PROTECT AI

BUILDING A SAFER AI-POWERED WORLD

Hearing on

Considering DHS' and CISA's Role in Securing Artificial Intelligence

The Subcommittee on Cybersecurity and Infrastructure Protection

December 12, 2023, at 10:00 a.m.
Cannon House Office Building
Washington, D.C.

Testimony of Ian Swanson
Chief Executive Officer
Protect AI, Inc.

Good morning members of The Subcommittee on Cybersecurity and Infrastructure Protection. I want to start by thanking the Chairman and Ranking Member for hosting this important hearing and inviting me to provide testimony.

My name is Ian Swanson, and I am the CEO of Protect AI. Protect AI is a cybersecurity company for artificial intelligence (AI), that enables organizations to deploy safe and secure AI applications. Previously in my career, I was a worldwide leader of AI/ML at Amazon Web Services and Vice President of Machine Learning at Oracle. Protect AI was founded on the premise that AI security needed dramatic acceleration. When I first started Protect AI, we had to convince industries that the need for security of AI was necessary. Now, industries and governments are openly talking about this need, and shifting the conversation from education of AI security to building security into AI. Against the backdrop of regulation, more front-page headlines on AI/ML security risks, and proliferation of AI/ML enabled tech to deliver business value, the recognition for securing AI/ML applications has never been greater.

AI is the development of computer systems or machines that can perform tasks that typically require human intelligence. These tasks can include things like understanding natural language, recognizing patterns, making decisions, and solving problems. AI encompasses machine learning (ML), which, according to Executive Order 14110 is "a set of techniques that can be used to train AI algorithms to improve performance on a task based on data." A ML model is an engine that can power an AI application and differentiate AI from other types of software code. For many companies and organizations, AI is the vehicle for digital transformation and ML is the powertrain. As such, a secure ML model serves as the cornerstone for a safe AI application, ensuring reliability and security akin to how robust software frameworks and high-grade hardware fortify an organization's technology ecosystem. This ML model, in essence, is an asset as indispensable as any other technology asset, such as databases, cloud computing resources, employee laptops and workstations, and networks. AI/ML assets have numerous challenges in developing, deploying, and maintaining it securely. These include:

- **Limited Transparency in the Operations of AI/ML Applications:** The complex nature of AI/ML algorithms leads to challenges in transparency, making it difficult to perform audits and investigative forensics of these systems.

- **Security Risks in AI/ML's Open Source Assets:** AI/ML technologies often depend on open-source software, which, while fostering innovation, also raises concerns about the security and reliability of these foundational elements.

- **Distinct Security Needs in AI/ML Development Process:** The process of developing AI/ML systems, from data handling to model implementation, presents unique security challenges that differ markedly from traditional software development.

- **Emerging Threats Unique to AI/ML Systems:** AI/ML systems are susceptible to novel forms of cyber threats, such as algorithm tampering and data manipulation, which are fundamentally different from conventional cybersecurity concerns.

- **Educational Gap in AI/ML Security Expertise:** There is a critical need for enhanced training and expertise in AI/ML security. This gap in specialized knowledge can lead to vulnerabilities in crucial AI/ML infrastructures.

Based on my experience and first-hand knowledge, millions of ML models are currently operational nationwide, not only facilitating daily activities but also embedded in mission-critical systems and integrated within our

physical and digital infrastructure. These models have been instrumental for over a decade in areas such as fraud detection in banking, monitoring energy infrastructure, and enhancing cybersecurity defenses through digital forensic analysis. Recognizing and prioritizing the safeguarding of these assets by addressing their unique security vulnerabilities and threats, is vital for this nation and any organization striving to excel in the rapidly advancing field of AI which impacts all elements of the American economy today, and into the future.

US businesses and the United States Government use a significant number of machine learning (ML) models for critical processes, ranging from defense systems to administrative task acceleration. Given the importance of these systems to a safe, functioning government, we pose a critical question: If this committee were to request a comprehensive inventory of all ML models in use in an enterprise or a USG agency, detailing their stages in the life cycle (including experimentation, training, or deployment), the data they process, and the personnel involved (both full time employees, government personnel, and contractors), would any witness, business, or agency be able to furnish a complete and satisfactory response?

Secure AI and ML requires oversight and understanding of an organization's deployments. However, many deployments of AI and ML are highly dispersed and can be heavily reliant on widely used open-source assets integral to the AI/ML lifecycle. This situation potentially sets the stage for a major security vulnerability, akin to the 'SolarWinds incident', posing a substantial threat to national security and interests. The potential impact of such a breach could be enormous and difficult to quantify.

Our intention is not to alarm but to urge this committee and other federal agencies to acknowledge the pervasive presence of AI in existing US business and government technology environments. It is imperative to not only recognize but also safeguard and responsibly manage AI ecosystems. This includes the need for robust mechanisms to identify, secure, and address critical security vulnerabilities within US businesses and the United States Federal Government's AI infrastructures.

Qualcomm[1], McKinsey & Company[2], and PwC[3] have shared analysis that AI can boost the US GDP by trillions of dollars. We must protect AI commensurate with the value it will deliver. To help accomplish this, AI manufacturers and AI consumers alike should be required to see, know, and manage their AI risk:

- **See.** AI/ML systems are fragmented, complex and dynamic. This creates hidden security risks that escape your current application security governance and control policies. Manufacturers and consumers of AI must put in place systems to provide the visibility they need to see threats deep inside their ML systems and AI Applications quickly and easily.

- **Know.** The rapidly evolving adoption of AI/ML adds an entirely new challenge for businesses to ensure their applications are secure and compliant. Safeguarding against a potential "SolarWinds" moment in ML is business critical. Manufacturers and consumers of AI need to know where threats lie in their ML system so they can pinpoint and remediate risk. They must create ML Bill of Materials, scan, and remediate their AI/ML systems, models, and tools for unique and novel vulnerabilities.

[1] Qualcomm: The generative AI economy: Worth up to $7.9T. Available at https://www.qualcomm.com/news/onq/2023/11/the-generative-ai-economy-is-worth-up-to-7-trillion-dollars

[2] McKinsey and Company: The economic potential of generative AI: The next productivity frontier. Available at https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

[3] PwC: PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution. Available at https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

- **Manage.** AI/ML security vulnerabilities are difficult to remediate. When operational, technological, and/or reputation security risks are identified that could harm customers, employees, and partners, the business must quickly respond and mitigate them to reduce incident response times. Manufacturers and consumers of AI/ML should create documented policies to help improve security postures, employ incident response management processes, enforce human-in-the-loop checks, and meet existing and future regulatory requirements.

Yes, I believe that the government can help set policies to better secure artificial intelligence. Policies will need to be realistic in what can be accomplished, enforceable, and not shut down innovation or limit innovation to just large AI manufacturers. Against this backdrop, the DHS and CISA play a crucial role in fortifying the security of AI applications.

In the past year, CISA has published two important documents with regard to Securing Artificial Intelligence: "Secure by Design" and the "CISA Roadmap for Artificial Intelligence". The Secure by Design document provides a clear articulation of the "Secure by Design" approach, which is a classic and well understood methodology for software resilience. I applaud the work by CISA and support the three "Secure by Design" software principles that serve as their guidance to AI/ML software manufacturers **1/** Take ownership of customer security outcomes, **2/** Embrace radical transparency and accountability, and **3/** Build organizational structure and leadership to achieve these goals. CISA advancing the "Secure by Design" methodology should help foster widespread adoption. Manufacturers of AI/ML must take ownership for the security of their products and be held responsible, be transparent on security status and risks of their products, and build in technical systems and business processes to ensure security throughout the ML development lifecycle - otherwise known as MLSecOps. While "Secure by Design" and the "CISA Roadmap for Artificial Intelligence" are a good foundation, it can go deeper in providing clear guidance on how to tactically extend the methodology to AI/ML.

I recommend the following 3 starting actions to this committee and other US government organizations, including CISA, when setting policy for secure AI/ML:

1. **Create an MLBOM standard in partnership with NIST and other USG entities.** The development of a Machine Learning Bill of Materials (MLBOM) standard, in partnership with NIST and other U.S. government bodies, is critical to address the unique complexities of AI/ML systems, which are not adequately covered by traditional Software Bill of Materials (SBOM). An MLBOM would provide a more tailored framework, focusing on the specific data, algorithms, and training processes integral to AI/ML, setting it apart from conventional software transparency measures.

2. **Invest in protecting the AI/ML open source software ecosystem.** Per a 2023 study by Synopsis Corporation[4], nearly 80% of AI/ML, Analytics, and Big Data systems use open source software. To protect this, CISA and DHS can mandate and direct other federal agencies to rigorously enforce and adhere to standardized security protocols and best practices for the use and contribution to open source AI/ML software, ensuring a fortified and resilient national cybersecurity posture. The committee should help expand Senate Bill 3050, which includes a proposition and directive on the requirement for AI/ML bug bounty programs in foundational artificial intelligence models being integrated into Department of Defense missions and operations, and be inclusive of all AI/ML assets.

[4] Synopsis Corporation: 2023 Open Source Security and Risk Analysis Report. Available at https://www.synopsys.com/software-integrity/resources/analyst-reports/open-source-security-risk-analysis.html

3. **Continue to enlist feedback and participation from technology startups.** It took a startup in the form of OpenAI to open the eyes of the world to the power and potential of AI. As such, when Congress and other authorities look to regulate AI, it is important to have a broad set of innovative opinions and solutions, and prevent only large enterprises from dominating the conversation, ensuring diverse and forward-thinking perspectives are included in shaping future AI policy and regulation.

In closing and as previously stated, I agree with and support the three principles in CISA's "Secure by Design." However, as mentioned in that document, *"some secure by design practices may need modification to account for AI-specific considerations."* To that end, we realize AI/ML is different from typical software applications and these principles will need to be continuously refined. I welcome the opportunity to propose ideas and solutions that will help drive government and industry adoption of MLSecOps practices, which can be enhanced by new technical standards and sensible governance requirements. I and my company, Protect AI, stands ready to help maintain the global advantage in technologies, economics, and innovations that will ensure the continued leadership of the United States in AI for decades to come.

Thank you, Mr. Chairman, Ranking Member, and the rest of the committee, for the opportunity to discuss this critical topic of security of artificial intelligence. I look forward to your questions.

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

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.

Testimony
Mr. Timothy O'Neill
Vice Presidnet
Chief Information Security Officer & Product Security
Hitachi Vantara
FOR A HEARING
BEFORE THE UNITED STATES HOUSE OF REPRESENTATIVES
Committee on Homeland Security
Subcommittee on Cybersecurity and Infrastructure Protection
"Considering DHS' and CISA's Role in Securing Artificial Intelligence"
D**ecember 12, 2023**
Washington, D.C.

Good morning. Thank you, Chairman Garbarino, Ranking Member Swalwell, and the members of the subcommittee for inviting me here today.

My name is Tim O'Neill and I am the Vice President, Chief Information Security Officer & Product Security, at Hitachi Vantara. Hitachi Vantara is a subsidiary of Hitachi, Limited, a global technology firm founded in 1910 and focused on creating a sustainable society via data and technology. We co-create with our customers to leverage information technology (IT), operational technology (OT), and our products and services to drive digital, green, and innovation solutions for their growth. Our regional subsidiary was established in the U.S. in 1959 and for over 30 years we have heavily invested in U.S. research & development through our 24 major R&D centers that are supporting high-skilled jobs in manufacturing and technology. Our commitment to the U.S. is demonstrated by the establishment of our digital business unit's global headquarters in Santa Clara, California, and we now employ over 16,000 in the U.S. in 30 states and across 60 group companies. North America is our second largest market, representing 17% of our global revenue.

Because of our heavy focus on the intersection of IT and OT technology, one of our major areas of business development and research has been in the industrial Artificial Intelligence (AI) area. This use of AI is often overlooked in favor of conversations about generative AI and ChatGPT; however, industrial AI has the potential to significantly enhance the productivity of U.S. manufacturing and create working environments that benefit employees assembling products. Our co-created AI solutions can address challenges in factories, from the quality of products to the productivity of workers, and respect and address worker concerns on health, safety, discrimination and bias, privacy, and security.

Today's AI systems are tools that workers can use to enhance their job performance. Programs are predicting possible outcomes based on the data being given to them and what the program has been trained to understand as the most likely scenario. That is true of a predictive maintenance solution Hitachi may create for a client to help them more quickly ascertain the likely cause of a breakdown, or of a generative AI system that is predicting what the next sentence could be for a maintenance manual. The system cannot think for itself, and thus humans are necessary to confirm the AI's outcomes or make the ultimate decision. It is like a piece of software that we would use in our jobs to perform a calculation, but just as in the case of an Excel document that is running a formula on a group of cells, it is important for the user to ensure the formula is correct.

The U.S. government has taken a number of positive steps over the last 5 years to promote and further the development of AI. The previous administration laid the foundation with their request to the stakeholder community asking how AI could be used in the federal government. This set the course for the AI standards work that we have seen from the National Institute of Standards and Technology (NIST). The Biden Administration has continued that work with their Blueprint for an AI Bill of Rights, and now this AI Executive Order. We encourage the U.S. to further the development of AI via engagement with international standards setting bodies as well as by reaffirming the U.S.'s commitment to digital trade standards, digital trade titles in treaties like the ones found in the United States-Mexico-Canada Agreement (USMCA), and promotion of digital trade policies in international trade settings.

The AI EO speaks frequently to the necessity of secure AI systems. CISA's core mission focuses on cyber treaties and cybersecurity, making them the obvious agency to take the lead in implementing this part of the EO. As an example, CISA's work on ransomware and the on-going updates and alerts of ransomware attacks has been vital to informing businesses and stakeholders and helping them identify,

defend, and recover from attacks. This same type of threat assessment can be provided for AI. CISA is integral to supporting and providing resources for other agencies on cyber threats and security as those agencies focus on their roles in implementing the EO; this mission is vital to the federal government and where CISA is by far the expert. Other agencies should turn to CISA for this threat identification and cyberthreat detection support.

We applaud the CISA team for their excellent outreach to stakeholders and private industry to understand the implications of security threats and help carry out solutions in the marketplace. Their outreach to the stakeholder community is a model for other agencies to follow. As CISA's expertise lies in assessing the cyber landscape, they are best positioned to support the AI Executive Order and help further development of AI innovation in the U.S. It is also important that CISA recognize the potential benefits AI could pose to critical infrastructure systems to help them identify possible attacks or defend against cyber or physical attacks, and not just on the ways AI could make them vulnerable to failure.

There is great potential for CISA to work across agencies to support or augment their AI work and provide insight into cybersecurity guidance and/or threat identification. CISA is also discouraged against creating separate frameworks, processes, or testbeds and instead should work collaboratively across the federal government to utilize the resources other agencies have, have already, or are currently creating. Manufacturers, especially those who are making products for critical infrastructure industries, have been engaged with their respective agencies and are assisting in the development of AI systems. While some manufacturers may not have engaged with CISA as they implement technology solutions in their operations, as CISA coordinates across agencies to implement the EO, it can broaden its reach to educate all on the crucial role cybersecurity plays in core IT and AI processes.

As an example, the Department of Energy's Cybersecurity, Energy Security, and Emergency Response office (CESER) is charged with overseeing cybersecurity in the electric grid, and thus manufacturers of components for the grid have worked with and continue to engage with CESER. CISA is best served by working with CESER on electric grid AI security issues versus creating a new regime that may duplicate existing work. We envision that CISA would continuously update CESER of threats or security concerns—ongoing or new—that could be used to attack the energy grid, and work with the office to develop guidance to direct manufacturers on how to mitigate potential threats in the manufacturing process.

The Department of Energy (DOE) and the National Science Foundation (NSF) are tasked with creating testing environments for AI systems, including testbeds. CISA, therefore, should avoid creating testbeds and instead work with the DOE and NSF on securing testing environments, including how they are accessed and used, to support their integrity and mitigate potential data manipulation which could compromise the subsequent testing or training of AI systems. CISA should also guide DOE and NSF on specific needs within those testbeds or testing environments to challenge the cyber resiliency of AI systems. This requires CISA's unique expertise, which agencies can lean on versus creating redundant processes or procedures for developers. Continuous evaluation of AI models for CISA should only be focused on the evolving cybersecurity threat landscape.

NIST is tasked with creating and promoting guidelines, standards, and best practice development. To date, it already has a well-established Cybersecurity Framework, the Secure Software Development Framework, and now the AI Risk Management Framework. CISA should encourage use of those existing documents and focus on additional frameworks to address gaps specific to their cybersecurity mandate. There is no need for CISA to create its own risk management or analytical framework for assessing AI

systems. Rather, the agency must work with NIST to promote awareness of emerging threats and ensure that frameworks and testing environments are regularly updated to address them.

Some manufactured products, including Hitachi-produced railcars and energy grid equipment, have been, and will be, in the market for decades. As new technology is incorporated into new products —for example, to assist in creating predictive maintenance schedules to anticipate failures before they happen, or create guided repair solutions to fix equipment issues faster—the future cybersecurity landscape needs to be better understood. CISA can, for instance, facilitate understanding around protecting assets when there are multiple versions of a technology in use at the same time. We believe that CISA's intention to create SBOM toolchains, and the desire to provide information on how AI fits into the Secure by Design program, are valuable avenues to pursue. Manufacturers of AI-enabled equipment, developers of AI programs, and deployers of AI systems must determine mitigation measures to keep the security of their equipment intact throughout its lifecycle. CISA can thus help develop threat assessment guidelines, and the necessary mitigation efforts, to guard against legacy technology becoming a possible gateway for bad actors.

Hitachi certainly supports CISA's ongoing cybersecurity work. CISA's Roadmap for AI has very meaningful areas that can help promote the security aspects of AI usage. We strongly recommend that CISA avoid duplicating the current or tasked work of other agencies as that could create multiple layers that manufacturers would then have to navigate. Such a multi-layered approach would create more harm than good and divert from CISA's well-established and much appreciated position as a cybersecurity leader. It could also create impediments for manufacturers, especially small and medium sized enterprises, from adopting AI systems that would enhance their workers' experience and productivity, improve factory safety mechanisms, and improve the quality of products for customers.

Thank you for your time today. I am happy to answer your questions.

House Homeland Security Committee
Subcommittee on Cybersecurity and Infrastructure Protection

"Considering DHS' and CISA's Role in Securing Artificial Intelligence"
December 12, 2023

Written Statement of
Alex Stamos
Chief Trust Officer
SentinelOne[1]

Chairman Garbarino, Ranking Member Swalwell, and Members of the Subcommittee, thank you for having me here today to discuss the challenges and opportunities presented by artificial intelligence and machine learning. These world-changing technologies have the potential to impact nearly every aspect of our lives, and they are likely to continue to scale at lightning speeds. This subcommittee, and policymakers at all levels of government, face the challenging task of matching the pace of innovation with thoughtful policies to harness the positive aspects of AI while minimizing its dangers. In the context of cybersecurity, AI and machine learning provides attackers and scammers with a powerful new tool that can probe for weaknesses, and make ransomware targeting more convincing and effective, among other dangers. But, used properly, these technologies also give defenders new resources that can make security technologies more effective and intuitive, while helping to ameliorate cyber workforce shortages.

I am currently the Chief Trust Officer of SentinelOne, a company that uses AI to help defend small to large enterprises, governments and nonprofits around the world. I am also a lecturer in the Computer Science and International Relations departments at Stanford University, where I teach classes in cybersecurity and online safety that include the creation of new AI tools by my students. I previously served as the Chief Information Security Officer at two large public companies, Facebook and Yahoo, and have consulted with hundreds of companies around the world both before and after serious cybersecurity incidents. I just finished a two-year term as a member of the DHS Cybersecurity Advisory Committee, am currently a member of the Aspen Institute U.S. Cybersecurity Working Group and also advise the NATO Cybersecurity Center of Excellence.

In my testimony, I will draw on my personal experience as a career cybersecurity professional to lay out a brief picture of the current security environment, with a focus on the ransomware threat, as well as some thoughts on how we can harness the power of AI in a safe way. I will also offer my thoughts on how we can build off of recent federal policy efforts like President Biden's AI Executive Order[2] to create an effective and sustainable framework for the safe use of AI in the public and private sectors.

[1] SentinelOne
[2] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence | The White House

The Current Situation in the Field

Over the last two decades I have helped investigate and respond to dozens of attacks against American businesses. Before addressing how AI could impact cybersecurity I wanted to offer a handful of observations from the past year:

- **Cyber-extortion is a massive risk for companies of all sizes.** While we do continue to see interesting and important intrusions from state-sponsored actors, the baseline risk for every company in the United States, no matter their size or industry, are the professional extortion groups - cybercriminals.

- **Extortionists are getting bold and inventive.** Extortion groups are regularly demanding massive ransoms, in the range of $40-60 million. When the victim (appropriately) attempts to negotiate this to a more reasonable level, threat actors use text messages to employees, emails to vendors and customers, ACH theft from the bank accounts of counterparties, and even the threat of Securities and Exchange Commission (SEC) investigation[3] to try to drive negotiations forward.

- **The current sanction regime has made paying more complicated but not less logical**. The cybercrime wave has created a niche industry of companies that specialize in tracking extortion groups. During several recent incidents, my clients were told by these specialists that, from the decision to pay the ransom being made, it would take five to seven days for the ransom payment to reach the threat actor. Most of that time is spent with sanctions compliance work, and given that these groups don't operate on layaway, the delay makes the strategy of paying to speed up recovery of systems less effective.

- **The SEC is creating new requirements that confuse cyber reporting**. In 2022, Congress passed the Cyber Incident Reporting for Critical Infrastructure Act[4] (CIRCIA) to standardize the process of reporting intrusions to the US Government. Via CIRCIA, Congress specifically directs CISA to be the focal point of cyber incident reporting in the US Government. The SEC has ignored Congress' will and imposed new reporting requirements for public companies that do not consider the difficult tradeoffs involved in public disclosure. While it is important that public companies are honest with investors, the requirement to file statements during the opening hours of a response and negotiation period gives the attackers more leverage and distracts from key response steps during the period when containment is almost never guaranteed. Threat actors have noticed and have used threats of SEC reporting to gain leverage, as previously referenced.

- **Many companies are vulnerable due to their traditional Microsoft architecture, and upgrading is extremely expensive.** Microsoft continues to dominate the enterprise information technology stack, with many organizations still running the same traditional on-premise Active Directory infrastructure that Microsoft recommended for years. Unfortunately, professional attackers have become extremely adept at finding and exploiting the common weaknesses in this kind of corporate network. More modern designs for Windows networks now exist, but generally require companies to subscribe to monthly Microsoft cloud services that many organizations find prohibitively expensive. The cost of Microsoft's licenses continue to slow down the adoption of modern technologies, and are also related to the forensic challenges faced by multiple Government agencies that struggled with

[3] Ransomware group reports victim it breached to SEC regulators | Ars Technica
[4] Cyber Incident Reporting for Critical Infrastructure Act of 2022 (CIRCIA) | CISA

investigating the breach of Microsoft's systems due to the lack of logging in their base cloud subscriptions[5].

- **Legal risks bend companies away from smart, transparent responses.** The first call by a company during a breach is to outside counsel, and due to privilege concerns the cyber-lawyers are represented on every single call or email thread. I have worked with some excellent attorneys on breaches, but the over-legalization of executive decision-making is keeping companies from making smart, ethical, and transparent decisions because doing so might increase their risk of Department of Justice (DOJ), SEC or shareholder action in the future. I once worked a breach where there were four law firms on every call, representing various parties at the company, which did not engender long-term, transparent decisions from the executive team.

- **It has become very hard to hire qualified Chief Information Security Officers (CISOs)**. There is a massive deficit of security leadership with the technical and leadership skills necessary to guide large enterprises through a cyber crisis. Recent actions by the SEC to lay the blame for systemic security failures on the CISO are exacerbating this problem[6], and I personally know two well-qualified people who have passed up promotions to CISO roles due to the personal risk they would be taking.

The Impact of AI on Cybersecurity[7]

I mention these facts because I expect that the AI revolution that we are just beginning to witness will have massive impacts on the struggle to secure US businesses from attacks, and that the basic roles played by security operators will look quite different in only a few years. This is mostly a good thing! As you can tell from my observations above, while great strides have been taken by Congress, the Executive Branch and companies across the nation, I am overall quite pessimistic about the current state of cybersecurity in the United States. One of the major drivers of our challenges is a lack of qualified individuals compared to the huge number of organizations that require them. While other industries rightfully fear AI replacing the jobs of humans, I am hopeful that the next several years will lead to AI developments that help close the massive gap in cybersecurity skills while leaving plenty of high-paying jobs for humans supervising AI agents.

Some of the benefits for defenders will include:

- Automated agents that can **sort through petabytes of security events and provide real-time visibility[8]** across a huge network. Our industry has done a great job of creating a huge amount of security telemetry from the tens of thousands of computers and other devices in a typical corporate network, but we have yet to put the ability to understand that data into the hands of your typical IT team.

- **AI-operated security operations centers** (SOC), where the difficult 24x7 work of responding to security alerts will be left in the hands of computers while humans are woken up to provide oversight and to double-check the decisions of the AI agents. AI-enabled investigations will be much faster

[5] Microsoft under fire after hacks of US State and Commerce departments | Reuters
[6] Cyber Chiefs Worry About Personal Liability as SEC Sues SolarWinds, Executive - WSJ
[7] In this testimony I will restrict myself to discussing the impact of AI on the traditional information security field. I also have concerns around the impact AI could have on the manipulation of the US public by our foreign adversaries which I discussed in my testimony to the bipartisan Senate AI Forum. Alex Stamos Statement - AI Insight Forum on Elections and Democracy
[8] SentinelOne is one of several companies working to deploy LLMs and other AI models to this end: Purple AI | Empowering Cybersecurity Analysts with AI-Driven Threat Hunting, Analysis & Response - SentinelOne

and simpler for defenders, allowing them to make plain-English queries like “Show me all the computers that spoke to our secure network in the last eight hours” instead of struggling to get the exact syntax right on a search like:
`ip.addr in {10.10.0.0 .. 10.10.0.254, 192.168.1.1..192.168.1.50}`

- Real-time **analysis of unknown binaries, user behaviors and potentially malicious scripts** in a manner that most IT workers can understand. “Figure out what this potentially malicious piece of code does” used to be a question answered by a highly-skilled individual with a disassembler and debugger, and only the most highly resourced security teams can have such professionals as full-time staff. AI systems that can supplement these skill sets and provide plain-English explainability of complex programs will be hugely beneficial to defenders.

- More **flexible and intelligent response automation**. Many security coordination tools require a huge amount of effort to initially configure and are based upon fragile, human written rulesets. AI systems that respond to attacks in ways not fully foreseen by human defenders are both a scary idea and also likely necessary to cope with future attacks.

- Software development tools that point out insecure coding patterns to software developers in real-time, well before such bugs can make it into production systems. Reducing security flaws upstream is a much cheaper solution to our overall software security challenges than trying to patch bugs later.

It is also likely, however, that AI will be useful to attackers in several ways:

- AI could help attackers **sort through the billions of exposed services** they regularly scan to **automatically exploit** and install malware after the release of new flaws. This already happens, using human-written scripts, but AI could become a competitive advantage for groups that are able to use it to move faster and automate currently manual exploitation steps. Ultimately, speed kills in cyber, and AI may give attackers a new advantage.

- We will start to see regular **exploit creation via binary analysis**. Just as it requires specialized skills to analyze advanced malware it also requires specialized skills to write it, and there has already been research into using AI to create stable exploit code just through analyzing vulnerable programs with minimal human guidance.

- Smart malware that operates free of human direction or Command and Control (C2). AI could create new opportunities for criminal organizations to create **smart malware that operates behind air gaps**[9] or moves through networks intelligently, choosing the correct exploits and escalation paths without human intervention.

- Large Language Models are already **automating the work of social engineering** and ransom negotiations. Transformer tools are actively being used by cyber criminals to write more effective communications, including random demand emails, overcoming prior limitations in their grasp of the English language.

[9] The best example of malware with this capability is Stuxnet, which clearly required large amounts of intelligence around the design of the Natanz facility. Smart malware that does not require this kind of pre-existing knowledge is a goal of attackers and a nightmare for defenders.

It is quite possible that we are moving towards a world where the "hands on keyboard" actions currently performed by human attackers and defenders are fully automated, while small groups of experienced people supervise the AI agents that are automatically exploiting networks or fighting back against those exploits. Defenders may currently have an advantage in this space, as there has already been a decade of investment and research by security vendors into the defensive application of AI, however, we should not expect it to take long for attackers to catch up. That will be true for both the groups that hack for money and those who work for America's adversaries.

The Near Term AI Policy Landscape

President Biden's AI Executive Order gave broad responsibilities to the Department of Homeland Security (DHS) and CISA, in particular, to aid the implementation of responsible, safe use of AI. The Order tasks CISA with developing guidance for critical infrastructure operators, and collaborating with public and private stakeholders to develop policies and practices around the use of AI[10].

This is a critical mission, and just one of many that CISA has, and will continue to perform. The creation of a defense-only, non-regulatory agency that can support and partner with US companies was a great step by the 115th Congress and President Trump, and Congress should continue to ensure that CISA has the resources it needs to carry out this mission in an effective, responsive, and timely way. As cyber incident reporting requirements are built out pursuant to CIRCIA, Congress should continue to support CISA as the focal point for these reports, as well as response and remediation, and should work to de-conflict the various reporting requirements being invented by agencies outside of Congress' direct recommendations.

As AI technologies evolve, it is important for policymakers to adopt nimble policies and safeguards made in careful collaboration with the private sector, and civil society groups representing a broad cross section of the country. As lawmakers carry out this vital but difficult mission, it is important that every effort is made to nurture and harness the positive benefits of AI, especially in the realm of security. Too many regulatory discussions around AI assume that only a handful of large American companies will dominate the space and can be utilized as chokepoints for preventing the malicious use of AI or spread of fundamental capabilities. This point of view is misguided and has led to warped regulatory priorities in the European Union and elsewhere.

The truth is that the AI genie is out of the bottle. There will be no reversing the spread of fundamental knowledge around modern AI techniques around the world. My Stanford students regularly use or even create new AI models as part of their classwork, and the amazing advances in open-source foundation models has demonstrated the capability of crowds of people to compete with US tech giants.

The spread of AI into every corner of personal and enterprise computing is inevitable. Congress should focus on encouraging responsible, thoughtful applications of these technologies and on maintaining the competitiveness of American champions instead of trying to control the spread of AI knowledge. America's adversaries, and cyber criminals at home and abroad are sure to use these capabilities at every opportunity. It is critical that new regulations around the use of AI, however well intentioned, don't hinder the ability of defenders to innovate and deploy these technologies in a beneficial way.

Thank you again for having me here today. I look forward to your questions.

---

[10] CISA's initial output on this topic has been published in tandem with the UK NCSC.