained human. Therefore, applications of AI must be designed to detect social engineering attack attempts as effectively as people can.

Business applications of AI are cyberattack targets, too. Beyond attempts to compromise AI-based cybersecurity, hackers will attempt to penetrate an organisation's AI-based business applications to induce unfavourable price changes, for example, or even to originate and send communications that damage the organisation's reputation.

How Do We Secure AI Technology?

Any enterprise planning to deploy AI-powered applications for cybersecurity or business should engage security professionals in the early stages. Security professionals will want to assess any new AI system for vulnerabilities, including in the new business processes that arise from the system's use. Specific consideration for the issues around AI should be included while conducting threat modelling for an AI-based system. Enterprises applying AI need to anticipate how hackers might apply their own AI systems to defeat it, and build in appropriate defences.

Traditional cybersecurity technology is not equipped for the new classes of attack that come with AI systems. For example, hackers will engage in data poisoning, in which they inject their own data samples for consumption as the system is retrained, to disrupt feedback mechanisms or alter classification of data to favour the hacker. Adversarial input is another new class of attack, whereby hackers develop inputs designed especially to be misclassified, such as malicious documents that avoid antivirus detection, or emails that get around spam filters. Hackers will also probe an AI system to duplicate its underlying model, in a class of attack called model stealing. Once stolen, the AI model could be merely used or sold as stolen property, or it can be examined to detect further vulnerabilities for hackers to exploit. Any AI system should include mechanisms to defend against these methods. Beyond that, AI models should be carefully verified and monitored to continually enhance robustness.

Enterprises that deploy AI-powered applications for cybersecurity or business should engage security professionals in the early stages to assess vulnerabilities, considering the fact that hackers are now similarly armed. The assessment should also include any new business processes that arise from the use of AI.

– Madhumita Bhattacharyya, Managing Director, Protiviti

Conclusion

AI is already a competitive differentiator for many companies. And while hackers are deploying AI themselves, AI is a particularly valuable tool in the arsenal of cybersecurity teams. The best strategy for an organisation is to embrace the potential benefits of AI and find opportunities to implement AI-based business solutions and cybersecurity applications of AI while also considering how AI might be used against the enterprise and integrating defensive thinking within its applications. For any new AI system, cybersecurity and data privacy risks should be considered from the outset and kept top-of-mind throughout the system's life cycle, as new threats and new cyber defence opportunities arise. These measures will ensure AI implementations realise promised advantages, while diminishing potential risk.

How Protiviti Can Help

Anyone deploying AI — to support their line of business or in the cybersecurity function — should engage qualified professionals early to evaluate current and new system risk, to think innovatively about anticipating new threats, and to secure AI systems against all the new threats that accompany them. Protiviti is a leading provider of cybersecurity and risk management consulting services, including advising clients on how to implement and secure AI effectively in their business. Our team of AI experts performs threat modelling to identify the full spectrum of possible threats, identifies vulnerabilities through reviews of the code itself, and assists with system architecture and implementation. We also perform "red team" exercises to expose vulnerabilities in an AI system and achieve a specific outcome.