A **CENTER FOR DIGITAL GOVERNMENT** THOUGHT LEADERSHIP PAPER | IBM

Safeguarding Government in the Era of GenAI

An action plan to ensure your agency can mitigate evolving risks

Generative AI (GenAI) is changing the security landscape for governments.

The challenge for policymakers, agency leaders, technologists, security professionals and regulators is to mitigate GenAI risks while encouraging innovation around this powerful technology.

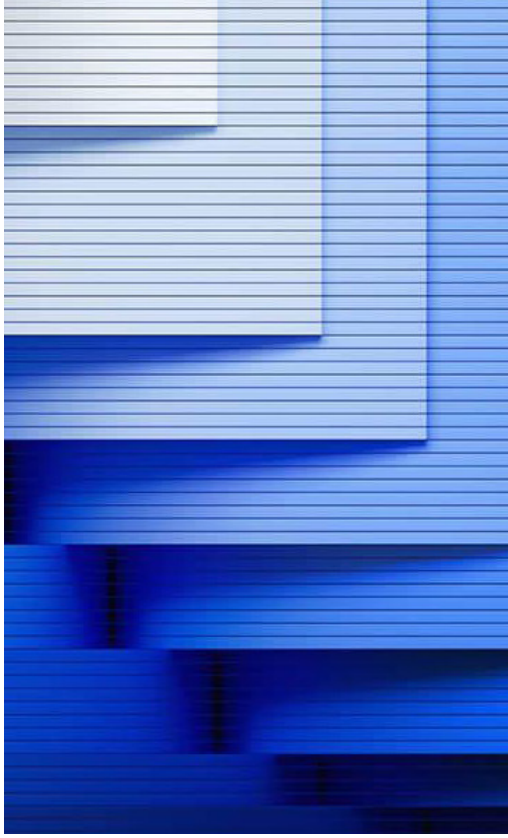
GenAI will be a valuable tool for helping governments drive automation, boost productivity and surface insights that support more impactful constituent service. However, GenAI comes with concerns that are distinct from those posed by traditional AI. Hallucinations, misinformation and data poisoning are just some of them.

“This is a big business risk that we as security professionals must put guardrails around,” says Chip Crane, security leader for the Americas at IBM.

Governments need to define effective control environments for GenAI. But current security approaches are often fragmented and misaligned. State and local agencies must enact more robust security governance, particularly around how data is deployed within AI systems.

“Organizations need to double down on data security and data governance,” says Joe Daw, principal security architect for the public sector at IBM. “We can’t protect all things equally, so we need to focus on what we really care about, and that’s our data.”

In addition, security practices will need to constantly evolve to balance GenAI benefits with potential pitfalls.



Understanding GenAI Cyber Risks

GenAI creates a range of security issues — both as a powerful new weapon for cyber attackers and a useful tool for agencies that requires appropriate safeguards.

One of the characteristics that makes Gen AI so perilous is how it enables hackers to execute continuous attacks with greater speed, accuracy and specificity.

Crane says this has contributed to the emergence of adversarial AI, which uses machine learning models to perpetrate personalized attacks on targets both large and small. A broad range of attacks fall under this category, including sophisticated phishing attacks that encompass conversational phishing, advanced malware and deepfakes.

“Think of a threat that is tailored precisely to you,” he says. “This is a huge challenge, because even the best of us can let our guard down for just a moment and get sucked into a conversation that we believe is with a human that really isn’t.”

GenAI can generate human-like texts, images and other content to manipulate users, causing them to click a link or engage with malicious content and unwittingly provide unauthorized access to

Agencies must enact more robust security governance, particularly around how data is deployed within AI systems.

systems. Hackers are also using AI to level up malware attacks, automatically producing complex code that can evade detection and exploit network vulnerabilities.

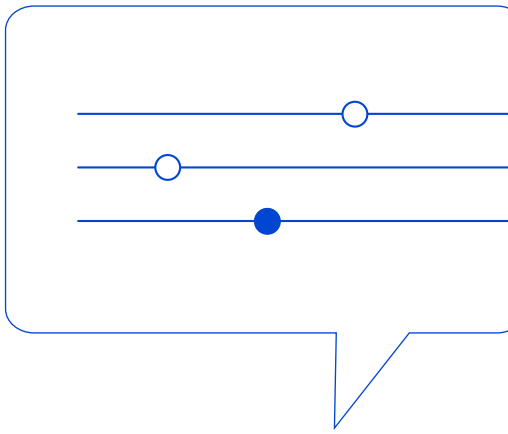
Data poisoning is another risk, says Deborah Snyder, a Center for Digital Government senior fellow and former chief information security officer for New York State.

"Adversaries can use GenAI techniques to poison and manipulate training data that's used by AI systems," Snyder says, which can lead to data bias and inaccurate predictions or decisions.

Beyond its potential for weaponization, AI also poses some inherent risks that organizations need to address. The first is privacy. AI models are trained on data, and organizations must prevent the inadvertent release of confidential information. AI models and algorithms also rely on programming code, which hackers can exploit for malicious purposes.

In this environment, governments can't afford to leave anything to chance.

"We have to close every hole because adversaries will use AI to keep pinging us and working us until they find whatever little gap that we leave," Crane says. "We have to be diligent at a level that we've only talked about in the past."



GenAI can generate human-like texts, images and other content to manipulate users.

An Action Plan for Addressing GenAI Risks

State and local governments must understand their vulnerabilities and execute a nimble cybersecurity strategy to keep pace with GenAI threats. This requires agencies to build on existing security controls, embrace integration for more effective security orchestration, and educate and train staff on GenAI risks.

Here are some best practices for refining your security strategy:

→ **Assess your risk:** Understand your current AI tools and the risks associated with them. Also inventory your security solutions to understand their capabilities and their effectiveness against AI-related threats.

Map where highly sensitive data resides within your ecosystem and identify which AI systems can access this information. Conduct penetration testing and create a capabilities matrix to uncover vulnerabilities.

"You need to understand the gaps you have, and then ask yourself, 'What do I need to do to get prepared?'" Daw says.

→ **Adopt an open, integrated security approach:** No single tool will completely protect agencies from GenAI risks. Integration is key. You'll need an open, integrated security apparatus that

leverages advanced capabilities across your ecosystem of security solutions. Look for solutions and services with these characteristics and capabilities:

- **Open source technology:** Consider platforms that leverage open source models to provide industry-leading, AI-enabled security capabilities from best-of-breed technologies.
- **Indemnification:** Advanced security tools will increasingly incorporate GenAI. Make sure these solutions indemnify your agency against potential legal liabilities around copyright and intellectual property violations. Indemnified models foster trust and transparency, giving organizations confidence they can safely leverage third-party AI models to strengthen security.
- **Out-of-the box security:** Prebuilt solutions that target ransomware and other threats combine security technologies and monitoring services to provide affordable protection for agencies with smaller budgets and IT teams.
- **Advanced threat protection and response:** Robust authentication and authorization tools that employ Zero-Trust principles and continuous monitoring help agencies mitigate GenAI risks. "The swifter you are in detecting anomalous behavior, the quicker you are to respond and recover from those incidents," Snyder says.
- **Threat intelligence feeds:** Third-party threat intelligence provides a holistic, real-time view of the threat landscape. IBM, for example, monitors GenAI threats and ranks them according to "Exploit Difficulty and Potential Impact" to help organizations prioritize their defense.

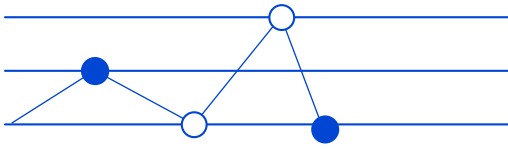
→ **Strengthen data governance:** Data is both foundational to AI and a crown jewel hackers seek to exploit. State and local governments must evolve their data management processes and risk mitigation measures to enact more comprehensive governance.

"Organizations often forget about the importance of comprehensive data governance and data protection — formal processes and risk mitigation measures related to data," Snyder says. "Do a data inventory. Questions to ask include: What data do we collect and retain? What are we doing to protect our valuable and sensitive data? What controls are in place, such as encryption, network isolation and access controls? Who can use the data and for what purpose? Where is the data shared? Where is it leaving our custody into the hands of others? What regulatory compliance and privacy concerns do we need to address?

"Good, comprehensive data governance goes a long way," she says.

Data governance policies also need to incorporate AI governance frameworks that emphasize accountability, fairness and transparency. IBM recently released the "IBM Framework for Securing Generative AI,"[1] which includes three key pillars: securing the data, securing the AI model and securing the usage of GenAI.

Data is both foundational to AI and a crown jewel hackers seek to exploit.



This framework provides a good launching point to craft your own internal policies and guidelines for the responsible use of AI.

→ **Keep humans in the loop:** Humans must oversee decisions and outputs of GenAI systems. Consider designating members of your security team and other key stakeholders as backstops to ensure the outputs of AI systems are reliable and explainable. Crane says analysts in a jurisdiction's security operations center are well-positioned to review AI system outputs.

Snyder adds that human-in-the-loop feedback will be vital as the use of Gen AI grows within government.

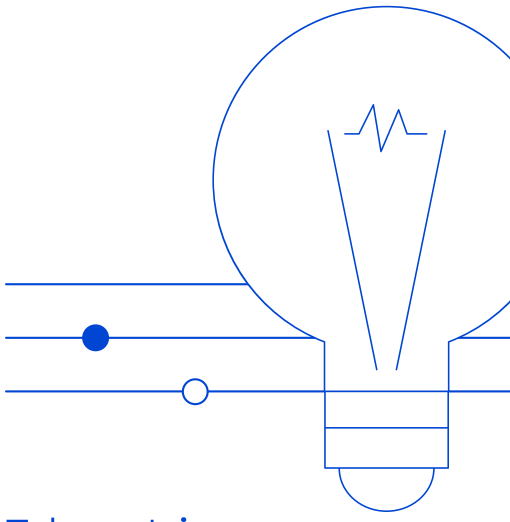
"I'm a firm believer that human oversight will remain critical in ensuring the reliability, safety and ethical use of AI, especially now as things are still in flight. Gen AI technologies are still evolving," she says.

→ **Build an education curriculum:** Technology is only one piece of the puzzle for addressing GenAI risks. Governments also need to educate employees about GenAI.

Education programs help governments expand their capacity to manage GenAI risks. These efforts build cyber awareness among business users, ensure internal cyber teams can evaluate and mitigate risks, and help key stakeholders make informed decisions about GenAI implementation.

→ **Stay flexible:** GenAI won't remain static, so your agency's security program needs to evolve as new GenAI threats emerge.

Conduct regular security audits and AI code reviews to identify new vulnerabilities and proactively address these risks. This is where threat intelligence feeds and knowledge-sharing with peer organizations are invaluable. In addition, work with your vendors to understand their security posture and to ensure their AI tools don't introduce vulnerabilities.



Education programs help governments expand their capacity to manage GenAI risks.

An Agile and Proactive Approach

GenAI is new, but you don't need to completely rewrite your security approach to address GenAI threats. Double down on what already works and incorporate new tools that specifically address GenAI-related vulnerabilities.

"In the short to medium term, we don't believe generative AI is dramatically going to change the types of attacks we tend to see," Crane says. "But it will dramatically scale cybercrimes and lower the barrier to entry for lower-skilled attackers using AI to perform attacks."

GenAI will amplify the frequency and size of cyberattacks, which means agencies should focus on fundamental security principles while also developing more agile and proactive defenses. A flexible, multifaceted security strategy will protect mission-critical systems and assets from future risks — from GenAI or other next-generation technologies.

1. https://www.ibm.com/blog/announcement/ibm-framework-for-securing-generative-ai/

This piece was written and produced by the Center for Digital Government Content Studio, with information and input from IBM.

CENTER FOR DIGITAL GOVERNMENT

Produced by the Center for Digital Government
The Center for Digital Government, a division of e.Republic, is a national research and advisory institute on information technology policies and best practices in state and local government. Through its diverse and dynamic programs and services, the Center provides public and private sector leaders with decision support, knowledge and opportunities to help them effectively incorporate new technologies in the 21st century.

www.centerdigitalgov.com

IBM

Sponsored by IBM
IBM is a leading provider of global hybrid cloud and AI, and consulting expertise. We help clients in more than 175 countries capitalize on insights from their data, streamline business processes, reduce costs and gain the competitive edge in their industries. More than 4,000 government and corporate entities in critical infrastructure areas such as financial services, telecommunications and healthcare rely on IBM's hybrid cloud platform and Red Hat OpenShift to affect their digital transformations quickly, efficiently and securely. IBM's breakthrough innovations in AI, quantum computing, industry-specific cloud solutions and consulting deliver open and flexible options to our clients. All of this is backed by IBM's long-standing commitment to trust, transparency, responsibility, inclusivity and service.

www.ibm.com

ADOBESTOCK/AI GENERATED