RE: Response to the Office of Science and Technology Policy "Request for Information: National Priorities for Artificial Intelligence" (OSTP-TECH-2023-0007-0001)

To Whom It May Concern:

Palantir Technologies Inc. ("Palantir") is a US-based software company that builds platforms to enable public, private, and non-governmental organizations to integrate, analyze, and collaborate on their data in a secure and privacy-protective way. We are proud to make software that enables the institutions that serve our societies to use their data responsibly and effectively.

Palantir was founded in 2003 on the conviction that it is essential to preserve fundamental principles of privacy and civil liberties while using data. It is for this reason that Palantir established one of the world's first Privacy & Civil Liberties Engineering ("PCL") teams more than a decade ago, specifically to focus on the development of privacy-protective technologies and to foster a culture of responsibility around their development and use. Our response to this Request for Information ("RFI") is based on insights gathered over 20 years of experience building technology to uphold and enforce ethical and accountable practices in the use of our software products, including Artificial Intelligence ("AI") enablement tools and platforms. Palantir has contributed extensively to the conversation on the rise and appropriate use of AI technology through multiple public forums and responses to Requests for Comments ("RFCs"). In this document, we reference many of our previous responses, sometimes quoting directly from them and, at other times, re-contextualizing the ideas expressed and applying them to this specific RFI.

Most prominently, we have submitted two responses to the National Telecommunications and Information Administration ("NTIA") and one to the Federal Trade Commission ("FTC"). The first was in response to the NTIA's Privacy, Equity, and Civil Rights ("PECR") RFC,[1] and the second was in response to the NTIA's AI Accountability Policy ("AIAP") RFC.[2] The third response referenced is to the FTC's Advanced Notice of Proposed Rulemaking on Commercial Surveillance and Data Security.[3] In addition to these domestic responses, we also contributed oral and written responses to the United Kingdom's House of Lords in their Inquiry on AI in Weapons Systems.[4] We have also published reflections on the UK House of Lords testimony, as well as video clips of our direct feedback.[5]

We are grateful to the OSTP for the opportunity to contribute to this important policy discussion. We welcome any request for clarification and look forward to the OSTP's final report on these critical issues.

Sincerely,

Courtney Bowman	**Arnav Jagasia**
Global Director of Privacy and Civil Liberties Engineering, Palantir Technologies	Privacy and Civil Liberties Engineering Lead, Palantir Technologies

[1] https://www.regulations.gov/comment/NTIA-2023-0001-0020

[2] https://www.regulations.gov/comment/NTIA-2023-0005-1360

[3] https://www.regulations.gov/comment/FTC-2022-0053-0702

[4] https://www.palantir.com/assets/xrfr7uokpv1b/T6XBvRNbtgOysf4XuYqpH/1ba005ae5b469eb47ea484ad34dadcea/Palantir_Submission_to_the_HL_AI_in_Weapons_Systems_Committee.pdf

[5] Appearance at UK House of Lords Committee on AI in Weapon Systems. PALANTIR BLOG. (2023), https://blog.palantir.com/appearance-at-uk-house-of-lords-committee-on-ai-in-weapon-systems-2354862a6641

Responses to RFI Questions

Protecting rights, safety, and national security

1. What specific measures – such as standards, regulations, investments, and improved trust and safety practices – are needed to ensure that AI systems are designed, developed, and deployed in a manner that protects people's rights and safety? Which specific entities should develop and implement these measures?

In our NTIA AIAP RFC response, Palantir discussed the structure that effective and actionable AI standards should take to ensure that AI systems are designed, developed, deployed, and regulated in a manner that protects people's rights and safety. Echoing those sentiments, we believe that the primary focus of AI development and regulation should be context- and domain-specific since the nature and the scale of risks attached to AI technologies will vary profoundly between domains and use cases. Attempting to address every risk (often subtle and socio-technical), across all sectors simultaneously in overarching legislation is a nonviable task. While general principles may be reasonably established to cut across sectors and span most applications of AI, AI standards and regulation will be most meaningful if they are established on a per-industry basis.

Both the benefits and risks of AI technology will depend on the particular environment in which it operates. Since AI technology is embedded within specific contexts, there will always be unique socio-technical harms that may require mitigation or intervention attuned to those domain considerations. The focus of AI regulation and accountability should therefore be on holistic, sectoral assessments and standards that address questions, such as:

- How does AI embed in/augment/replace specific, existing industry applications?
- How do AI capabilities fit within a broader digital infrastructure (e.g., as components of complex systems)?
- What outcomes do AI capabilities drive, broadly conceived?
- How might established objectives for industry applications of technologies change with the introduction of AI components (both functioning and potentially malfunctioning)?
- How can novel AI applications best take into account the unique standards, norms, and histories of particular industries?

The context-dependent nature of both deploying and assessing the success and failure of AI capabilities may make the prospect of AI rulemaking and guidance seem difficult. However, as we have discussed in the AIAP RFC response, accountability measures can be reasonably organized to address this context dependency head-on. By improving context sharing and reducing (to the greatest extent possible) information asymmetry, AI systems can be made to better accommodate the equities of all stakeholders affected by their deployment, thus helping to mitigate potential harms to people's rights and safety. To help achieve responsible outcomes, we reiterate the following, actionable recommendations from the AIAP RFC response[6]:

- **Provide model documentation:** All AI models used in AI systems should have documentation that, at minimum, provides details about how the model was trained (including an outline of data used in training), what its intended use is, and what its known areas of limitation are. Additional documentation parameters should be adopted in accordance with the type of model, its domain of use, and other identified sensitivities.
- **Encourage strong security controls to constrain model misuse:** Depending on the nature and intended use, AI systems should — to the extent possible — be designed with strong security

[6] https://www.regulations.gov/comment/NTIA-2023-0005-1360

controls in mind. For example, highly sensitive use cases may opt for "fail-closed" and "Zero Trust"[7] architectures (i.e., by default, no one has access to anything; all access must be granted in accordance with domain-specified authorizations for use).

- **Require tools for model monitoring and observability, especially in consequential domains:** AI models used in production settings should produce telemetry and audit logs so that systems engineers, governance teams, product managers, and others can understand how the models are being used. This is a critical component of the model lifecycle, as monitoring informs better problem definition, model development, and testing & evaluation (T&E) strategy, in addition to being a key lever for accountability.

The above mechanisms, among others described at greater length in our AIAP RFC response, will help mitigate the potential downstream negative consequences that AI may have on people's rights and safety.

In addition to these more granular mechanisms, Palantir has previously addressed overarching structural challenges associated with AI implementation. In our FTC Rulemaking response,[8] Palantir publicly addressed high-level aspects of frameworks that we consider most relevant to Responsible AI. These include AI systems that meet the following criteria:

- **Reliable (Safe, Secure, Resilient, Robust)**: AI systems should be built with capabilities for assessing the safety, security, and effectiveness of models throughout their entire lifecycles. AI systems should also be designed to mitigate or reduce the potential impact of accidents and other unintended harmful behavior,[9] as well as provide capabilities for assessing and minimizing adversarial attempts to either degrade models or undermine the privacy (and other rights) of individuals whose data might have been used to train the models.
- **Traceable (Auditable, Governable)**: AI systems should provide the capabilities to understand relevant development processes, data sources, and the provenance of all data used for model development. AI systems should also provide transparent access to auditable standard operating procedures (SOPs), design guidelines, and appropriate documentation.[10]
- **Accountable (Liable, Responsible)**: Accountability has widely been cited as an important consideration for the development of algorithms and models.[11] In order to put accountability into practice, there should be a clear definition of the roles and workflows for people responsible for the different parts of the AI system. Moreover, such systems should allow for both third-party oversight and internal audits.[12]

[7] *Building Software for a Zero Trust World*. PALANTIR BLOG. (2023), https://blog.palantir.com/building-software-for-a-zero-trust-world-61d440e5976e

[8] https://www.regulations.gov/comment/FTC-2022-0053-0702

[9] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané. *Concrete Problems in AI Safety*. (Jul. 25, 2016), https://arxiv.org/abs/1606.06565.

[10] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru. *Model cards for model reporting*. PROC. CONF. FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 220. (Jan. 2019), https://dl.acm.org/doi/abs/10.1145/3287560.3287596.

[11] Maranke Wieringa. *What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability*. PROC. CONF. FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 1. (Jan. 2020), https://dl.acm.org/doi/abs/10.1145/3351095.3372833.

[12] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, Parker Barnes. *Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing*. PROC. CONF. FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY. (2020), https://dl.acm.org/doi/abs/10.1145/3351095.3372873.

- **Human-centered (Participatory, Socially Beneficial)**: AI systems should benefit individuals, society, and the environment overall. They should not erode trust, and should augment, not replace, human decision-making. For uses of automation that impact individuals' privacy and civil liberties, in particular, the goal of AI systems should be to enhance the context and quality of human judgment.
- **Scoped (Problem-driven, Reproducible, Rigorous)**: AI systems should be developed and deployed with clear specifications for intended scope or domain of use. For some classes of AI technologies, this may mean a narrowly-defined and appropriately-scoped purpose specification, as well as an explicit representation that the model may be ill-suited for applications outside that scope. It may also imply measures to rigorously ensure that model results can be reproduced for a given modeling problem.[13] For other classes of models – such as so-called "general purpose AI" (which, nonetheless, are subject to identifiable limitations in model applicability and reliability) – problematic domains or categories of use should be plainly articulated, and measures should be implemented to constrain, contextualize, control, or otherwise limit risk of misapplication.

2. How can the principles and practices for identifying and mitigating risks from AI, as outlined in the Blueprint for an AI Bill of Rights and the AI Risk Management Framework, be leveraged most effectively to tackle harms posed by the development and use of specific types of AI systems, such as large language models?

Palantir has elsewhere provided recommendations on how to effectively apply principles similar to those outlined in the Blueprint for an AI Bill of Rights and the AI Risk Management Framework.[14] Our recommendations include:

- Ensure that AI accountability and risk mitigation processes cover the entirety of a fully-integrated AI system, and not just component AI tools.
- Ensure that AI accountability and risk mitigation processes are applied throughout the entire model lifecycle, with a specific focus on clarity of model applicability and model maintenance. This also includes extending AI model lifecycles to ensure they are accounting for the unique characteristics of specific AI systems.
- Ensure that the teams applying these principles represent diverse disciplines and roles, as model management and development is a multi-disciplinary, multi-stakeholder process.
- Invest in foundational data platforms for, *inter alia*, the development, deployment, and use of AI that have capabilities for version control, branching, security controls, and other data and security protection primitives, as these primitives are often crucial for operational use of AI in many consequential settings. For more on the importance of creating a foundational data infrastructure to support AI innovation, see Palantir CTO Shyam Sankar's testimony[15] to the U.S. Senate Armed Services Subcommittee on Cybersecurity.
- Combine technical tools — including those described above — with organizational approaches that promote governance and AI ethics. As discussed in our PECR RFC response[16], we have

[13] Sayash Kapoor and Arvind Narayanan. *Leakage and the Reproducibility Crisis in ML-Based Science*. (2022), https://arxiv.org/pdf/2207.07048.pdf.

[14] See for example our whitepaper, *AI On RAILs, A Responsible AI Lifecycle Framework*. (2023), https://www.palantir.com/assets/xrfr7uokpv1b/4nVc0FDbOrqeVHUZQdIcwZ/21b4e3f13479ecf87c4da4fcc0e8c1a0/RAILS_Whitepaper-FINAL-.pdf

[15] *Testimony Before the Senate Armed Services Subcommittee on Cybersecurity: Statement by Shyam Sankar*. https://www.armed-services.senate.gov/download/sankar-statement?download=1

[16] https://www.regulations.gov/comment/NTIA-2023-0001-0020

advocated for creating organizational data governance bodies and implementing codes of conduct as means for reducing data harms.

4. What are the national security benefits associated with AI? What can be done to maximize those benefits? 5. How can AI, including large language models, be used to generate and maintain more secure software and hardware, including software code incorporating best practices in design, coding and post deployment vulnerabilities? 7. What are the national security risks associated with AI? What can be done to mitigate these risks? [Note: The response below addresses Questions 4, 5, and 7 together.]

While the potential national security benefits associated with AI may be too numerous to be thoroughly accounted for here — and many such benefits may yet to be conceived — we highlight the most significant benefits below:

- AI and AI-enabled software can improve the speed, efficacy, and security of real-time data integration and information sharing. This capability can help create a unique "decision advantage" for policymakers and warfighters, which in turn may help prevent conflict and mitigate risks as they arise.
- AI and AI-enabled software can maximize the safety, security, effectiveness, efficiency, and precision of legacy defense systems, as well as underpin the development and use of future platforms and weapons systems.
- Large Language Models (LLMs) can lower the barrier to entry for the Defense/Intel workforce to use advanced software by allowing users to conduct analysis and give computational commands through natural language (i.e., not code). As such, LLMs allow more government personnel to incorporate advanced AI-enabled models in their workflows without the need to invest in and rely on the upgraded computer programming proficiency of each end-user.
- AI and LLMs can assist with the code review process to help identify and fix security issues before and during software deployment. AI can also serve as a useful guide to writing secure code by notifying developers of errors being made in real time.
- AI and AI-enabled software can facilitate prototyping and experimentation of new operational workflows at the battalion, or even company echelons; improved document search functions for subject matter experts working on massive data systems; speedy battle damage assessments; creative assistance for operational planning; and general wargaming tasks.
- AI and AI-enabled software can help strengthen an adherence to Just War and International Humanitarian Law (IHL) principles by improving the speed, clarity, and accuracy of battlefield situational awareness[17] — potentially limiting civilian harm through reduced targeting errors — as well as improving the strength of post-engagement investigations into potential IHL violations.

Just as AI offers a plethora of National Security benefits, so too does it pose National Security risks. While there are too many considerations to enumerate here, we identify several prominent risks and associated risk mitigation techniques in the table below:

[17] *AI, Automation, and the Ethics of Modern Warfare*. PALANTIR BLOG. (2023), https://blog.palantir.com/ai-automation-and-the-ethics-of-modern-warfare-df1f0b212397

Risk	Mitigation Technique
Lack of clarity in terms of who is ultimately responsible for the outcomes of an AI system. This is a general risk but one that is exacerbated in consequential domains like National Security and Defense.	Invest in foundational platforms for AI and tools for data governance, transparency, and security. To use AI operationally is to contend with the full operating context and lifecycle of AI systems. These kinds of foundational investments are necessary for AI to operate both effectively and responsibly.
Underemphasizing AI maintenance and sustainability. National Security and Defense challenges are dynamic and continuously changing. AI solutions can become stale or be erroneously repurposed.	Focus on building capabilities for long-term model maintenance and monitoring; invest in software systems that consider the end-to-end AI system that can help users react to changes in data and context; move contracts from defining a fixed capability to requiring updates over time.
AI is fielded without the proper testing, training, or iteration in realistic usage scenarios, which in turn leads to failures in production and deployment.	Following the development process pursued by the Algorithmic Warfare Cross-Functional Team — perhaps the largest and most successful U.S. Government AI project — employ a "field-to-learn" methodology to develop AI against our nation's most pressing national security problems. Funding opportunities for responsibly-constructed, "field-to-learn" experiments is an effective way to expose technologists, ethicists, policy-makers, and AI users to the specific challenges of AI deployment and use, which is necessary to create and test accountability mechanisms that address the operational and real-world challenges of AI technologies.
Algorithm vulnerability to non-authorized and/or foreign influence and interference.	Focus on building data provenance and security control features to help detect and inform inappropriate access, as well as create monitoring systems with human oversight to check algorithm outputs and verify the basis for a prediction or other algorithmic output before critical decisions are made.

Table 1: AI National Security Risks and Mitigation Techniques

Advancing equity and strengthening civil rights

12. What additional considerations or measures are needed to assure that AI mitigates algorithmic discrimination, advances equal opportunity, and promotes positive outcomes for all, especially when developed and used in specific domains (e.g., in health and human services, in hiring and employment practices, in transportation)?

AI and Machine Learning (ML) models are developed from real-world data, meaning that algorithms may reflect current and historical structural shortcomings — such as biases and discriminations — in ways that can harden inequities by race and ethnicity, gender identity, sexual orientation, disability, age, social class, and geography.

We believe that an ethical approach to AI must deal with the risks associated with the disproportionate impact of a system's empirical outcomes. Evaluating how AI outcomes may be skewed across different population segments — including groups representing sensitive or vulnerable categories (e.g., age, gender, race, ethnicity) — can assist in further system refinements aimed at optimizing outcomes on appropriate evaluation parameters. One particular challenge in this regard is that societal notions of fairness may not always cleanly translate into mathematical correctness in the form of specific optimization parameters such as (in the context of binary classifier identification) True Positive, False Positive, Positive Predictive Value, False Discovery Rate, or any of an expansive slew of potential performance evaluation measures. Programs using AI that seek to encode a single canonical measure for evaluating and assessing determinations of model fairness must grapple with both this normative ambiguity (i.e., that there may be irreducible qualitative aspects of fairness) and also the demonstrable mathematical impossibility of being able to achieve all (or even most) measures of fairness at all times.[18] We believe that programs using AI that seek to produce ethically defensible outcomes must forthrightly evaluate and understand these trade-offs. Ultimately, an AI program should be intentional and explicit in determining which optimization measures are employed, as well as document the consequences of those decisions.

Ensuring positive outcomes must be underpinned by applying accountability to systems that consistently exhibit undesirable bias. From our experience, the form of accountability most relevant to the use of AI technologies is never "one-size-fits-all." Rather, accountability invariably relates to the context of use and its corresponding risk profile. Accountability simply does not make sense as a technological or procedural abstraction. For this reason, in defining context-specific accountability mechanisms, it is important to recognize that those contexts will also intersect closely with other intrinsic areas of concern or risk (including privacy, civil liberties, fundamental rights, sustainability, equity, inclusion, diversity, etc.) which should be directly factored into the chosen or mandated accountability mechanism. As written in our AIAP RFC response, "[o]ur recommendation is not to come up with a law that diverges from existing, well-established frameworks. To the contrary, there is a benefit in the (soft) harmonization of existing AI regulations, which will likely be most effective if it is encouraged at the federal level. Among others, federally-imposed regulations could contribute to reducing the cost of compliance and more effectively enforcing principles of AI accountability."[19]

Echoing other sentiments expressed in the PECR RFC response, "technologists, social scientists, and policymakers can best advance the responsible use of technology, while safeguarding against harms especially to marginalized populations, by focusing on building fault tolerance and resilience into technology systems."[20] By building systems with these priorities in mind, we can innovate and make progress while simultaneously taking account of and acting appropriately in light of risks.

On this note, there are already pre-existing domestic and international laws from which any measures taken by the Federal government can draw inspiration.[21] The General Data Protection Regulation (GDPR) has set the standard of data protection in the European Union (EU) and in other jurisdictions (including an ever-growing number of US state-level consumer privacy legislative acts) due to its extraterritorial reach. As the first modern comprehensive privacy law, the GDPR has proved useful in ensuring the accountability of data processing systems, including AI systems. Although the GDPR is focused on regulating the processing of personal data and not AI, the use of AI may involve the processing of personally identifiable information (PII), and so the GDPR can be directly relevant to the implementation of AI tools. The GDPR's most impactful provisions are those related to individual rights. For instance, an individual has a right to be informed about the processing of her data, can request deletion of her information, and has the right to access her personal data stored or processed by AI systems. The GDPR

[18] *Palantir Technologies' Approach to AI Ethics*. (2023), https://www.palantir.com/pcl/palantir-ai-ethics/.
[19] https://www.regulations.gov/comment/NTIA-2023-0005-1360
[20] https://www.regulations.gov/comment/NTIA-2023-0001-0020
[21] We focus our remarks here on already enacted regulation but recognize that the forthcoming EU AI Act will likely carry additional implications for advancing equity and civil rights/liberties in the use of AI technologies with transatlantic reach. Here we focus on GDPR as one example of international regulation providing a baseline for accountability considerations in all forms of data processing systems, including AI technologies.

also requires organizations to provide individuals with information about the logic involved in automated decision-making, as well as the significance and consequences of such processing.

Beyond these provisions on data subject rights, the GDPR improves accountability of AI systems by requiring notifications to the end-users of changes to the subprocessors of PII. To the extent that AI providers act as subprocessors of personal information, end-users need to be informed about the potential passing of their personal information to such third parties. Finally, the GDPR's basic principles of data minimization, purpose limitation, and lawfulness of data processing remain unchanged in the AI context. Palantir's position on this has been publicly held for some time, most recently in the FTC Rulemaking response.[22] AI systems should be designed in a way that minimizes the sharing of PII (e.g., AI outputs should, to the extent possible, be limited to non-PII, even when a model is trained on personal data) and ensure a legal basis of that sharing (e.g. legitimate interest, a research exemption, or valid consent).

Specifically in the realm of health information, and as we have written in the PECR RFC response[23], transparency and trust in data are two critical considerations when working with sensitive health information like PII. First, the development of AI systems that handle sensitive health information should adhere to the principle of transparency by requiring a diverse breadth of stakeholders to have input in its development and use. Having advocates play impactful roles in the development of an AI system and ensuring visibility into data quality is particularly helpful for marginalized groups, who may otherwise lack access or the means to determine how their data is used. Second, AI models that deal in health data are going to be scrutinized as to the accuracy and completeness of the data. Especially when AI systems process sensitive information, it may be wise to design AI systems such that AI models are subject to the same security control mechanisms and restrictions as human users. For example, granular access control systems may be applied to better support use limitation principles by guarding against the misuse and repurposing of data, while still allowing for legitimate uses to proceed.

Promoting economic growth and good jobs

19. What specific measures – such as sector-specific policies, standards, and regulations – are needed to promote innovation, economic growth, competition, job creation, and a beneficial integration of advanced AI systems into everyday life for all Americans? Which specific entities should develop and implement these measures?

Promoting AI innovation should be contingent on preventing damaging outcomes or uses of AI technology. Generally speaking, AI is most effective and responsible when employed to assist and enhance human execution and decision-making, rather than replacing it. As should be the case with all technologies, the impact of AI should be in elevating humanity, not in undermining, endangering, or replacing it.

In the same vein, Palantir supports the Bipartisan Policy Center's (BPC's) recommendation that, "Preparing the workforce of the future and managing the rise of AI in an inclusive manner can help us best capture the potential of the new technology while softening the problems. The American worker can thrive in the AI-driven economy, but policymakers should help them prepare to reach their full potential."[24]

To ensure that AI has a positive-long-term impact on the American economy, and that the rise of AI technology does not disproportionately harm certain vulnerable populations and economic sectors, Government and Industry leaders must work hand-in-hand. The regulations and guidelines that the U.S.

[22] https://www.regulations.gov/comment/FTC-2022-0053-0702
[23] https://www.regulations.gov/comment/NTIA-2023-0001-0020
[24] *AI and the Workforce*. BIPARTISAN POLICY CENTER. https://bipartisanpolicy.org/report/ai-the-workforce/

Government and private sector agree upon should determine the optimal development and deployment of AI technology into the workforce, as well as the safety net and retraining mechanisms that need to be present in response to potential job losses and setbacks.

In terms of more specific measures that can simultaneously promote innovation, economic growth, competition, job creation, and the beneficial and responsible integration of advanced AI systems into everyday life for all Americans, we emphasize that such measures are best set by entities at the sector- and industry-level. As such, sector-specific agencies should be given the tools, resources, and expertise-based discretion to directly carry out their own programs for responsible and effective innovation and job creation, in alignment with the unique objectives determined through sector-specific evaluations. In short, the agencies and entities most likely to create and promote impactful measures are those that understand their domains best — these are the entities that should be most empowered to develop and implement measures since high-level top-down prescriptions are unlikely to be as effective.

23. How can the United States ensure adequate competition in the marketplace for advanced AI systems?

The United States can ensure adequate competition in the marketplace for advanced AI systems by supporting the important role of the private sector in the development of this critical technology. As Palantir CTO Shyam Sankar identified in his testimony to the U.S. Senate Armed Services Subcommittee on Cybersecurity:

"[W]e believe that the Department of Defense must recognize that while there are some cases where it makes sense to build in-house, it is often more efficient and effective to acquire AI capabilities from the commercial sector. The bleeding edge of AI development is happening in America's robust marketplace of commercial firms. Instead of the government insisting on building in-house (which stands in direct competition with American businesses), or itself trying to serve as a systems integrator, the choice to buy commercial solutions will lead to faster, more cost effective, and sustainable advancement of AI capabilities for America's warfighters. Furthermore, the acquisition of commercially available AI capabilities will allow the Department of Defense to progress to the 'field-to-learn' stage of AI development from the start, instead of waiting years to develop certain capabilities in-house."[25]

The above statement is not only true for ensuring a robust marketplace for Defense AI. It holds equal merit across the entire spectrum of Government agencies, services, and missions.

Innovating in public services

24. How can the Federal Government effectively and responsibly leverage AI to improve Federal services and missions? What are the highest priority and most cost-effective ways to do so? 25. How can Federal agencies use shared pools of resources, expertise, and lessons learned to better leverage AI in government? 26. How can the Federal Government work with the private sector to ensure that procured AI systems include protections to safeguard people's rights and safety? 27. What unique opportunities and risks would be presented by integrating recent advances in generative AI into Federal Government services and operations? 28. What can state, Tribal, local, and territorial governments do to effectively and responsibly leverage AI to improve their public services, and what can the Federal Government do to support this work? [Note: The response below addresses Questions 24-28 together.]

The use of AI opens numerous avenues for the public sector at-large and the Federal Government, in particular, to improve public services. We advocate for an operational, "field-to-learn" approach to AI

[25] *Testimony Before the Senate Armed Services Subcommittee on Cybersecurity: Statement by Shyam Sankar*. https://www.armed-services.senate.gov/download/sankar-statement?download=1

deployment which provides a framework that enables technical innovation, as well as legal and ethically accountable boundary-setting. As we have previously noted, it does so by better exposing technologists, ethicists, policy-makers, and AI users to the real-world challenges of AI deployment and use, as opposed to more theoretical musings that, while interesting, are often untethered from the reality of both the technology and operational setting. Those benchmark realities are essential components of constituting AI accountability ecosystems that work (i.e., that address the practical and real challenges of AI technologies).

Public entities — from the Federal Government to state, Tribal, local, and territorial (STLT) entities — can ensure they are effectively and responsibly leveraging AI to improve their services through two steps: 1) Laying the data-based[26] foundations[27] for AI success; and 2) Using that foundation to deploy AI towards projects that directly serve the public interest.

Overall, AI technology is only as effective and powerful as the data on which it is trained, as well as the execution of tasks it is assigned. Considering this, public entities will benefit most from AI by first building the necessary data infrastructure — including version control, reproducibility, and security among others — for AI technology to flourish. A strong data foundation facilitates greater adherence to accountability and regulatory mechanisms by enabling the auditing, governance, and traceability of data, tools, and processing within an AI system. Leading with a robust and secure data foundation will ensure that AI technology is working with accurate data and that human users are able to more effectively oversee AI activities and make necessary corrections.

[26] *Trust in Data (Palantir Explained, #4)*. PALANTIR BLOG. (2023), https://blog.palantir.com/trust-in-data-palantir-explained-4-c2adcdc31325

[27] *Taking Your Data Science Models to the Next Level*. PALANTIR BLOG. (2022), https://blog.palantir.com/taking-your-data-science-models-to-the-next-level-149d9c4269ec

References

AI and the Workforce. Bipartisan Policy Center. https://bipartisanpolicy.org/report/ai-the-workforce/

AI On RAILs, A Responsible AI Lifecycle Framework. (2023), https://www.palantir.com/assets/xrfr7uokpv1b/4nVc0FDbOrqeVHUZQdIcwZ/21b4e3f13479ecf87c4da4fcc0e8c1a0/RAILS_Whitepaper-FINAL-.pdf

AI, Automation, and the Ethics of Modern Warfare. Palantir Blog. (2023), https://blog.palantir.com/ai-automation-and-the-ethics-of-modern-warfare-df1f0b212397

Appearance at UK House of Lords Committee on AI in Weapon Systems. Palantir Blog. (2023), https://blog.palantir.com/appearance-at-uk-house-of-lords-committee-on-ai-in-weapon-systems-2354862a6641

Building Software for a Zero Trust World. Palantir Blog. (2023), https://blog.palantir.com/building-software-for-a-zero-trust-world-61d440e5976e

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané. *Concrete Problems in AI Safety*. (Jul. 25, 2016), https://arxiv.org/abs/1606.06565.

Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, Parker Barnes. *Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing*. Proc. Conf. Fairness, Accountability, & Transparency. (2020), https://dl.acm.org/doi/10.1145/3351095.3372873.

Maranke Wieringa. *What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability*. Proc. Conf. Fairness, Accountability, & Transparency 1. (Jan. 2020), https://dl.acm.org/doi/abs/10.1145/3351095.3372833.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru. *Model cards for model reporting*. Proc. Conf. Fairness, Accountability, & Transparency 220. (Jan. 2019), https://dl.acm.org/doi/abs/10.1145/3287560.3287596.

Palantir Technologies. *Palantir Technologies' Approach to AI Ethics*. (2023), https://www.palantir.com/pcl/palantir-ai-ethics/.

Sayash Kapoor and Arvind Narayanan. *Leakage and the Reproducibility Crisis in ML-Based Science*. (2022), https://arxiv.org/pdf/2207.07048.pdf.

Taking Your Data Science Models to the Next Level. Palantir Blog. (2022), https://blog.palantir.com/taking-your-data-science-models-to-the-next-level-149d9c4269ec

Trust in Data (Palantir Explained, #4). Palantir Blog. (2023), https://blog.palantir.com/trust-in-data-palantir-explained-4-c2adcdc31325

Prior responses to similar RFIs, RFCs, or other requests:

FTC Rulemaking response: https://www.regulations.gov/comment/FTC-2022-0053-0702

NTIA AIAP response: https://www.regulations.gov/comment/NTIA-2023-0005-1360

NTIA PECR response: https://www.regulations.gov/comment/NTIA-2023-0001-0020

Testimony Before the Senate Armed Services Subcommittee on Cybersecurity: Statement by Shyam Sankar. https://www.armed-services.senate.gov/download/sankar-statement?download=1

UK HOL response:
https://www.palantir.com/assets/xrfr7uokpv1b/T6XBvRNbtgOysf4XuYqpH/1ba005ae5b469eb47ea484ad34dadcea/Palantir_Submission_to_the_HL_AI_in_Weapons_Systems_Committee.pdf