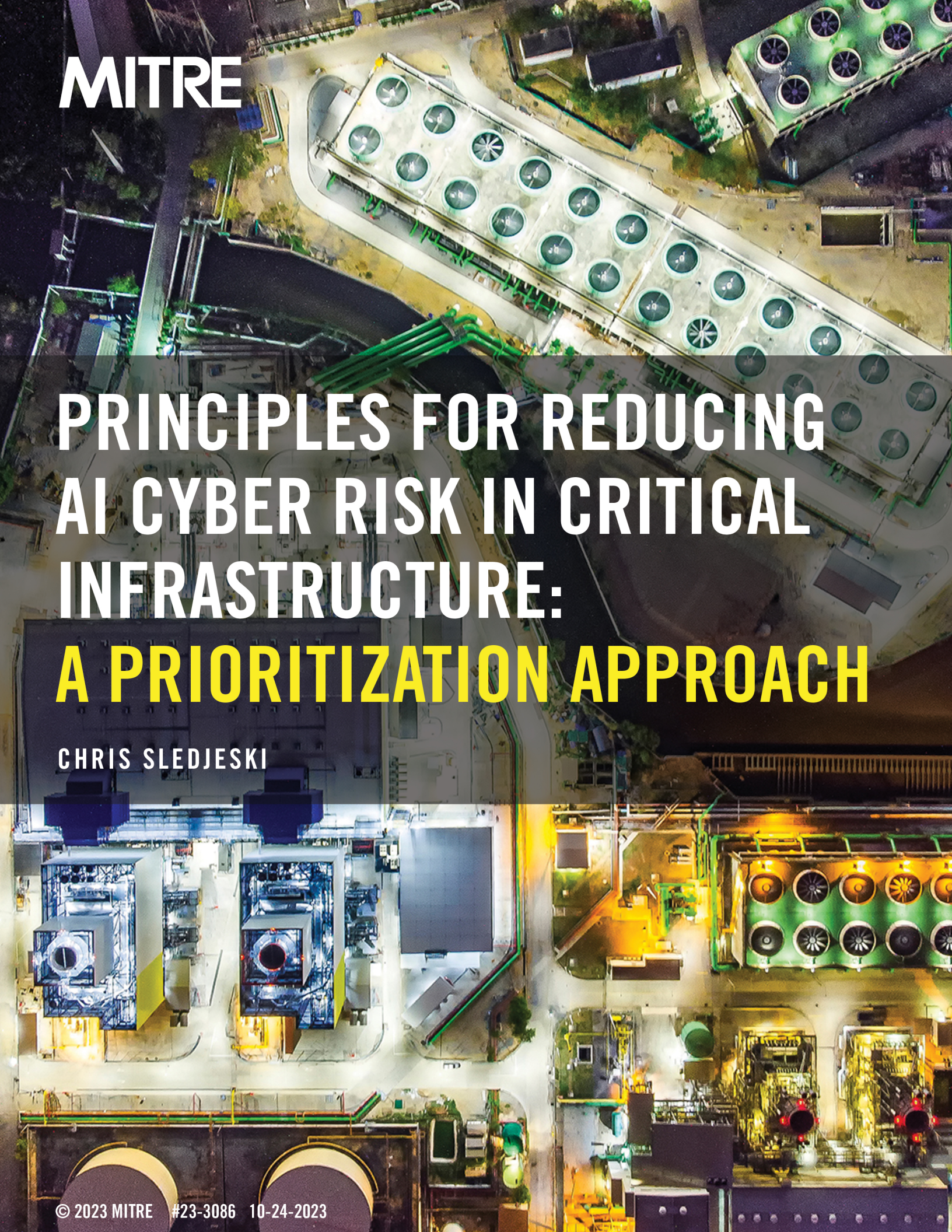

MITRE

PRINCIPLES FOR REDUCING AI CYBER RISK IN CRITICAL INFRASTRUCTURE: A PRIORITIZATION APPROACH

CHRIS SLEDJESKI

 #23-3086 10-24-2023

Contents

Executive Summary **1**
Introduction **1**
Historical Context: Recent Models for Improving Critical Infrastructure Cybersecurity ... 1
Recommendations **3**
Scope cybersecurity priorities for high-risk functions enabled by AI. 3
Define a level of applicability for cybersecurity that addresses the source of risk. 5
Define a threshold where AI functions could generate unacceptable consequences. 7
Consider a focus on foundation models and algorithms in high-risk applications. 7
Leverage the public-private partnership mindset. 8
Conclusion **9**
About the Author **11**
Endnotes **12**

Executive Summary

Artificial Intelligence (AI) brings many benefits, but disruption of AI could, in the future, generate impacts on scales and in ways not previously imagined. These impacts, at a societal level and in the context of critical infrastructure, include disruptions to National Critical Functions.[a] A prioritized risk-based approach is essential in any attempt to apply cybersecurity requirements to AI[b] used in critical infrastructure[c] functions. The topics of critical infrastructure and AI are simply too vast to meaningfully address otherwise.

The National Institute of Standards and Technology (NIST) defines cyber secure AI systems as those that can "maintain confidentiality, integrity and availability through protection mechanisms that prevent unauthorized access and use."[1] Cybersecurity incidents that impact AI in critical infrastructure could impact the availability, reliability, and safety of these vital services.[2] High-risk applications in critical infrastructure of particular concern include "safety-critical cyber-physical systems—those that … create the opportunity for injury or death to people, the loss or damage of equipment or property, or environmental harm … due to the scale and speed [AI] enables."[3]

This paper was prompted by questions presented to MITRE about to what extent the original NIST Cybersecurity Risk Framework, and the efforts that accompanied its release, enabled a regulatory approach that could serve as a model for AI regulation in critical infrastructure. The NIST Cybersecurity Risk Framework was created a decade ago as a requirement of Executive Order (EO) 13636.[4] When this framework was paired with the list of cyber-dependent entities identified under the EO, it provided a voluntary approach for how Sector Risk Management Agencies (SRMAs) prioritize and enhance the cybersecurity of their respective sectors.

An important insight from this history is to scope what is in bounds early and decisively, based on the risk level of critical infrastructure functions enabled by AI, and the potential for unacceptable outcomes. To do this, it is important to define a level of unacceptable consequences before deciding (1) whether to apply AI to a critical infrastructure function

---

[a] National Critical Functions are the functions of the government and the private sector so vital to the United States that their disruption, corruption, or dysfunction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof (see https://www.cisa.gov/national-critical-functions-set).

[b] The term "AI" is used interchangeably in the paper with several categories of AI except where otherwise specified. The AI ecosystem can be divided into three broad categories: (1) engineered systems that use AI as a component or subsystem; (2) AI as an augmentation of human capabilities; and (3) AI operating autonomously under its own agency. AI functions for critical infrastructure are generally captured, at the moment, by categories 1 and 2 (see https://www.mitre.org/news-insights/publication/sensible-regulatory-framework-ai-security).

[c] The term "critical infrastructure" has the meaning provided in Section 1016(e) of the USA Patriot Act of 2001 (42 U.S.C. 5195c(e)), namely systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters (see https://www.cisa.gov/sites/default/files/2023-01/ppd-21-critical-infrastructure-and-resilience-508_0.pdf).

and (2), where AI is applied, what lengths to take to ensure its cybersecurity. Risk mitigation efforts for AI in critical infrastructure should be directed toward the features or behavior that are causing and/or elevating the risk versus issuing wholesale requirements. This helps to ensure that limited resources are applied with the greatest risk reduction effect. Generally, any AI regulation should account for use context and favor existing domain-specific regulations where applicable rather than wholesale requirements.[5]

Another insight is that while some companies may proactively and unilaterally implement more rigorous cybersecurity postures than any regulatory floor would require, this is the exception more than the rule. Though there have been improvements to critical infrastructure cybersecurity, a decade of a voluntary cybersecurity adoption approach did not result in implementation of a common minimum set of cybersecurity practices across critical infrastructure sectors.[6,7] As such, an unknown level of cybersecurity risk remains.[8]

In any case, government should make its risk reduction priorities and desired end states transparent and keep a close partnership with SRMAs and the private sector to aid in the adoption and efficacy of any new cybersecurity requirements for AI in critical infrastructure functions. There is also a window of opportunity to establish or better align Information Sharing and Analysis Centers (ISACs) around AI concerns in critical infrastructure[d] to elevate sector-specific concerns with AI-specific vulnerabilities and incidents with the government through trusted channels.

[d] ISACs collect, analyze, and disseminate actionable threat information to their members and provide members with tools to mitigate risks and enhance resiliency.

Introduction

Movement toward enhanced cybersecurity requirements in U.S. civilian critical infrastructure has been long sought. As Congress debates how to regulate artificial intelligence (AI)[e] and its associated subdisciplines, such as machine learning (ML),[f] insights from the path to improve critical infrastructure cybersecurity can be leveraged to enhance the adoptability and efficacy of recommendations or regulations for cybersecurity of AI in critical infrastructure applications.

Cybersecurity concerns for critical infrastructure came to lawmakers with urgency given foreseeable risks, and it is proving the same for AI.[9] As OpenAI's CEO accurately stated in his May 2023 testimony to Congress, "If this technology goes wrong, it can go quite wrong."[10] The National Institute of Standards and Technology (NIST) argues that "risks from AI-based technology can be bigger than an enterprise, span organizations, and [potentially] lead to societal impacts."[11] AI brings many benefits, but disruption of AI could, in the future, generate impacts on scales and in ways not previously imagined.

Interest is growing for Congress to adopt cybersecurity requirements for AI use in critical infrastructure functions. These discussions include enhanced cybersecurity requirements for the currently concentrated number of companies developing AI hardware, software, or firmware, as well as those critical infrastructure entities employing these technologies in potentially high-risk applications.[g]

Historical Context: Recent Models for Improving Critical Infrastructure Cybersecurity

Unable to generate a consensus required for a minimum set of cybersecurity requirements for critical infrastructure a decade ago,[12] Presidential Policy Directive (PPD) 21,[13] Executive Order (EO) 13636,[14] and the NIST Cybersecurity Risk Framework[15] were released in 2013 and 2014, respectively, to provide voluntary guidance for implementing cybersecurity in critical infrastructure. PPD 21 assigned the U.S. Department of Homeland Security (DHS) (acting through what is known today as the Cybersecurity and Infrastructure Security Agency [CISA]) responsibility for leadership and interagency coordination of voluntary public-private

[e] AI is the capability of a device to perform functions that are normally associated with human intelligence, such as reasoning, learning, and self-improvement (see https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf).

[f] ML is a branch of AI and computer science that focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy (see https://www.nccoe.nist.gov/ai/adversarial-machine-learning).

[g] See NIST AI Risk Management Framework 1.0, Appendix B (p. 38), for a more expansive list of AI-specific risks that are "new" or "increased."

partnerships across 16 designated critical infrastructure sectors.[h] Sector-Specific-Agencies—now called Sector Risk Management Agencies (SRMAs)—are assigned to coordinate, organize activities, support incident management, provide technical assistance, and support DHS's infrastructure prioritization statutory requirements.[16] Some SRMAs had preexisting regulatory authorities, but no new authorities were granted under the PPD. As SRMAs have matured, perhaps unsurprisingly, there is a strong correlation between the sectors that had a history of security regulation and the sectors that have made the most progress in implementing at least minimum cybersecurity requirements—such as the electricity subsector.

EO 13636 was the first time the government defined cyber-dependent entities that, if disrupted by a cybersecurity incident, could result in catastrophic impacts on economic security, national security, and public health.[17] This was an important scoping decision that, when combined with the associated threshold criteria, dramatically narrowed the number of entities under consideration and, more importantly, provided the government with an initial risk-based focus for a finite pool of resources and attention.

The process of scoping required a series of direct engagements with SRMAs and infrastructure operators to leverage their expertise in determining which entities met the criteria of being cyber dependent and would be capable of generating catastrophic effects if disrupted through a cyber incident. As a requirement of EO 13636, NIST was tasked to create a Cybersecurity Risk Framework.[18] This framework was outcome focused, meaning it set the desired end state across a variety of cybersecurity considerations, but it was meant to guide versus prescribe specific solutions.[19] Specifically, the NIST Cybersecurity Risk Framework provided a prioritized, flexible, repeatable, performance-based, and cost-effective approach, including information security measures and controls, to help owners and operators of critical infrastructure identify, assess, and manage cyber risk.[20]

A voluntary approach did not result in the wide-scale adoption of a minimum set of cybersecurity practices across critical infrastructure sectors, as evidenced by the most recent call for minimum cybersecurity requirements from the Federal Chief Information Security Officer (CISO)[i] in August 2023.[21] The NIST Cybersecurity Framework and the prioritization efforts around EO 13636 did provide a guide for scoping priorities and activities in those sub-sectors that introduced formal cybersecurity requirements, such as the North American Reliability Corporation (NERC) Critical Infrastructure Protection

[h] The 16 sectors defined under PPD 21 are Chemical; Commercial Facilities; Communications; Critical Manufacturing; Dams; Defense Industrial Base; Emergency Services; Energy; Financial Services; Food and Agriculture; Government Facilities; Healthcare and Public Health; Information Technology; Nuclear Reactors, Materials, and Waste; Transportation Systems; and Water and Wastewater.

[i] The Federal CISO is responsible for strengthening cybersecurity at all federal agencies, which detrimentally rely on mostly privately owned and operated critical infrastructure to perform their missions.

standards for bulk electric system cybersecurity,[22] and the more recent cybersecurity directive from the Transportation Security Administration (TSA) for pipelines.[23,24,25]

A similar voluntary risk framework was recently issued for AI: the NIST Artificial Intelligence Risk Framework. Today, both frameworks, in addition to EO 13636, provide a means to begin to scope and prioritize the focus for cybersecurity of AI in high-risk critical infrastructure functions at least at an entity (company or system) and outcome level. These frameworks and the recommendations that follow may not be revolutionary, but they carry forward still relevant concepts from the National Infrastructure Protection Plan's Risk Management Framework[26] for the critical infrastructure cycle of continuous improvement (e.g., set objectives, identify infrastructure, assess risks, prioritize, implement programs, measure effectiveness).

Recommendations

Scope cybersecurity priorities for high-risk functions enabled by AI.

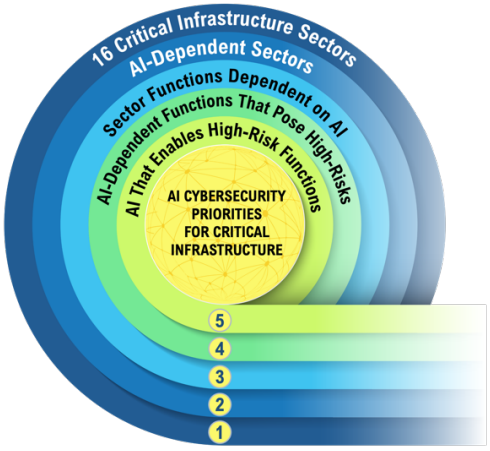

Figure 1. Identifying AI Cybersecurity Priorities for Critical Infrastructure Working from the Universe of Critical Infrastructure down to AI Applications that Enable High-Risk Functions in AI-Dependent Sectors

A crucial first in the EO 13636 process was to scope down what is critical infrastructure. Critical infrastructure is always context based—critical for who, in what way, and in what context? The U.S. government has, for example, an established statutory context from which it perceives criticality, defined as "systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters."[27]

Critical infrastructure is always context based—critical for who, in what way, and in what context?

EO 13636 sought to refine (1) which parts of the existing critical infrastructure sectors were dependent on cyber and (2) where the sector dependence on cyber could generate catastrophic damage to economic security, national security, or public health if disrupted. The same approach could be applied for AI-dependent critical infrastructure. A first step would be to identify what infrastructure is dependent on AI for functions which, at least at first glance, could generate an unacceptable level of risk if their availability, integrity, or confidentiality is disrupted through a cybersecurity incident.[j]

Figure 1 depicts a prioritization process very similar to that used in EO 13636 but in the context of AI cybersecurity for critical infrastructure. Beginning in the outer band, from the universe of potential critical infrastructure, one identifies sectors that are dependent on AI and what types of functions within those sectors are enabled by AI. Identifying AI adoption by sector is probably easiest derived as a qualitative answer from SRMA expertise because quantitative sector-specific adoption data is not yet well developed. Information in the public record indicates that sectors such as Energy, Healthcare and Public Health, Transportation, Financial Services, Critical Manufacturing, and the Defense Industrial Base are employing AI to varying degrees. SRMAs can also be helpful in identifying lesser-known sector-specific AI applications.

From an initial set of sector use, one would further down-select to those functions capable of generating high risks if disrupted. The threshold of "high risk" in a national context could follow the definition given by the U.S. government for critical infrastructure, which accounts for risks that could be capable of "debilitating disruptions to the economy, public health and safety, or national security."[28] One would then identify the essential hardware, software, and/or firmware to which cybersecurity requirements could be applied to reduce the risk to critical infrastructure operations.

EO 13636 leveraged a series of in-person engagements with SRMAs as a shortcut to answer questions on technology risk to sector functions. Not surprisingly, infrastructure entities from most sectors that DHS defines as lifeline sectors[k] figured prominently on the list of sectors whose dependency on cyber could potentially generate catastrophic consequences if disrupted. Entities chosen for the EO13636 Section 9 list generally used cyber to enable their vital functions, and the functions could not be

[j] The NIST AI Risk Management Framework 1.0 provides guidelines that could be leveraged in combination with statutory definitions of critical infrastructure for defining risk levels.

[k] There are four designated lifeline functions—transportation, water, energy, and communications—which means that their reliable operations are so critical that a disruption or loss of one of these functions will directly affect the security and resilience of critical infrastructure within and across numerous sectors (see https://www.cisa.gov/sites/default/files/publications/Guide-Critical-Infrastructure-Security-Resilience-110819-508v2.pdf).

practically performed without the availability and integrity of cyber. Most[l] lifeline sectors will probably leverage AI for at least some high-risk applications. There will also be new considerations, depending on adoption trends, like certain AI-enabled functions in Healthcare and Public Health, Financial Services, and Critical Manufacturing that may not have previously met a high-risk threshold.

In short, a significant amount of scoping work has occurred through the efforts to identify high-risk cyber-dependent infrastructure sectors that can accelerate the prioritization process for AI cybersecurity in critical infrastructure.

Define a level of applicability for cybersecurity that addresses the source of risk.

Where should cybersecurity requirements be applied if they are desired? The Section 9 list under EO 13636 sought to designate entities—a company- or system-level designation. The benefit of this approach was that it quickly generated a high-priority engagement list. But the list on its own was insufficient to identify the appropriate level for cybersecurity enhancements.

Applied to AI cybersecurity, an entity-level analysis identifies companies or systems that produce the AI hardware, software, and/or firmware that, if disrupted through cyber means, could generate unacceptable levels of risk. While extending software engineering best practices for the software aspect of AI at designated entities might be necessary, it is insufficient to account for all cyber risk drivers because there are unique risks that AI itself introduces, where applied and in interaction with other subsystems.[m]

Based on a risk evaluation, AI cybersecurity requirements may be needed on one or more levels to address the drivers of the cyber risk. AI hardware, software, or firmware that enables high-risk functions in critical infrastructure is a potential focal area for a minimum level of cybersecurity. An important sub-focus of cybersecurity for AI in critical infrastructure is identifying where disruption of AI functions can scale disruptive impacts.

An important sub-focus of cybersecurity for AI in critical infrastructure is identifying where disruption of AI functions can scale disruptive impacts.

Depending on the true drivers of cybersecurity risk in a unique AI-enabled function in critical infrastructure, cybersecurity requirements may be needed in the high-risk application hardware, software, and/or firmware; the production environments that produce it; the products that incorporate it; and/or the critical infrastructure services that

[l] Some lifeline sectors may opt not to employ or to delay the use of AI in certain high-risk applications. Some lifeline sectors may not decide to use AI for high-risk functions simply due to the cost or complexity relative to the size of their operations.

[m] NIST AI 100-1 AI Risk Management Framework Appendix B describes the differences between AI and traditional software risk in detail (see https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf).

connect to it. Also, high-risk critical infrastructure AI functions may be found on information technology[n] networks or operational technology[o] networks.

High-risk AI applications in critical infrastructure of particular concern are "safety-critical cyber-physical systems—those that … create the opportunity for injury or death to people, the loss or damage of equipment or property, or environmental harm … due to the scale and speed [AI] enables."[29] Such AI disruptions might occur through techniques such as the Evade the ML Model (AML.T0015), Denial of ML Service (AML.T0029), Erode ML Model Integrity (AML.T0031), and System Use for External Effect (AML.T0048).

Over time, cybersecurity requirements for AI used in critical infrastructure may need to consider more precise levels of analysis (e.g., product class, model type, foundational or proprietary source code). The lifecycle and key dimensions of AI systems include applications, data, models, and outputs, which are developed in production environments, integrated into product solutions, and often feature a regular set of communications to regularly update and debug the solutions where implemented.[30] Some cybersecurity considerations for AI in critical infrastructure may not be known until AI is employed in the real world. To account for AI post-deployment concerns, it will be important to incorporate feedback from operators of critical infrastructure.

High-risk AI functions in critical infrastructure will rely heavily on the availability and integrity of data and information services. This is particularly true in the case of safety-critical cyber-physical infrastructure functions.

High-risk AI functions in critical infrastructure will rely heavily on the availability and integrity of data and information services. This is particularly true in the case of safety-critical cyber-physical infrastructure functions.

Confidentiality is a factor in AI cybersecurity in critical infrastructure, but only inasmuch as the loss of such confidentiality could be leveraged to generate catastrophic impacts on the economy or national security. Confidentiality is also a feature of AI cybersecurity that will likely be addressed by other standards and requirements such as privacy laws.

---

[n] NIST 800-128: "Any services, equipment, or interconnected system(s) or subsystem(s) of equipment, that are used in the automatic acquisition, storage, analysis, evaluation, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information."

[o] NIST SP 800-37 Rev 2: "Programmable systems or devices that interact with the physical environment (or manage devices that interact with the physical environment). These systems/devices detect or cause a direct change through the monitoring and/or control of devices, processes, and events. Examples include industrial control systems, building management systems, fire control systems, and physical access control mechanisms."

Define a threshold where AI functions could generate unacceptable consequences.

It is probably easiest to align a threshold for unacceptable consequences with those applied based on an existing statutory definition. The threshold values for what is unacceptable could be calibrated based on potential harms (e.g., lives lost, economic impacts, loss of National Critical Functions) that could then be applied to determine cutoff points for cybersecurity requirements.

In April 2021, the European Union proposed the AI Act, which provides a tiering structure for AI applications. As an example, "AI applications that pose an 'unacceptable risk' would be banned; high-risk applications in such fields as finance, the justice system, and medicine would be subject to strict oversight."[31] MITRE considers AI risks such as "damage, harm, or loss to human life, health, property or the environment."[32]

There are historical threshold values for what the U.S. government considers to be critical or high risk.[p] While any number set for these thresholds is arbitrary to a certain extent, using the precedent of existing threshold values for what the government considers critical at least provides consistency.

Cybersecurity scenarios used to evaluate where AI functions could generate unacceptable consequences in critical infrastructure applications do not have to be highly detailed to evaluate whether AI applications fall within the order of magnitude envisioned by the threshold (nor will such complete data be consistently available to establish "clean" quantitative thresholds). Much of the work done under EO 13636 to determine cyber-dependent infrastructure entities capable of causing unacceptable risk used a set of broad "working definition"-type scenarios or simply a broad-based acknowledgment that cyber was able to impact the availability, integrity, or confidentiality of an AI-enabled function in a conceivably scalable manner.

Consider a focus on foundation models and algorithms in high-risk applications.

Within AI, there are a variety of cybersecurity issues that could be addressed, but in the context of critical infrastructure and high-consequence scenarios, many will center on issues with adversarial attacks, data security, model security, and the transparency of the models to evaluation.[q] U.S. Army leadership, for example, has recently begun to push for more transparency in models to "rule out risks like Trojans,

---

[p] For example, those threshold values set internally by DHS for its National Critical Infrastructure Prioritization Program or EO13636 Section 9 lists. These values are controlled but unclassified information.

[q] MITRE ATLAS™ (Adversarial Threat Landscape for Artificial-Intelligence Systems) is a knowledge base of adversary tactics, techniques, and case studies for ML systems based on real-world observations, demonstrations from ML red teams and security groups, and the state of the possible from academic research (available at https://atlas.mitre.org/).

triggers, poison data sets, or prompting of unintentional outcomes."[33]

Cybersecurity requirements for AI in critical infrastructure cannot address everything all at once. High-risk algorithms or widely used foundation models represent only part of the overall AI risk, but they have asymmetric potential for disruptive scaling if exploited through AI cybersecurity vulnerabilities.[34,35] As a hypothetical, consider the disruptive cyber scaling potential if there were an exploitable vulnerability discovered in a foundation model that is used in dozens of industries globally.

High-risk algorithms or widely used foundation models represent only part of overall AI risk, but they have asymmetric potential for disruptive scaling if exploited through AI cybersecurity vulnerabilities.

AI cybersecurity risks for critical infrastructure functions come at least in part from the massive scale and complexity of interactions these models could have, and also from a lack of traditional code development scrutiny.[36] Evaluation of models would not be an easy endeavor, as there is much entropy in models and algorithms and not all models are equally consequential. To aid this process, it may be necessary to develop an ontology of specific AI models used in critical infrastructure functions and their associated risks at a useful level of abstraction to help with model risk evaluation.

This may also be an opportunity to adapt federal government critical infrastructure plans to address increased risk due to AI-enabled scale and speed in critical functions. But there are also clearly opportunities to use AI for defense—for example, to reduce risk through automated red teaming.

Leverage the public-private partnership mindset.[r]

A reoccurring lesson learned in attempts to improve cybersecurity in critical infrastructure is the need for ongoing and transparent dialogue with industry in the development of standards and requirements. Government must come prepared with reasonable but meaningful cybersecurity outcomes to protect public interests when it engages the private sector. If it does not, additional requirements unrelated to the desired outcomes could develop, or worse yet, the government's desired outcomes could be lost in discussions.

The private sector is, in most cases, the maker of AI and cybersecurity solutions, as well as the operator of critical infrastructure, and it is an invaluable resource for innovating on how to achieve AI cybersecurity outcomes. A robust public-private dialogue will ensure that requirements are technically credible and can be applied to result in a set of outcomes that are beneficial to both the public and industry. Where outcomes must be delineated by government, there are almost always more ways than one to reach them, and some ways are decidedly more efficient.

[r] The public-private partnership mindset is an increased level of private-sector participation that aids the effectiveness and efficiency of implementing solutions.

Congress should also consider incentives or rewards for companies that do go above and beyond any minimum requirement.

A robust public-private dialogue will ensure that requirements are technically credible and can be applied to result in a set of outcomes that are beneficial to both the public and industry.

Improving the cybersecurity of AI use in critical infrastructure applications requires a public-private partnership. The government should also consider implementing a plan for how it would better support entities that utilize AI in high-risk functions, including, but not limited to, enhanced information sharing, joint research, risk mitigation guidance, assessment tools, and cost sharing where appropriate. Government also has a role in bringing the best of its research and development and standards bodies to aid this process and to provide the right incentives.

There is also a window of opportunity to establish and/or better align Information Sharing and Analysis Centers (ISACs).[s] To ensure cross-sector collaboration as well as sector-specific expertise, these could include both broad AI topic ISACs that look at overarching threats and vulnerabilities (e.g., validating foundational models) and more narrowly focused sector-specific ISACs that address unique concerns (e.g., evaluating vulnerabilities in AI models used in the Financial Services sector). ISACs play an important role in providing situational awareness and information sharing within their sectors. In the context of AI and critical infrastructure, ISACs could play an important role in elevating concerns with AI vulnerabilities and incidents with the government through trusted channels.

Conclusion

AI brings many benefits, but disruption of AI could, in the future, generate impacts on scales and ways not previously imagined. These risks, at a societal level and in the context of critical infrastructure, include risks to National Critical Functions. A prioritized risk-based approach is essential to success in any attempt to apply cybersecurity to AI for critical infrastructure. Scope what is in bounds early and decisively, based on the risk level of functions enabled by AI and the potential for debilitating impacts on public health and safety, the economy, or national security. Identify and focus risk mitigation on behavior or features that are causing and/or elevating the risk versus wholesale requirements. Also identify where disruption of AI functions can scale disruptive impacts across critical infrastructure. High-risk algorithms or widely used foundation models represent only part of overall AI risk, but they have asymmetric potential for disruptive scaling if exploited through AI cybersecurity vulnerabilities. While extending software engineering best practices for the software aspect of AI at defined entities might be necessary, it is insufficient to account for all cyber risk drivers because there are unique risks

[s] ISACs collect, analyze, and disseminate actionable threat information to their members and provide members with tools to mitigate risks and enhance resiliency.

that AI itself introduces, where applied and in interaction with other subsystems. Finally, a close partnership with SRMAs and the private sector will aid in the adoption and efficacy of any cybersecurity requirements applied to AI use in critical infrastructure high-risk functions.

About the Author

Chris Sledjeski is a Senior Principal and Intelligence Analyst in MITRE's Cyber Infrastructure Protection Innovation Center. He previously served the U.S. Department of Energy's Office of Intelligence and Counterintelligence, the Defense Intelligence Agency, and the U.S. Department of Homeland Security. He continues to provide assessments on cyber threats to critical infrastructure systems for multiple U.S. government agencies.

Acknowledgments

The author would like to thank Charles Clancy, Ph.D., Douglas Robbins, Ozgur Eris, Ph.D., Lashon Booker, Ph.D., Mark Bristow, Katie Enos, Duane Blackburn, Guido Zarella, Chuck Lewis, Brian Abe, EJ Hillman, and Mary Bruzzese for their thoughtful input and review of this document.

About MITRE

MITRE'S mission-driven teams are dedicated to solving problems for a safer world. Through our public-private partnerships and federally funded R&D centers, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

Endnotes

[1] NIST, “Artificial intelligence risk management framework,” Jan. 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [Accessed Aug. 29, 2023].

[2] C. Clancy, D. Robbins, O. Eris, L. Booker, and K. Enos, “A sensible regulatory framework for AI security,” MITRE, Jun. 14, 2023. [Online]. Available: https://www.mitre.org/news-insights/publication/sensible-regulatory-framework-ai-security [Accessed Sep. 18, 2023].

[3] Ibid.

[4] The White House, “E.O. 13636: Improving critical infrastructure cybersecurity,” Feb. 12, 2023. [Online]. Available: https://www.energy.gov/ceser/improving-critical-infrastructure-cybersecurity-eo-13636 [Accessed Aug. 29, 2023].

[5] C. Clancy, D. Robbins, O. Eris, L. Booker, and K. Enos, “A sensible regulatory framework for AI security,” MITRE, Jun. 14, 2023. [Online]. Available: https://www.mitre.org/news-insights/publication/sensible-regulatory-framework-ai-security [Accessed Sep. 18, 2023].

[6] GAO, “Critical Infrastructure Protection: Agencies Need to Assess Adoption of Cybersecurity Guidance.” [Online]. Available: https://www.gao.gov/products/gao-22-105103 [Accessed Sep. 1, 2023].

[7] Lawfare, “A Review of NIST’s Draft Cybersecurity Framework 2.0,” Sep. 13, 2023. [Online]. Available: https://www.lawfaremedia.org/article/a-review-of-nist-s-draft-cybersecurity-framework-2.0 [Accessed Sep. 13, 2023].

[8] GAO, “High-Risk Series: Federal Government Needs to Urgently Pursue Critical Actions to Address Major Cybersecurity Challenges.” [Online]. Available: https://www.gao.gov/products/gao-21-288 [Accessed Sep. 13, 2023].

[9] E. Strickland, “The who, where, and how of regulating AI,” *IEEE Spectrum*, Jun. 14, 2023. [Online]. Available: https://spectrum.ieee.org/ai-regulation-worldwide [Accessed Aug. 29, 2023].

[10] CBS, ”Father of Chat GPT: AI Could Go Quite Wrong,” May 16, 2023. Available: https://www.cbsnews.com/news/sam-altman-senate-chatgpt-ai-could-go-quite-wrong/ [Accessed Sep. 1, 2023].

[11] NIST, “Artificial intelligence risk management framework,” Jan. 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [Accessed Aug. 29, 2023].

[12] M. Teplinsky, “A review of NIST’s Draft Cybersecurity Framework 2.0,” Lawfare, Sep. 13, 2023. [Online]. Available: https://www.lawfaremedia.org/article/a-review-of-nist-s-draft-cybersecurity-framework-2.0 [Accessed Sep. 13, 2023].

[13] The White House, “PPD21: Critical infrastructure security and resilience,” Feb. 12, 2013. [Online]. Available: https://www.cisa.gov/sites/default/files/2023-01/ppd-21-critical-infrastructure-and-resilience-508_0.pdf [Accessed Aug. 29, 2023].

[14] The White House, “E.O. 13636: Improving critical infrastructure cybersecurity,” Feb. 12, 2023. [Online]. Available: https://www.energy.gov/ceser/improving-critical-infrastructure-cybersecurity-eo-13636 [Accessed Aug. 29, 2023].

[15] NIST, “Framework for improving critical infrastructure Cybersecurity,” Apr. 16, 2018. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf [Accessed Aug. 29, 2023].

[16] CISA, “Sector Risk Management Agencies.” [Online]. Available: https://www.cisa.gov/stopransomware/sector-risk-management-agencies [Accessed Aug. 29, 2023].

[17] The White House, "E.O. 13636: Improving critical infrastructure cybersecurity," Feb. 12, 2023. [Online]. Available: https://www.energy.gov/ceser/improving-critical-infrastructure-cybersecurity-eo-13636 [Accessed Aug. 29, 2023].

[18] Ibid.

[19] NIST, "Framework for improving critical infrastructure Cybersecurity," Apr. 16, 2018. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf [Accessed Aug. 29, 2023].

[20] The White House, "E.O. 13636: Improving critical infrastructure cybersecurity," Feb. 12, 2023. [Online]. Available: https://www.energy.gov/ceser/improving-critical-infrastructure-cybersecurity-eo-13636 [Accessed Aug. 29, 2023].

[21] Meritalk, "Federal CISO Calls for Minimum Cyber Regulations for CI," Aug. 29, 2023. [Online]. Available: https://www.meritalk.com/articles/federal-ciso-calls-for-minimum-cyber-regulations-for-ci/ [Accessed Sep. 1, 2023].

[22] NERC, "Reliability standards." [Online]. Available: https://www.nerc.com/pa/Stand/Pages/ReliabilityStandards.aspx [Accessed Aug. 29, 2023].

[23] TSA, "TSA revises and reissues cybersecurity requirements for pipeline owners and operators," Jul. 21, 2022. [Online]. Available: https://www.tsa.gov/news/press/releases/2022/07/21/tsa-revises-and-reissues-cybersecurity-requirements-pipeline-owners [Accessed Aug. 29, 2023].

[24] D. P. Pekoske, "Pipeline cybersecurity: Protecting critical infrastructure," TSA, Jul. 27, 2021. [Online]. Available: https://www.tsa.gov/news/press/testimony/2021/07/27/pipeline-cybersecurity-protecting-critical-infrastructure [Accessed Aug. 29, 2023].

[25] J. Marron, A. Gopstein, and D. Bogle, "Benefits of an updated mapping between the NIST Cybersecurity Framework and the NERC Critical Infrastructure Protection Standard," NIST, Sep. 29, 2021. [Online]. Available: https://doi.org/10.6028/NIST.CSWP.21 [Accessed Aug. 29 2023].

[26] DHS, "National infrastructure protection plan," 2013. [Online]. Available: https://www.dhs.gov/sites/default/files/publications/National-Infrastructure-Protection-Plan-2013-508.pdf [Accessed Sep. 19, 2023].

[27] Public Law 110-53, "Implementing recommendations of the 9/11 Commission Act 2007." [Online]. Available: https://www.congress.gov/110/plaws/publ53/PLAW-110publ53.htm [Accessed Sep. 10, 2023].

[28] DHS, "Critical Infrastructure Sectors." [Online]. Available: https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors [Accessed Sep. 1, 2023].

[29] C. Clancy, D. Robbins, O. Eris, L. Booker, and K. Enos, "A sensible regulatory framework for AI security," MITRE, Jun. 14, 2023. [Online]. Available: https://www.mitre.org/news-insights/publication/sensible-regulatory-framework-ai-security [Accessed Sep. 18, 2023].

[30] Ibid.

[31] IEEE Standards Association, "THE IEEE AI Impact Use Cases Initiative." [Online]. Available: https://standards.ieee.org/industry-connections/ai-use-cases-initiative/ [Accessed Aug. 29, 2023].

[32] MITRE, "A Sensible Regulatory Framework for AI Security," Jun. 14, 2023. [Online]. Available: https://www.mitre.org/news-insights/publication/sensible-regulatory-framework-ai-security [Accessed Sep. 18, 2023].

[33] C4ISRNet, "US Army may ask defense industry to disclose AI algorithms." [Online]. Available: https://www.c4isrnet.com/artificial-intelligence/2023/05/31/us-army-may-ask-defense-industry-to-disclose-ai-algorithms/ [Accessed Aug. 29, 2023].

[34] P. Laplante and B. Amaba, "Artificial intelligence in critical infrastructure systems," *Computer*, Oct. 2021. [Online]. Available: https://www.computer.org/csdl/magazine/co/2021/10/09548022/1x9TFbzhvTG [Accessed Aug. 29, 2023].

[35] NIST, "Artificial intelligence risk management framework," Jan. 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [Accessed Aug. 29, 2023].

[36] Ibid.