IRON MOUNTAIN®

White paper

How generative AI is reshaping security

Summary

When it comes to security, is GenAI a tool for good or for evil? Enterprise security teams are grappling with what this emerging technology means for their organisations. And they increasingly find that generative AI brings both opportunity and risk.

Contents

GenAI: A new tool for everything

Without a doubt, one of the most important recent technological innovations is generative artificial intelligence (genAI). When ChatGPT launched at the end of November 2022, it ushered in a new era of computing.

You can use GenAI to write an article, or to output the code for your latest software update. It can generate images that look like realistic photos, or create videos and audio that are almost indistinguishable from the real thing. It can scour the web and serve up answers to nearly any question you have. And we've just scratched the surface of what this new technology might be capable of.

Organisations are understandably excited about the opportunities that genAI represents. An AWS and MIT survey found that 80% of chief data officers believe genAI will transform their organisations, and 45% said their companies had already adopted it widely.

"Generative AI has the potential to be the most disruptive technology we've seen to date," says Steve Chase, U.S. Consulting Leader for KPMG. "It will fundamentally change business models, providing new opportunities for growth, efficiency, and innovation, while surfacing significant risks and challenges. For leaders to harness the enormous potential of generative AI, they must set a clear strategy that quickly moves their organisation from experimentation into industrialisation."

Against this backdrop, managers are feeling the pressure to start using genAI as quickly as possible. But some are understandably hesitant. For all its potential benefits, generative AI also presents some significant risks.

Many security leaders are still struggling to formulate a plan for if, or how, they will handle generative AI. In the AWS and MIT survey, 16% of chief data officers said their companies had completely banned the use of genAI, and only 6% said they were using genAI in production.

The rest were experimenting at the individual, team, or organisational level, trying to better understand the capabilities of this tool. Ultimately, that's what genAI is – a tool. And like any tool, it can be used for good or for evil.

From a cybersecurity perspective, it's clear that generative AI has major implications for three different groups of people: security teams, employees at large, and bad actors inside, or outside the organisation.

Security teams + genAI = better protection

Security vendors have been integrating AI and machine learning capabilities into their products for many years, with great benefit for their customers. The advent of genAI seems to be accelerating that trend by improving the capabilities of security software.

Generative AI tools seem particularly good at detecting threats and identifying attacks. According to the analysts at Bain, "Threat identification holds the greatest potential for generative AI to improve cybersecurity . . . Generative AI is already helping analysts spot an attack faster, then better assess its scale and potential impact. For instance, it can help analysts more efficiently filter incident alerts, rejecting false positives. Generative AI's ability to detect and hunt threats will only get more dynamic and automated."

A separate study from IBM revealed that, "Organisations with extensive use of both AI and automation experienced a data breach lifecycle that was 108 days shorter compared to studied organisations that have not deployed these technologies (214 days versus 322 days)." However, the same IBM study found that 40% of organisations had not yet deployed security AI and automation.

Some security teams are also using generative AI to enhance zero-trust strategies. GenAI can help create risk profiles for different endpoints. And its pattern-matching capabilities can help detect any anomalous events.

GenAI can also augment staff capabilities in other ways. For example, it can help research emerging threats or potential vendors. It can analyse historical data to find patterns, assist with the writing of reports, and craft policies designed to prevent or mitigate future security incidents.

In short, generative AI can enhance the capabilities of existing staff, helping them to become more efficient and more effective, ultimately providing better security for the organisation.

But unfortunately, not all the potential impacts of generative AI are completely positive.

Employees + genAI = Increased risk

A lot of the experimentation with genAI is happening outside the security department. Employees from just about every team in an organisation are likely to try out the technology. Unfortunately, no one really knows how much employees are using generative AI, or even which tools they are using. In much the same way that organisations have long had "shadow IT," they now have "shadow AI."

That's a problem because generative AI not only offers tremendous benefits, it also poses significant risks. And one of the biggest potential problems is data leaks.

GenerativeAI tools are based on large language models (LLMs). These LLMs ingest a large volume of text, which they then use to predict the next word when generating new text. One of their most helpful use cases is in ingesting existing text and improving it.

However, if employees input personally identifiable information, proprietary code, or other company secrets into the LLM, the tool can store that data and output it for another person.

Analysts at PwC explain, "GenAI applications could exacerbate data and privacy risks; after all, the promise of large language models is that they use a massive amount of data and create even more new data, which are vulnerable to bias, poor quality, unauthorised access and loss."

Another potential problem is that beginner developers can sometimes use chatbots to write insecure code. "In the context of cybersecurity, we should expect that inexperienced programmers will turn to predictive language model tools to assist them in their projects when faced with a difficult coding problem," InfoWorld writes. "While not inherently negative, issues can arise when organisations do not have properly established code review processes and code is deployed without vetting."

Humans could easily leak company secrets, or write insecure code without using GenAI, of course. But this new tool represents a new vector for data leakage — one that is particularly difficult to secure. That means the workforce will need more education about how to use it securely. Security and risk management teams will need ways to monitor employee use of tools to help ensure that they aren't unnecessarily exposing the organisation to increased risk.

Bad actors + genAI = Potential disaster

The most significant implication of genAI for enterprise security is the possibility that bad actors will use generative AI to commit crimes. The same tool that makes it easier for security teams to find attacks also makes it easier for cyberattacks to find new ways to launch those attacks.

In fact, evidence that criminals are using genAI is already surfacing. According to Bain, "Mentions of generative AI on the dark web proliferated in 2023. It's common to see hackers boasting that they're using ChatGPT."

For some wanna-be hackers, generative AI makes it easier to get started as a cybercriminal. Instead of learning to write their own code, they can just have one of the AI coding tools write it for them. The tools do have some safeguards – you can't just ask ChatGPT to write you some malware, for example. But you don't have to be too clever to find workarounds to some of these barriers. And the cybercriminals are sharing their methods with others.

These newbie hackers aren't the biggest threat to corporate networks, however. Enterprise security tools should be able to catch most of the attacks coming from amateurs. A much bigger threat is the potential for more realistic phishing attacks.

Most people know not to trust an email with lots of misspelled words. And corporate education initiatives have done a good job at training people to double-check out-of-the-ordinary emails.

But what about when those emails sound exactly like every other email? Or what if it had an attached video or voice memo that looked and sounded exactly like your boss?

PwC writes, "The most immediate risk to worry about? More sophisticated phishing. More compelling, custom lures used in chats, videos, or live generated 'deep fake' video or audio, impersonating someone familiar, or in a position of authority."

Those same tools could also be used to damage your company's reputation online. Bad actors can create fake images, audio, or video, and post it on social media. Or they could threaten to post it online in hopes of extorting a ransom.

GenAI tools

Companies are adding generative AI capabilities to a wide variety of products, and new GenAI startups are cropping up every day. Some more well-known GenAI applications include the following:

ChatGPT – the breakthrough large language model from OpenAI, which answers questions and carries on conversations

GitHub Copilot – AI coding assistant that calls itself "the world's most widely adopted AI developer tool"

Copy.ai – writing tool designed for tasks like blogs and marketing content

Scribe – a writing assistant that specialises in creating documentation and guides

Bing – Microsoft's search engine that now incorporates responses from GPT-4

Bard – Google's alternative to ChatGPT and Bing

Dall-E2 – OpenAI's tool for creating photo-realistic images from text

Synthesia – AI platform that transforms text into realistic videos

Rephrase.ai – test-to-video platform with stock and custom avatars

Bardeen – workflow automation tool that handles tedious work tasks

Murf.ai – audio generator for creating voice-overs based on real human voices

Designs.ai – graphic design tool for creating logos, videos, ads, and more

Looking beyond deep fakes, generative AI can also be used to generate malicious code. For example, security researchers at HYAS used genAI to create a new kind of polymorphic malware that they dubbed "Black Mamba." While it looks like benign code, Black Mamba actually rewrites itself at runtime to become a malicious keylogger that exfiltrates data via Microsoft Teams. And because the malware continuously rewrites itself, it evades even strong cybersecurity defenses.

"Using these new techniques, a threat actor can combine a series of typically highly detectable behaviors in an unusual combination and evade detection by exploiting the model's inability to recognise it as a malicious pattern," HYAS explained. "This problem is compounded when artificial intelligence is at the helm and driving cyberattacks, as the methods it chooses may be highly atypical compared to those used by human threat actor counterparts. Furthermore, the speed at which these attacks can be executed makes the threat exponentially worse."

As frightening as Black Mamba is, security researchers say that it's probably not the worst threat posed by generative AI. That title might belong to indirect prompt injection, a type of attack that subverts the popular generative AI tools by feeding them malicious data through seemingly ordinary websites.

Wired reported, "In one experiment in February, security researchers forced Microsoft's Bing chatbot to behave like a scammer. Hidden instructions on a web page the researchers created told the chatbot to ask the person using it to hand over their bank account details. This kind of attack, where concealed information can make the AI system behave in unintended ways, is just the beginning."

Almost certainly, cybercriminals are — right this minute — hard at work thinking of other ways to use genAI as an attack method. Enterprise IT and security leaders are aware of the threat, but so far, have done little to combat it. A McKinsey study found that 53% of those surveyed believed genAI posed a security risk, but only 38 percent are working to mitigate that risk.

That begs the question, what should they be doing?

Options for action

If genAI were only a threat, the response would be obvious: organisations would lock down their systems to prevent employees from accessing generative AI. They would employ the strictest methods possible to prevent and mitigate attacks that incorporated GenAI capabilities. But genAI isn't just a threat. It's also an opportunity.

Smart security teams are looking for ways to incorporate this tool into their organisations to help them reach their goals, while combating the threat. With that in mind, consider the following options for action.

1. Continue existing security measures. The good news is that existing cybersecurity tools provide a measure of security against genAI threats. If you already have robust security measures in place, you are well on your way to being prepared for the new world of generative AI.

2. Improve protection for your AI models. As your organisation expands its use of AI, your models represent a very attractive target. PwC notes, "GenAI adds a valuable asset for threat actors to target — and for your organisation to manage. They could manipulate AI systems to make incorrect predictions or deny service to customers." With that in mind, the firm recommends, "Your proprietary language and foundational models, data and new content will need stronger cyberdefense protections."

3.Add genAI and automation to your security arsenal. Organisations that deploy both AI and automation as part of their defenses find malware much more quickly than those that do not. In addition, GenAI can also make your security team more productive in a myriad of other ways. By incorporating these new tools into your ongoing efforts, you will be better prepared to thwart attacks that attempt to use generative AI against you.

4. Educate your staff. Because generative AI is so new, research is changing all the time. Encourage your teams to stay up to date on the latest information. InfoWorld advises, "Most important, at this time, it would be wise to rethink your employee training to incorporate guidelines for the responsible use of AI tools in the workplace. Your employee training should also account for the AI-enhanced sophistication of the new social engineering techniques."

5. Monitor regulations. It isn't just the research or information you need to monitor — you also need to carefully track government responses to the new tools. Gartner notes, "The EU AI Act and other regulatory frameworks in North America, China, and India are already establishing regulations to manage the risks of AI applications."

It adds, "Be prepared to comply, beyond what's already required for regulations such as those pertaining to privacy protection."

6. Write and implement policies. Your employees are already using generative AI. But if you're like most organisations, you probably haven't yet implemented any rules around what you believe is appropriate for your workplace. According to McKinsey, "Just 21 percent of respondents reporting AI adoption say their organisations have established policies governing employees' use of genAI technologies in their work." That's a problem because, as PwC states, "Without proper governance and supervision, a company's use of generative AI can create or exacerbate legal risks."

7. Formalise risk management. Risk is inherent in every business. But smart companies carefully choose which risks they are willing to take, and mitigate against the potential dangers. That process is very difficult unless you have a formal risk management process. Gartner forecasts, "By 2026, AI models from organisations that operationalise AI transparency, trust and security will achieve a 50% improvement in terms of adoption, business goals and user acceptance." If your organisation doesn't currently have a formal risk management process, or if you don't think your current processes are adequate for the challenge of generative AI, get help from an outside partner, like the risk management consultants at Iron Mountain.

8. Improve data management. You can also help protect your organisation against the dangers of generative AI by effectively governing and backing up your data. Following best practices for data management makes it much less likely that you will experience a data breach. And if bad actors do infiltrate your systems, having adequate backup and disaster recovery mechanisms in place can protect against data loss. Again, you might want to seek help from a data management vendor like Iron Mountain to make sure you are prepared for generative AI threats.

9. Evaluate vendors carefully. As you look for generative AI tools or seek help related to cybersecurity and data protection, make sure that you evaluate any outside companies you work with. With new technology, it's tempting to rush to deploy new technologies. While you shouldn't delay, you do want to take enough time to be sure that you can trust the partners that you choose to help you integrate genAI into your workflows, and to protect you from genAI-related risk.

About Iron Mountain

Iron Mountain Incorporated (NYSE: IRM), founded in 1951, is the global leader for storage and information management services. Trusted by more than 225,000 organisations around the world, and with a real estate network of more than 98 million square feet across more than 1,400 facilities in over 60 countries, Iron Mountain stores and protects billions of valued assets, including critical business information, highly sensitive data, and cultural and historical artifacts. Providing solutions that include information management, digital transformation, secure storage, secure destruction, as well as data centres, cloud services and art storage and logistics, Iron Mountain helps customers lower cost and risk, comply with regulations, recover from disaster, and enable a more digital way of working.

IRON MOUNTAIN®

+44 (0) 1782 654 710 | ironmountain.com/en-gb

R.O.I. 1800 732 673 | N.I. +44 (0) 1782 654 710 | ironmountain.com/en-ie