How CISOs can balance the risks and benefits of AI

Rapid growth and development of AI is pushing the limits of cybersecurity and CISOs must take charge now to be ahead of a range of risks including data leak, compliance and prompt injection attacks.

The rapid pace of change in AI makes it difficult to weigh the technology's risks and benefits and CISOs should not wait to take charge of the situation. Risks range from prompt injection attacks, data leakage, and governance and compliance.

All AI projects have these issues to some extent, but the rapid growth and deployment of generative AI is stressing the limits of existing controls while also opening new lines of vulnerability.

If market research is any indication of where the use of AI is going, CISOs can expect 70% of organizations to explore generative AI driven by the use of ChatGPT. Nearly all business leaders say their company is prioritizing at least one initiative related to AI systems in the near term, according to a May PricewaterhouseCoopers' report.

The reason for the investment boom isn't just defensive. Goldman Sachs predicts that generative AI could raise the global GDP by 7%. According to McKinsey, the top AI use cases are in customer operations, marketing and sales, R&D, and software engineering. In software, for example, a survey by global strategy consulting firm Altman Solon shows that nearly a quarter of tech firms are already using AI for software development, and another 66% are likely to adopt it within the next year.

AI-driven cyberattacks

According to Gartner, 68% of executives believe that the benefits of generative AI outweigh the risks, compared with just 5% who feel the risks outweigh the benefits. However, executives may begin to shift their perspective as investments deepen, said Gartner analyst Frances Karamouzis in the report. "Organizations will likely encounter a host of trust, risk, security, privacy and ethical questions as they start to develop and deploy generative AI," she said.

One of the newest risks is that of prompt injection attacks, a brand-new threat vector for organizations. "It's a new attack vector, a new compromise vector, and legacy security controls aren't good enough," says Gartner analyst Avivah Litan. In other cases, chatbot users have been able to see others' prompts, she says.

Many public instances of "jailbreaking" ChatGPT and other large language models have been seen tricking it into doing things they're not supposed to do -- like writing malware or providing bomb-making instructions. Once enterprises start rolling out their own generative AIs, such as for customer services, jailbreaks could allow bad actors to access others' accounts or perform other harmful actions.

Earlier this year, OWASP released its top ten list of most critical vulnerabilities seen in large language models, and prompt injections were in first place. Attackers could also exploit these models to execute malicious code, access restricted resources, or poison training data. When companies deploy the models themselves they have the ability to put firewalls around the prompts, and observability and anomaly detection around the prompt environment. "You can see what's going on, and you can build controls," Litan says.

That's not necessarily the case for third-party vendors. Even if a vendor has top-notch security controls on the training data initially used to create the model, the chatbots will need access to

operational data to function. “The legacy protection controls aren’t applicable to the data going into the model and to the prompt injections,” Litan says. “It really makes sense to keep all this on premise -- but you have to put the protections in place.”

Mitigating data exposure risk from using AI

Employees love ChatGPT, according to a Glassdoor survey of more than 9,000 US professionals that found 80% were opposed to a ban on the technology. But ChatGPT and similar large language models are continuously trained based on their interactions with users. The problem is that if a user asks for help editing a document full of company secrets, the AI might then learn about those secrets -- and blab about them to other users in the future. “Those are very valid, realistic concerns,” says Forrester Research analyst Jeff Pollard.

“We’ve seen doctors taking patient information and uploading it to ChatGPT to write letters to patients,” says Chris Hodson, CSO at Cyberhaven.

Platforms designed specifically for enterprise use do take this issue seriously, says Forrester’s Pollard. “They’re not interested in retaining your data because they understand that it’s an impediment to adoption,” he says.

The most secure way to deploy generative AI is to run private models on your own infrastructure. However, according to Altman Solon, this isn’t the most popular option, preferred by only 20% of companies. About a third are opting for deploying generative AI by using the provider’s own environment, leveraging public infrastructure. This is the least secure option, requiring the organization to place a great deal of trust in the generative AI vendor.

The biggest share of enterprises, 48%, are deploying in third-party cloud environments, such as virtual private clouds. For example, Microsoft offers secure, isolated ChatGPT deployments for enterprise customers in its Azure cloud. According to Microsoft, more than 1,000 enterprise customers were already using ChatGPT and other OpenAI models on the Azure OpenAI Service in March, and the number grew to 4,500 by mid-May. Companies using the service include Mercedes-Benz, Johnson & Johnson, AT&T, CarMax, DocuSign, Volvo, and Ikea.

AI risks in governance and compliance

The record-breaking adoption rate of generative AI is far outpacing companies’ abilities to police the technology. “I know people who are saving large amounts of time every week in their jobs and nobody in those organizations knows about it,” says Gerry Stegmaier, a partner focusing on cybersecurity and machine learning at global law firm Reed Smith LLP. “Businesses are getting radical individual productivity improvements today at the individual employee level -- but are for the most part not aware of the productivity gains by their employees.”

According to a Fishbowl survey released in February, 43% of professionals have used tools like ChatGPT, but nearly 70% of them did so without their boss’ knowledge. That means enterprises may be taking on technical debt in the form of legal and regulatory risk, says Stegmaier -- debt that they don’t know about and can’t measure.

A recent report by Netskope, based on usage data rather than surveys, shows that ChatGPT use is growing by 25% month-over-month, with 1% of all employees using ChatGPT daily. As a result, about 10% of enterprises are now blocking ChatGPT use by employees.

The lack of visibility into what employees are doing is just half the battle. There's also the lack of visibility into laws and regulations. "Large enterprises need a certain amount of predictability. And right now the uncertainty is immeasurable," Stegmaier says.

There's uncertainty about intellectual property and the training data that goes into the models, uncertainty about privacy and data security regulations, and new legal and compliance risks are emerging all the time. For example, in June, OpenAI was sued for defamation and libel after it said that a radio host had embezzled funds. OpenAI or other companies sued for what their chatbots tell people may or may not be liable for what the bots say. It depends on how product liability laws apply. "If the money is real enough, people get creative in their legal theory," Stegmaier says.

There have been some changes regarding whether software is a product or not, and the potential implications could be monumental, according to Stegmaier. There's also the potential of new laws around data privacy, including the EU's AI Act. But he doesn't expect similar laws in the United States in the near future because it's hard to get consensus, but the FTC has been issuing statements regarding AI. "AI is very sexy for consumers, sexy for business, sexy for regulators," he says. "When all three of those things come together there's often a tendency for new enforcement or new regulatory activity to happen."

To stay ahead, he recommends that organizations ramp up their generative AI learning curve so that they can apply their current best-practice tools, including privacy by design, security by design, and anti-discrimination principles. "With respect to generative AI, 'run fast and break stuff' is not going to be an acceptable strategy at scale, especially for large enterprises," Stegmaier says.

Unfortunately, depending on the deployment model, companies may have little or no visibility into what's going on with generative AI -- even if they know it's happening. For example, if an employee asks ChatGPT's help writing a letter to a customer, then ChatGPT will need to get some information about the customer, at the very least while coming up with its answer. That means that, for some period of time, the data will be on OpenAI's servers. Plus, if the employee has ChatGPT save the conversation history, the data will remain on those servers indefinitely.

The data movement issue is particularly important in Europe and other jurisdictions with data residency laws, says Carm Taglienti, distinguished engineer at Insight Enterprises. "You don't really understand where it goes," he says. "You don't know what operations occurred on the data you submitted. Once it's out of your control, it's a vulnerability." He recommends that organizations give some serious thought to the controls they need to have in place if they plan to use generative AI. One place to start is NIST's AI Risk Management Framework, he suggests.

Ideally, enterprises should think about governance issues before they pick a platform. However, according to a KPMG survey, only 6% of organizations have a dedicated team in place to evaluate the risk of generative AI and implement risk migration strategies. In addition, only 35% of executives say their company plans to focus on improving the governance of AI systems over the next 12 months, and only 32% of risk professionals say they're now involved in the planning and strategy stage of applications of generative AI. Finally, only 5% have a mature responsible AI governance program in place -- though 19% are working on one and nearly half say they plan to create one.

What should enterprises do first? "My recommendation is that CSOs immediately begin educating their employees on the potential risks of generative AI usage," says Curtis Franklin, principal analyst for enterprise security management at Omdia. "It's unlikely they'll be able to stop it, but they need to let their employees know that there are risks associated with it. That's immediate. By the time you finish reading this article you should be thinking about how to do that."

The next step is to form a committee that involves stakeholders from different business units to look at how generative AI can legitimately be used in the organization -- and to begin balancing these benefits against the risks. “You should have a risk-based framework on which to make decisions about how you’re going to use it and protect the organization against potential misuse or poor use,” Franklin says.