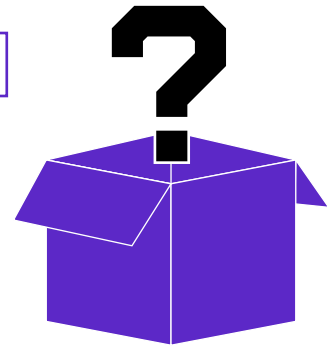


AI and Cybersecurity: Working With the 'Known Unknowns'

As agencies grow more familiar with artificial intelligence (AI), they see its cybersecurity implications more clearly. Fears of world-ending machines are giving way to more focused concerns — and the threats aren't always what agencies had expected.

In a recent GovLoop roundtable, government and industry experts talked about the cybersecurity and AI issues that worry them most. Here are highlights from their conversation.



PARTICIPANTS

Taka Ariga

Chief Data Officer, Acting Chief AI Officer, Director of Enterprise Data, U.S. Office of Personnel Management (OPM)

Thomas Dempsey

Branch Chief, Cyber Risk Management, U.S. Customs and Border Protection (CBP)

Abigail Haddad

Machine Learning Engineer, AI Corps, U.S. Department of Homeland Security (DHS)

Amy S. Hamilton

Visiting Faculty Chair, from the U.S. Department of Energy to Department of Defense National Defense University, College of Information and Cyberspace

Lucy Hyde

Data Scientist, U.S. Customs and Border Protection (CBP)

Rudolf Rojas

Information Technology Management/Systems Analysis, U.S. Department of Agriculture (USDA)

Brian Tirsch

Chief Security Architect, Microsoft Federal

DATA LEAKAGE

One longstanding concern with generative AI (GenAI) is data leakage — when using AI exposes data it shouldn't. That can happen when employees enter confidential data into a publicly available AI program that may retain those inputs and spit them out later in answer to other users' queries.

DHS' Abigail Haddad calls it "shadow IT" and said that "when you're not giving access to tooling [employees] want to do their jobs, they're going to commercial tools where we don't want them putting data," she said.

Data also leaks when an AI application, such as a customer-side chatbot, is designed to search an agency's databases for answers and accesses confidential information. "We will have to think really hard about making sure we're not giving these tools access to [data] that they shouldn't have," Haddad said.

But what do you do when you don't even know the AI is there? Hidden IT is a worry for Taka Ariga at OPM. "I'm not too concerned about the AI applications that we know about," he said, "but given the hype cycle of AI, every application now has an AI label on top of it."

Worse, the features are often turned on without notifying the user, which raises potential cybersecurity concerns, he said. And because "AI" often is more a marketing label than a real functionality, agencies must figure out whether risk is real.

Possibly the worst leak risk related to GenAI is prompt-hacking, in which bad actors find ways to ask for answers they shouldn't get, warned Microsoft's Brian Tirsch. Last year, there were several reports of users tricking ChatGPT into revealing how to make a bomb, despite guardrails that should have prevented it. In theory, those techniques could be used against agency chatbots for more sophisticated attacks.

And more sophisticated attacks are coming — including those assisted by AI. "A lot of nation states, like Russia, Iran, China and North Korea, are using bots to really hammer us," said Rudolf Rojas of USDA.

DATA GOVERNANCE

All of this underscores the fact that safely using AI calls for state-of-the-art data governance, and agencies are working to make sure that the right protections are in place as they make AI applications accessible to employees.

“One of the most important things is nailing down your data within your organization, keeping it structured and confined so you have good data that you’re feeding these models,” said Rojas.

That also includes identifying and isolating the data AI can reach. “People now focused on preparing for AI are mostly doing data labeling and classification protections,” said Tirch. Where agencies haven’t kept up best practices, the push for AI provides new incentives. “AI has forced data governance to be where it was supposed to have been for a long time,” he said.

Securing AI, Lucy Hyde of CBP noted, also calls for securing the whole chain of data. “AI exists in the context of everything that came before it. You can’t just secure your model. You need to secure critical infrastructure, your training model, and [out to] your edge endpoint user.”

WHAT IS GenAI GENERATING?

For CBP’s Thomas Dempsey, checking what AI generates includes focusing on classification. You may provide the AI with data at one classification status, but the material it generates could turn out to require a higher level. “Are there checks in place to assure that doesn’t happen?” he asked. “If we don’t have those things in place, we can see the threat landscape just blow up. The worst thing to do in cybersecurity is not know what’s going on in your environment.”

That leads to a related issue — retaining the skills, such as coding, needed to make those checks. “It comes back to user training,” Dempsey said, and having a test environment “to make sure when we deploy [AI], it’s using what we intended it to.”

Amy Hamilton, from the National Defense University, cautioned against over-reliance on GenAI. “I was talking to a colleague who said, ‘I don’t do any coding now without GenAI.’ And the question then is, ‘could you code without [it]?’ Can you make sure there’s not something getting written into that code that is going to be dangerous?” she said.

“If you’re so dependent on GenAI that you’re forsaking your understanding of the math you need, there is no possible way you will ever be able to speak to cybersecurity,” added Hyde. “If you don’t understand those mathematical concepts, the statistics involved, how are you ever going to protect it?”

THE NEED FOR INTERDISCIPLINARY TEAMS

Agencies are also discovering the need for interdisciplinary teams. “We need to expand the knowledge base of traditional cyber folks into more data-type teams,” Tirch said. “Really, cybersecurity is data science.” But teams need to reach beyond that. “You’ll find on the team, it’s not your cyber analyst, it’s the average user [who finds the weaknesses] because a lot of [the threat] is social engineering,” he said.

Individuals need to develop multi-disciplinary skills as well. “In today’s world, we need a T-shaped resource,” said Ariga. The vertical is your identity as a data scientist, cybersecurity, cloud or other specialist. The horizontal component is “understand[ing] how data works and how models work. Otherwise, you end up creating just a bunch of silos.”

All speakers agreed that there’s no going back. “We must know how to use these [AI] tools,” said Hamilton. “And we have to start looking at the tools to make sure they’re doing what we want.”

Learn more: microsoft.com/government

