Workshop Report

Securing Critical Infrastructure in the Age of AI

Authors

Kyle Crichton*
Jessica Ji*
Kyle Miller*
John Bansemer*
Zachary Arnold
David Batz
Minwoo Choi
Marisa Decillis
Patricia Eke
Daniel M. Gerstein
Alex Leblang
Monty McGee
Greg Rattray
Luke Richards
Alana Scott

***Workshop Organizers**

CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

October 2024

This workshop and the production of the final report was made possible by a generous contribution from the Microsoft Corporation. The views in this document are strictly the authors' and do not necessarily represent the views of the U.S. government, the Microsoft Corporation, or of any institution, organization, or entity with which the authors may be affiliated.

Reference to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply an endorsement, recommendation, or favoring by the U.S. government, including the U.S. Department of the Treasury, the U.S. Department of Homeland Security, and the Cybersecurity and Infrastructure Security Agency, or any other institution, organization, or entity with which the authors may be affiliated.

Executive Summary

As artificial intelligence capabilities continue to improve, critical infrastructure (CI) operators and providers seek to integrate new AI systems across their enterprises; however, these capabilities come with attendant risks and benefits. AI adoption may lead to more capable systems, improvements in business operations, and better tools to detect and respond to cyber threats. At the same time, AI systems will also introduce new cyber threats that CI providers must contend with. Last year's AI executive order directed the various Sector Risk Management Agencies (SRMAs) to "evaluate and provide … an assessment of potential risks related to the use of AI in critical infrastructure sectors involved, including ways in which deploying AI may make critical infrastructure systems more vulnerable to critical failures, physical attacks, and cyber-attacks."

Despite the executive order's recent direction, AI use in critical infrastructure is not new. AI tools that excel in prediction and anomaly detection have been used for cyber defense and other business activities for many years. For example, providers have long relied on commercial information technology solutions that are powered by AI to detect malicious activity. What has changed is that new generative AI techniques have become more capable and offer novel opportunities for CI operators. Potential uses include more capable chatbots for customer interaction, enhanced threat intelligence synthesis and prioritization, faster code production processes, and, more recently, AI agents that can perform actions based on user prompts.

CI operators and sectors are attempting to navigate this rapidly changing and uncertain landscape. Fortunately, there are analogues from cybersecurity that we can draw on. Years ago, innovations in network connectivity provided CI operators with a way to remotely monitor and operate many systems. However, this also created new attack vectors for malicious actors. Past lessons can help inform how organizations approach the integration of AI systems. Today, risk may arise in two ways: from AI vulnerabilities or failures in systems deployed within CI and from the malicious use of AI systems against CI sectors.

This workshop report provides technical mitigations and policy recommendations for managing the use of AI in critical infrastructure. Several findings and recommendations emerged from this discussion.

- Resource disparities between CI providers within and across sectors have a major impact on the prospects of AI adoption and management of AI-related risks. **Further programs are needed to support less well-resourced providers**

with AI-related assistance, including financial resources, data for training models, requisite talent and staff, forums for communication, and a voice in the broader AI discourse. **Expanding formal and informal means of mutual assistance** could help close the disparity gap. These initiatives share resources, talent, and knowledge across organizations to improve the security and resiliency of the sector as a whole. They include formal programs, such as sharing personnel in response to incidents or emergencies, and informal efforts such as developing best practices or vetting products and services.

- There is a recognized need to integrate AI risk management into existing enterprise risk management practices; however, ownership of AI risk can be ambiguous within current corporate structures. This risk was referred to by one participant as the AI "hot potato" being tossed around the C-suite. **A clear designation of responsibility for AI risk within the corporate structure is needed.**

- Ambiguity between AI safety and AI security also poses substantial challenges to operationalizing AI risk management. Organizations are often unsure how to apply guidance from the National Institute of Standards and Technology's recently published AI risk management framework alongside the cybersecurity framework. **Further guidance on how to implement a unified approach to AI risk is needed**. Tailoring and prioritizing this guidance would help make it more accessible to less well-resourced providers and those with specific, often bespoke, needs.

- While there are well-established channels for cybersecurity information sharing, there is no analogue in the context of AI. **SRMAs should leverage existing venues, such as the Information Sharing and Analysis Centers, for AI security information sharing**. Sharing AI safety issues, mitigations, and best practices is also critical, but the channels to do so are unclear. Clarity on what constitutes an AI incident, which incidents should be reported, the thresholds for reporting, and whether existing cyber-incident reporting channels are sufficient would be valuable. To promote cross-sector visibility and analysis that spans both AI safety and security, the **sectors should consider establishing a centralized analysis center for AI safety and security.**

- Skills to manage cyber and AI risks are similar but not identical. The implementation of AI systems will require expertise that many CI providers do not currently have. As such, **providers and operators should actively upskill their current workforces** and seek opportunities to cross-train staff with

relevant cybersecurity skills to effectively address the range of AI- and cyber-related risks.

- Generative AI introduces new issues that can be more difficult to manage and that warrant close examination. CI providers should **remain cautious and informed before adopting newer AI technologies, particularly for sensitive or mission-critical tasks**. Assessing whether an organization is even ready to adopt these systems is a critical first step.

Table of Contents

Introduction

In October 2023, the White House released an Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Section 4.3 of the order specifically focuses on the management of AI in critical infrastructure and cybersecurity.[1] While regulators debate strategies for governing AI at the state, federal, and international levels, protecting CI remains a top priority for many stakeholders. However, there are numerous outstanding questions on how best to address AI-related risks to CI, given the fractured regulatory landscape and the diversity among the 16 CI sectors.

To address some of these questions, the Center for Security and Emerging Technology (CSET) hosted an in-person workshop in June 2024 that brought together representatives from the U.S. federal government, think tanks, industry, academia, and five CI sectors (communications, information technology, water, energy, and financial services). The discussion was framed around the issue of security in CI, including the risk from both AI-enabled cyber threats and potential vulnerabilities or failures in deployed AI systems. The intention of the workshop was to foster a candid conversation about the current state of AI in critical infrastructure, identify opportunities and risks—particularly related to cybersecurity—presented by AI adoption, and recommend technical mitigations and policy options for managing the use of AI and machine learning in critical systems.

The discussion focused on CI in the United States, with some limited conversation on the global regulatory landscape. This report summarizes the workshop's findings in four primary sections. The **Background** section contains CSET research on the current and potential future use of AI technologies in various CI sectors. The **Risks, Opportunities, and Barriers** section addresses these issues associated with AI that participants raised over the course of the workshop. The third section, **Observations**, categorizes various themes from the discussion, and the report concludes with **Recommendations**, which are organized by target audience (government, CI sectors, and individual organizations within both the sectors and the AI industry).

Background

In preparation for this workshop, CSET researchers examined the reports submitted by various federal departments and agencies in response to the White House AI executive order, section 4.3. These reports provided insight into how some CI owners and operators are already using AI within their sector, but it was sometimes unclear what types of AI systems CI providers were employing or considering. For example, the U.S. Department of Energy (DOE) summary report overviewed the potential for using AI-directed or AI-assisted systems to support the control of energy infrastructure, but it did not specify whether these were generative AI or traditional models. This was the case for many of the sources and use cases assessed for the background research, spanning information technology (IT), operational technology (OT), and sector-specific use cases. This ambiguity reduces visibility into the current state of AI adoption across the CI sectors, limiting the effectiveness of ecosystem monitoring and risk assessment.

This section summarizes CSET's preliminary research for the workshop and provides examples of many of the current and potential future AI use cases in three sectors—financial services, water, and energy—based on federal agency reporting.

Research Methodology

The U.S. Department of Homeland Security (DHS) recently released guidelines for CI owners and operators that categorize over 150 individual AI use cases into 10 categories.[2] While the report encompassed all 16 CI sectors, the use cases were not specified. To identify AI use cases for the sectors that participated in the workshop, we assessed reports from the U.S. Department of the Treasury (financial services), DOE (energy), and the U.S. Environmental Protection Agency (EPA, water). We also examined the AI inventories for each department and agency, but they only included use cases internal to those organizations, not the sectors generally.

The Treasury and DOE reports were written following the AI executive order, were relatively comprehensive, and considered many AI use cases.[3] Further use cases in the finance and energy sectors were pulled from nongovernmental sources (e.g., the *Journal of Risk and Financial Management* and Indigo Advisory Group).[4] The EPA sources were dated and lacked details on AI use cases.[5] To identify more use cases in the water sector, we assessed literature reviews from *Water Resources Management* (a forum for publications on the management of water resources) and *Water* (a journal on water science and technology).[6] Although we primarily focused on sources covering U.S. CI, some research encompassed CI abroad. A full list of sources can be found in Appendix A.

The Current and Future Use of AI in Critical Infrastructure

We classify AI use cases in CI into three broad categories: **IT**, **OT**, and **sector-specific use cases**. IT encompasses the use of AI for "traditional" cybersecurity tasks such as network monitoring, anomaly detection, and classification of suspicious emails. All CI sectors use IT, and therefore they all have the potential to use AI in this category. OT encompasses AI use in monitoring or controlling physical systems and infrastructure, such as industrial control systems. Sector-specific use cases include the use of AI for detecting fraud in the financial sector or forecasting power demand in the energy sector. These broad categories provide a shared frame of reference and capture the breadth of AI use cases across sectors. However, they are not meant to be comprehensive or convey the depth of AI use (or lack thereof) across organizations within sectors.

When discussing use cases for CI, we consider a broad spectrum of AI applications. While newer technologies such as generative AI (e.g., large language models) have recently been top of mind for many policymakers, more traditional types of machine learning systems, including predictive AI systems that forecast and identify patterns within data (as opposed to generating content), have long been used in CI. The various AI systems present differing opportunities and challenges, but generative AI introduces new issues that can be more difficult to manage and that warrant close examination. This includes difficulties in interpreting how models process inputs, explaining their outputs, managing unpredictable behaviors, and identifying hallucinations and false information. Even more recently, generative models have been used to power AI agents, enabling these models to take more direct action in the real world. Although these systems are still nascent, their potential to automate tasks—whether routine work streams or cyberattacks—deserves close watching.

Themes in AI-CI use cases from the reports examined include:

- Many IT use cases employ AI to supplement existing cybersecurity practices and have commonalities across sectors. For example, AI is often used to detect malicious events or threats in IT, be it at a financial firm or water facility. Some AI IT use cases, such as scanning security logs for anomalies, go back to the 1990s. Others have emerged over the past 20 years, such as anomalous or malicious event detection. New potential use cases have surfaced with the recent advent of generative AI, such as mitigating code vulnerabilities and analyzing threat actor behavior.

- Based on reported use cases, there are no explicit examples of generative AI being used in OT. While some applications of traditional AI are being used, such as in infrastructure operational awareness, broader adoption is still fairly limited. This is in part due to concerns over causing errors in critical OT. However, future use cases are being actively considered, such as real-time control of energy infrastructure with humans in the loop.

- Many sector-specific AI use cases seek to improve the reliability, robustness, and efficiency of CI. However, they also raise concerns about data privacy, cybersecurity, AI security, and the need for governance frameworks to ensure responsible AI deployment. It can be more challenging to implement a common risk management framework for these use cases because they are specialized and have limited overlap across sectors.

- AI adoption varies widely across CI sectors. Organizations across each sector have varying technical expertise, funding, experience integrating new technologies, regulatory or legal constraints, and data availability. Moreover, it is not clear whether certain AI use cases were actively being implemented, considered in the near term, or feasible in the long term. Many of the potential AI use cases highlighted in relevant literature are theoretical, with experiments conducted only in laboratory, controlled, or limited settings. One example is a proposed intelligent irrigation system prototype for efficient water usage in agriculture which was developed using data collected from real-world environments, but not tested in the field.[7] The feasibility of implementing these applications in practice and across organizations is currently unclear.

- The depth of AI use across organizations within sectors is difficult to assess. There are thousands of organizations across the financial, energy, and water sectors. It is unknown how many organizations within these sectors are using or will use AI, for what purposes, and how the risks from those different use cases vary.

Figure 1 aggregates all AI use cases identified in the preliminary research.* Each sector is divided into IT, OT, and sector-specific use cases and subdivided into current/near-term and long-term use cases.

Figure 1. Examples of AI Use Cases in Critical Infrastructure by Sector

		Financial Sector	Energy Sector	Water Sector
Operational Technology (OT) Security	Current / Near-Term	*[None identified]*	- Detect and diagnose malicious events - Detect and diagnose anomalous events (non-malicious) - Enhance operational awareness in infrastructure - Implement predictive maintenance - Detect faults - Baseline events and incidents	*[None identified]*
	Future	*[None identified]*	- Use active controls (i.e., control energy system operations) - Manage and control grids	- Secure conditional access points - Manipulate ICS inputs and generate unexpected behaviors that create a wide array of outcomes
Information Technology (IT) Security	Current / Near-Term	- Incorporate anomaly detection and behavior analysis AI methods into endpoint protection, intrusion detection/prevention, data-loss prevention, and firewalls - Classify suspicious emails - Comprehend malicious code	- Detect and diagnose malicious events - Detect and diagnose anomalous events (non-malicious) - Baseline events and incidents	*[None identified]*
	Future	- Analyze threat actor behaviors and streamline alerts, investigations, and responses - Identify and mitigate code vulnerabilities	*[None identified]*	- Authenticate and authorize users - Detect threats and potential attack vectors - Take immediate actions against cyber events - Identify new threat profiles and vectors - Identify viruses, malware, ransomware and malicious code and phishing - Firewall probing and attack optimization
Sector-Specific	Current / Near-Term	- Detect fraud - Improve high-frequency trading (e.g., process real-time market data to identify optimal buy/sell times) - Use large language models and Retrieval Augmented Generation (RAG) with proprietary data - Generate code	- Employ modeling and simulation (e.g., transmission line ampacity) - Use image recognition to inspect infrastructure - Forecast energy demand and supply - Manage assets (e.g., wildfire detection and vegetation management) - Prioritize work orders and optimize workflow	- Monitor and improve forecasts for water consumption - Facilitate data collection in wastewater management, improved dam operation safety, and food risk mitigation in cities - Monitor water pollution
	Future	- Devise more intricate and lucrative trading strategies by implementing AI in sentiment analysis and market prediction - Scrutinize market data and forecast trends (e.g., predicting asset prices) - Enhance risk assessment and mitigation (e.g., credit risk assessments)	- Improve system planning (e.g., changing equipment) - Forecast (e.g., weather conditions) - Generate scenarios (e.g., generate synthetic system data to train operators and enable testing of protection measures) - Improve or expand consumer services - Create digital twins of physical systems to simulate, monitor, and optimize performance	- Estimate waste - Assess wastewater quality and treatment - Improve water infrastructure resiliency - Forecast (e.g., flood risks) - Implement predictive water supply maintenance - Examine future water quality patterns

Source: CSET (See Appendix A).

* The sources examined during our preliminary research did not contain any current, near-term, or future examples of AI use cases in financial sector OT, current or near-term examples of AI use cases in water sector OT or IT, nor any future AI use cases in energy sector IT.

Risks, Opportunities, and Barriers Associated with AI

As evidenced by the wide range of current and potential use cases for AI in critical infrastructure, many workshop participants expressed interest in adopting AI technologies in their respective sectors. However, many were also concerned about the broad and uncharted spectrum of risks associated with AI adoption, both from external malicious actors and from internal deployment of AI systems. CI sectors also face a variety of barriers to AI adoption, even for use cases that may be immediately beneficial to them. This section will briefly summarize the discussion concerning these three topics: risks, opportunities, and barriers to adoption.

Risks

AI risk is twofold, encompassing both malicious use of AI systems and AI system vulnerabilities or failures. This subsection will address both of these categories, starting with risks from **malicious use**, which several workshop participants raised concerns about given the current prevalence of cyberattacks on U.S. critical infrastructure. These concerns included how AI might help malicious actors discover new attack vectors, conduct reconnaissance and mapping of complex CI networks, and make cyberattacks more difficult to detect or defend against. AI-powered tools lower the barrier to entry for malicious actors, giving them a new (and potentially low-cost) way to synthesize vast amounts of information to conduct cyber and physical security attacks. However, the addition of AI alone does not necessarily present a novel threat, as CI systems are already targets for various capable and motivated cyber actors.[8] Most concerns about AI in this context centered on its potential to enable attacks that may not currently be possible or increase the severity of future attacks. A more transformative use of AI by attackers could involve seeking improved insights as to what systems and data flows to disrupt or corrupt to achieve the greatest impact.

Generative AI capabilities are currently increasing threats to CI providers in certain cases. These threats include enhanced spear phishing, enabled by large language models. Researchers have observed threat actors exploring the capabilities of generative AI systems, which are not necessarily game-changing but can be fairly useful across a wide range of tasks such as scripting, reconnaissance, translation, and social engineering.[9] Furthermore, as AI developers strive to improve generative models' capabilities by enabling the model to use external software tools and interact with other digital systems, digital "agents" that can translate general human instructions into executable subtasks may soon be used for cyber offense.

The other risk category participants identified was related to **AI adoption**, such as the potential for data leakage, a larger cybersecurity attack surface, and greater system complexity. Data leakage was a significant concern, regarding both the possibility of a CI operator's data being stored externally (such as by an AI provider) and the potential for sensitive information to accidentally leak due to employee usage of AI (such as by prompting an external large language model).

Incorporating AI systems could also increase a CI operator's cybersecurity attack surface in new—or unknown—ways, especially if the AI system is used for either OT or IT. (A use case encompassing OT *and* IT, which are typically strictly separated with firewalls to limit the risk of compromise, would increase the attack surface even further.) For certain sectors, participants pointed out that even mapping an operator's networks to evaluate an AI system's usefulness—and subsequently storing or sharing that sensitive information—could present a target for motivated threat actors. CI operators face more constraints than organizations in other industries and therefore need to be extra cautious about disclosing information about their systems. Newer AI products, especially generative AI systems, may also fail unexpectedly because it is impossible to thoroughly test the entire range of inputs they might receive.

Finally, AI systems' complexity presents a challenge for testing and evaluation, especially given that some systems are not fully explainable (in the sense of not being able to trace the processes that lead to the relationship between inputs and outputs). Risks associated with complexity are compounded by the fact that there is a general lack of expertise at the intersection of AI and critical infrastructure, both within the CI community and on the part of AI providers.

Opportunities

Despite acknowledgment of the risks associated with the use of AI, there was general agreement among participants that there are many benefits to using AI technologies in critical infrastructure.

AI technologies are already in use in several sectors for tasks such as anomaly detection, operational awareness, and predictive analytics. These are relatively mature use cases that rely on older, established forms of AI and machine learning (such as classification systems) rather than newer generative AI tools.

Other opportunities for AI adoption across CI sectors include issue triage or prioritization (such as for first responders), the facilitation of information sharing in the cybersecurity or fraud contexts, forecasting, threat hunting, Security Operations Center

(SOC) operations, and predictive maintenance of OT systems. More generally, participants were interested in AI's potential to help users navigate complex situations and help operators provide more tailored information to customers or stakeholders with specific needs.

Barriers to Adoption

Even after considering the risk-opportunity trade-offs, however, several participants noted that CI operators face a variety of barriers that could prevent them from adopting an AI system even when it may be fully beneficial.

Some of these barriers to adoption are related to hesitancy around AI-related risks, such as data privacy and the potential broadening of one's cybersecurity attack surface. Some operators are particularly hesitant to adopt AI in OT (where it might affect physical systems) or customer-facing applications. The trustworthiness—or lack thereof—of AI systems is also a source of hesitancy.

Other barriers are due to the unique constraints faced by CI operators. For instance, the fact that some systems have to be constantly available is a challenge unique to CI. Operators in sectors with important dependencies—such as energy, water, and communications—have limited windows in which they can take their systems offline. OT-heavy sectors also must contend with additional technical barriers to entry, such as a general lack of useful data or a reliance on legacy systems that do not produce usable digital outputs. In certain cases, it may also be prohibitively expensive—or even technically impossible—to conduct thorough testing and evaluation of AI applications when control of physical systems is involved.

A third category of barriers concerns compliance, liability, and regulatory requirements. CI operators are concerned about risks stemming from the use of user data in AI models and the need to comply with fractured regulatory requirements across different states or different countries. For example, multinational corporations in sectors such as IT or communications are beholden to the laws of multiple jurisdictions and need to adhere to regulations such as the European Union's General Data Protection Regulation (GDPR), which may not apply to more local CI operators.

Finally, a significant barrier to entry across almost all sectors is the need for workers with AI-relevant skills. Participants noted that alleviating workforce shortages by hiring new workers or skilling up current employees is a prerequisite for adopting AI in any real capacity.

Observations

Throughout the workshop, four common trends emerged from the broader discussion. Different participants, each representing different sectors or government agencies, raised them at multiple points during the conversation, an indicator of their saliency. These topics include the disparities between large and small CI providers, the difficulty in defining lines between AI- and cyber-related issues, the lack of clear ownership over AI risk within an organization, and the challenges posed by fractured regulation and guidance. In the following sections, we examine these observations and highlight the issues raised during the workshop.

Disparities Between and Within Sectors

CI in the United States covers many different organizations and missions, ranging from nationwide banks to regional electric utilities to local water providers that may serve only a few thousand residents. The wide gap in resources across CI providers, falling roughly along the lines of large organizations and small providers, was repeatedly raised throughout the workshop. This disparity can exist between sectors, such as between the comparatively better-resourced financial services sector and the less well-resourced water sector, and within sectors, such as between major banks and regional lenders.

These resource disparities between providers impact cybersecurity and the prospects of AI adoption within CI in several ways.

- **Financial resources:** Differences across and within sectors in available monetary resources to implement AI have led and likely will continue to lead to the concentration of AI adoption among the most well-financed organizations. As such, the numerous potential benefits of AI discussed previously will likely be out of reach for many small providers without financial or technical assistance.

- **Talent:** Closely related to the issue of adequate funding is the limited technical expertise that different providers have on staff or have the ability to hire. Workers with AI and cybersecurity skills are already scarce. The competitive job market, and higher salaries for these positions, make it difficult for smaller providers to attract requisite talent. Some sectors, such as IT and finance, already have large technical staffs and are well positioned to incorporate and support new AI talent compared to organizations in the manufacturing, electric, or water sectors, which typically have more limited IT operations and staff.

- **Data:** The ability to produce or obtain large amounts of data for use in AI applications can be a substantial challenge for small providers. The size of the organization and scale of operations is only one aspect of the problem. Small utilities often operate older or bespoke OT systems that generate limited data or lack digital output. Making bespoke data usable for AI applications is often costly and time-consuming. Furthermore, many of these systems are configured to fit the unique needs of the provider, which may prevent the generalization of models trained on data from the same machines or devices deployed in other environments.

- **Forums:** Methods of communication and coordination between organizations within sectors vary widely. While trusted third parties—such as the Sector Coordinating Councils and Information Sharing and Analysis Centers (ISACs)—exist in most sectors, certain sectors have additional forums to facilitate collaboration, sharing of threat information, and the development of best practices, all of which play a key role in the adoption of new technology such as AI. Examples of well-established forums for collaboration include the Financial and Banking Information Infrastructure Committee and the Cyber Risk Institute in the financial services sector and the Electricity Subsector Coordinating Council's Cyber Mutual Assistance Program in the energy sector. The Cybersecurity and Infrastructure Security Agency (CISA), Sector Risk Management Agencies (SRMAs), and the sectors themselves will need to identify, deconflict, and potentially expand existing forums to manage emerging AI risks and security issues. This could also include additional cross-sector coordination.[*]

- **Voice**: Smaller organizations within sectors face many obstacles in contributing to the formation of best practices and industry standards. The absence of input from these groups risks the development of AI standards that do not account for resource constraints and, lacking appropriate guidance on prioritizing practices, can be difficult or infeasible for smaller organizations to implement.

Despite all these challenges, there are compelling reasons to pursue AI applications even for smaller, less well-resourced organizations and sectors. Of the many potential benefits afforded by AI, the use of this technology for anomaly and threat detection is particularly impactful and, in the context of CI, vitally important. Smaller providers can ill afford to be left behind in adopting AI for cyber defense, especially given the

[*] The recently formed DHS AI Safety and Security Board could serve as another forum as its roles and responsibilities are further delineated.

potential threat posed by faster, subtler, and more sophisticated AI-enabled cyberattacks. Solutions offered as a service or that work to tailor AI for bespoke applications would help lower these barriers and enable the use of sector or organizational datasets—once properly formatted for AI training—to support IT or OT security tasks.

Unclear Boundary Between AI and Cybersecurity

Distinguishing between the issues related to AI and cybersecurity, as well as the overlap between the two, was a common challenge identified across sectors. In general, this challenge reflects the underlying ambiguity between AI safety and AI security—two academic disciplines that have developed separately, but both of which are needed for robust AI risk management.[10] This ambiguity arose in three contexts: risk, incidents, and the workforce.

- **Risk:** Determining whether a given risk associated with an AI system is an AI risk, which would fall under the purview of the National Institute of Standards and Technology's AI Risk Management Framework (AI RMF), or a cybersecurity risk, which would align with NIST's Cybersecurity Framework 2.0 (CSF), is not abundantly clear. This ambiguity raises the question whether this explicit distinction needs to be made at all, yet the existence of separate frameworks and the division of risk ownership within corporate structures—both discussed in detail later in this report—seems to demand this distinction be made. Take, for example, the question of whether issues of bias and fairness are an AI risk, a cyber risk, or both. This may largely depend on the context of the application in which AI is being used and how it pertains to the critical function of the provider. For example, bias and fairness surrounding the use of AI in decisions regarding credit scores, a critical function in the financial sector, presents a risk that spans across safety and security. This presents a serious challenge for organizations attempting to clearly divide AI and cybersecurity—or, alternatively, AI safety and AI security—risk management responsibilities. As the AI RMF acknowledges, "Treating AI risks along with other critical risks, such as cybersecurity and privacy, will yield a more integrated outcome and organizational efficiencies." However, it was clear during the discussion with workshop participants that further guidance on how to implement a unified approach to risk is needed.

- **Incidents:** There is similar ambiguity surrounding what qualifies as a cyber incident, an AI incident, a safety incident, an ethical incident, or some combination of these. While there are clear requirements and channels for

cyber-incident reporting, which could possibly cover AI-related cyber incidents, it is unclear if and how information related to non-cyber AI incidents should be shared. Furthermore, the analogues between cyber and AI incidents are not perfect. For example, some AI incidents may not have easily defined remediations or patches, as has been noted in other research.[11] This suggests that remediation efforts for AI incidents will need additional mitigation strategies.[12] Defining the range of AI-related incidents and what subset falls under existing reporting requirements would be valuable. For AI incidents that are not covered by existing requirements, the benefit to sharing information as it pertains to AI-related failures, mitigations, and best practices was widely recognized by workshop participants. However, there was disagreement as to whether this information sharing should be done through formal channels, with explicit reporting requirements, or informal channels such as the AI Incident Database or other proposed public repositories.[13] Clarity on what constitutes an AI incident, which incidents should be reported, the thresholds for reporting, and whether existing cyber-incident reporting channels are sufficient would be valuable. Ongoing work at CISA, through the Joint Cyber Defense Collaborative (JCDC), aims to provide further guidance later this year.[14]

- **Workforce:** Projecting what workforce CI organizations will need to leverage AI and meet the challenges posed by AI-enabled threats is difficult. It is unclear if AI risk management will require personnel with AI-specific skills, cybersecurity experts with specialization or cross-training in AI risk, or a completely new set of personnel with both AI and cybersecurity expertise. Some aspects of traditional cybersecurity best practices such as authentication and data protection also apply to managing AI risk. However, the design and implementation of AI systems requires unique expertise that many CI providers may not have in their current cyber workforce. At a minimum, the AI and cybersecurity experts in an organization will need some cross-training to collaborate effectively and speak a common language to address the full range of AI and cyber risks.

Challenges in AI Risk Management

As AI applications become more prevalent in CI, sectors and organizations must manage the attendant risk. Participants noted the need to integrate AI risk management into existing processes at many organizations. Yet, at the same time, ownership of AI risk can be ambiguous within current corporate structures. It was referred to by one participant as the AI “hot potato” being tossed around the C-suite.

Today, AI risk management does not neatly fall under any single corporate leadership position, such as the chief information security officer, the chief technology officer, the chief information officer, or the chief data officer. Aspects of AI, and its related risk, often span the responsibilities of these different roles. While the need to include AI risk management into the overall enterprise strategy is clear, who owns AI risk within the organization is anything but. For example, Govern 2.1 of the NIST AI RMF states that "roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks are documented and are clear to individuals and teams throughout the organization," but the details on which actors should be directly responsible are limited.[15] Some organizations are approaching this challenge by appointing a new chief AI officer, while others have rolled it into the responsibilities of a chief resilience officer. However, the most common—albeit potentially less permanent—solution has been for organizations to share the responsibility across roles or to "dual-hat" an existing officer, typically the chief data officer.

While organizations within and outside of CI are grappling with how to manage risks posed by AI, these challenges may be particularly acute within the CI sectors. Many CI providers have a "compliance culture" due to the high degree of regulation they face and the essential services they manage, such as providing clean water or keeping the lights on. Therefore, regulatory requirements and resulting organizational policies are often written in a binary manner—either the organization does or does not meet the given requirement. However, the same approach does not apply well in the context of AI. The output of AI models is inherently probabilistic: a system will or will not produce a certain outcome with probability of *n*. This is at odds with policies and requirements under a compliance-oriented regime that specify a system will (a 100 percent likelihood) or will not (a 0 percent likelihood) do something with complete certainty. As such, AI risk management demands a "risk-aware culture" in which the focus is on reducing the likelihood of harm rather than meeting a checklist of requirements. These differences in risk management cultures may affect the safe and secure adoption of AI in many CI sectors.

Fractured Guidance and Regulation

A commonly expressed concern during the workshop was that many CI providers are struggling to operationalize AI risk management. In addition to the resource constraints discussed earlier, two key factors contribute to this problem: fractured guidance and regulation.

- **Guidance:** There are a multitude of overlapping frameworks that pertain to AI, cybersecurity, and privacy. These include NIST's AI RMF (and subsequent

"Playbook" and draft "Generative AI Profile"), CSF, and Privacy Framework; the Federal Trade Commission's Fair Information Practice Principles; and a variety of standards from the International Organization for Standardization. Understanding how these frameworks work together, which set of guidance is applicable where, and how to operationalize recommended practices for a given AI use case represents a substantial hurdle for organizations. Participants noted two key challenges related to this issue.

- First, each respective framework presents numerous recommended practices to implement, and, when combined, the scope of those recommendations can become burdensome, even for well-resourced organizations. The lack of general guidance on how to prioritize among the multitude of recommended practices, particularly when facing resource constraints, and the lack of guidance tailored to specific sectors were highlighted as major obstacles to operationalizing recommended practices. Participants noted that community profiles, like those produced to accompany the CSF, were helpful additions to the high-level guidance. However, these profiles take time to develop, and currently there are no finalized profiles for the AI RMF. With the rapid pace of AI development and the push for adoption, there may be an important role for trusted third parties to move faster in addressing this guidance gap.

- Second, the ambiguity at the intersection of these overlapping frameworks makes it challenging for organizations to interpret what guidance applies where. For example, the core activities in both the cybersecurity and privacy frameworks include a protect function ("Protect" and "Protect-P," respectively), which covers recommended safeguards and security measures. Yet, the AI RMF does not have a protect function. While organizations can draw on security practices from the CSF, analogues from cybersecurity—such as red-teaming—do not always translate directly to the context of AI.[16] Furthermore, these measures may not protect against the range of vulnerabilities unique to AI systems.[17] The ambiguity and potential gaps that arise at the intersection of these frameworks make it difficult to piece together how they should be applied in concert. As a result, CI providers looking to implement safe and secure AI systems face the challenge of trying to deconflict implementation guidance from a patchwork set of frameworks, technical research reports, and industry practices. Distilling this information requires time and expertise that many organizations, particularly less well-resourced ones, cannot afford without assistance. Ongoing efforts within

NIST, such as the Data Governance and Management Profile, are likely to help in this regard and were deemed a high priority by participants.[18]

- **Regulation:** Concerns over the fractured regulatory environment regarding data protection and cybersecurity, and the potential for a similar governance regime for AI, pose another major barrier for CI providers in adopting AI systems. With the lack of overarching federal regulation for privacy or cybersecurity, a patchwork of requirements has been made at the state level that various CI providers must comply with. Furthermore, some CI providers have a global presence and are impacted by international regulations as well, notably the European Union's GDPR and the more recent Artificial Intelligence Act. The lack of harmonization between these different regulations poses a compliance risk for organizations seeking to implement AI systems, particularly those that may be customer facing or that train on consumer data.

Recommendations

The following recommendations stem from discussions held during the workshop and are designed to provide an array of policy options for governing the future use of AI in critical infrastructure. They are divided into four subsections by stakeholders at different levels of governance: (1) **cross-cutting** recommendations that apply to all actors at the intersection of AI and critical infrastructure; (2) recommendations for **government** actors to consider; (3) recommendations for CI **sectors**; and (4) recommendations for individual **organizations**, encompassing both CI operators and AI developers and deployers.

Cross-Cutting Recommendations

The following recommendations apply to all stakeholders within the critical infrastructure and AI ecosystems:

- **Participate in information sharing.** The sharing of best practices, threat information, and incidents is critical to maintaining the safety and security of AI systems employed in CI. While the specific channels for sharing AI security versus AI safety information are unclear, the need for information sharing across both domains is paramount.
    - **SRMAs should leverage existing venues for AI security information sharing.** Current ISACs provide a natural forum for additional collaboration on AI-enabled threats and security vulnerabilities in AI systems. The JCDC could potentially aid in these efforts as well. Less clear are the mechanisms for sharing information on AI safety risks that do not pertain to security. Channels for sharing AI safety information—such as cases of incorrect output, bias, or failures discovered in a given AI model—could be incorporated into existing ISACs or instituted separately. Integrating AI safety communication into the existing ISACs could reduce overhead, prevent redundancy, provide more holistic insight for managing risk, and alleviate the ambiguity between AI safety and security discussed previously. On the other hand, creating separate information-sharing centers for AI safety could provide more tailored intel, help reduce the volume of information to process, and maintain the security-focused mission of the ISACs[*]. An example of a sector-specific

[*] Separate information sharing channels for AI safety could potentially fit into or complement the AI Safety Institute as it continues to develop and gains capacity.

safety center (not focused on AI) is the Aviation Safety Information Analysis and Sharing operated by MITRE.

- **The CI sectors should consider establishing a centralized analysis center for AI safety and security.** High-level visibility into AI use across the CI sectors is vital to managing overarching risk. This includes identifying where and how AI is being used, developing best practices, and assessing AI safety and security information—whether shared through the same or different channels. To promote cross-sector information sharing and analysis that spans both AI safety and security, we recommend the creation of a centralized AI safety and security analysis center. The establishment of a National Critical Infrastructure Observatory, as recommended in a recent report from the President's Council of Advisors on Science and Technology, would create one potential home for this cross-sector center.[19]

- **CI operators and providers should share information on AI-related incidents, threats, vulnerabilities, and best practices.** Defining AI incidents and sharing relevant information when they occur, whether there are cybersecurity implications or not, will be critically important to identify new vulnerabilities and harms. For this information to be useful, providers need to ensure that they are collecting relevant data and audit logs to assess what led up to the incident occurring, how the incident unfolded, and what efforts were undertaken afterward to identify the source of the issue and remedy it going forward. We note that there is currently little guidance on communicating AI incidents, and the sooner guidance can be released the better. As discussed above, determining the communication channels to use for information sharing and to whom that information is sent is an important prerequisite.

  CI providers should also take proactive steps to share information on observed threats, vulnerabilities discovered, and industry best practices related to AI use and deployment. Furthermore, the sharing of sector-specific data, including training data for AI systems, could help CI providers. While there may be a tendency to retain data for proprietary reasons or risk of liability, a collaborative approach would help benefit organizations within each sector, particularly smaller providers who may not generate the requisite volume of data for AI use cases. An initial step could be prioritizing efforts to share data for AI applications that facilitate or protect critical services such as predictive maintenance and cyber

defense. Data sharing in these areas is likely more feasible, as incentives align across organizations, and is potentially very impactful.

- **Develop the workforce.** Participants universally agreed that hiring and developing AI talent is a crucial prerequisite for effectively adopting AI in critical infrastructure systems.
    - **Federal and state government organizations should fund training programs and other workforce development initiatives.** As mentioned above, workforce capacity—and the lack thereof—was a theme throughout the entire discussion. Some participants recommended that policymakers consider funding workforce development initiatives explicitly aimed at improving capacities within the CI sectors.
    - **CI sectors should coordinate workforce development efforts and develop sector-specific requirements.** The sectors should play an important intermediary role in the design and implementation of AI training programs. This starts with identifying the specific AI talent needs within their sector and developing requirements that help inform the design of the training programs. In addition, the CI sectors should take a leading role in coordinating the implementation of these programs and prioritizing where resources for workforce development are needed most.
    - **CI operators and providers should actively upskill their current workforces.** Developing requisite AI talent will remain a large undertaking, and one way to partially address the demand is to upskill existing staff. One aspect of this upskilling may be training individual workers capable of deploying and managing AI systems for organizations that operate them. Another may include promoting general AI literacy among staff on the proper use of AI-enabled tools as well as risks posed by AI-enabled threats, such as sophisticated spear-phishing attacks. Of particular note, CI providers should ensure that their staff are aware of the risk of including proprietary or sensitive information in prompts sent to third-party AI services.

Responsible Government Departments and Agencies

Specific recommendations for relevant government actors include:

- **Harmonize regulation relevant to CI.** Participants expressed confusion and uncertainty about patchwork security and data protection requirements,

particularly at the state level. Regulatory harmonization would help CI operators chart a path forward and better evaluate risks associated with AI adoption. Some participants also expressed a desire for harmonization efforts to apply to any future AI-specific legislation, both at the state and federal levels.

- **Work with sector partners to tailor and operationalize guidance for each sector.** Government guidance aimed at the CI sectors can be difficult to operationalize because of its generality. Inherently, developing guidance that applies to everyone runs the risk of fitting no one. This is particularly salient within the CI sectors, where providers often operate bespoke and specialized systems. Guidance tailored to specific sectors and additional guidance on operationalization would benefit many operators. For example, prior to the release of the NIST CSF, NIST had released a version of the cybersecurity framework specifically targeted at improving CI cybersecurity.[20] Similar tailoring of guidance related to AI—at a level more specific than existing resources such as the *NIST AI RMF Playbook*—may be helpful to CI operators, especially those who are under-resourced.[21] The NIST "Generative AI Profile" and the *Secure Software Development Practices for Generative AI and Dual-Use Foundation Models* are examples of such tailored guidance.[22] Developing AI profiles specific to CI sectors, similar to existing cybersecurity ones, would also help advance safe and secure adoption.

- **Support relevant infrastructure to test and evaluate AI systems.** As mentioned above, the practice of AI model evaluation remains immature. Many evaluations are conducted by model developers without a mechanism to independently evaluate the results. Importantly, however, the role third parties will play in evaluations remains unclear. Organizations such as the NIST AI Safety Institute could play a leading role in future model evaluations but will need additional resourcing in the form of funding and personnel. Third-party auditing of models and assessments against defined benchmarks could provide CI operators additional confidence in the safety and security of these models. Ideas discussed at the workshop included using existing test beds for CI or designing test beds exclusively for AI testing and evaluation, which could allow for continued research on how models behave in deployed environments. Additionally, further research into risk evaluation metrics for AI is needed, as well as a shared understanding of how cybersecurity test and evaluation practices can be used to test network infrastructures deploying AI technologies. Ideally, these resources should be accessible to all CI sectors.

- **Expand ecosystem monitoring efforts.** A continuation and expansion of efforts to identify AI use cases being deployed across CI sectors is critical for maintaining ecosystem-wide visibility and assessing overall risk. In conducting our background research on how sectors are using current AI applications, we found that many reported use cases lacked important details needed to assess risk, such as how the organization used the AI system (e.g., experiments in a laboratory or deployed in production) and what type of model they used. Future visibility efforts, such as the annual cross-sector risk assessments conducted by CISA, should collect and report these details.

Sectors

Recommendations for the CI sectors as a whole include:

- **Develop best practices.** Establishing and sharing best practices around the implementation and use of AI within a given sector is critical to operationalizing AI safety and security guidance. The CI sectors should facilitate the development and coordination of these tailored best practices, ensuring that providers both small and large can provide input into the process.

- **Expand and support mutual assistance.** To help address the disparities between and within sectors, workshop participants recommend expanding both informal and formal means of mutual assistance. These initiatives help share resources, talent, and knowledge across organizations in an effort to improve the security and resiliency of the sector as a whole. An example of formal mutual assistance is the Electricity Subsector Coordinating Council's Cyber Mutual Assistance Program, which connects a network of cybersecurity professionals who provide support to participating providers in the case of a cyber emergency. Informal mutual assistance often results from the efforts of large providers that have spillover or secondary benefits for smaller providers. Some examples could include the development of industry standards and the vetting of products and service providers. To address the issue of smaller providers not having a voice in some of these informal practices, larger organizations and sector-coordinating bodies should work to gather and incorporate input from smaller providers as a part of these processes.

Organizations

We break down the recommendations for individual organizations into those directed toward CI providers and those for AI developers.

Critical Infrastructure Operators

Recommendations for providers and organizations within CI sectors include:

- **Integrate AI risk management into enterprise risk management.** To address AI risk properly, it must be fully integrated into existing enterprise risk management practices. Organizations should develop these practices based on NIST's AI RMF and utilize tools such as NIST's recently released Dioptra assessment platform.[23] However, as noted previously, further tailored guidance is needed on how to integrate the AI RMF recommendations into existing practices, which are often based on the CSF and Privacy Framework.

- **Designate clear ownership over AI risk management.** While perspectives differed on who within the corporate structure should own AI risk, it is clear that integrating AI risk into enterprise risk management is dependent on defining ownership clearly. Since issues related to AI risk span many of the responsibilities of the standard corporate leadership positions, one option could be establishing a new chief AI officer role. If organizations are reluctant to create new positions, they may need to consider leveraging an existing internal forum or creating a new one that brings together the relevant organizational leaders to assess AI risk. However, for many smaller providers or organizations looking to deploy AI in a very narrow scope, assigning ownership of AI risk to an existing officer—or specific board member, for local providers—is likely preferable.

- **Remain cautious and informed before adopting newer AI technologies, particularly for sensitive or mission-critical tasks.** On the positive side, older machine learning techniques have been extremely beneficial in cybersecurity, particularly anomaly and threat detection. However, for newer AI technologies such as generative AI, participants were in favor of sectors adopting a cautious and measured approach to adoption. Tools like MITRE's AI Maturity Model can be helpful for providers to assess their ability and readiness to adopt AI systems.[24]

AI Developers

AI developers are producing new generative AI capabilities almost daily. These products have the potential to assist CI providers in the operation of their systems and their sector-specific use cases. However, many CI operators do not have the AI expertise to make informed risk management decisions. To assist CI operators, AI developers should:

- **Engage in transparency best practices.** This includes publishing information about models in the form of model cards or "nutrition labels," similar to what has been proposed for Internet of Things devices.[25] Participants also noted that increased information on training data provenance, which most AI developers currently do not provide, would be beneficial to evaluate risk associated with an AI system. Transparency on the results of model evaluations (for safety, security, or otherwise) and model vulnerabilities would also be valuable.

- **Improve trust by developing methods for AI interpretability and explainability.** While the two terms are sometimes used interchangeably, *interpretability* generally refers to the ability to mechanistically analyze the inner workings of a model's decision-making process, while *explainability* refers to providing post hoc explanations for a model's behavior. While methodologies for both interpretability and explainability would help improve trust in AI systems, interpretability may be particularly important for logging and verification. Meanwhile, a lack of explainability is a major deterrent for CI operators considering adopting AI systems, especially for OT or customer-facing use cases. While these are evolving fields of research and participants acknowledged that there is currently no magic bullet for explainable or interpretable AI, continued investment in these fields could be beneficial for improving operators' trust in AI systems.

Authors

The four workshop organizers are listed first, followed by coauthors listed alphabetically by last name.

Kyle Crichton – Research Fellow, CSET
Jessica Ji – Research Analyst, CSET
Kyle Miller – Research Analyst, CSET
John Bansemer – Senior Fellow and the Director of the CyberAI Project, CSET
Zachary Arnold – Analytic Lead for the Emerging Technology Observatory, CSET
David Batz
Minwoo Choi – PNC Financial Services
Marisa Decillis – Pacific Northwest National Laboratory
Patricia Eke – Microsoft Corporation
Daniel M. Gerstein
Alex Leblang
Monty McGee – Director of Cyber Partnerships & Engagement, Edison Electric Institute
Greg Rattray – Chief Strategy and Risk Officer, Andesite
Luke Richards – Pacific Northwest National Laboratory
Alana Scott – Ericsson

Acknowledgments

We would like to thank several participants in the roundtable who contributed greatly to the discussion but were unable to participate in the writing process: Benjamin Amsterdam, Dakota Cary, Miles Martin, Kat Megas, Nikhil Mulani, Matt Odermann, Noah Ringler, Dr. Jonathan Spring, and Martin Stanley.

Document Identifier: doi: 10.51593/20240032

Appendix A: Background Research Sources

Government / Intergovernmental
DHS, "FACT SHEET: DHS Facilitates the Safe and Responsible Deployment and Use of Artificial Intelligence in Federal Government, Critical Infrastructure, and U.S. Economy," April 2024.
DHS, *Mitigating Artificial Intelligence (AI) Risk: Safety and Security Guidelines for Critical Infrastructure Owners and Operators*, April 2024.
Treasury, *Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector*, March 2024.
DOE, *Potential Benefits and Risks of Artificial Intelligence for Critical Energy Infrastructure*, April 2024.
T. Boe et al., "Application of Artificial Intelligence in EPA Homeland Security," EPA, May 2023.
Government Accountability Office (GAO), *Artificial Intelligence: Fully Implementing Key Practices Could Help DHS Ensure Responsible Use for Cybersecurity*, February 2024.
DHS, "Artificial Intelligence Use Case Inventory," last modified August 2024.
DOE, "Agency Inventory of AI Use Cases," accessed May 2024.
Treasury, "Artificial Intelligence (AI) Use Cases," May 2023 and August 2022.
EPA, "EPA Artificial Intelligence Inventory," last modified November 2023.
Vida Rozite, Jack Miller, and Sungjin Oh, "Why AI and Energy are the New Power Couple," International Energy Agency, November 2023.
Science / Academia / Nongovernmental Organizations / Federally Funded Research and Development Centers / Industry
Christopher Sledjeski, "Principles for Reducing AI Cyber Risk in Critical Infrastructure: A Prioritization Approach," MITRE, October 2023.
Almando Morain et al., "Artificial Intelligence for Water Consumption Assessment: State of the Art Review," *Water Resources Management*, April 2024.
Daniel M. Gerstein and Erin N. Leidy, "Emerging Technology and Risk Analysis: Artificial Intelligence and Critical Infrastructure," RAND, April 2024.
Ahmed E. Alprol et al., Artificial Intelligence Technologies Revolutionizing Wastewater Treatment: Current Trends and Future Prospective," *Water*, January 2024.

Matt Mittelsteadt, "Critical Risks: Rethinking Critical Infrastructure Policy for Targeted AI Regulation," Mercatus Center, March 2024.
Tobias Sytsma et al., "Technological and Economic Threats to the U.S. Financial System: An Initial Assessment of Growing Risks," RAND, July 2024.
Mohammad El Hajj and Jamil Hammoud, "Unveiling the Influence of Artificial Intelligence and Machine Learning on Financial Markets: A Comprehensive Analysis of AI Applications in Trading, Risk Management, and Financial Operations," *Journal of Risk and Financial Management*, October 2023.
Indigo Advisory Group, "Utilities & Artificial Intelligence—A New Era in the Power Sector," Medium, May 2024.
Maeve Allsup and Laura Weinstein, "Seven Ways Utilities Are Exploring AI for the Grid," *Latitude Media*, October 2023.
Documents Mentioned During Workshop
Office of the National Cyber Director, *Summary of the 2023 Cybersecurity Regulatory Harmonization Request for Information*, June 2024.
CISA National Risk Management Center, "National Critical Functions: An Evolved Lens for Critical Infrastructure Security and Resilience," April 2019.
MITRE, "The MITRE AI Maturity Model and Organizational Assessment Tool Guide," November 2023.
NIST, *NIST AI RMF Playbook*, Trustworthy & Responsible AI Resource Center, accessed May 2024.
Treasury, "The Financial Services Sector's Adoption of Cloud Services," May 2024.
Hong He Fei and Jiang Yun, "Intelligent Operation Robot Becomes a 'Little Expert' in Substation Inspection," ETO Scout, June 2024.

Endnotes

[1] Exec. Order No. 14110, 3 CFR (2023), www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[2] DHS, *Mitigating Artificial Intelligence (AI) Risk: Safety and Security Guidelines for Critical Infrastructure Owners and Operators* (Washington, DC: DHS, 2024), www.dhs.gov/sites/default/files/2024-04/24_0426_dhs_ai-ci-safety-security-guidelines-508c.pdf.

[3] Treasury, *Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector* (Washington, DC: Treasury, 2024), https://home.treasury.gov/system/files/136/Managing-Artificial-Intelligence-Specific-Cybersecurity-Risks-In-The-Financial-Services-Sector.pdf; Office of Cybersecurity, Energy Security, and Emergency Response, *Potential Benefits and Risks of Artificial Intelligence for Critical Energy Infrastructure* (Washington, DC: DOE, 2024), www.energy.gov/sites/default/files/2024-04/DOE%20CESER_EO14110-AI%20Report%20Summary_4-26-24.pdf.

[4] Muhammad El Hajj and Jamil Hammoud, "Unveiling the Influence of Artificial Intelligence and Machine Learning on Financial Markets: A Comprehensive Analysis of AI Applications in Trading, Risk Management, and Financial Operations," *Journal of Risk and Financial Management* 16, no. 10 (October 2023): 434, https://doi.org/10.3390/jrfm16100434; Indigo Advisory Group, "Utilities & Artificial Intelligence—A New Era in the Power Sector," Medium, May 15, 2024, https://medium.com/@indigoadvisory/utilities-artificial-intelligence-research-brief-0890e4ec5533.

[5] T. Boe et al., "Application of Artificial Intelligence in EPA Homeland Security," EPA, July 26, 2022, https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=357842&Lab=CESER.

[6] Almando Morain et al., "Artificial Intelligence for Water Consumption Assessment: State of the Art Review," *Water Resource Management* 38 (2024): 3113–3134, https://doi.org/10.1007/s11269-024-03823-x; Ahmed E. Alprol et al., "Artificial Intelligence Technologies Revolutionizing Wastewater Treatment: Current Trends and Future Prospective," *Water* 16, no. 2 (January 2024): 314, https://doi.org/10.3390/w16020314.

[7] Ashutosh Bhoi et al., "IoT-IIRS: Internet of Things Based Intelligent-Irrigation Recommendation System Using Machine Learning Approach for Efficient Water Usage," *PeerJ Computer Science* 7:e578 (June 21, 2021), https://doi.org/10.7717/peerj-cs.578.

[8] CISA, "#StopRansomware: Ransomware Attacks on Critical Infrastructure Fund DPRK Malicious Cyber Activities," February 9, 2023, www.cisa.gov/news-events/cybersecurity-advisories/aa23-040a; CISA, "PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure," February 7, 2024, www.cisa.gov/news-events/cybersecurity-advisories/aa24-038a.

[9] Microsoft Threat Intelligence, "Staying Ahead of Threat Actors in the Age of AI," *Microsoft Security Blog*, February 14, 2024, www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/.

[10] Xiangyu Qi et al., "AI Risk Management Should Incorporate Both Safety and Security," arXiv preprint arXiv:2405.19524 (2024), https://arxiv.org/abs/2405.19524.

[11] Micah Musser et al., "Adversarial Machine Learning and Cybersecurity: Risks, Challenges, and Legal Implications" (CSET, April 2023), https://cset.georgetown.edu/publication/adversarial-machine-learning-and-cybersecurity/.

[12] Andrew Lohn and Wyatt Hoffman, "Securing AI: How Traditional Vulnerability Disclosure Must Adapt" (CSET, March 2022), https://cset.georgetown.edu/publication/securing-ai-how-traditional-vulnerability-disclosure-must-adapt/.

[13] Office of Senator Mark R. Warner, "Warner, Tillis Introduce Legislation to Advance Security of Artificial Intelligence Ecosystem," news release, May 1, 2024, www.warner.senate.gov/public/index.cfm/2024/5/warner-tillis-introduce-legislation-to-advance-security-of-artificial-intelligence-ecosystem.

[14] CISA, "CISA, JCDC, Government and Industry Partners Conduct AI Tabletop Exercise," news release, June 14, 2024, www.cisa.gov/news-events/news/cisa-jcdc-government-and-industry-partners-conduct-ai-tabletop-exercise.

[15] NIST, Artificial Intelligence Risk Management Framework (Washington, DC: Department of Commerce, 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[16] Jessica Ji, "What Does AI Red-Teaming Actually Mean?," CSET, October 24, 2023, https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/.

[17] MITRE, "MITRE ATLAS," 2024, https://atlas.mitre.org/.

[18] NIST, *NIST Joint Frameworks Data Governance and Management Profile Concept Paper* (Washington, DC: Department of Commerce, 2024), www.nist.gov/system/files/documents/2024/06/18/DGM%20Profile%20Concept%20Paper%20%2806.18.24%29.pdf.

[19] President's Council of Advisors on Science and Technology, *Strategy for Cyber-Physical Resilience: Fortifying Our Critical Infrastructure for a Digital World* (Washington, DC: Executive Office of the President, 2024), www.whitehouse.gov/wp-content/uploads/2024/02/PCAST_Cyber-Physical-Resilience-Report_Feb2024.pdf.

[20] NIST, *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1* (Washington, DC: Department of Commece, 2018), https://csrc.nist.gov/pubs/cswp/6/cybersecurity-framework-v11/final.

[21] NIST, *NIST AI RMF Playbook* (Washington, DC: Department of Commerce, 2024), https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.

[22] NIST, *AI Risk Management Framework*; Harold Booth et al., *Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile* (Washington, DC:

NIST, 2024), www.nist.gov/publications/secure-software-development-practices-generative-ai-and-dual-use-foundation-models-ssdf.

[23] NIST, "What Is Dioptra?," 2024, https://pages.nist.gov/dioptra/.

[24] MITRE, "Artificial Intelligence Maturity Model" (MITRE, 2024), www.mitre.org/news-insights/fact-sheet/artificial-intelligence-maturity-model.

[25] Margaret Mitchell et al., "Model Cards for Model Reporting," arXiv preprint arXiv:1810.03993 (January 14, 2019), https://arxiv.org/abs/1810.03993; NIST, *Report for the Assistant to the President for National Security Affairs (APNSA) on Cybersecurity Labeling for Consumers: Internet of Things (IoT) Devices and Software* (Washington, DC: Department of Commerce, 2022), www.nist.gov/system/files/documents/2022/05/24/Cybersecurity%20Labeling%20for%20Consumers%20under%20Executive%20Order%2014028%20on%20Improving%20the%20Nation%27s%20Cybersecurity%20Report%20%28FINAL%29.pdf.