Openpolicy

August 12, 2024
To: Departmental Offices, Department of the Treasury
Via Federal eRulemaking Portal

OpenPolicy's comment on
Department of the Treasury's Request for Information on Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector

OpenPolicy appreciates the opportunity to provide feedback to the US Department of the Treasury's RFI on the use of Artificial Intelligence (AI) in the financial services sector. OpenPolicy is a technology company seeking to democratize the ability of innovative companies of all sizes to engage with policymakers and provide feedback on relevant policy deliverables. OpenPolicy is further engaged in emphasizing the need to use AI, innovation, and technology to foster open policymaking and broader use of technology to streamline compliance and governance automation. We anticipate active engagement with the Department of the Treasury as this RFI develops and are ready to offer additional input as necessary.

We remain eager to support the implementation of the AI Executive Order and its objectives. We are committed to supporting the use and implementation of this essential document designed to guide organizations designing, developing, deploying, or using AI systems to help manage AI risks and promote trustworthy and responsible development and use, actively engaging with NIST and other implementing agencies alongside our participation in the NIST AI Safety Institute Consortium.

We believe that an open and collaborative policymaking process is essential for advancing the successful implementation of the AI Executive Order (AI EO). The involvement of innovative companies, particularly startups that are at the forefront of developing cutting-edge AI security and safety solutions, is crucial to this endeavor. These companies are responsible for creating many technologies that support the **AI EO's requirements** and the **relevant OMB and NIST guidelines,** including frameworks for measuring, testing, and auditing AI, securing data, and facilitating AI's secure and responsible adoption. We are committed to actively collaborating with these forward-thinking communities to ensure that their expertise and innovations contribute meaningfully to the Treasury's AI initiatives.

These technologies evaluate the measurements, considerations, and development of AI software, including data sourcing, designing, training, fine-tuning, and evaluation perspectives that guide and lead the industry, which OpenPolicy actively collaborates with

these communities: Armis, HiddenLayer, Lasso, Kiteworks, HUMAN, Finite State, Cranium, InfoSec Global, Merlin Cyber, and Vaultree.

Given AI's integration into the financial sector, the Treasury's exploration of **how AI impacts opportunities and risks,** particularly regarding bias, discrimination, and privacy, is crucial. The Treasury's pivotal role in **promoting responsible innovation and competition within the financial sector concerning AI, including recommendations for legislative, regulatory, or supervisory enhancements, is of utmost importance.** AI offers both opportunities and challenges to the security and resiliency of the financial services sector, and thus, we highlight the importance of **secure collaboration tools**, such as secure streaming technologies that permit editable file access without actual data transfer, ensuring data remains within the owner's secure environment. The inclusion of advanced version control and auditing capabilities is crucial for effectively tracking document changes and access histories in collaborative scenarios.

Dynamic and secure access management controls should also be prioritized, enabling **real-time permission adjustments, access revocation, and detailed activity monitoring.** By focusing on **technology adaptability and future-proofing,** the Treasury can ensure that its approach to AI and digital rights management remains effective and resilient in the face of future security challenges.

Furthermore, the Treasury's initiatives would greatly benefit from aligning with the **governance and compliance strategies outlined in NIST's July AI RMF (600-1)[1] and NIST SSDF SP 800-218A[2] publications**. Establishing transparency policies and processes for documenting the origin and history of training data and generated data for Generative AI (GAI) applications directly supports the emphasis on governance and compliance, ensuring that **data privacy risks are proactively addressed,** which reflects a commitment to advancing digital content transparency while balancing proprietary considerations.

Additionally, including policies to evaluate GAI's risk-relevant capabilities and the robustness of safety measures, both before deployment and on an ongoing basis **(as mentioned in NIST AI RMF, section GV-1.2-002)**, underscores the necessity of continuous risk evaluation and the implementation of AI security measures, which aligns with the need for **real-time threat detection and scalable risk management strategies,** ensuring that

[1] See NIST Trustworthy and Responsible AI NIST AI 600-1 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
[2]See NIST Special Publication 800 NIST SP 800-218A Secure Software Development Practices for Generative AI and Dual-Use Foundation Models An SSDF Community Profile https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218A.pdf

financial organizations are equipped to manage the growing reach and complexities of generative AI technology.

Bolstering oversight with independent evaluations or assessments of GAI models, where the type and robustness of evaluations are proportional to the identified risks, will ensure that large, small, and medium-sized entities can manage risks effectively, leveraging both human and automated processes for continuous improvement. By integrating these objectives into this RFI and subsequent AI-related initiatives, **the Treasury can align its efforts with broader cybersecurity and regulatory frameworks, setting a standard for proactive, ongoing risk management and governance in the financial services sector**.

- **AI risks and enhancing AI security (under questions 4B, and 9)**

OpenPolicy commends the Treasury for its leadership in addressing the use of AI in the financial services sector through this RFI. We are particularly encouraged by the emphasis on AI security, data confidentiality, and the importance of governance in these areas. The focus on eliciting specific, actionable responses from stakeholders is critical in shaping effective initiatives for evaluating and auditing AI systems, especially when addressing potential opportunities and risks such as bias, third-party personnel and software, autonomous agents, privacy, data security, bypass, extraction, and other baseline security concerns.

However, we believe that a significant gap remains in enhancing **the measurability of AI risks and understanding AI security and safety outcomes within the financial sector**. As AI systems become more sophisticated, the tactics and techniques employed by malicious actors to exploit vulnerabilities within these models are also evolving. These threat vectors are further amplified as AI is increasingly integrated into larger software development life cycles and critical financial technology use cases. Traditional cybersecurity and risk management methods—such as threat monitoring, detection, and response—are often too slow and manual to effectively counter these rapidly emerging threats. These conventional approaches were not designed with AI in mind and therefore lack the capability to address the complex and multifaceted attack techniques and vectors associated with AI. This presents a substantial challenge for the financial sector, where the stakes are high, and the impact of security breaches can be severe.

The risks associated with AI within the financial sector are far from isolated; they become significantly amplified when AI is integrated with broader datasets, exposed to software vulnerabilities, targeted by IoT threats, or subjected to bot-scaled attacks. The specific context in which AI operates—whether related to particular data models or in

conjunction with bot interactions—introduces complex risks that necessitate a more comprehensive and nuanced approach. Therefore, it is crucial to adopt a **broader risk framework** that addresses these extended threats and the heightened risks AI presents across all technology paradigms to better equip the financial services sector to effectively navigate and mitigate the complex and evolving risks associated with AI.

This approach aligns with the latest **NIST AI RMF 600-1 July revision**, which emphasizes the need for continuous risk evaluation and robust AI security measures. The inclusion of policies to assess GAI's risk-relevant capabilities and the strength of safety measures, both before deployment and on an ongoing basis (GV-1.2-002), underscores the importance of **real-time threat detection and scalable risk management strategies.** This alignment with the Treasury's priorities ensures organizations are prepared to manage generative AI technology's growing scope and complexity.

Accordingly, we encourage the Treasury to identify, map, and measure the various threats through the suggested measurement actions under NIST AI RMF, specifically **MS 2.7: AI system security and resilience, as outlined in the Measure function.** This includes sections **MS-2.7-001 and MS-2.7-005,** which address various AI and GAI risks with significant implications for the financial services industry, such as data privacy, information integrity, information security, value chain, and component integration. These actions align with the dynamic nature of AI technology and its integration with other technological paradigms, as well as the governance approaches outlined in other NIST frameworks, such as the Cybersecurity Framework (CSF 2.0). This comprehensive approach will help ensure that AI risks are effectively managed and that the **financial services sector remains resilient** in the face of evolving threats.

It is imperative that the implementation and development of the Treasury's RFI align with ongoing revisions to other **NIST guidelines, controls,** and recent agency best practices. A coherent approach between **the Cyber EO (14028), AI EO, and related efforts will ensure that AI governance remains consistent** and robust and extends to a comprehensive risk assessment framework that considers both technical and non-technical risks, including data privacy, algorithmic bias, fairness, explainability, and potential impacts on underserved communities. Furthermore, it is essential to acknowledge AI's interconnected nature with data and ensure secure and safe data sharing and posture management. Alignment with data security measures, including those that protect against **cryptography risks,** is vital for protecting data sets within the financial sector.

- **AI Managment and Testing Requirements**

In accordance with the Treasury's objectives for this RFI on AI management and testing, particularly as outlined in **question 7,** it is imperative for financial institutions to integrate automated tools with the nuanced expertise of human testers in AI red teaming to manage AI software risks effectively. This approach is not merely beneficial but essential, as it combines broad flaw detection by general red teamers with the specialized insights of experts who can identify more nuanced issues. Such a collaborative environment leverages the collective knowledge of the community, fostering practices like bias bounties and vulnerability disclosures, which are critical for uncovering flaws that might otherwise go unnoticed. This approach directly supports the **Treasury's interest in understanding the governance structures and risk management frameworks** that financial institutions expect to apply in the development and deployment of AI, ensuring that AI technologies are both secure and resilient against emerging threats.

The Treasury should consider adopting a unified approach to AI testing, building on established methodologies such as **MITRE-ATLAS, and clearly distinguishing between cybersecurity and AI-specific terms, particularly while still waiting for the upcoming NIST guideline on AI Red Teaming framework.** This approach would support the practical evaluation of AI systems by developing and differentiating testing methodologies for both security and "non-security harms," such as bias and discrimination. Additionally, establishing distinct definitions and testing methodologies for traditional cybersecurity threats versus AI algorithmic flaws, will ensure a comprehensive evaluation of AI systems. Establishing standards and language for "AI red-teaming" within the financial sector will further distinguish between adversarial and non-adversarial testing approaches, enhancing the efficacy and trustworthiness of AI technologies.

Given the dynamic nature of AI threats and risks, a proactive approach to security is crucial, where continuous monitoring enables the detection of anomalies and potential threats as they occur, allowing immediate action to mitigate risks before they can cause significant harm. The Treasury should underscore the need for continuous monitoring and evaluation of AI systems, which includes **real-time detection and alerts, monitoring AI data flows to prevent unauthorized disclosures, and enforcing comprehensive software protection policies.** These measures are vital for maintaining a robust software security framework for AI models, ensuring that security breaches are promptly identified and responded to, thereby reducing the window of opportunity for malicious actors.

To further align with the Treasury's objectives outlined in the RFI, particularly regarding the **governance structure and risk management frameworks** that financial institutions are expected to apply in the development and deployment of AI, it is critical to emphasize continuous **monitoring and evaluation of AI systems**, including model validation and

human rights impact assessment. Integrating AI red teaming with continuous testing and enhancing model validation through automated tools are key strategies that can significantly bolster the security and trustworthiness of AI technologies. **Red teaming,** which involves simulating attacks to uncover weaknesses in AI systems, combined with ongoing testing, ensures that these systems remain secure over time. These proactive measures are essential for addressing the complex and evolving threats faced by AI systems, ensuring their safe and ethical deployment within the financial sector.

Additionally, **incorporating non-invasive, software-based measures that supports transparency and adaptability can further safeguard AI models without compromising their functionality or data integrity.** These actions are crucial not only for protecting AI software and systems against threats but also for maintaining operational efficiency, which aligns with the Treasury's approach and commitment to ensuring secure and responsible AI deployment in the financial services sector. By adopting these practices, financial institutions can better manage the development and validation of AI models, address potential gaps in human capital, and effectively mitigate risks related to AI explainability and bias, thereby enhancing the overall governance and risk management frameworks applied to AI technologies.

Governance and automation are central to this approach. We encourage the Treasury to leverage technology and automation tools to ensure that AI adopters' practices align with a responsible and secure AI posture. Specifically, solutions that dynamically assess AI system risks against established controls are vital for managing the complex and evolving landscape of AI threats. **Such measures are critical as financial institutions increasingly use AI across various domains, including product offerings, risk management, capital markets, customer services, regulatory compliance, and marketing.**

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

Respectively,

/s/ Michelle Sahar
Michelle Sahar
Cybersecurity Policy Director, OpenPolicy