hackerone

June 4, 2025

VIA ELECTRONIC SUBMISSION
whs.mc-alex.ad.mbx.eosd-psb-branch-mailbox@mail.mil
leanne.m.condren.civ@mail.mil

Re: Request for Information (RFI) – Software Fast Track (SWFT) Automation & Artificial Intelligence (AI)

- Company name, DUNS/UEI, CAGE Code, and point of contact:
  - HackerOne Inc.
  - DUNS/UEI, CAGE Code: 079498330, 7JUM3
  - Point of Contact: Ilona Cohen, policy-team@hackerone.com.

- Capability statement demonstrating relevant experience:
  - HackerOne is a global leader in offensive security solutions. Our HackerOne Platform combines AI with the ingenuity of the largest community of security researchers to find and fix security, privacy, and AI vulnerabilities across the software development lifecycle. The platform offers bug bounty, vulnerability disclosure, pentesting, AI red teaming, and code security. We are trusted by industry leaders like Amazon, Anthropic, Crypto.com, General Motors, GitHub, Goldman Sachs, Uber, and the U.S. Department of Defense. HackerOne consistently advocates for the widespread adoption of cybersecurity measures that have proven effective at addressing unmitigated vulnerabilities in both commercial and government contexts. This advocacy extends to the realm of AI, where we conduct AI security testing and algorithmic reviews to help reduce unintended outputs from AI. As the demand for secure AI grows, HackerOne assists enterprises in navigating the complexities of building and deploying AI models securely.

- Recommended contract types, contract vehicles (e.g., GSA MAS, GWACs, etc.), and business size status

- Feedback on the feasibility of the work described

- Potential risks, challenges, or innovations

- Past performance examples (if applicable)

- Rough Order of Magnitude (ROM) cost info

HackerOne Inc. (HackerOne) respectfully submits the following comments in response to the Department of Defense (DoD) Chief Information Officer's (CIO) Request for Information

(RFI) regarding the SWFT External Assessment.[1] HackerOne appreciates the opportunity to provide input and commends the DoD for its proactive efforts in enhancing its cybersecurity posture and promoting collaborative engagement with stakeholders.

2. What are potential challenges in the implementation of automation or AI for high trust situations related to cybersecurity authorization official responsibilities?

One of the primary challenges in implementing automation or artificial intelligence (AI) in high-trust cybersecurity contexts – such as those involving Authorization to Operate (ATO) decisions – is the inherent unpredictability and complexity of AI systems. These systems are intended to influence, or even drive, key risk determinations but still may produce unintended outputs. This creates a potential trust gap for authorization officials, who are responsible for certifying the security of systems based on inputs that may not be fully accurate.

In mission-critical environments where the consequences of incorrect or unreliable outputs can be severe, a high degree of confidence in the performance and reliability of AI tools is essential. However, traditional validation or testing methods may fail to detect certain vulnerabilities. To mitigate these risks, AI red teaming must be incorporated as a core component of any AI model deployment intended to support or inform authorization responsibilities.

AI red teaming is a structured, adversarial testing process designed to uncover flaws and vulnerabilities that could lead to misuse, failure, or harmful unintended consequences. AI red teaming typically takes an adversarial approach, mimicking attackers' attempts to challenge the model's security, performance, reliability, and alignment with model guidelines. AI red teaming can evaluate both functional vulnerabilities (e.g., if security flaws can be exploited) and systemic risks (e.g., how the model might generate harmful outputs). Model developers use red teaming both before an AI system is made available on the marketplace or deployed in critical environments as well as on an ongoing basis post-deployment as technology and attacker TTPs evolve.

By proactively identifying weaknesses that internal teams or automated scans may miss, red teaming helps ensure that AI systems used in high-trust cybersecurity decisions are resilient, secure, and reliable. We strongly recommend that DoD require AI red teaming, both pre- and post-deployment, for any system influencing authorization outcomes, especially those that touch on national security, public safety, or critical infrastructure.

4. **What are the considerations that DoD should prioritize when evaluating automation and AI solutions for DoD-led SWFT risk assessments and determinations?**

As the DoD increasingly leverages automation and AI to support SWFT risk assessments and security determinations, it is essential that these systems are not only effective, but also secure, transparent, and accountable.

[1] Department of Defense (DOD) Chief Information Officer (CIO), Request for Information, Software Fast Track (SWFT) Artificial Intelligence, June 4, 2025, https://sam.gov/opp/7ca9ff30bad5407db7de079f7bf397c0/view.

A critical consideration is the potential for vulnerabilities in the programming and design of these tools. Like all software, AI systems can contain security issues, misconfigurations, or systemic weaknesses that may be exploited or that could lead to inaccurate or misleading assessments. As highlighted previously, these issues are particularly concerning when AI systems are used to influence authorization outcomes.

To address this, we strongly recommend that the DoD also require any provider of an AI or automated system used in SWFT evaluations to maintain a vulnerability disclosure policy (VDP). As a "see something, say something" approach, a VDP provides a structured and secure channel for third parties, including security researchers and system users, to report vulnerabilities or unintended outcomes. This includes not only security issues, but also model behaviors that could affect the accuracy or integrity of risk evaluations.

VDPs are widely recognized as a best practice and are consistent with long-standing federal initiatives to strengthen software security. The Internet of Things Cybersecurity Improvement Act of 2020 mandated VDPs for information systems across civilian federal agencies.[2] That same year, the Office of Management and Budget's Memorandum M-20-32[3] and the Department of Homeland Security's Binding Operational Directive 20-01[4] reinforced this approach by requiring federal agencies to implement VDPs as part of their baseline cybersecurity practices. Building on this foundation, the Federal Contractor Cybersecurity Vulnerability Reduction Act, which was reintroduced earlier this month, proposes to extend this requirement to all U.S. government contractors.[5]

* * *

HackerOne appreciates the opportunity to submit comments in response to this RFI. As AI and automation continue to reshape DoD operations, AI red teaming and VDPs are two critical mechanisms for ensuring that these systems remain secure, transparent, and reliable.

We welcome continued collaboration with the Department of Defense and stand ready to support its efforts in operationalizing secure AI solutions and safeguarding national security. If we can be of additional assistance, please contact Ilona Cohen at policy-team@hackerone.com.

[2] Public Law No: 116-207, https://www.congress.gov/116/plaws/publ207/PLAW-116publ207.pdf.
[3] OMB memorandum 20-32, https://www.whitehouse.gov/wp-content/uploads/2020/09/M-20-32.pdf
[4] DHS BOD 20-01, https://www.cisa.gov/news-events/directives/bod-20-01-develop-and-publish-vulnerability-disclosure-policy
[5] S.1899, Federal Cybersecurity Vulnerability Reduction Act, 119th Cong., https://www.congress.gov/bill/119th-congress/senate-bill/1899.