5 security risks of Generative AI and how to prepare to address them.

Here are the perspectives from leaders in the Generative AI field – both AI application developers and cybersecurity professionals – on the security risks posed by AI!

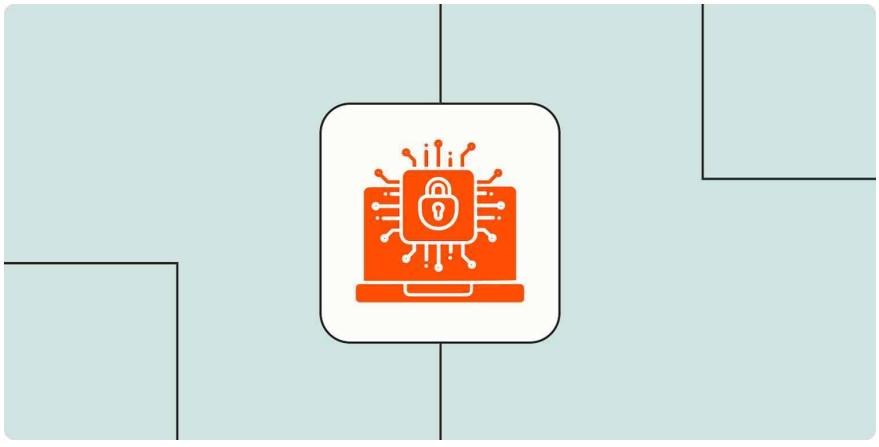
Artificial intelligence (AI) is everywhere. Thousands of new AI applications are launched every day. And there are constant reminders that if you don't keep up with AI trends, you'll fall behind. But don't let the pressure to jump on the "AI train" make you ignore the real cybersecurity risks.

We are integrating AI into browsers, email, and document management systems. We are giving it autonomy to act on our behalf as a virtual assistant. We are sharing personal and work-related information with it. All of this is creating some new cybersecurity risks and amplifying the risks of traditional cyberattacks.

Here are the perspectives from leaders in the Generative AI field – both AI application developers and cybersecurity professionals – on the security risks posed by AI!

Risks from Generic AI

First, let's consider five cybersecurity risks associated with using Generative AI tools.



1. Poor development process

The speed at which companies can deploy Generative AI applications is unprecedented in the software development world. Conventional software development controls and lifecycle management may not always be fully implemented.

According to Adrian Volenik, founder of aigear.io, *"It's incredibly easy to disguise an AI application as a genuine product or service, when in reality, it's created in just one afternoon with no oversight or concern for user privacy, security, or even anonymity* . "

Greg Hatcher, founder of White Knight Labs, a cybersecurity consulting firm, agrees:

"There's an AI craze. On Twitter and LinkedIn, there are countless messages promoting new AI applications that someone wrote the night before. And you have to use them or you'll be left behind. But it's just a scam. They're just using HTML and JavaScript to create the user interface, but under the hood, it's just ChatGPT."

2. Increased risk of data leaks and identity theft.

When we share personal or business information with any software application, we trust that the company handles the information responsibly and has robust protections against cyberattacks. However, with Generative AI tools, we may inadvertently share more than we realize.

Ryan Faber, founder and CEO of Copymatic, warns: *"AI applications are actually accessing our users' data to extract crucial information to enhance the user experience. The lack of proper processes for how data is collected, used, and disposed of raises some serious concerns."*

3. Poor security within the AI ??application itself.

Adding any new application to the network creates new vulnerabilities that can be exploited to gain access to other areas of your network. Generative AI applications pose particular risks because they contain complex algorithms that make it difficult for developers to identify security flaws.

Sean O'Brien, founder of the Yale Privacy Lab and a lecturer at Yale Law School, stated:

"AI is not yet sophisticated enough to understand the complex nuances of software development, which makes its code vulnerable. Research assessing the security of code generated by GitHub Copilot shows that nearly 40% of top AI proposals, as well as 40% of all AI proposals, lead to code vulnerabilities. Researchers also found that small, non-semantic changes such as comments can affect code security."

O'Brien shared several examples of how these risks could manifest:

1. If AI models can be tricked into misclassifying dangerous input as safe, an application developed using this AI could execute malware and even bypass security controls to grant the malware higher privileges.
2. AI models lacking human oversight can be vulnerable to data poisoning attacks. If a chatbot developed with artificial intelligence is asked about the menu of a local restaurant, where to download a privacy-respecting web browser, or the most secure VPN to use, users could be redirected to fake websites containing ransomware .

4. Data leaks expose confidential company information.

If you've been using AI tools for a while, you probably know the role of writing a good prompt in achieving quality results. You provide the AI ??chatbot with background information and context to get the best response.

However, you might be sharing proprietary or confidential information with an AI chatbot—and that's not good. Research conducted by Cyberhaven, a data security company, shows that

1. 11% of the data that employees paste into ChatGPT is confidential information.
2. 4% of employees have pasted sensitive data onto it at least once.

Employees are sharing company intellectual property, sensitive strategic information, and customer data.

Dennis Bijker, CEO of SignPost Six, an internal risk consulting and training firm, reiterated:

"The most concerning risks for organizations are data privacy and intellectual property leaks. Employees could share sensitive data with AI-powered tools, such as ChatGPT and Bard. Think about the potential trade secrets, confidential information, and customer data fed into these tools. This data could be stored, accessed, or misused by service providers."

5. Using malicious Deepfakes

Voice and facial recognition are being used more and more as a security measure for access control. AI presents an opportunity for malicious actors to create deepfakes to bypass those security systems.

1. This dangerous app can "penetrate" women's clothing in just seconds thanks to deepfake technology.
2. Deepfake – a nightmare for women when photos posted online turn into nude images.
3. Deepfake impersonates a bank executive and steals $35 million from the bank.

How to enhance security in the age of AI.

It will take some time to get used to these new risks, but there are a few places to start.

Research the company behind the app.

You can assess an app's reputation and track record by looking at its other tools and services. But don't assume that a familiar name guarantees an acceptable level of security.

For example, Hatcher said, *"It's not necessarily safer to choose a big name. Some companies protect personally identifiable information and customer data better. In fact, it's very difficult to get a malicious iOS app onto Apple's App Store. But it's quite easy on the Android Play Store and Google Play."*

You also need to consider the company's privacy policies and features. Information you share with the AI ??tool can be added to its large language model (LLM) , meaning that information may appear in responses to other people's questions.

Hatcher recommends that you ask the company to provide a certificate stating that the app's security has been audited by verified third parties.

Train employees on how to use AI tools safely and correctly.

You already have a sound social media usage policy in place for your employees, and you should train them on good cybersecurity practices. The widespread use of Generative AI tools means adding some new policies and training topics to that framework. These could include:

1. What they can and cannot share with Generative AI applications.
2. An overview of how LLMs work and the potential risks of using them.
3. Only allow the use of approved AI applications on company devices.

Consider using a security tool designed to prevent excessive sharing.

As the production of Generative AI tools continues to grow, we will soon see a growing collection of cybersecurity tools specifically designed to address their vulnerabilities. LLM Shield and Cyberhaven are two tools designed to help prevent employees from sharing sensitive or proprietary information with Generative AI chatbots. This article does not endorse any particular tool, but simply wants to let you know that this market exists and will continue to grow.

You can also use network auditing tools to monitor which AI applications are currently connecting to your network.