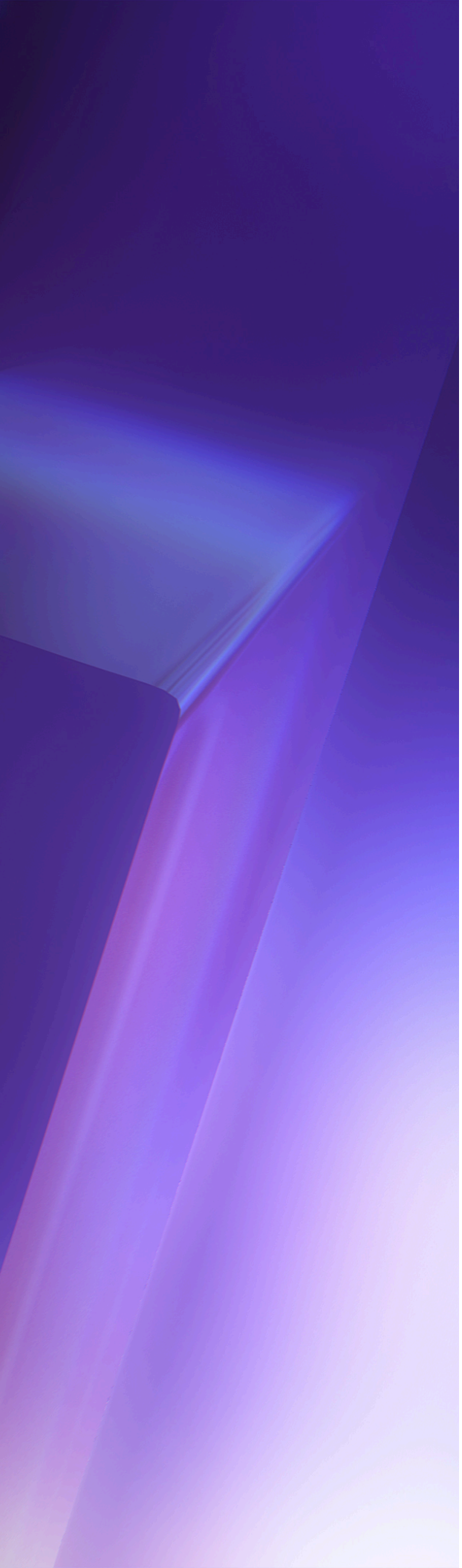


NIST AI Executive Order Request for Information (RFI)

SentinelOne Response to Docket No. 231218-0309

Submitted By Christopher Krebs, Chief Intelligence and Policy Officer

February 2, 2024

Table of Contents

Introduction

SentinelOne appreciates the opportunity to respond to the National Institute of Standards and Technology Request for Information Related to NIST's Assignments under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence.

As a leading provider of AI-enabled cybersecurity solutions to global clients across industries–including dynamic small and medium sized business and Fortune 100 multinational corporations–we wish to offer our perspective and specific recommendations to inform NIST's ongoing efforts in this critical technology domain.

First, we **provide our assessment of the impact of emerging AI technologies on cybersecurity for both offensive and defensive purposes**. While current effects are nascent, we expect these technologies to become increasingly used by both malign actors and network defenders - effectively we are entering the era of AI vs. AI.

It is critical that federal policy enable rather than hamstring research and development efforts that help American firms and government agencies keep pace with rapidly moving threats. We describe one example of AI-enabled cybersecurity technology, Purple AI, that we developed to drive industry innovation and stay ahead of the risk curve.

We then provide our **observations of AI risk management approaches across the client industries we serve**. Here we summarize three different stances being adopted by firms with various risk tolerances, market incentives, and industry considerations. We also provide a summary of the analytic framework we developed to advise firms on establishing and maintaining enterprise-wide AI risk management processes and tools.

Finally, we offer **three specific recommendations** that encourage NIST to emulate its successful approach with the Cybersecurity Framework. Our recommendations emphasize the importance and value of common ground truths and lexicon, industry-specific framework profiles, and a focus on voluntary, risk-based guidance.

SentinelOne looks forward to engaging with NIST and other federal agencies as they tackle this challenging but strategically critical policy area.

Impact of AI on Cybersecurity

Emerging Artificial Intelligence (AI) systems are force enablers for both offensive and defensive cyber operations. SentinelOne assesses that threat actors, including state and non-state groups, are using AI to augment existing tactics, techniques, and procedures (TTPs) and improve their offensive effectiveness. We also see the rapid emergence of industry technologies that leverage AI to automate detection and response and improve defensive capabilities.

Cyber Threat Actor Use of AI

SentinelOne has observed uptake and integration of AI tools across the cyber threat landscape to serve the malign interests of both non-state criminal groups and state-direct actors. While these findings broadly apply to the cybersecurity implications of AI, we have witnessed the security challenges facing AI developers themselves, including the leading frontier model labs.

Given the increasing strategic importance of these high-value technologies (embedded as they are in simple model weight files), these organizations will have to continually uplevel their security posture. The effect of a successful penetration (cyber and/or human-enabled) could have implications for geopolitical competition as well as cybercriminal proliferation risks.

1. Up-leveling of cybercriminal capabilities

The fallibility of humans remains a persistent vulnerability even as security tools grow increasingly sophisticated. As a result, social engineering is a component of the vast majority of cyberattacks. While ChatGPT brought discussions about generative AI to the forefront this year, the use of these tools for social engineering is not new.

- In 2019, attackers used AI voice technology to spoof the voice of a UK-based energy company's chief executive to scam another higher-up out of $243,000.[1]
- By 2022, two thirds of cybersecurity professionals reported that deepfakes were a component of attacks they had investigated the previous year.[2]
- In 2023, the business software development company Retool shared details of a social engineering attack they suffered which used generative AI. According to their postmortem: "The caller claimed to be one of the members of the IT team, and deepfaked our employee's actual voice. The voice was familiar with the floor plan of the office, coworkers, and internal processes of the company."[3]

[1] "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case", WSJ, August 30, 2019, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402
[2] "Global Incident Response Threat Report", VMware, August 8, 2022, https://news.vmware.com/releases/vmware-report-warns-of-deepfake-attacks-and-cyber-extortion
[3] "When MFA isn't actually MFA", Retool, September 13, 2023, https://retool.com/blog/mfa-isnt-mfa

Proliferating multimodal generative AI tools will broaden access to capabilities that accelerate and enable malicious social engineering attacks. For example, Microsoft's VALL-E model can create a voice deepfake based on just three seconds of sampled audio.[4] Attackers can find such samples of executive voices from talks and interviews or collect samples from IT employees' voice mails.

Note that Microsoft released an open source version of the VALLE–X Text-to-Speech model in September 2023. This came on the heels of Meta's open source Llama-v2 release and was followed by Mark Zuckerberg's recent commitment to develop open source artificial general intelligence models going forward.[5] The net impact of this open-sourcing trend will be to broaden access to close-to-frontier AI capabilities with less application controls or policy constraints on users. While some firms pursue closed-source proprietary development, powerful industry players see a market incentive to open-source their models to capture market share, achieve network effects, and gain indirect influence over downstream model integrators and product innovation.

These models allow attackers to turn text to speech (and others text to video) to create convincing spoofs of trusted individuals in an organization. In Retool's case, attackers convinced an employee over a phone call to provide a MFA code. With this information, the attacker added their own device to the employee's account giving them full access to MFA codes. In this case, MFA was a hindrance more than a deterrent. The added "benefit" of authentication token sync in Google authenticator means users' tokens appear on any authenticated device (including the attackers').

Lesser skilled, opportunistic hackers and hacktivists will likely leverage these tools to significantly increase the effectiveness of their target reconnaissance, phishing, and exfiltration activities given the lower barrier to entry. This will increase the scale and frequency of access operations and successful compromise of devices and accounts.[6]

More capable, organized cybercriminal groups will see less relative benefit in the near term, but will drive the innovation frontier in AI-enabled tools that will quickly proliferate across the "ransomware-as-a-service" ecosystem. In particular, while cybercriminal social engineering will

[4] VALL-E (X), Microsoft, https://www.microsoft.com/en-us/research/project/vall-e-x/
[5] "Mark Zuckerberg's new goal is creating artificial general intelligence", The Verge, January 18, 2024, https://www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview
[6] The near-term impact of AI on the cyber threat", National Cyber Security Centre, January 24, 2024, https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat

get a boost, their use of AI tools to rapidly ingest and assess large quantities of compromised data will help them quickly identify high-value assets, enhancing the value and impact of ransomware attacks.[7]

Social engineering remains a critical threat even as a company's security infrastructure gets more sophisticated. In a joint cybersecurity alert, the Cybersecurity and Infrastructure Security Agency (CISA), the National Security Agency (NSA), and the Federal Bureau of Investigation (FB)I cited two recent examples of synthetic media to target organizations, describing how:

> "In May 2023, an unknown malicious actor targeted a company for financial gain using a combination of synthetic audio, video, and text messages. The actor, impersonating the voice of a company executive, reached a member of the company using a poor quality audio call over WhatsApp. The actor then suggested a Teams meeting and the screen appeared to show the executive in their office. The connection was very poor, so the actor recommended switching to text and proceeded to urge the target to wire them money."[8]

Generative AI will require a posture shift by organizations accustomed to status quo verification tools and processes. Threat models need to be agile and adapt to emerging tactics, and companies need to continuously communicate with staff to warn them when traditional verification, such as recognizing someone's voice, is no longer sufficient. Realizations like these are prompting firms to re-examine their MFA tools and procedures, internal access controls, and update phishing training programs.

Many firms recognize that they require layered defenses that not only detect such attacks but give their security teams a chance to mitigate those that slip around end-point detections. The increasing proliferation of AI-enabled offensive tools in the hands of cybercriminals will require an equivalent response by the cybersecurity industry and security organizations to keep pace.

2. Observed interest in supporting nation-state operations

Highly capable state threat actors are best placed to fully leverage the AI frontier for advanced cyber operations, but these effects will remain hard to discern and attribute. The UK's National Cyber Security Centre found in a recent assessment that:

> "AI is likely to assist with malware and exploit development, vulnerability research and lateral movement by making existing techniques more efficient. However, in the near

[7] Ibid.
[8] "Contextualizing Deepfake Threats to Organizations", CISA, September 12, 2023, https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF

> term, these areas will continue to rely on human expertise, meaning that any limited uplift will highly likely be restricted to existing threat actors that are already capable. AI has the potential to generate malware that could evade detection by current security filters, but only if it is trained on quality exploit data. There is a realistic possibility that highly capable states have repositories of malware that are large enough to effectively train an AI model for this purpose."[9]

Technical indicators that an attacker is using AI to enable a state operation may be sparse for some time. Instead, AI tools may improve offensive operations in a way not easily observed by the defender. This is owing to *how* adversaries are considering using AI for offense.

Case Study: China's Application of AI for Offensive and Defensive Cyber Activities

Public evidence indicates that some universities connected to People's Republic of China (PRC) security services host research institutes and PhDs working on applying AI to "APT attack and defense". Among the topics covered by some of these schools include using AI to improve the pace at which software vulnerabilities are discovered–a capability that would improve PRC operational tempo, but would not be easily discernible as an impact of AI by the defenders.

Similarly, China has begun hosting competitions to automate vulnerability discovery, exploitation, and patching–another process that would improve operational efficiency but go unseen by the defenders. Finally, it is clear that the PRC has built a cyber range with significant computational resources, ties to the security services, and an interest in automating attack path decision making with AI.

None of the technologies being researched by actors in the PRC and covered here would provide technical indicators that AI was used to enable the attack. Instead, vulnerabilities discovered and exploited, and the attack paths taken by attackers, will continue to look "normal." Evidence of AI in offensive operations may only be discernable in the operational pace and efficiency of operations–analysis that would require more complete knowledge of PRC operations than any one company may have.

The impact of China's efforts will be to accelerate the pace and effectiveness of their overall cyber operations. This will exacerbate the existing significant challenge the U.S. and its allies already face in confronting broad-scale and aggressive PRC cyber activity. It should motivate a sense of urgency in driving development and adoption of AI-enabled defensive tools and capabilities by public and private organizations across the Western world.

[9] The near-term impact of AI on the cyber threat", National Cyber Security Centre, January 24, 2024, https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat

Cyber Defenders' Use of AI

The cybersecurity industry is moving apace to develop and deploy defensive technology and processes that leverage emerging AI capabilities. There are four major areas of current development that align to the NIST Cybersecurity Framework core functions:

1. **Identify**: AI may augment continuous monitoring of network traffic and user behavior for anomalies, enhance user authentication and access control, and support vulnerability identification via automated scans and threat modeling.
2. **Protect**: AI-enabled email filtering and endpoint security may help detect and block phishing attempts and known malware and AI tools can help enforce access controls and zero-trust architectures.
3. **Detect**: AI-driven security data lake solutions may prove more adaptable, responsive, and cost effective to a rapidly changing security environment than legacy security information and event management (SIEM) tools, helping detect and categorize potential threats and augment human analysts.
4. **Respond**: AI tools may help automate threat hunting and incident response by prioritizing security incidents and accelerating forensic and attribution tasks.

SentinelOne has learned lessons in developing our own AI-enabled cyber defense capabilities. Increasingly, we recognize that analysts can get overwhelmed by alert fatigue, forcing responders to spend valuable time synthesizing complex and ambiguous information. Security problems are becoming data problems. Given this, we designed our Purple AI system to take in large quantities of data and use a generative model to use natural language inputs, rather than code, to help human analysts accelerate threat-hunting, analysis and response.

Purple AI is an example of how the cybersecurity industry is integrating generative AI into solutions that allow defenders, like threat hunters and SOC team analysts, to leverage the power of large language models to identify and respond to attacks faster and easier. Using natural language conversational prompts and responses, even less-experienced or under-resourced security teams can rapidly expose suspicious and malicious behaviors that hitherto were only possible to discover with highly-trained analysts dedicating many hours of effort.

With products like Purple AI, analysts can get rapid, accurate and detailed responses to any question, in any language, that otherwise would have required hours of research and multiple queries – not to mention years of analyst experience – to obtain an answer. These tools will

allow threat hunters to ask questions about specific, known threats and get fast answers without needing to create manual queries around indicators of compromise.

For example, the analyst could use a prompt such as “Is my environment infected with SmoothOperator?”, or “Do I have any indicators of SmoothOperator on my endpoints?” to hunt for a specific named threat. In response, these tools will deliver results along with context-aware insights based on the observed behavior and identified anomalies within the returned data. Suggested follow up questions and best next actions are also provided. With a single click of a button the analyst can then trigger one or multiple actions, while continuing the conversation and analysis.

Purple AI streamlines threat investigations by comprehending data and cybersecurity concepts. It swiftly identifies event sequences and offers insights with recommendations, reducing the need for manual analysis. This significantly boosts analysts' efficiency, empowering them to handle a larger number of alerts in less time.

While this innovation has been successful, we still observe several challenges in driving AI-enabled cybersecurity tools at the scale and pace needed to keep up with the threat:

1. **Data**: These tools rely on high-quality and diverse data, but ensuring data quality and availability can be problematic. Inadequate historical data for training and the risk of adversarial attacks affecting data integrity pose significant challenges.
2. **Accuracy and Trust**: Maintaining the balance between false positives and false negatives is crucial, as is achieving transparency in AI models. The challenge lies in managing the trust and reliance on AI-generated alerts while also ensuring that AI's decision-making process is understandable and justifiable to human analysts with different levels of experience.
3. **Integration and Infrastructure**: Integrating AI solutions with existing systems and infrastructure can be complex. Resource and scalability constraints, coupled with the resource-intensive nature of AI deployment, add to the challenges. Ensuring seamless compatibility and efficient resource utilization is critically important.
4. **Regulatory Considerations**: AI cyberdefense tools raise inevitable privacy concerns and require strict compliance with data protection regulations that span different jurisdictions.
5. **Human Factors**: Effective management of AI systems demands human expertise and continuous training. The ever-evolving landscape of cyber threats and the need to align AI practices with organizational values and ethical norms introduce further complexities.

Industry Approaches to AI Risk Management

SentinelOne is increasingly asked by our clients for security and risk management advice as they develop and deploy AI-enabled systems and products. Amid the swirl of rapid technology progress, rising consumer expectations, and evolving regulatory regimes, firms are taking one of three stances: No-Go, Go Slow, YOLO ("You Only Live Once").

Observed Industry Approaches to AI Risk Management

The **No-Go** firms (an increasingly small cohort) have prohibited by policy (with limited technical visibility and patchwork security controls) personnel from using third-party AI tools, like ChatGPT, Claude, and Bard. Many are curious about (and even eager to develop or deploy) these capabilities but are opting for fully vetted and compliant solutions with limited features and functionality that they can either run on-prem or in a trusted cloud instance with strong data/access controls and application support. Others are abstaining completely, for now, given their role in critical infrastructure, operational environments, and other regulated arenas where AI product risks impinge on life safety and homeland security.

The **Go-Slowers** are dipping their toe in the water, standing up AI centers of excellence, convening tiger teams, and spinning up pilot projects to explore model and application development (in a controlled and limited manner). Here, corporate leadership has put up hard-to-enforce but easy-to-reference policy and governance guardrails on external tools. Small teams of engineers and ambitious managers are developing narrow applications to save admin cost or launch new services with either in-house fine-tuned models or open-source tools. However, enterprise scale deployment is delayed by executive, legal, and security trepidation.

Bitten by the entrepreneurial spirit, some adventurous firms are "**YOLOing**", trying to beat their competitors, impress investors, and outmaneuver regulators. Moving fast and breaking things is the mantra of the silicon valley society which birthed these tools, and many are smitten by the exponential potential. There is an intense arms-race dynamic among those industries where the returns to "first-mover advantage" are extreme and regulatory costs or negative externalities can be managed (or ignored).

Regardless of the above approach, organizations are concerned about "Shadow AI." Many leaders suspect that staff or contractors are using these systems against policy and outside of the visibility and control of the information security and risk management departments. This can

expose the firms to data leaks, privacy violations, regulatory non-compliance, and security breaches. Even with proper governance and policy regimes in place, technical monitoring and security tooling is required. As open-source and low-cost proprietary solutions continue to accelerate in performance, the incentive will grow for staff who can't use compliant tools to find loopholes or evade policy. This will create growing risk for organizations at every scale.

How SentinelOne Advises Firms On AI Risk Management

We tell firms to focus on six areas of AI risk management, with specific considerations for each:

(1) **Regulatory & Compliance**

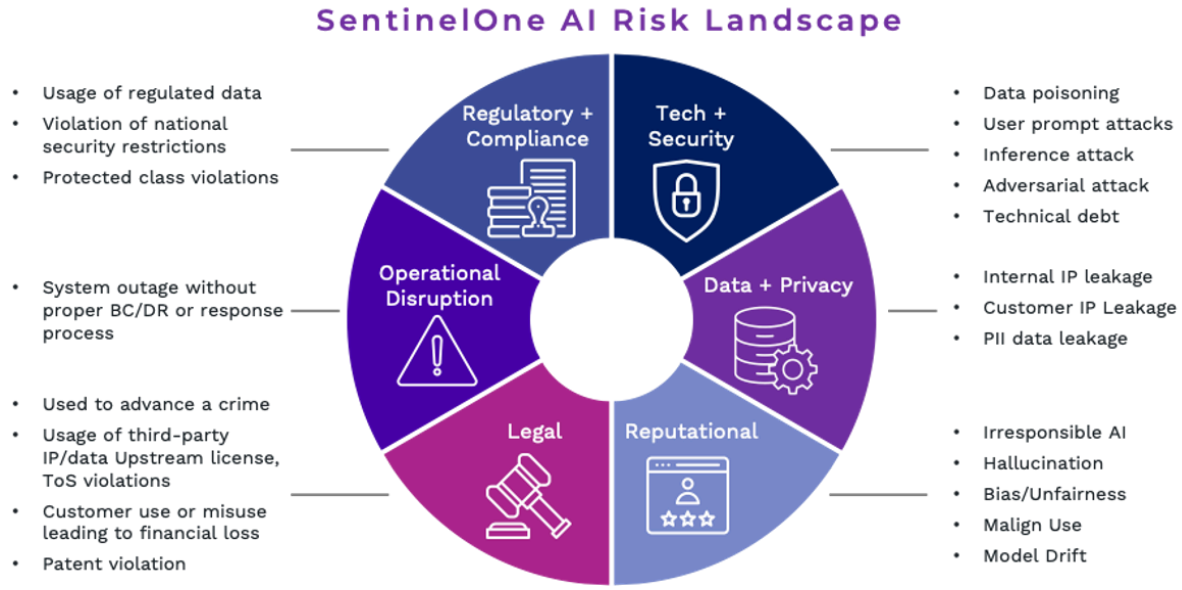
(2) **Technology & Security**

(3) **Data & Privacy**

(4) **Reputational**

(5) **Legal**

(6) **Operational Disruption**



We have found that those organizations that already have effective cross-functional teams to coordinate infosec, legal, and enterprise technology responsibilities are well positioned to manage AI integration. An area of common challenge, however, is the relationship between AI development engineers, product managers, trust & safety, and infosec teams. In many firms, for example, it is not clear who owns model poisoning/injection attacks, prompt abuse, or corporate data controls, among other emerging AI security challenges. Further, the shifting landscape of third-party platform integrations, open-source proliferation, and DIY capability sets hinder technology planning, security roadmaps, and budgets.

Generative AI Risks to Enterprise Consumers

Most firms are finding themselves at the far end of the AI value chain, consuming third-party AI products and/or internally developed capabilities. Critical risks here encompass the obvious data confidentiality breaches, but also less appreciated upstream license/ToS/IP violation risks, trust & safety issues, and potential operational disruptions from premature integrations.

Risks from Enterprise AI Development

Very few firms are willing or able to train their own foundation models, but many are fine-tuning state-of-the-art LLMs and custom developing their own applications. The risk surface varies widely by industry and use-case, but novel pitfalls emerge when these applications (which may be quite powerful) are not properly tested or secured. Current DevSecOps methodologies need to be adapted and expanded to fit the use-case, regulatory, and security paradigm of each firm.

Mapping, Measuring, and Managing Risk in AI Systems

We are advising our clients to take a crawl, walk, run approach (correlated to their broader Don't Go, Go Slow, or YOLO AI strategy). This starts by conducting an inventory of current and anticipated use-cases. Already we've seen enterprises consider: local training/usage & inference, information retrieval, writing assistance (multi-language), code review & generation, contract review & generation, document and template creation, business intelligence, financial projection & modeling, data analysis, marketing & advertising copy/imagery, work product design, corporate communications, customer support, and more.

Getting visibility and oversight of these proliferating use-cases and applications is critical. Firms must then put in place a comprehensive and dynamic risk governance, monitoring, and control system to assess and mitigate security and compliance risks. This will require increased cross-functional collaboration as well as flexible business processes, analytics, and tooling.

While industry recognizes these processes are needed (and are beginning to implement them internally), the technical tooling is lagging. By historical analogy, AI enterprise risk management is redolent of the early era of cloud adoption ten years ago. Then, as now, a new technical capability drove incentives for many firms to rapidly transform their enterprise architecture and business processes. This "move to the cloud" generated novel risks and security challenges that created a market for novel technical products and services including cloud access security

brokers (CASB). We will see a similar dynamic take place in the AI enterprise security space and expect new security management solutions to quickly emerge.

Third-party AI dependency and risk management

As the marketplace for AI applications and services rapidly develops, some first-mover firms will accrue a critical mass of many firms' data and product dependencies. The efficiency and capability advantages will drive many to deeply integrate their infrastructure, models, and/or applications into their corporate business processes and services. Some third-party AI services and applications could become systemically, even existentially, critical to U.S. businesses.

Given the pace of technology change there is also strong potential for novel solutions to emerge outside of current incumbent offerings. At this point, dissatisfied with their current provider's lagging performance, many firms might look for alternatives in the marketplace, only to find out that they are too structurally and technologically dependent to switch over. This form of walled-garden lock-in could have serious safety and security implications.

Companies are still early to understand and adopt AI. Right now, crucial acquisition and product development decisions (or mistakes) are being made that either prevent or ensure AI third-party dependency and lock-in. The terms and conditions of those contracts and the confidentiality and security of the technical implementation will tightly constrain future risk management options.

While baseline concerns are shared by many organizations, we have observed how different industries face unique challenges and need to develop tailored risk management approaches. As a result, **we recommend NIST follow a similar approach as with the implementation of the Cybersecurity Framework and develop industry-specific AI Framework Profiles**. These Framework Profiles would align RMF components with specific regulatory and operational requirements and enable organizations to establish a roadmap for reducing AI risk that is well aligned with organizational and sector goals, considers legal/regulatory requirements and industry best practices, and reflects risk management priorities and risk tolerance.

AI Safety and Security

While specific tools and technical processes will differ across industry and use-case, we observe common approaches to AI safety and security taking shape. These approaches combine traditional cybersecurity red teaming methodologies with evolving AI-specific techniques that ensure system outputs and performance are trustworthy, safe, and secure.

AI security and safety assurance involves a much broader approach than conventional information security practices (e.g., penetration testing) and must incorporate assessments of model fairness/bias, harmful content, and misuse. Any AI risk mitigation framework should encourage firms to deploy an integrated suite of tools and processes that address:

- **Cybersecurity** (e.g., compromised confidentiality, integrity, or accessibility);
- **Model security** (e.g., poisoning, evasion, inference, and extraction attacks); and
- **Ethical practice** (e.g., bias, misuse, harmful content, and social impact).

This requires security practices that emulate not only malicious threat actors but also how normal users may unintentionally trigger problematic outputs or leak sensitive data. To do this effectively, an **AI safety and security team** requires a mix of cybersecurity practitioners, AI/ML engineers, and policy/legal experts to ensure compliance and user trust.

There will be a need for specific practices and tools for specialized use-cases. For example, the production of synthetic media may require embedded **digital watermarking** to demonstrate provable provenance and traceability of training data to avoid copyright liability.

Also, as **AI agents** become more powerful and prevalent, a much larger set of legal, ethical, and security considerations will be raised regarding what controls are in place to govern the behavior of such agents and constrain their ability to take independent action in the real world (e.g., access cloud computing resources, make financial transactions, register as a business, impersonate a human, etc.).

Further, the implications for **geopolitical competition and national security** will become increasingly important as great powers race to capture strategic advantage. Working at the frontier of these technologies will involve inherent risk and U.S. adversaries may accept a higher risk tolerance in order to leap ahead. International standard setting and trust-building measures will be necessary to prevent a race-to-the-bottom competitive dynamic and security spiral.

To manage and mitigate these risks, we will need common and broad guardrails but also specific best practices and security tools calibrated to different industries. These should be based on the nature of the use-case, operational scope, scale of potential externalities, effectiveness of controls, and take into account a cost-benefit balance between innovation and risk. Given the pace of change, maintaining this balance will be an ever evolving effort.

Recommendations

We offer a few key recommendations to guide NIST as it carries out its responsibilities under the Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. These recommendations are informed by our perspective as cybersecurity technology and service providers as well as our experience advising firms across several industries with AI capabilities integration and risk management.

From our vantage point, while the frontier labs will remain at the forefront of technical capability development and strategic risk, businesses will be the contact point between these emerging tools and the economy and society at large. Therefore, AI risk evaluation frameworks and assessment policies should keep this balance of equities in mind. To that end, we recommend the following:

1. NIST should align and cross-walk the Cyber Security Framework and AI Risk Management Framework to ensure a set of common ground truths.

The CSF established key cybersecurity definitions that have proven useful in developing robust cybersecurity programs. The established terms such as “events”, “incidents”, and types of “risks” provide the groundwork for common understanding. The AI RMF should leverage these terms and ensure no terms conflict between documents where possible. This will ensure compatibility and allow for industry efficiency and transparency in organizational risk management approaches.

The AI RMF should also incorporate the CSF’s assumptions and stakeholder perspectives spanning the diversity of private sector entities, business models, and operational requirements. In particular, the CSF’s structure and implementation guidelines provide an applicable starting point for the AI RMF. The CSF’s Framework Core Functions, for example, outline cybersecurity outcomes at the highest levels. While the application of these terms to AI risk will look different, they successfully capture intended outcomes.

NIST’s AI RMF should maintain consistent key definitions and terms to facilitate ease of implementation. The firms we advise often look towards industry practices at competing firms to model their cybersecurity. Having a common set of terms and standards will enable companies to understand and strive towards industry benchmarks that both set the floor and raise the ceiling across diverse sectors.

2. **NIST should follow the CSF model and develop framework profiles for various sectors and subsectors, including profiles that range from SMBs to large enterprises.**

NIST should develop, with industry input, industry-specific AI Risk Framework Profiles, starting with the 16 critical infrastructure sectors defined by CISA. These Framework Profiles would help organizations in these diverse sectors align their organizational requirements and objectives, risk appetite, and resources against the desired outcomes of the RMF and its core components. Such an AI Risk Profile would give firms in a given industry:

1. A method to identify opportunities for developing and deploying AI-enabled systems, services, and products across relevant use cases;
2. An evaluation schema to assess their ability to maintain and monitor these systems across their enterprise environment at their acceptable risk level; and
3. A standardized approach to prepare and execute an AI risk management plan for ongoing assurance of the safety, security, and trustworthiness of these systems.

When refined and combined with further NIST red team assessment tools and other guidance, these Profiles may be used to identify opportunities for improving AI risk management posture by comparing a “Current” Profile with a “Target” Profile given specific business needs and security objectives. While specific and detailed, these Profiles would provide a voluntary risk-based approach for managing AI activities and reduce risk to enterprise systems, customers, partners, and society at large.

3. **In keeping with the successful CSF approach, NIST should seek to maximize voluntary adoption of its guidance that addresses the societal impacts and negative externalities from AI that pose the greatest risk – prescriptive regulation is a domain best suited for congressional and executive action, given the larger national security and economic considerations at issue.**

We encourage NIST to construct the AI RMF on a similar basis as the CSF’s guiding principles. In particular, the CSF’s voluntary nature, focus on societal externalities (versus normal business risks), and industry input helped accelerate broad adoption across many organizations. Sound guidance, common lexicon, and useful materials that reflect the aggregated wisdom of diverse industry participants are most likely to be picked up on a sustainable basis.

In contrast, poorly defined mandatory regulations can have unintended consequences for security. For example, the EU's Cyber Resilience Act (CRA), while well intentioned, requires a third-party risk assessment for certain high risk manufacturing. However, Europe does not have the capacity to perform third-party assessments at sufficient scale and with sufficient rigor and quality. The result is large bottlenecks in supply chains and industry resistance.[10]

In the balance between requiring specific measures and promoting voluntary implementation of best practices, NIST should heed the warnings of the EU's CRA and lean towards the latter approach.

By issuing clear, compelling, and useful AI risk management guidance and assessment frameworks, NIST can help foster the conditions for public and market scrutiny of industry approaches. Existing regulations and compliance requirements on data security and cybersecurity combined with AI-specific risk management guidance should drive adoption of industry practices.

In fact, NIST's prior approach with the CSF successfully introduced rules of the road while remaining flexible with the pace of technological change and the rapid churn of diverse business models in a globally competitive market. NIST should follow a similar approach here.

[10] "The Cyber Resilience Act as it stands risks creating COVID-style supply chain disruptions", DigitalEurope, November 6, 2023, https://dd80b675424c132b90b3-e48385e382d2e5d17821a5e1d8e4c86b.ssl.cf1.rackcdn.com/external/ceo-letter-on-cyber-resilience-act-1.pdf

Conclusion

Generative AI has opened the door to new and increasingly sophisticated threats to the modern enterprise. Adversaries are using AI to execute machine speed attacks with less dwell time, and we are seeing a higher volume of automated, simultaneous incursions. Deepfakes–both voice and video–are being used to destabilize trust and run scams. Phishing is more sophisticated.

Adversaries are moving beyond breaking and entering, using AI to tailor what people consume and execute full human compromises. Disinformation and influence operations have emerged as a new kind of warfare. The ability to turn an employee into an insider is an increasing risk.

Hackers have also figured out how to use AI to observe and predict how defenders will respond to their malware evasion techniques and adjust them on the fly. We are observing a proliferation of adaptive malware, polymorphic malware and autonomous malware propagation.

All of this has made cyberspace an increasingly dangerous and difficult environment to defend. But for all the evil it can do, AI can be used as a force for good. Machine-generated attacks require machine-generated responses, and AI is a tool that can bring order to chaos.

With AI, enterprises can detect and prevent threats with speed and efficiency and help secure a broader range of assets better than humans can. These tools aren't limited by staff constraints in the Security Operations Center or the expertise of their threat hunters. Instead, they can watch in real time, at scale, and help defend environments against novel attacks.

Security today isn't just about threat detection and prevention. It's about gaining visibility and insight into data across the entire enterprise and transforming it into decisive action to protect business. What organizations need is a unified platform to secure the enterprise that provides:

- Real time-autonomous response
- Real time, automated, environment-wide immunization from detected threats
- Context enrichment and visibility
- Generic anomaly detection and baseline monitoring
- Automatic investigation
- Proactive risk identification and soft/blind spots
- Scaled, one-to-many actions to manage increasing data and device proliferation

We appreciate the opportunity to share with NIST our perspective on both AI risk and opportunity. We encourage further industry engagement to ensure the U.S. keeps its innovative edge in AI-enabled cybersecurity to stay ahead of increasingly capable and malign threats.

SentinelOne Background

SentinelOne (NYSE:S) is pioneering autonomous cybersecurity to prevent, detect, and respond to cyber attacks faster and with higher accuracy than ever before. SentinelOne's Singularity™ Platform detects, prevents, and responds to cyber attacks at machine speed, empowering organizations to secure endpoints, cloud workloads, containers, identities, and mobile and network-connected devices with speed, accuracy and simplicity. Over 11,500 customers, including Fortune 10, Fortune 500, and Global 2000 companies, as well as prominent governments, trust SentinelOne to secure the future today.

To learn more, visit www.sentinelone.com