Government Acquisitions | NVIDIA | DELL Technologies

Cyber and AI: Balancing Innovation and Responsibility

Cyberattacks are increasing, putting the nation at risk and pushing the government to find new, advanced ways to defend systems and networks.

In FY 2023, federal agencies reported 32,211 attacks on critical infrastructure and devices, up 9.9% from the previous year. Most attacks follow standard tactics, but threat actors increasingly use artificial intelligence and generative AI to make incidents harder to track, automating and customizing attacks for a more targeted approach.

Generative AI models, for example, allow adversaries to create sophisticated phishing attacks using social media accounts. The quick pace of AI development also makes securing networks more challenging. AI lets adversaries hit more people faster and more pervasively; this doesn't make the adversaries smarter — they just have better tools.

To counter this, federal agencies can use adversarial AI to detect and defend against these threats but first, they must work with industry to define what secure AI models look like.

TRANSPARENT AI

According to Dr. Art Villanueva, Federal Chief AI Technology Strategist for Dell Technologies, fair AI starts with transparency. Large language models are "black boxes" with minimal visibility because of their complexity. To alleviate this issue, AI needs to be auditable, users must conceptually understand how their tools work, and engineers must employ systems thinking.

"That includes implementing a rigorous validation and verification process," Villanueva said. "You have specific needs and requirements for the AI system, including following safety. It is incumbent on the engineer to guarantee that whatever the AI does and how it performs has traceability to the requirements."

IMPLEMENTING GUARDRAILS AND COLLABORATIVE EFFORTS

Agencies will likely use pre-built AI models from third parties, so visibility into those models and how they integrate into the network is critical. Establishing guardrails around those models will help direct them within set restrictions, said Ryan Simpson, chief technologist at NVIDIA.

"How can we constrain that model through a guard railing system, either to prevent users from asking questions that they shouldn't or prevent the model from responding in ways you don't want it to?" he said.

When behavior is detached from the AI component, it's easier to set human constraints and a "checks-and-balance" system before the input or output goes to the AI.

"That allows us to think of the AI model as an interchangeable component," Simpson said.

IT teams can build tests to ensure guardrails perform correctly, regardless of the model. Think of it as zero trust for AI — don't inherently trust the model's behavior. Build safety guardrails around it to direct the flow of AI tools securely without requiring teams to control every output.

Unbiased AI is also determined by its users. Having a diverse set of users and understanding the user base can help shape guidelines.

“It really is understanding the use case and then the unbiased bounds within that use case,” Villanueva said. “We have to define a set of rules and then build systems to monitor and enforce those rules.”

So, how should those rules be defined, and by whom?

INDUSTRY AND GOVERNMENT COLLABORATION

These are the questions the industry and government are still tackling, and it’s where partnerships like those between GAI, NVIDIA, and Dell Technologies come into play.

Collaborating with industry partners while working with a federal end customer provides greater technical expertise and understanding of those partners’ technologies.

“Being able to work with them and the customer in combination ensures that the outcome that we’re orienting everything around is the right north star,” said Robin Braun, vice president of AI and data strategy at GAI. This way, everyone involved is represented while tackling these larger, complex challenges.

AI isn’t a one-provider solution. As agencies implement AI tools into their technology stack, that provider base will grow as use cases expand. Industry must work together to ensure these tools cooperate, remain within defined guidelines and continue to operate securely.

From a cybersecurity standpoint, “the second you solve a problem, your adversaries are going to change their approach,” Villanueva said. “It’s not a one-stop solution, and having a deep understanding of your customer is critical to that evolutionary process.”

That’s why industry partners work together to develop AI-driven cybersecurity solutions. Dell Technologies, for example, enables AI capabilities for customers through what it calls AI factories. It’s a framework to accelerate AI adoption by focusing on use cases and data, fostering a broad ecosystem of partners.

“No AI solution can be solved and developed by a single partner or a single capability,” said JP Marcelino, business development lead for AI/ML/DA, Digital Engineering for Dell Technologies.

Dell Technologies focuses on a broad portfolio of infrastructure and hardware solutions that bring AI to where the data resides. This way, agencies can bring in expertise from other industry partners, like NVIDIA or GAI, to form an end-to-end AI software ecosystem.

Connect with the experts at NVIDIA, GAI and Dell Technologies to start your journey toward responsible AI.