Openpolicy

April. 29, 2024

To: Office of Management and Budget (OMB), *Atten: Samantha Hubner, OFCIO_AI@OMB.eop.gov*

OpenPolicy comments on

Request for Information (RFI) Related to OMB's Responsible Procurement of Artificial Intelligence in Government

Overview

OpenPolicy appreciates the opportunity to provide feedback on the OMB RFI under the AI EO 14110. As a technology company, OpenPolicy aims to empower innovative businesses of all sizes to interact with policymakers and contribute to policy development. We are dedicated to promoting the use of AI, innovation, and technology to enhance open policymaking and facilitate the integration of technology into compliance and governance processes. We look forward to a continuous collaboration with the OMB as this RFI develops and are ready to offer additional input as necessary. We remain eager to support the EO and its implementation, as illustrated by our involvement with the ***NIST AI Safety Institute*** Consortium and continued support of the White House AI Executive Order ("AI EO").[1]

We believe that the open and collaborative nature of the policymaking dialogue is essential to further the implementation of the AI EO, which can be significantly contributed by the participation of innovative companies and, specifically, "startup" companies that develop cutting-edge AI security and safety solutions and AI-enabled solutions. Indeed, many, if not most, of the technologies used to support the requirements of the AI EO and relevant OMB and the underlying NIST guidelines referred to, evaluate the measurements, testing, and audibility of AI, data and security posture, and facilitate the secure adoption of AI and sharing of data, more broadly, are developed by such innovative companies and startups – ***these are the communities OpenPolicy collaborates with***. The involvement of these companies in the process is further essential to reduce to "lock-in" effect referred to in the RFI

[1] [1] White House, "What are they Saying", https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/31/what-they-are-saying-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ (Oct. 31, 2023).

and support the involvement of a diverse supply of AI solutions to the U.S. government.

OpenPolicy applauds OMB for its leadership in implementing the AI EO requirements. We are especially satisfied with the emphasis on managing the performance and risks of AI and metrics development to enable performance-based procurement of AI, which should include AI security and safety, AI red teaming, governance and automation in these areas, as well as the need for specific, actionable responses from stakeholders to shape initiatives for evaluating and auditing AI systems. We are further encouraging OMB to focus on the need to diversify the nature of AI solutions used by agencies, as well ensuring their security and safety.

Recent reports have illustrated that extended reliance on particular solutions, that may have vulnerabilities, can expose the government to extended risk. We encourage OMB to continue to survey diverse solutions for AI security and saftey and remain available to engage on this topic.

Today, we believe that one of the most critical gaps is fostering enhanced measurability of AI risk, and AI security and safety outcomes, a better understanding of the holistic risk of AI, and cultivating a ***technology-scaled/centric*** approach to AI risk management and controls alignment/compliance. Moreover, it is key to acknowledge the interconnected nature of AI ***with data***, and the need to facilitate secure and safe data sharing and posture management, as well as ensuring alignment with data security measures that are essential for data set protection, such as measures to protect from ***cryptography risks***. We expand on these topics below.

The current RFI's emphasis on managing AI risks, while important, does not fully capture the broader spectrum of threats associated with AI, including risks emerging from the interaction of AI with extensive datasets, software, IoT, or large-scale bot operations, as well as the nuanced risk environments in which AI operates—such as specific data, models, or bot interactions. We recommend that OMB expands its risk framework to more comprehensively address these varied threats and the increased risk landscape, acknowledging how AI amplifies risks across all technological paradigms. OMB asks about measures to address the ownership and security of data and other information created, used, processed, stored, maintained, disseminated, disclosed, or disposed of by a contractor or

subcontractor on behalf of the Federal Government. Cutting-edge solutions are already used by the government to ensure unstructured content is being shared with governance measures to ensure data at use protection and such an approach can be levarged in the context of data to AI usage (intake and output of model).

Risk measurement and controls must be scaled by technology and automation to scale compliance and governance

OpenPolicy encourages OMB to identify further opportunities to utilize technology and automation tools to guarantee that agency adoption practices are in line with a proactive, secure and safe approach to AI. Simply establishing AI security and risk measurement controls is not enough; there must also be a scalable strategy to ensure these controls are effectively implemented- as they seek to adopt controls. Put simply, it is not sufficient to simply lay down controls for AI security and risk measurement - if we do not have a scaled approach to ensure such controls are adopted. A risk-based approach can only be scaled, given the nature of AI threats constantly evolving, by leveraging solutions to dynamically assess the risk of the system/model vis-a-vis the controls used. Governance based on documentation alone, has proven is not sufficient. The approach for the control measurement, under the AI EO and RFI, consistent with the direction of NIST CSF 2.0 - ***needs to include a key pillar of governance which is supported by automated means.***

Such an approach can benefit from tools like OSCAL for ensuring control mappings as more controls are added, and the ongoing implementation of the Cyber Executive Order and other government initiatives.

More robust measurements scaled by run-time monitoring of the risk and controls applied, alongside third-party tooling and testing, are needed to scale compliance, and compliance measurements are needed to ensure agency preparation, alongside, of course, ample budget and workforce. Reporting mechanisms of such solutions should be included in agency assessment, in a similar manner, to the cyber self-attestation form used under Cyber EO 14028.

Further, these measures could also foster more accurate and scaled techniques for model validation, human rights impact assessments, and controls for public application presentation which are crucial aspects and practices of risk management strategies and secure development and deployment of AI, and

considered in the AI OMB memo. Third-party evaluations should be also used to validate agency self-attestation.

Risk management, governance, and assessment methods that can adapt with controls released can enable broader monitoring agency implementation.

Automation can further support the complex goal of aligning the implementation of the AI EO, with the Cyber Executive Order with the ongoing implementation and upcoming revisions of FedRamp FISMA 2024 priorities related to IoT and OT protections (where AI can be leveraged), and NIST documents such as CSF 2.0 and NIST RMF /SSDF revisions. A further focus on leveraging automation and measurable compliance (e.g., schemes like OSCAL) can advance risk governance in this respect and increase the ability of NIST to scale agency measurements, as agencies and auditors can collaborate with solutions providers to automate control measurements.

Our ecosystem of innovative companies specializing in AI security and risk mitigation, is eager to work with the administration to explore how technology and automation can best support the deployment of AI, security, and privacy-related controls and their effective enforcement and measurement. OpenPolicy is also eager to work with OMB to explore how control parsing can further better public-policy engagement on specific controls.

Alignment with other ongoing initiatives and languages

The implementation and development of the RFI deliveries should align with upcoming revisions to other guidelines and controls and recent agency (e.g, NIST RFI, CISA "Secure by Design") published best practices and foster further coherence between the Cyber EO (14028), AI EO, and related efforts (such as CMMC) controls' deployment and standards by NIST, as it relates to guidelines development and enforcement efforts. Furthermore, as part of the future and forthcoming revision of the White House ***Zero-Trust Framework***, the threats and protections of AI must be considered. Beyond this effort, the entire body of cybersecurity controls that apply to agencies needs to be considered in this context, or else the critical work done under this RFI, we fear, may not achieve its full goals.

It is key that the RFI consider the need for coherence between the implementation

of relevant cyber requirements and newly introduced AI cybersecurity requirements and controls under the ongoing implementation of the Cyber EO (14028) while we await final details of implementation related to AI EO, to ensure agencies and contractors are implementing the relevant controls, and not further cultivate adherence to past methods, that may expose federal systems to threats.

As an additional example, further consideration should be given to whether the Federal Zero Trust strategy needs revisions to align further with the newly developed AI requirements.

Standards, implementation and proposed controls

1. AI system and models development

Under the section of *Managing the Performance and Risks of AI* and as mentioned in the context of the OMB AI memo, AI EO and other related publications, agencies must assess potential risks, document stakeholders impacted by AI, and consider failure modes; they should also implement automated tools and technology contextualization techniques to identify and prioritize potential risks associated with AI applications, monitor AI models for drifts in performance, analyze AI models, data sets and maintain data privacy within their AI models and intended use cases to identify potential biases, security vulnerabilities, and ethical concerns. Further in order to support how agencies structure their procurements to reduce the risk that an AI system or service, by using noninvasive scaled technology that analyze the way data interacts with their used model to detect suspicious patterns that might indicate an attack would allow a practical solution to agencies easily integrated and secure with existing systems.

To further enhance the security of AI technologies, adopting guidelines and best practices that mandate the protection of AI models and their data, including implementing solutions to detect adversarial attacks, ensuring data privacy in compliance with regulations such as GDPR and CCPA and emerging AI regulations,, and monitoring model performance is key as well as enabling run-time detection, as we explain below. Furthermore, non-invasive, software-based measures can support transparency, and adaptability, with minimal disruption, thereby safeguarding AI models without compromising their functionality or data integrity.

2. Implementing real-time detection for immediate threat notification and response, and adoption a full life cycle risk management approach to AI threats and LLM

In addressing the risks of synthetic content and broader AI and LLM threats, it's essential to set up stringent security governance. This includes data classification and access control, especially for LLMs. An integrated approach to managing LLM threats should encompass the entire model and AI lifecycle. It should also incorporate strategies for real-time detection and alerts, monitor LLM data flows to prevent unauthorized disclosures, and enforce extensive protection policies. These policies should include monitoring the risk posture through event monitoring, analysis, user behavior context, and anomaly and threat detection.

This approach should encompass detailed data transmission tracking and stringent security measures, providing a holistic defense against LLM threats and supporting secure and trustworthy AI development and deployment lifecycle at different levels of the AI system and in different modes of model deployment as the section address. It should also include governance and auditing of which data sets are unstructured content and data is being used by am AI systems that can build on SP 800-171 controls.

3. Red Teaming and Testing Teams Integration

In the quest to manage AI risk, integrating automated tools with the nuanced understanding of human testers in AI red teaming efforts is crucial. This approach includes enlisting both general red teamers for broad flaw detection and specialists for pinpointing more nuanced issues such as misinformation and the implications of deepfakes. By fostering a collaborative environment that draws on the community's collective wisdom, thorough evaluation, varied skill sets, and encouraging practices like bias bounties and vulnerability disclosures, NIST can uncover otherwise unnoticed flaws.

Further, OMB can consider developing a unified ***approach for AI testing,*** building on approaches such as MITRE-ATLAS, clearly distinguishing between cybersecurity and AI-specific terms to support the practical evaluation of AI systems. This includes

developing and differentiating testing methodologies for security and "non-security harms", such as bias and discrimination, and establishing terminology for algorithmic flaws. Distinct definitions and testing methodologies for traditional cybersecurity threats versus AI algorithmic flaws, ensuring comprehensive evaluation of AI systems. Establishing clear standards and language for "AI red-teaming" can further distinguish between adversarial and non-adversarial testing approaches, enhancing the efficacy and trustworthiness of AI technologies.

4. Governing information-sharing among agencies, vendors, and the public

Fostering open channels for sharing information on AI risks, including non-security issues like bias, underlines the importance of community-wide cooperation in enhancing AI system trustworthiness, guided by frameworks like NIST's AI Risk Management Framework and CSF 2.0 more broadly, with the focus on coordinated vulnerability disclosure like processes (bug bounty, red-teaming and vulnerability disclosure programs) that span both security and broader AI-harms. That said, recommending such processes should be done in conjunction with allowing, and developing, better protections for third-party testing and via contractual "safe harbors" and legislative protections, for those invited to test these systems.[2]

5. Automated risk strategies

A key focus should be given to establishing transparent and automated risk reporting, testing, and mitigation strategies to communicate identified risks and mitigation strategies to relevant stakeholders, including agency leadership, policy experts, and affected communities, to foster trust and accountability in AI cross-sector participation and ecosystem. Measurement of agency posture should be tied to accountability measures to support more robust adoption of controls and foster transparency and interpretability that could be used and elevated through automation of software-based approaches to securing AI models and identifying risks with minimal disruption to existing model workflows or data pipelines.

This strategy should encompass software tools that support the inventory of AI assets including datasets, models, experiments, and deployments in line with OMB

[2] See, ***OpenPolicy comments on the DMCA 1201 proceeding***, https://downloads.regulations.gov/COLC-2023-0004-0064/attachment_1.pdf. See also HPC comments on this proceeding and DMCA 1201 proceeding.

requirements, and to support a diverse innovative community and security concerns, founded on a model agnostic approach that also encompasses traditional predictive AI solutions. It should also integrate these assets into development pipelines while enabling effective monitoring and alerts for major adversarial threats such as data poisoning, model backdoors, model evasion, and model inversion. Additionally, it involves enhancing transparency in compliance reporting and improving visibility and sharing within the supply chain.

As OMB RFI Implementation evolves, we look forward to discussing these proposals with OMB and are available for any questions. We remain excited to collaborate with OMB to increase engagement with innovative companies.

Respectively,

Dr. Amit Elazari, CEO & Co-Founder, OpenPolicy