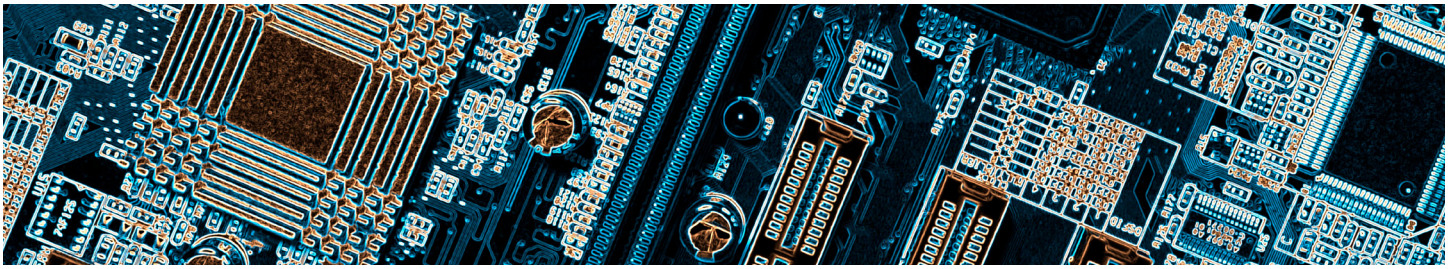

How to harness the potential — and reduce the risks — of AI

Artificial intelligence (AI) and machine learning capabilities have taken a big jump forward, yet true intelligence is still to come. Understanding the technology's potential cyber threat is key to protecting your business, now and in the future.

Key takeaways

- Businesses that leverage AI tools face important questions about how to use the technology responsibly while managing the cyber risk that these tools present.
- Criminals and other malicious actors are adopting AI tools to make current tactics for cyber intrusion even more effective.
- A cyber-aware approach to AI tools depends on effective human oversight of AI tools and their outputs.

The technology that sits under the umbrella term of artificial intelligence has enormous potential. While much of that potential has yet to be realized, AI is already in evidence every day, affecting how we work, how we consume and package information and how we think about digital security. Many industries are realizing benefits as these tools remove friction from transactions, improve fraud detection, enable chatbots and eliminate repetitive tasks from employee workloads.

Like every powerful tool that preceded it, AI presents an array of opportunities, options, and risks. But its recent expansion has brought many hypothetical situations into the real world in the relative blink of an eye. Every business that leverages AI tools faces important questions about how to use the technology responsibly while managing the very real cyber risk that these tools present.

Every business that leverages AI tools faces important questions about how to use the technology responsibly while managing the very real cyber risk that these tools present.

The best time to start considering the security implications of AI is now. Proactive enterprises will consider safety when establishing protocols for using AI tools, while

simultaneously asking how bad actors might manipulate the same tools to infiltrate company systems or create cyber incidents. Otherwise, businesses may find it challenging to keep up with what these tools and their users (good and bad actors alike) deliver, and how they do it.

Fortunately, understanding the opportunities of AI tools and hedging against their risks does not require a brand-new approach to security. Foundational principles and best practices of cyber hygiene are not obsolete. But as new tools and use cases arrive, businesses and institutions will need to think creatively about their organizations and prioritize a dynamic, enterprisewide approach to security.

What are the risks associated with AI at this early adoption stage?

Assessing the capabilities of cyber criminals and nation-states is always challenging for security experts, since adversaries rarely broadcast their tools and capabilities, and their claims about intrusions are not always truthful. But speculating about the effectiveness of large language models (LLMs) available on the dark web skips over a key fact: Bad actors are always looking to manipulate public LLMs with fake data or prompt injections, or to target private LLMs by unauthorized access to corporate systems.

So while it's not impossible that bad actors might create an even more powerful AI tool (or one with few or no restrictions on the

How to implement AI responsibly

Every business and industry will need to set standards and expectations around responsible AI use. They also will need to calculate its specific risk and usage protocols. But some general principles will apply to most businesses, whether they are using these tools or just adapting to the AI environment:

Focus on realistic threats. At this stage, there are many claims made about AI's capabilities; some are exaggerated. Correspondingly, some concerns about AI's downsides, while realistic in theory, do not apply to its current stage of development. Companies should not lose sight of immediate concerns (e.g., how an AI model is using their data) by focusing too much on a possible future (e.g., sentient or malicious AI).

Generating value from AI tools is dependent on responsible and cyber-secure use. Businesses should realize the value of acquiring, nurturing and encouraging employees who have aptitude for using AI tools and understanding their potential risks.

Approach AI tool adoption with caution. Rapid scaling of AI tools may not be a wise course for every organization. Legal ramifications of data sharing, penalties for issuing misinformation and exposure of company assets are risks that may motivate a business to put tighter restrictions on the use of AI models until more robust security controls are in place.

Recognize that every AI tool expands the risk of cyber events. AI models store vast amounts of information, and information is always a lure for cyber criminals. The models could be stolen or corrupted by external actors or insiders, leading to data loss or corruption, privacy breaches or reputational damage.

Take extra steps to verify identity. Scammers and bad actors will attempt to create more convincing deepfakes based on publicly available or stolen information and personal identifiers, such as voice recordings and video. Eventually, a text message or even video conferencing feed from an account holder may in fact be a simulation. To adjust to this reality, organizations can train their employees about deepfake threats and develop protocols for verifying the identity of the person in question when receiving an unusual request.

Continue to educate employees around the essentials of cyber hygiene. AI models help people accomplish tasks more quickly and efficiently. This applies to scammers and other bad actors as well as legitimate employees and partner organizations. AI models will not eliminate phishing, credential theft, account takeovers or other established cyber-crime tactics. Rather, they will often make these crimes harder to spot.

content it generates), risk assessment should focus on legitimate, widely available tools.

The proliferation of LLMs has lowered the bar of entry for bad actors. Creative prompt injection techniques could bypass an LLM's safeguards to rapidly generate phishing email text or imperfect malware code, which could be refined and deployed by criminals who lack top-tier skills.

While it's not impossible that bad actors might create an even more powerful AI tool (or one with few or no restrictions on the content it generates), risk assessment should focus on legitimate, widely available tools.

As a result, more bad actors will be able to generate convincing email text and social engineering campaigns. By leveraging AI tools to create audio and visual simulations known as deepfakes, it will be easier to create voicemails that sound like a boss, client, family member or friend. After stealing account credentials, a criminal could use an LLM to create messages that impersonate the account holder to make requests for bank account information or instruct others to initiate illegitimate payments.

Another factor to consider is that human error (e.g., clicking a malicious link, weak passwords, insufficient data protection) is still at the root of most cyber incidents. In their current form, AI tools will not change that reality. What they can do is make it easier for bad actors to trick unsuspecting targets. This reality makes a strong argument for increased human oversight and extra layers of approval and verification of sources, transaction details and many other business functions.

How can businesses balance AI functionality and risk?

Fortunately, companies that incorporate AI tools into their workflows can mitigate potential risks while benefiting from their capabilities.

A key point in any conversation about AI adoption is the impact the technology will have on cybersecurity. AI tools are excellent at pattern recognition, but when they become powerful enough for security teams to make better inferences about anomalous activity within networks, detecting and remediating cyber intrusions should become faster and more comprehensive.

However, businesses are a long way from being able to "set and forget" AI security tools — or any AI tool that supports business operations or processes. Organizations of all sizes still depend on people who can create new applications and use cases for AI and, importantly, provide necessary checks on the tools' performance while evaluating and managing their cybersecurity risk. The relationship between AI and humans is still a coexistence, with humans, for now, in the critical decision-making role.

AI's potential is almost incalculable, but a measured, rational approach to adopting it and defending against its misuse can benefit a business's bottom line and security posture. However, companies should recognize that managing AI cyber risk will never be a purely technological challenge. Like the advent of the internet, cloud computing and mobile devices, these models will demand a combination of the right security investments and a sensible, informed approach that spans the entire organization. ■