ARTIFICIAL INTELLIGENCE AND THE FUTURE OF RISK

As companies speed toward investment and adoption of generative artificial intelligence (GenAI) applications, government and industry increasingly recognize that adopting the technology also creates disruptions. A new report by the National Institute of Standards and Technology (NIST) emphasized that cybersecurity is one of those disruptive influences at stake.

AI Adoption is Rapidly Changing the Face of Business

The shift toward AI is happening quickly. A little more than a year ago, Microsoft introduced ChatGPT, which easily demonstrated for consumers how GenAI can generate new, unique content from previously learned data. Within months of its introduction, ChatGPT registered more than 100 million unique monthly users.

ChatGPT is not an isolated example. Today, GenAI tools assist businesses by responding to questions in multiple languages, as well as producing images, video and audio that can be indistinguishable from human-produced content. As the implementation of GenAI continues to transform industries, the growth and speed of take-up should also accelerate. However, much like the emergence of the internet, the promise of AI will also generate new risks to be managed.

GenAI and Cybersecurity

NIST's report, *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations,*[1] notes that like other online systems, the confidentiality, integrity or availability of AI-related applications can be subject to conventional security compromises or attacks against platforms that support them. For GenAI, NIST cautions that users should also consider the threat of "abuse violations," where "an attacker repurposes a GenAI system's intended use to achieve their own objectives."

In their basic form, abuse violations can simply leverage GenAI as a tool to carry out cyberattacks efficiently. For example, IBM researchers recently demonstrated how AI can generate phishing emails with click-through rates comparable to human-generated phishing, but at a fraction of the time and effort.

However, the potential for leveraging GenAI goes deeper. Developers attempt to avoid bad outcomes from GenAI usage with rule-based limitations on content that the technology can generate. These limitations rely on the GenAI algorithm's ability to scrutinize the intent of a request.

Success for the bad actor relies on crafting a prompt to avoid this scrutiny. For example, instead of asking the application to "write a phishing email," a threat actor might succeed by requesting "5 examples of phishing emails for my security team to review."

This risk of abuse attacks can escalate quickly with an attack known as "indirect prompt injections." For these attacks, threat actors will bury a malicious instruction within other data, such as an external website or a software program. After a user's benign initial request, GenAI performs a data search, interfaces with the tainted third-party data, and discovers the malicious prompt. This second, embedded prompt instructs the AI ¬technology to disregard the initial request and carry out some unpredicted action, such as bypassing security measures, exploiting vulnerabilities or injecting malware into code.

This could prove problematic for cyber defenders. Given the vast sea of data available to GenAI, indirect prompt injections present the potential for a seemingly eternal attack surface. AI developers often struggle to define how their technology operates, but soon it may become an even bigger challenge for cyber defenders to discover where the algorithm failed.

1. https://csrc.nist.gov/pubs/ai/100/2/e2023/final

Catching Up with the Consequences

While industries race toward adoption of GenAI, government and the tech sector have quickly pursued a flurry of initiatives to address AI risks. Recently, dozens of international governments joined together for a statement, known as the Bletchley Declaration, which affirms responsible development of AI technology as a priority. Several nations that joined the declaration—including the European Union, China, Australia and Japan—have already introduced laws or regulations to oversee AI development.

From the United States, President Biden signed an Executive Order (EO) last November on Safe, Secure, and Trustworthy Artificial Intelligence (AI) to steer AI development. The EO, in part, directs federal agencies to create standards to develop AI securely, to audit AI threats against critical infrastructure, and to build out cybersecurity programs for AI. While these efforts should support responsible AI development, the creation of standards and regulations will also increase regulatory risk for AI developers and users.

Looking Forward

As AI technologies become increasingly sophisticated, use and misuse of the technology will also impact cybersecurity risk. "This technology is still emerging and should only be deployed with abundance of caution," NIST warns in its report. Ultimately, AI likely will streamline operations and help businesses defend against cyber-attacks, while also creating the opportunity for disruption of those same networks.

How Guy Carpenter Can Help

The Guy Carpenter Cyber Center of Excellence is staying abreast of both the development of GenAI on both technology and regulatory fronts, to help our clients better understand and more effectively mitigate the potential risk that comes with this emerging technology. We work with our clients in understanding their cyber risk objectives and concerns, in order to create a tailored reinsurance solution to meet their unique needs.

About Guy Carpenter

Guy Carpenter & Company, LLC is a leading global risk and reinsurance specialist with 3,500 professionals in over 60 offices around the world. Guy Carpenter delivers a powerful combination of broking expertise, trusted strategic advisory services and industry-leading analytics to help clients adapt to emerging opportunities and achieve profitable growth. Guy Carpenter is a business of Marsh McLennan (NYSE: MMC), the world's leading professional services firm in the areas of risk, strategy and people. The company's more than 85,000 colleagues advise clients in over 130 countries. With annual revenue of $23 billion, Marsh McLennan helps clients navigate an increasingly dynamic and complex environment through four market-leading businesses including Marsh, Mercer and Oliver Wyman. For more information, visit www.guycarp.com and follow us on LinkedIn and X.