Debbie Taylor Moore
Vice President and Senior Partner, Consulting Cybersecurity Services
IBM

House Subcommittee on Cybersecurity and Infrastructure Protection
House Committee on Homeland Security
Considering DHS' and CISA's Role in Securing Artificial Intelligence

December 12, 2023

Introduction

Chairman Garbarino, Ranking Member Swalwell and distinguished members of the subcommittee, I am honored to appear before you today to discuss the important topic of cybersecurity and its relationship to and with AI.

My name is Debbie Taylor Moore, and I am VP and Senior Partner for IBM Consulting. I lead the Quantum Safe and Secure AI consulting practice for North America, including the delivery of security consulting services to commercial critical infrastructure and government clients. During my 20+ year career in cybersecurity, I have had the great privilege to participate and witness first-hand, the impact of successful public and private sector partnership. With each innovation we have risen to the occasion and asked ourselves the difficult questions: "how to optimize the promise, while minimizing the peril of technology advancement?" I have also collaborated with the Department of Homeland Security (DHS) since its inception as a federal contractor, a woman-owned small business at an early-stage start-up, and a fortune 100 executive, to today, working at the intersection of security and emerging technology for IBM.

Let me ground my testimony at the outset on three foundational points.

First, AI is not intrinsically high-risk, and like other technologies, its potential for harm is expressed in both how it is used, and by whom. AI risk is not a new story – we've been here before, as any new powerful technology poses both risks and benefits. Like then, we provide appropriate guardrails and accountability for our technology.

Second, the economic potential for AI is phenomenal. Yet, industry needs to hold itself accountable for the technology it ushers into the world. That is part of the reason that IBM recently signed onto the White House Voluntary AI Commitments to promote the safe, secure, and transparent development and use of generative AI (foundation) model technology.

Third, the government has a critical role to play, in collaboration with industry and all stakeholders. The White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence ("EO on AI") assigns DHS and its Cybersecurity and Infrastructure Security Agency (CISA) with tasks to ensure agencies and critical infrastructure providers understand what is needed to deploy AI safely and securely in executing their missions. It also tasks DHS to continue to work with industry through a soon to be developed AI Safety and Security Advisory Board. This subcommittee's hearing and oversight of the implementation of the EO on AI is a critical part of this dialogue.

My testimony will raise awareness and share how organizations today are: A) utilizing AI to improve security operations; B) promoting the trustworthy and secure use of AI broadly; and C) protecting AI in critical infrastructure. Lastly, I will share recommendations.

A. AI for Security

In my work with clients in the public and private sector, I see how deploying AI is helping to enable cybersecurity defenders more effectively and efficiently do their job. AI systems are proving to be security assets that industry is using to bolster existing security best practices regardless of critical infrastructure designation. AI can help to:

- Improve speed and efficiency. When AI is built into security tools, cybersecurity professionals can identify and address, at an accelerated rate, the increasing volume and velocity of threats. For example, machine learning can be used to identify and analyze patterns and key indicators of compromise. Over time the system trains itself on the data it collects, reducing the number of false positives, honing-in on the incidents which require human intervention and investigation. This form of augmentation helps Security Operation Centers personnel who can be overwhelmed by the sheer number of events. In certain cases, IBM’s managed security services team used these AI capabilities to automate 70% of alert closures and speed up their threat resolution timeline by more than 50% within the first year of operation.

- Contextual awareness. Providing context from multiple sources delivers insights, prioritization and offers recommendations for security analysts to follow to remediate issues. For example, generative AI can confidentially and comprehensively answer questions and render responses which make it possible for a junior analyst to achieve higher level skills and complete complex tasks above and beyond current proficiency.

- Improve resilience and response time. For example, AI leverages machine learning algorithms to predict future risk and to develop a consistent risk profile and set of potential actions based on historical data. This predictive modeling helps organizations anticipate problems and proactively address them, reducing mean time to resolution and costs. IBM’s Cost of a Data Breach 2023 report found that using AI was the single most effective tool for lowering the cost of a data breach. The average cost of a data breach is $4.5M dollars; up 15% over the previous year.

B. Promoting the Trustworthy and Secure use of AI

At IBM, we recognize that the use of AI and large language models in an application or system may increase the overall attack surface which must be protected, and that traditional security controls alone, may not be sufficient to mitigate risk(s) associated with AI. That is why we are proud to help clients deploy Trustworthy AI, ready for enterprise use – which means it is fair, transparent, robust, explainable, privacy-protecting, and secure – now and in the future.

Here are examples at how we implement Trustworthy AI practices, including security, at three key touchpoints in client engagements:

First, data -- we use data that is curated, protected, and trusted. Our guardrails help ensure data quality, compliance, and transparency. Data ownership is also extremely important. Our clients trust that their data will not be used by someone else. And we help clients to protect training and sensitive data from theft, manipulation, and poisoning, and compliance violations and to employ zero-trust access management policies and encryption.

Second, AI models -- securing the model development stage is paramount, as new applications are being built in a brand-new way, often introducing new, exploitable vulnerabilities for attackers to use as entry points to compromise AI, introducing the risk of supply chain attacks, API attacks and privilege escalations. For example, we help clients:

- **Secure the usage of AI models** themselves, by implementing security controls for privileged access management, preventing/detecting data leakage, and preventing/detecting new attacks like poisoning (where you control a model by changing the training data), extraction (where you steal a model by using queries), or evasion (where you change the model behavior by changing the input).

- **Secure against new AI generated attacks**, by helping them monitor for malicious activity like using AI to rapidly generate new malware, or to mutate existing examples to avoid detection. Also help clients detect highly personalized phishing attacks and impersonation.

- **Employ red-team testing**: as attack surfaces of AI will continually be uncovered, we are committed to and invested in discovering these to stay ahead of the adversary. We do comprehensive security assessments which simulate a layered attack on an organization's physical systems, data, applications, network and AI programs and assets. Expanding far beyond a routine penetration test or vulnerability assessment, red teaming seeks to offer a learning opportunity while evaluating an organization's response in a crisis. It mimics the tactics, techniques and procedures of known threat actors and helps the organization to identify gaps and improve its security posture. Participation is encouraged across multi-stakeholders and domains.

Third, AI pipeline -- we give clients the tools to extend governance, trust, and security across the entire AI pipeline. Even the most powerful AI models cannot be used if they are not trusted – especially in mission-critical industries. That is why we are creating and using AI governance toolkits to help make them more transparent, secure, and free of bias. Instilling trust in AI is key for AI to be deployed safely and widely. Security, too, must be extended to the inferencing and live use stage of the AI pipeline, to protect against prompt injections, model denial of service, model theft risks, and more, as discussed further below.

C. Protecting AI in Critical Infrastructure

Critical infrastructure underpins the economic safety and the physical well-being of the nation. Adversaries have worked for years to disrupt, exploit, and undermine the safety and security of power grids, air and land transportation systems, telecommunications, and financial networks. Further, we recognize that highly capable AI models that are not developed and deployed with responsible guardrails can today, and could in the future, be modified by bad actors to pose safety risks to these networks from adversarial attacks to deep fakes giving false instructions to undermine industrial control systems.

By "breaking" AI models we can better understand, assess, and clearly define the various levels of risk that governments and critical infrastructure alike need to manage.

Let me explain. To address the security risk of an AI system, we can "breakdown" AI to learn of its potential weaknesses. In addressing security, to protect a system — whether software or hardware — we often tear it down. We figure out how it works but also what other functions we can make the system do that it wasn't intended to. Then, we address appropriately – from industrial/military grade strength defense mechanisms to specialty programs built to prevent or limit the impact of the unwanted or destructive actions. We, collectively as industry and critical infrastructure providers, have the tools to do this – and in many cases are already doing this. We also have the governance and compliance know-how to enforce.

Here are two examples from IBM efforts.

- Through security testing, we discovered that there are ways for adversaries to get a train to derail from its tracks. That know-how allowed us to create preventative ways to stop it from happening in a real-world instance. Same with ATM machines being compromised to eject unsolicited cash. And so forth.

- IBM X-Force research illustrated months ago how an attacker could hypnotize large language models like ChatGPT to serve malicious purposes without requiring technical tactics, like exploiting a vulnerability, but rather simple use of English prompts. From leaking confidential financial information and personally identifiable information to writing vulnerable and even malicious code, the test uncovered a new dimension to language learning models as an attack surface. It is important for government and critical infrastructure entities to recognize that AI adds a new layer of attack surface. We are aware of this risk and can create appropriate mitigation practices for clients before adversaries are able to materially capitalize on it and scale.

Further, the critical infrastructure ecosystem is also aware of the increased risk vectors that could be applied to critical infrastructure due to AI. Critical infrastructure providers are not only taking internal steps, or working with companies like IBM, to address this, but also working with the technology industry, government, and others to set and advance best practices and tools. Here are some examples:

- **Defcon red-teaming**. Thousands of offensive security professionals recently gathered in Las Vegas to attack multiple popular large language models in a bid to discover flaws and exploitable vulnerabilities that could serve malicious objectives or that could otherwise produce unreliable results, like bad math. Those “fire drills” – often called “red teaming” as discussed above – identified risks to be addressed before they could manifest into active threats.

- **Public-private “best practices.”** Government, working closely with industry, has published best practices, guidance, tools, and standards to help bolster our nation’s security. These include: NIST’s Secure Software Development Framework and CISA’s Software Bill of Materials as well as secure development best practices, emphasized in CISA’s Secure by Design Principles and subsequent Guidance to Secure AI Systems, to provide a path for AI models to be built, tuned, trained, and tested following safe and secure best practices.

- **Public-private collaboration and information sharing**. Collaboration vehicles for critical infrastructure providers, industry and government exist already. For example, IBM is pleased to partner, across verticals and industry through collaboration with the private sector led, Information Sharing and Analysis Centers (ISACs). The ISACs are critical collaborators for DHS and CISA to develop proactive, essential platforms to effectively communicate best practices, like those listed above, and outcome from the soon-to-be-launched NIST AI Safety Institute. This Institute will convene experts to set the guidelines for “red teaming” best practices and other similar AI safety standards. CISA has a role here, too. Just as CISA’s Secure Software by Design leveraged NIST’s Secure Software Development Framework, we see a role here for collaboration as well, which we discuss further in the next section.

Recommendations

Addressing the risks posed by adversaries around AI and critical infrastructure will require a combination of smart policy, tight collaboration, and efficient agency execution. Thankfully, the US government is aware that a multi-faceted, multi-stakeholder approach is needed evidenced from the US National Cybersecurity Strategy, the recent EO on AI, and this hearing.

We have a strong foundation to build on. What we need is urgency, accountability, and precision in our execution. Specifically, we encourage:

1. **CISA should accelerate existing efforts and broadened awareness, rather than reinventing the wheel**. CISA is "America's Cyber Defense Agency" chartered to help protect systems of sixteen (16) critical infrastructures sectors, the majority of which are owned and operated by the private sector. As it achieves its mission through partnerships, collaboration, education and raising awareness, as well as conducting risk assessments, risk management, and incident response and recovery, AI security should be embedded into the agencies' work as a top priority. We suggest that CISA:
   a. Execute on its Roadmap for AI. Published in November, this is a great first step. The Roadmap seeks to promote the beneficial uses of AI to enhance cybersecurity capabilities, protect the nation's AI systems from cybersecurity threats, and deter malicious actors' use of AI capabilities to threaten critical infrastructure. Critically it has a component that addresses workforce as well. We strongly support this and hope to see its timely execution.
   b. Elevate AI training and education resources from industry within CISA's own workforce and critical infrastructure that it supports. And, it should accelerate implementation of the National Cyber Workforce and Education Strategy. To help close the global AI skills gap, IBM has committed to training two million learners in AI by the end of 2026.
   c. Advance information sharing. CISA should leverage existing information sharing infrastructure that is sector-based to share AI information, such as potential vulnerabilities and best practices. Also, share outcomes from the NIST Safety AI Institute as well as threat intelligence, as appropriate, from National Security Agency with Federal Civilian Executive Branch Agencies and ISACs to ensure the broadest reach of AI information.
   d. Implement AI Governance. To improve understanding of AI and its risk, CISA needs to know where AI is enabled and in which applications. This existing "AI usage inventory" could be improved through common definitions of AI and its componentry. Ideally, this could then be leveraged to implement an effective AI governance system.
   e. Align efforts domestically, and globally, with the goal of widespread utilization of tools, rather than just their development. For example, encourage the tracking of security requirements, risks, and design decisions throughout the AI lifecycle. CISA has made progress here through its Secure by Design Principles and Guidelines for Secure AI System Development issued this year in collaboration with the UK and other governments across the globe. To increase utilization of these tools, guidance on execution is also important.

2. **The Department of Homeland Security should have a collaborative and strategic AI Safety and Security Advisory Board as directed by the EO on AI.** We recommend that it:

a. Ensure members are a diverse representation of critical infrastructure owners, technologists, security experts, and agency stakeholders to best determine scope of work and mission.

b. Collaborate with existing efforts to leverage learnings and outcomes from the National AI Advisory Committee, NIST AI Safety Institute, and CISA Cyber Safety Review Board. These Board and Committee outputs matter.

c. Rationalize the threat to minimize hype and disinformation. Attention should be directed towards addressing and mitigating material risks. This Advisory Board can help to identify best practices and guidance for securing AI for our government systems and critical infrastructure. Then, it can educate on that and how to address the new threats to our citizens, agencies, and critical infrastructure providers.

3. **The Department of Homeland Security should implement the directives from the EO on AI in a timely manner**. DHS is directed to study how to better use AI for cyber defense and to conduct operational pilots to identify, develop, test, evaluate, and deploy AI capabilities. These capabilities will aid in discovery and remediation of vulnerabilities in critical U.S.G. software, systems, and networks. This subcommittee can invite DHS to present any relevant findings and identify what would be needed to ensure interoperability and scale across government.

Conclusion

I will end where I started, addressing the risks posed by adversaries is not a new phenomenon. Using AI to improve security operations is also not new. Both will require focus on what we have already assembled. We do not need to re-invent the wheel. What we need is urgency, accountability, and precision in our execution.