hackerone

April 29, 2024

David A. Myklegard
Deputy Federal Chief Information Officer
Office of Management and Budget (OMB)

Christine J. Harada
Senior Advisor, Office of Federal Procurement Policy
Office of Management and Budget (OMB)

VIA ELECTRONIC SUBMISSION

Re: Responsible Procurement of Artificial Intelligence in Government

Dear Mr. Myklegard and Ms. Harada,

HackerOne Inc. (HackerOne) submits the following comments in response to the Office of Management and Budget's (OMB) Request for Information (RFI) on Responsible Procurement of Artificial Intelligence in Government.[1] HackerOne appreciates the opportunity to provide input, and we commend OMB for its openness in working with industry stakeholders on this important issue.

HackerOne is the global leader in human-powered security. We leverage human ingenuity to pinpoint the most critical security flaws across your attack surface to outmatch cybercriminals. HackerOne's Attack Resistance Platform combines the most creative human intelligence with the latest artificial intelligence to reduce threat exposure at all stages of the software development lifecycle. From meeting compliance requirements with pentesting to finding novel and elusive vulnerabilities through bug bounty, HackerOne's elite community of ethical hackers helps organizations transform their businesses with confidence. HackerOne has helped find and fix vulnerabilities for sector leaders including Coinbase, General Motors, GitHub, Goldman Sachs, Hyatt, PayPal, and the U.S Department of Defense.

In these comments, we focus on "AI red teaming," a form of testing that uses adversarial methods to identify flaws and vulnerabilities within AI systems. In the context of AI safety, AI red teaming can help prevent AI systems from generating harmful content, such as providing instructions on creating bombs or producing offensive language. For AI security, the practice can help prevent bad actors from abusing AI systems, including by compromising the confidentiality,

[1] Office of Management and Budget (OMB), Request for information: Responsible procurement of artificial intelligence in government. Mar. 29, 2024, https://www.federalregister.gov/documents/2024/03/29/2024-06547/request-for-information-responsible-procurement-of-artificial-intelligence-in-government.

integrity, or availability of data within an organizations' systems.[2] We believe that by facilitating the use of human powered AI red teaming, OMB and federal agencies can help ensure a more safe and secure deployment of AI systems within federal networks.

Below we provide more targeted responses to some questions listed in the RFI:

Question 5: ***What access to documentation, data, code, models, software, and other technical components might vendors provide to agencies to demonstrate compliance with the requirements established in the AI M-memo? What contract language would best effectuate this access, and is this best envisioned as a standard clause, or requirements-specific elements in a statement of work?***

To demonstrate compliance with the requirements in OMB's Memorandum on *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence* (AI M-memo), vendors should be required to disclose information about the methodologies used in their AI red teaming assessments.[3] This would allow federal agencies to better understand the risks associated with a specific AI use case and compare the risks associated with a different AI system in a particular context. Moreover, if publicly circulated, these methodologies and best practices would improve future AI red teaming assessments, contributing to the security and safety of both the federal and non-federal ecosystems.

Question 9: ***How might agencies structure their procurements to reduce the risk that an AI system or service they acquire may produce harmful or illegal content, such as fraudulent or deceptive content, or content that includes child sex abuse material or non-consensual intimate imagery?***

Before they can mitigate the risks associated with AI procurement and deployment, federal agencies must first be clear about the harms they seek to avoid. To increase clarity on this issue, federal agencies should identify and document specific types of harms and unacceptable bias elements. Vendors could then use this information when conducting AI red teaming and other assessment activities. OMB could require federal agencies to identify and categorize the types of harms they are most concerned with in the context of particular contracts and mission requirements.

To mitigate the risk of AI systems after procurement by federal agencies, OMB should require vendors to implement AI red teaming throughout each system's life cycle (i.e., development, deployment, and operation). AI red teaming should include a combination of human-powered and automated assessments where attackers are simulating a real-world environment.

---

[2] Michiel Prins et. al, An Emerging Playbook for AI Red Teaming With HackerOne, Apr. 1, 2024, https://www.hackerone.com/thought-leadership/ai-safety-red-teaming.

[3] See Office of Management and Budget (OMB), Memorandum on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, Mar. 28, 2024, https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

OMB should also seek to ensure that the vendors offering AI systems are prepared to accept unsolicited information about vulnerabilities, both to AI models and to company information systems, from good-faith security researchers. Therefore, OMB should require that vendors selling AI capabilities to federal agencies have a Vulnerability Disclosure Policy (VDP) in place. To ensure that federal agencies are also adequately equipped to collect such information, OMB should direct the National Institute of Standards & Technology (NIST) to update its Special Publication (SP) 800-216 on *Recommendations for Federal Vulnerability Disclosure Guidelines* to accommodate AI systems. [4]

Conclusion

HackerOne appreciates the opportunity to provide comments to this request for information. As the conversation around this topic continues to evolve, we would welcome the opportunity to further serve as a resource and ensure the safety, security and reliability of AI.

* * *

Respectfully Submitted,

Ilona Cohen
Chief Legal and Policy Officer
HackerOne

[4] See National Institute of Standards & Technology (NIST), SP 800-216 Recommendations for Federal Vulnerability Disclosure Guidelines, May 24, 2023, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-216.pdf