MITRE's Response to the OSTP RFI on a National Artificial Intelligence Strategy

July 7, 2023

For additional information about this response, please contact:

Duane Blackburn
Center for Data-Driven Policy
The MITRE Corporation
7596 Colshire Drive
McLean, VA 22102-7539

policy@mitre.org
(434) 964-5023

About MITRE

MITRE is a not-for-profit company that works in the public interest to tackle difficult problems that challenge the safety, stability, security, and well-being of our nation. We operate multiple federally funded research and development centers, participate in public-private partnerships across national security and civilian agency missions, and maintain an independent technology research program in areas such as artificial intelligence, intuitive data science, quantum information science, health informatics, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience. MITRE's approximately 10,000 employees work in the public interest to solve problems for a safer world, with scientific integrity being fundamental to our existence. We are prohibited from lobbying, do not develop or sell products, have no owners or shareholders, and do not compete with industry. Our multidisciplinary teams (including engineers, scientists, data analysts, organizational change specialists, policy professionals, and more) are thus free to dig into problems from all angles, with no political or commercial pressures to influence our decision making, technical findings, or policy recommendations.

MITRE has a 50-year history of partnering with federal agencies to apply the best elements of artificial intelligence (AI) and machine learning (ML) while developing and supporting ethical guardrails to protect people and their personal data. Our team's experience with the entirety of the AI/ML adoption and life cycle has strengthened our ability to anticipate and resolve future needs that are vital to the safety, well-being, and success of the public and the country.

Introduction and Overarching Recommendations

<u>Taking a Strategic Approach for the National AI Strategy</u>

MITRE applauds the administration's efforts to create a "cohesive and comprehensive" National AI Strategy that provides a whole-of-society approach to AI and that will focus on *both opportunities and risks*. While strategy and policy work on AI is far from novel, prior analyses have predominantly targeted a singular benefit/issue and/or focused on a singular use case. An overall strategy can serve to place those efforts into context but, more important, to also set holistic national goals and expectations (which will enable us to see which areas still require enhanced attention). To that end, MITRE recommends that the administration use a strategic planning framework consistent with the Government Performance and Results Act in developing and implementing this strategy (see Figure 1).

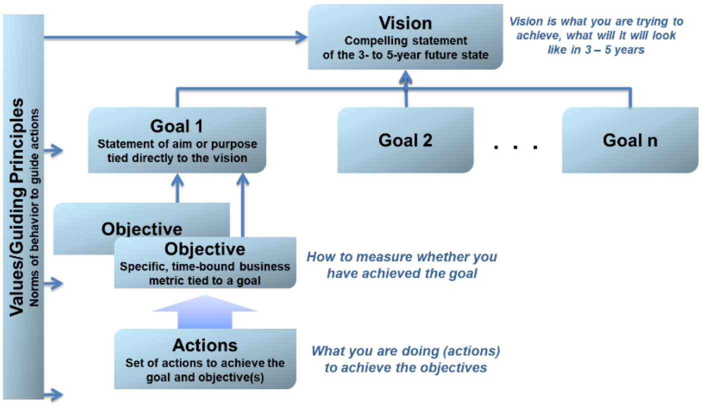
Vision
Compelling statement of the 3- to 5-year future state
Vision is what you are trying to achieve, what will it will look like in 3 – 5 years
Values/Guiding Principles
Norms of behavior to guide actions
Goal 1
Statement of aim or purpose tied directly to the vision
Goal 2
. . .
Goal n
Objective
Objective
Specific, time-bound business metric tied to a goal
How to measure whether you have achieved the goal
Actions
Set of actions to achieve the goal and objective(s)
What you are doing (actions) to achieve the objectives

Figure 1 – Strategic Planning Framework with Values/Guiding Principles

Such a structured planning framework provides:

- A universal and compelling vision for the future of AI in the United States
- A series of goals that collectively enables the vision to be met
- Subordinate objectives and strategies that are specific and time-bound, which both help to drive activities so that they successfully meet the goals and provide the Executive Office of the President the ability to measure progress
- A set of values and principles for AI that will help guide subsequent activities

Overcoming Slow AI Adoption

The application of AI technologies has tremendous potential to improve the efficiency and quality of services delivered across all sectors of society, yet AI adoption is not what it can be. There is a need to accelerate the assurance of AI technologies and overcome risk aversion to early AI adoption if our nation is to extract the maximum value out of AI. This national strategy can help overcome this issue by promoting AI assurance and accountability.

AI solution developers and adopters need safe harbor in laboratory and sandbox environments to develop and explore new AI applications. This involves working to advance context-sensitive, formative test and evaluation approaches to discover, chart, and mitigate emerging consequences of AI solutions as they are developed, especially when AI intersects with human values in ways that may be difficult to predict. "Assurance cases" can be instituted to document and provide the evidence and a compelling argument that an AI system satisfies certain critical assurance properties in specific contexts providing a level of transparency and justified confidence to both regulators and the public. It may be beneficial to envision a common foundational approach to AI assurance and accountability based on the National Institute of Standards and Technology (NIST) AI Risk Management Framework where sector regulators and AI adopters assess and mitigate risks tailored to specific AI use cases. Also, AI assurance and accountability cannot be "one and done." They must be executed along each step of the value chain and in a continuous feedback loop as the AI technology and its use evolves. With these best practices and due diligence in place, AI adopters can know how they will be held accountable and, having gone through an informed decision-making process, they can be provided protections for having done so even if the eventual outcome may entail the realization of certain risk. Risk identification and mitigation cannot eliminate all risk. However, excessive risk aversion stemming from lack of protections to those who exercise sound, informed decision-making processes can undermine our ability to maximize the potential benefits of AI. Moreover, while these measures are geared to mitigate risks, lessen risk aversion, and promote adoption, assured AI implementation will require significant investment. Underestimation of the cost required to deploy and sustain AI solutions is another hindrance to achieving rapid AI adoption with assurance and accountability.

A cohesive national strategy will require the federal government to provide educational materials and best practices on AI and its application to state and local government entities. These issues are complicated, and properly understanding them will be difficult for these entities, which often lack in-house expertise. For example, the decisions (and lack of evidence as their foundation) throughout the nation on face recognition over the past few years are a harbinger of bigger issues and inconsistencies to come in the broader AI domain.

MITRE's Approach to Answering the RFI's Questions

Developing a "cohesive and comprehensive" national strategy also requires considering matters from all angles (such as benefits vs. risks and opportunities vs. status quo) together. To support this analysis, MITRE is providing narratives that cross questions issued in the RFI. While doing so, we reference the original question numbers for traceability. We also acknowledge the administration's plans to study prior request for information (RFI) responses in developing this strategy and are therefore silent in this response on matters in which we have already provided input (see Appendix A).

RFI Response Narratives

Collaborating on AI (RFI Questions 11, 25, 26, and 28)

Collaboration is key to accelerating AI innovation and ensuring appropriate use throughout the nation, as well as to ensure the nation's future security and competitiveness.[1,2,3] There are many forms of collaboration that must take place (including interagency, public-private, international, cross-domain, and developer-user), and there must also be collaboration across these forms for the nation to succeed. It is imperative that the nation approach collaboration on AI issues strategically and holistically, rather than as individual and disconnected endeavors. The national strategy that this RFI is supporting must set the foundation, direction, and expectation to do so, both individually and collectively.

As discussed in this response's introduction, MITRE recommends that the National AI Strategy be based on a strategic planning framework that provides a universal vision for the future of AI in the United States, a series of goals that collectively enables the vision to be met, and

[1] C. Ford, et al. A "Horizon Strategy" Framework for Science and Technology Policy. 2021. MITRE, https://www.mitre.org/sites/default/files/2021-11/prs-21-1440-horizon-strategy-framework-science-technology-policy.pdf.

[2] Final Report. 2021. National Security Commission on Artificial Intelligence, https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

[3] Mid-Decade Challenges to National Competitiveness. 2022. Special Competitive Studies Project, https://www.scsp.ai/wp-content/uploads/2022/09/SCSP-Mid-Decade-Challenges-to-National-Competitiveness.pdf.

subordinate objectives and strategies for each goal. Necessary forms of collaboration can then be selected for each objective and strategy, with a purpose, timeline, and interdependencies identified for each. Doing so provides context and purpose for each collaboration, ensuring that it properly supports the national vision (rather than simply collaborating for collaboration's sake).

Collaboration Within the Context of Technological Evolution

Figure 2 shows typical technological evolution and adoption trends via an overlay of the hype cycle, adoption curve, and productivity curve. The types of entities most heavily involved (and for which we wish to drive collaboration) vary depending on the current stage of evolution, and the needs and desires that each has will change as the evolution advances.

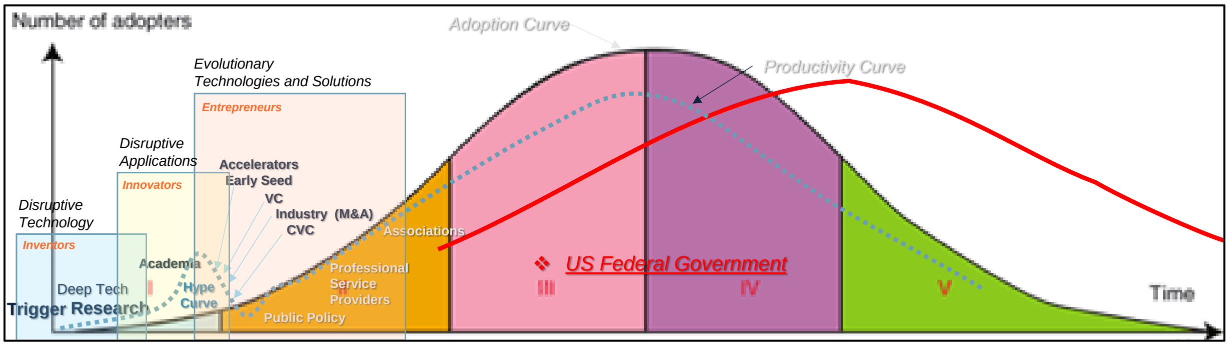

Figure 2 – Rogers Theory of Diffusion of Innovation Overlaid with the Gartner "Hype Cycle" Technology Productivity Curve

For advanced AI, we're still very much in the early stages of this evolution and attempting to strategically speed up the full transition from innovation (light blue) to early adoption (orange) while simultaneously taking steps to compress the future early adoption stage so that we move into initial majority use (pink). Activities and objectives, as well as their supporting collaborations, for innovation, early adoption, and majority use are necessarily different. If we're to successfully meet our enhanced AI evolution objectives, we need to simultaneously ensure that each collaboration can succeed as rapidly as possible and that hand-offs (or transitions) into later-stage collaborations are strategically executed. Such planning and execution need to happen at multiple levels, and by a variety of stakeholders, as discussed in the next section.

Collaboration Governance

The first step in implementing this national strategy is to recognize that it will predominantly be a voluntary collaboration spanning a wide range of entities, each with varying levels of commitment and resources. Participants must feel not only sufficient value to warrant their continued participation, but also that they have some input or influence over decisions made. That said, there must also be a strong leadership and coordination function to succeed.

MITRE and the Office of Management and Budget (OMB) previously encountered such a juxtaposition when designing the operating model of OMB's proposed Government Effectiveness Advanced Research (GEAR) Center. After reviewing multiple options, we settled on an approach that could also be effective in this context. This design had three components, to which we are adding a fourth for the AI space:

- A **Federal Government role** that leverages its unique ability to set a national vision, convene groups to work collaboratively, and to balance competing needs and equities for everyone's benefit. The federal government is also a huge source of data and piloting opportunities in a variety of contexts.

- A **Private Sector network of networks** to bring together and leverage private sector expertise and resources from throughout industry, academia, research organizations, non-profits, and the venture capitalist community.
- An **"Operator" entity** to serve as both a strategic and tactical coordinator and as a trusted third party between the government and private sector.
- The **Public**, which, through various engagement activities, can have roles in designing and implementing the strategy.

Federal Government. Even though the federal government would not be individually deciding and directing activities in this model, it still has critical leadership and support roles as the preeminent catalyst of the activity. It also has unique qualities that it can bring to bear across the collection of the group's activities, such as:

- It is the most influential determiner of national priorities.
- It is the nation's largest sponsor of research.
- It possesses huge amounts of data on a variety of matters.
- It has an unprecedented ability to convene executives.
- It has a breadth of piloting environments and opportunities.
- It has the largest and widest audience for publicizing the group's activities and its impact.

Federal activities should leverage shared pools of resources, expertise, and lessons learned. Various approaches to do so are highlighted in the MITRE document "Interagency S&T Leadership."[4]

Private Sector. Our analyses showed that success would depend on reaching large groups of thought leaders from throughout the extended private sector ecosystem (including the venture capital community) quickly, systematically, and strategically. Rather than taking a shotgun approach of targeting entities directly, the plan instead would be to predominantly identify existing networks (with diversity of thought, experiences, and geographic locations) to leverage and pull their members into the broader collaboration. We also recognized that each participant's role would vary by their level of commitment and involvement, and would generally fall into one of three categories:

- **Knowledge**. Provide subject matter experts to aid in strategic planning and to lead or participate in collaboration (the most participants).
- **Resources**. Capital investments and assets such as facilities, data, tools, and human capital to facilitate execution.
- **Governance**. Help shape the strategic direction of the collaboration and its supporting activities (the fewest participants).

An important note is that each private sector network, and individual, participant will need to feel there is sufficient value recovered from their investment(s) in the initiative's activities. This

[4] D. Blackburn. Interagency S&T Leadership. 2016. MITRE, https://www.mitre.org/sites/default/files/publications/pr-16-0916-interagency-s-and-t-leadership.pdf.

will vary by the category of their involvement and their individual areas of focus in their normal business.

Operator. The Operator provides an unbiased servant leadership role, providing the "nuts and bolts" that is required within a collaboration for it to succeed. Example activities include:

- Serving as the trusted third party between the government and private sector (and between competing private sector entities), driving them to consensus decisions and facilitating (or even managing, in some cases) agreed-on collaborations
- Being entrusted with data that cannot be widely shared but must be leveraged[5]
- Technical and operational expertise
- Centralized point of contact for the collective effort
- Meeting venue and facilitation, administrative assistance, reporting, and budgeting

Public. This group is the primary beneficiary of AI products and services, sometimes directly and other times indirectly. It is imperative that they see value in AI, that issues have been overcome or are being managed, and that they can influence how it is used. They thus need to be consulted with so that their needs and desires can play a role in designing and implementing the strategy. The public should be leveraged to help understand what use cases are appropriate and can also serve as an at-scale monitor for when things are going wrong.

To be effective, however, inputs from the public (like all others) will need to be evidence-based and sufficiently representative of the different groups that will be affected by AI, possibly in different ways. That is currently problematic as numerous instances of mis-, dis-, and malinformation about AI are having corruptive influences. Overcoming this influence will be required for most public input to be useful and will also benefit other aspects of the strategy.

International Collaboration

International collaboration poses an additional layer of challenges, and not only because our adversaries are also attempting to lead on AI. Some of our closest partner nations have differing views on privacy approaches and appropriateness of activities by the government versus the private sector.

In our ongoing collaborations with international AI institutions and leaders, MITRE has found it beneficial to initially focus on AI applications for societal good to provide a foundation to build on. Our collaborators have also emphasized the importance of not-for-profit and charity organizations in creating an internationally independent voice for responsible use and adoption of AI. By collaborating with international partners, the United States will be able to amplify the positive benefits of AI and mitigate harms by tapping into existing community networks to track adoption and assess impact.

[5] One example of such a role is MITRE's efforts within Aviation Safety Information Analysis and Sharing (ASIAS), a collaboration between the Federal Aviation Administration and the aviation community.

Advancing Equity & Human-Centered Solutions (RFI Questions 9, 10, and 12)

Considerable attention has been placed on identifying and mitigating ways that AI can amplify inequitable impacts. These well-founded concerns are the subject of growing regulatory, public policy, governance, and research focus, and remain open challenges to the successful adoption of AI that is consistent with our nation's values and laws. As MITRE has already provided input along these lines, and expects most respondents to this RFI to focus on this aspect, we're instead going to focus on opportunities for AI to enhance equity and human-centered solutions.

> *"Algorithms have the potential to help us to excise disparate treatment, to reduce discrimination relative to human decision-making, to limit disparate impacts, and also to predict much more accurately than humans can in ways that disproportionately benefit disadvantaged groups—what we call the 'disparate benefit' of algorithms."*[6]

A fundamental aspect of such analyses is recognizing the importance of making proper comparisons. Many do not, making the mistake of comparing AI capabilities and limitations to an idealized state rather than comparing them to the current state. This happens from both an accuracy perspective (something much more accurate than the status quo can be a significant improvement, even if it's not as accurate as desired) and from a bias perspective (many existing approaches have undetermined biases, which could be worse than AI-driven biases in a proposed new approach). We need to begin properly assessing the status quo when assessing AI options and impacts, not letting our desire to reach "perfect" blind us to solutions that are only "significantly better." The government can be a model for such practice going forward.

Example Domains

Healthcare. There is concrete evidence of disparate outcomes across all aspects of healthcare. Access to care, positive outcomes from medical interventions, and overall physical and mental health are measurably poorer for underserved populations, for a complex mix of reasons. AI can be leveraged to overcome such issues for the underserved without negatively impacting other patients.[7,8,9]

Access to justice. "(E)quality under the law is often elusive for moderate-income and poor individuals. Those who can't afford an attorney are at a disadvantage as compared to those represented by counsel and often disproportionately burden courts, agencies, and other institutions that must adjudicate their claims or defenses. Used appropriately, AI technology has

[6] J. Kleinberg, et al. Discrimination in the Age of Algorithms. 2019. arXiv, https://arxiv.org/abs/1902.03731. Last accessed July 3, 2023.

[7] E. Pierson, et al. An algorithmic approach to reducing unexplained pain disparities in underserved populations. 2021. Nature Medicine, https://pubmed.ncbi.nlm.nih.gov/33442014/. Last accessed July 3, 2023.

[8] J. Halamka, et al. Addressing racial disparities in surgical care with machine learning. 2022. NPJ Digital Medicine, https://pubmed.ncbi.nlm.nih.gov/36180724/. Last accessed July 3, 2023.

[9] C. Johnson-Mann, et al. Equity and artificial intelligence in surgical care. 2021. JAMA Surgery, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8273554/. Last accessed July 3, 2023.

the potential to reduce the disadvantages of self-representation by intermediating between litigants and rule-governed institutions."[10] Also: [11,12]

Access to credit. Using AI to more accurately recognize creditworthy borrowers who have been historically overlooked due to insufficient credit history, including underbanked individuals and credit invisibles, can provide benefits to underserved cohorts without having an impact on others. By leveraging alternative data sources, such as income, employment history, and "digital exhaust," AI can more accurately assess borrowers' credit scores.[13]

MITRE recommends sustained policy attention to identify and reduce barriers to adoption of new capabilities for enhancing equitable outcomes, science and technology (S&T) investments to further advance these capabilities, and broader communication about current opportunities to enhance equity using AI. In addition, the federal government can lead by example by having a sustained focus on advancing equity in administrative law, criminal justice, healthcare, and other consequential impacts at the federal level.

Understanding Bias Origins

Advances in computational social science present opportunities to better understand the origins, pathways, and secondary effects of various forms of bias and to identify good intervention points to address. MITRE recommends S&T investment to advance broader exposure to opportunities and results for policymakers. MITRE also recommends leveraging AI within social science analyses. This untapped area of research presents opportunities to better understand the origins, pathways, and secondary effects of various forms of existing bias and to identify the most-promising intervention points to consider.

Remember the Human Aspects

The equity impact (good or bad) of AI systems is not strictly about "what runs on electricity." Rather, AI is part of a complex consequential sociotechnical system,[14] which analyses must recognize. Humans are very involved in developing the systems, feeding them data, analyzing their output, and making actionable decisions—and humans are inherently flawed. Human activities can create multiple and significant issues within an AI system, yet the majority of policy discussion (and blame for incorrect results) to date has centered solely on the algorithm

[10] Workshop on Artificial Intelligence for Access to Justice. Karl Branting, https://www.karlbranting.net/ICAIL2023/. Last accessed July 3, 2023.

[11] K. Sonday. Tech-Enabled A2J: How tech is helping pro se litigants navigate the courts. 2020. Thomson Reuters, https://www.thomsonreuters.com/en-us/posts/legal/tech-enabled-a2j-pro-se-litigants/. Last accessed July 3, 2023.

[12] J. Snyder. RoboCourt: How artificial intelligence can help pro se litigants and create a "fairer" judiciary. 2022. Indiana Journal of Law and Social Equity, https://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=1136&context=ijlse. Last accessed July 3, 2023.

[13] J. Kleinberg, et al.

[14] From the NIST AI Management Framework (https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf): "AI systems are socio-technical in nature, meaning they are a product of the complex human, organizational, and technical factors involved in their design, development, and use. Many of the trustworthy AI characteristics—such as bias, fairness, interpretability, and privacy—are directly connected to societal dynamics and human behavior. AI risks—and benefits—can emerge from the interplay of technical aspects combined with socio-technical factors related to how a system is used, its interactions with other AI systems, who operates it, and the social context into which it is deployed."

and the data. The strategy being developed cannot forget the people aspect; they need proper training, guidance, and oversight, specific to their role within the system construct.

Cybersecurity (RFI Questions 5 and 6)

There are several ways that AI, including large language models (LLMs), can be used as part of a human-machine teaming approach to support the development of more secure software and hardware, including:

- LLMs can be trained on repositories of secure source code and in turn generate new code that, with human review, may be less vulnerable to various kinds of known attacks.
- AI can be leveraged to automate security testing to identify vulnerabilities that may not be discovered by programmers.
- AI can be deployed to continuously monitor software for vulnerabilities as new attack vectors are discovered over time.

Recent advances in LLMs also create new opportunities for applications of AI to cybersecurity problems, including:

- Automation support to human analysts, such as assisting intrusion analysts and incident responders in understanding observed behavior and identifying potential response options
- Assisting malware analysts in more rapidly decoding and explaining the functionality of malware samples undergoing static and dynamic analysis
- Assisting cyber threat intelligence analysts in identifying adversary tactics, techniques, and procedures in intelligence reporting and digital forensic artifacts along with forming inferences around adversary attribution, capability, and intent

While some preliminary research has been conducted in this area on an informal basis, there is inadequate published research into these potential applications. Investment in research and development of LLMs applied to these cybersecurity tasks is needed.

Regarding the use of AI to rapidly identify cyber vulnerabilities in existing critical infrastructure, AI could conceptually be used to identify critical nodes and support rapid, large-scale what-if scenarios to inform risk and gap identification that can guide resource allocations. However, this potential will not be realized until efforts to capture the details of individual critical infrastructure and the dependencies between different infrastructure instances become more complete.

National Security (RFI Questions 4 and 7)

Understanding specific benefits and risks of using AI within national security contexts requires accurately comparing anticipated technological capabilities against mission need. The specifics of this comparison quite often require delving into classified matters and are thus inappropriate for this document, but MITRE could discuss them in an appropriate setting.

At a high level, AI holds the potential to be transformational to our national security capabilities, and one of the main risks is that we will be slow to operationalize its use compared to our adversaries. AI has the potential to transform the way defense and intelligence agencies gather, analyze, and act on information as it can process vast amounts of information and can help detect threats much faster than current approaches. For example, AI can be used as a second set of eyes to review incoming data for patterns, trends, and anomalies. AI can be used to summarize available data, both raw and finished, to provide more complete, up-to-date, and timely foundational intelligence. Additionally, newer AI methods have shown promise as a guide in critiquing intelligence products by improving language and suggesting alternative hypotheses.

However, there is also a "significant separation between the people who have problems to solve and the people who understand and develop the technology to solve those problems … (t)oo many of our senior leaders and even operators have an unrealistic view of what AI can and can't do. Too many of our AI experts are naïve about the real problems"[15] and operational opportunities. This issue is then compounded by security classifications and special access limitations that are inherent to the national security space, as well as contracting rules that can be overly deliberative and time-consuming. While these concerns hold true for many technological innovations for national security, their impacts will be more pronounced for AI due to its unusual pace of advancement and potential impact. MITRE recommends accelerating AI assurance efforts to keep up with the pace of AI technology development.

Oversight/Regulatory (RFI Questions 1, 2, 3, 19, and 23)

Any attempt to secure or regulate a new technology should be informed by its **vulnerabilities**, **threats** that exploit those vulnerabilities either intentionally or unintentionally, and the ultimate **risk** of damage, harm, or loss of human life, health, property, or the environment.[16] The foundation of effective oversight and regulation in the AI sphere requires the following:

- A consistent definition of AI to best delineate and characterize what objects are being regulated
- A scalable regulatory instrument that governs AI
- A clear governance mechanism that offers a combination of options from voluntary self-regulation to government-mandated policies and procedures

MITRE previously offered comments regarding systematic approaches to oversight and regulation of AI in response to the National Telecommunications and Information Administration's RFI on AI Accountability.

[15] E. Niewood. Applying AI to the Right National Security Problems. 2022. Aerospace America, https://aerospaceamerica.aiaa.org/departments/applying-ai-to-the-right-national-security-problems/. Last accessed June 29, 2023.

[16] T. Clancy, et al. A Sensible Regulatory Framework for AI Security. 2023. MITRE, https://www.mitre.org/sites/default/files/2023-06/PR-23-1943-A-Sensible-Regulatory-Framework-For-AI-Security_0.pdf.

Appendix A – MITRE's Recent RFI Responses on AI

- The Office of Science and Technology Policy's (OSTP) RFI to the Update of the National Artificial Intelligence Research and Development Strategic Plan: https://www.mitre.org/sites/default/files/2022-03/pr-21-01760-16-mitre-response-ostp-rfi-national-artificial-intelligence-research-and-development-strategic-plan.pdf.
- The National Telecommunication and Information Administration's Request for Comment on AI Accountability Policy: https://www.mitre.org/sites/default/files/2023-06/PR-22-01891-21-MITREs-Response-to-the-NTIA-RFI-on-Artificial-Intelligence-Accountability.pdf.
- OSTP's and the National Science Foundation's RFI on Implementing Initial Findings and Recommendations of the National Artificial Intelligence Research Resource Task Force: https://www.mitre.org/sites/default/files/2022-07/pr-21-01760-25-response-mitre-ostp-rfi-implementing-findings-recommendations-national-artificial-intel.pdf.
- OSTP's RFI on Public and Private Sector Uses of Biometric Technologies: https://www.mitre.org/sites/default/files/2022-04/pr-21-01760-11-mitre-response-information-on-public-and-private-sector-uses-of-biometric-technologies.pdf.