

Building a Responsible AI

There's no question that AI is the topic of the moment. As agencies dig into its practical uses, they're developing a concept of "responsible AI" that allows them to take advantage of its benefits while minimizing its potential risks.

During a recent GovLoop roundtable, government and industry experts discussed the policies and practices that define responsible AI. Here are some of their thoughts.



Tim Ahrens

Division Chief, U.S. Department of State



Kristin Hempstead

Public Sector Lead for Advanced Compute Solutions, HP, Inc.



Howard Spira

Chief Information Officer, Export-Import Bank of the United States



Apostol Vassilev

Research Team Supervisor, Computer Security Division, National Institute of Standards and Technology (NIST)

Don't Expect One-and-Done

"There are multiple technology spaces where our adversaries are constantly adapting and evolving. AI is one example," said State's Tim Ahrens. "It's nearly impossible to tighten down the security so much that they're not going to get in, so it's more about awareness, continuous monitoring and having the right alerts in place across our systems."

For AI, that means keeping track of the prompts entered into the system and setting alerts for those that violate policies. And because agencies often give AI access to all their data from different sources to run analytics, responsible AI also involves making sure any sensitive data is protected.

"You need to be aware of how you're securing that data at rest, how you're monitoring and protecting it," Ahrens said. Even so, he added, you need systems in place to respond when — not if — there's an incident: "Let's be proactive and as quick [to react] as we can be."

Take Trust Seriously

"You cannot have responsible AI if you don't have trustworthy AI," said Apostol Vassilev. What makes trustworthy AI? "At NIST, we define it in terms of seven characteristics."

"First of all, it has to be valid. You don't want it to hallucinate and give you bogus information," he continued. "It has to be safe — not giving you directions that will harm you psychologically, physically or in any other way."

"It has to be secure and resilient, such that attackers cannot abuse it to change the behavior in a way that they want but you don't."

"It has to be transparent and explainable. You need to know how the information it generates is generated," Vassilev said. "It has to guard the privacy of people and organizations whose data is used to train the AI. And it has to have harmful bias mitigated, so you don't have preferential treatment of certain groups of people at the expense of others."

"If you have all these attributes, it [will] be trustworthy [and] then you can hope for responsible deployment," Vassilev said of AI.

But he offered a note of caution: "The trouble with these attributes ... is that they're not independent." Prioritizing one may mean less robustness in another — for example, maximizing validity and accuracy may come at a cost of diminishing the ability to block malicious use. "It's a tradeoff relationship," he said.

And tradeoffs among the seven attributes can be complex. "That's when people need to take over the process and decide which tradeoff they're comfortable with," he said. That can depend on laws, regulations or even local or agency customs.



Think of AI in Terms of Culture

“I think responsible AI ... is choices about producing technology that comports with a value system,” said Howard Spira from the U.S. Export-Import Bank. That goes beyond compliance with government guidance and requires fitting the technology to the culture where it will be used.

Banking is inherently conservative, noted Spira. Trust and confidence are core to the industry. Given the nature of banking, controls and cyber security have always been highly intertwined into the practice of bank technology. In many cases, the AI use cases are about how to protect against various forms of electronic fraud and theft.

The newest forms of AI (large language models and all their derivatives) have made industry professionals look very carefully at issues around identity and adjacent abuse cases. “AI is unique because the essence of AI is to demonstrate human-like behavior. In a world where more and more of our interactions are disintermediated with technology, do you really know who you are interacting with and their motivations? Can you trust that your resulting banking transaction hasn’t been swayed or modified by an interaction with AI?” he said.

Having worked internationally outside of the US and Europe, “responsible AI” can also mean different things in different places, Spira added. Looking at NIST’s seven characteristics, for example, “I don’t know that all societies have a consensus on what a ‘harmful bias’ looks like. There are several of the seven NIST characteristics whose outcomes are very tied into notions of culture,” he said.

Start at the Beginning

Responsible AI starts with good data, said HP’s Kristin Hempstead: “The data must be high quality, nonbiased and completely vetted.” It also needs to be specifically chosen for your purpose. “It’s imperative to start with a good pool of data, [one] that’s very specific to the planned end-to-end strategy,” she added.

That requires making everyone in the organization competent and comfortable using AI, so they can pose the right questions. “It can’t just be the AI team,” she explained. “It has to be everybody. You have to upskill your staff in understanding the good parts of AI, the bad parts of AI, and what they, as responsible users, should and shouldn’t be doing around the data and AI they are incorporating into their workflows.”

Hempstead also recommends starting small — but thinking big.

“Pilots and proof-of-concept [projects] are absolutely pivotal,” she said. “But ensure the [pilots] have scalability and adaptability to allow for the project to be rolled out to a wider group or across an agency, if the use case fits. Doing that proactive work in the beginning can save loads of time.”

It also helps to take advantage of what others are learning, Hempstead said. “Use your resources. Talk to other agencies. Make sure you understand all your agency’s regulations and requirements. Talk to industry experts. Talk to the private sector. Talk to other countries,” she said. “Everybody’s doing AI and they all can offer valuable insights that could help an agency with their AI journey.”

