



AI Cybersecurity Collaboration Playbook

Fact Sheet

The AI Cybersecurity Collaboration Playbook provides guidance to organizations across the AI community –including AI providers, developers, and adopters – for sharing AI-related cybersecurity information voluntarily with the Cybersecurity and Infrastructure Security Agency (CISA) and other partners through the Joint Cyber Defense Collaborative (JCDC). While focused on strengthening collaboration within JCDC, the playbook also identifies actionable information sharing categories applicable to broader critical infrastructure stakeholders and other sharing mechanisms. CISA encourages organizations to adopt the playbook’s guidance to enhance their own information-sharing practices, contributing to a unified approach to AI-related cybersecurity threats across critical infrastructure.

This playbook aims to:

- Facilitate collaboration between federal agencies, private industry, international partners, and other stakeholders to raise awareness of AI cybersecurity risks and improve the resilience of AI systems.
- Guide JCDC partners on how to voluntarily share information related to cybersecurity incidents and vulnerabilities associated with AI systems.
- Delineate information sharing protections and mechanisms.
- Outline CISA’s actions upon receiving shared information to strengthen collective defense.

AI safety topics, such as risks to human life, health, property, or the environment, are outside the intended scope of the JCDC AI Cybersecurity Collaboration Playbook. Stakeholders should address any risks or threats involving human life, health, property, or the environment in a timely and appropriate manner, in accordance with their own applicable process or procedures for such events. Similarly, issues related to AI fairness and ethics are also outside the scope of this playbook. This playbook does not create policies, impose requirements, mandate actions, or override existing legal or regulatory obligations. All actions taken under this playbook are

This document is marked TLP: CLEAR: Disclosure is not limited. For more information on the Traffic Light Protocol, see <https://www.cisa.gov/tlp>.





voluntary. This playbook will undergo periodic updates, evolving to address these challenges through active collaboration among government, industry, and international partners.

CISA's Information Sharing and Collaboration Through JCDC

Given the complexity of AI systems and the challenges in identifying security issues and their root causes, JCDC partners should share information early and often. Actively sharing information helps CISA and JCDC partners understand the current operating environment to make informed decisions about potential defensive actions. The playbook delineates CISA's information

sharing and collaboration process (Figure 1) amongst JCDC partners providing guidance on actionable information categories during routine operations and enhanced coordination supported by tables, checklists, and examples focused on:

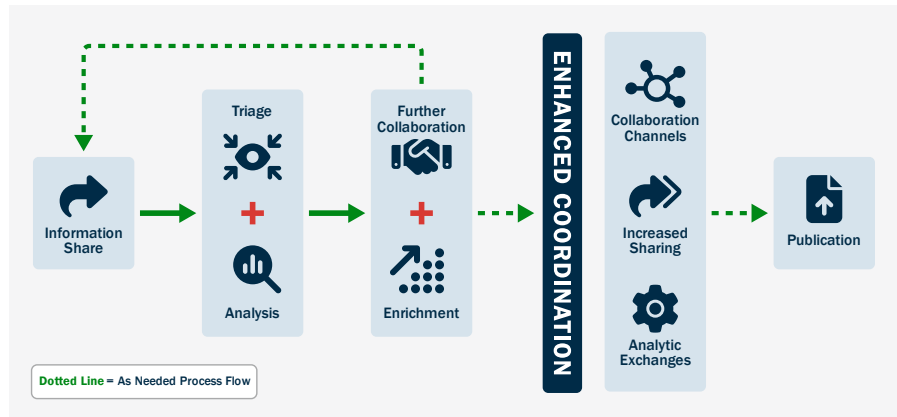


Figure 1: CISA Information Sharing and Collaboration Process

- **Proactive Information Sharing.** Consistently and proactively share information on malicious activity, trends, pre-release publications, and assessments to maintain situational awareness of the evolving landscape, enabling the early detection, identification, and remediation of critical threats.
- **Information Sharing Regarding an Incident or Vulnerability.** Actively share information regarding an AI cybersecurity incident or vulnerability to coordinate within JCDC (See [Tear Sheet](#)). Additionally, using CISA's web form to [voluntarily report a cyber incident](#) or a [vulnerability in a product or service](#) is a good way to provide all relevant information to CISA via an encrypted channel. If using the web form, JCDC partners should notify a JCDC representative via email.





- **Information Analysis and Operational Use.** CISA follows an approach to aggregate, validate, analyze, anonymize, and enrich, the data shared to determine appropriate defensive action(s). Adhering to the Traffic Light Protocol (TLP)¹ dissemination control marking, information may also be shared with industry, U.S. federal government, state, local, tribal, and territorial (SLTT), and international partners to support cyber defensive purposes.

Enhanced Coordination

CISA assesses information shared by partners to decide on defensive actions as the situation evolves. These actions can be executed individually or combined, depending on the nature of the identified threat or vulnerability. The process is inherently dynamic and involves collaboration among multiple stakeholders, often working simultaneously.

These actions include but are not limited to:

- Share information for detection and prevention purposes.
- Expose and disrupt adversary tactics and infrastructure.
- Coordinate on strategies to address malicious infrastructure.
- Identify and notify victim entities.
- Share detection capabilities.
- Produce and distribute relevant threat intelligence products.
- Offer proactive services and engagements.
- Assess evolving threats with responsive engagements.

Information Sharing Protections

By sharing information through JCDC, companies benefit from enhanced coordination, government support, and gain the ability to collaborate on AI cybersecurity issues within a trusted environment enabled by information sharing protections outlined in the Cybersecurity Information Sharing Act of 2015 (CISA 2015).

¹ Traffic Light Protocol (TLP) Definitions and Usage," <https://www.cisa.gov/news-events/news/traffic-light-protocol-ttp-definitions-and-usage>.





The Cybersecurity Information Sharing Act of 2015 (CISA 2015) (6 U.S.C. §§ 1501-1533) creates protections for non-federal entities to share cyber threat indicators and defensive measures for a cybersecurity purpose in accordance with certain requirements with the government and provides that they may do so notwithstanding any other law. Such protections include the non-waiver of privilege, protection of proprietary information, exemption from disclosure under the Freedom of Information Act (FOIA), prohibition on use in regulatory enforcement, and more.¹ CISA 2015 also requires DHS to operate a capability and process for sharing cyber threat indicators with both the federal government and private sector entities and provides for liability protection for information shared through this process. The statute also creates protections for cyber threat indicators and defensive measures shared in accordance with the statutory requirements with state, local, tribal, and territorial (SLTT) entities, including that the information shall be exempt from disclosure under SLTT freedom of information laws. CISA 2015 does not cover information shared that is not a cyber threat indicator or defensive measure, as defined by the law. **AI-related information is covered under the Act to the extent the information qualifies as a cyber threat indicator or defensive measure.** These aspects are further detailed in multiple guidance documents, especially the DHS-DOJ [Guidance to Assist Non-Federal Entities to Share Cyber Threat Indicators and Defensive Measures with Federal Entities under the Cybersecurity Information Sharing Act of 2015](#).

Call to Action

JCDC partners should integrate the playbook into their incident response and information-sharing processes, make iterative improvements as needed, and provide feedback to CISA through CISA.JCDC@cisa.dhs.gov. To strengthen collaboration and engagement, JCDC invites AI security specialists and stakeholders to consider the following actions:





- **Flag opportunities for technical exchanges** related to emerging threats, adversaries, or vulnerabilities affecting the AI community.
- **Identify priority issues for the AI community** to inform JCDC priorities and cyber defense activities.
- **Promote post-mortem analyses and knowledge sharing** to foster a proactive approach to addressing AI security challenges.
- **Become a JCDC partner** to engage in synchronized cybersecurity planning, cyber defense, and response. To learn more about JCDC, please visit CISA's [JCDC webpage](#) and email CISA.JCDC@cisa.dhs.gov.





Tear Sheet: Voluntary Information Sharing for AI Cybersecurity Incident and Vulnerability Reporting and Protections

CISA encourages JCDC partners to use this checklist to voluntarily share actionable information regarding an AI cybersecurity incident or vulnerability via email. Other stakeholders can share voluntary information with JCDC via CISA.JCDC@cisa.dhs.gov. While JCDC partners should follow the checklist, CISA welcomes any shared information, even if not all checklist points are met.

Checklist for Voluntary Information Sharing



<input checked="" type="checkbox"/> Description of the incident or vulnerability	<ul style="list-style-type: none"> <input type="checkbox"/> Is this information related to an incident, an attempted attack, scanning activity, or suspicious activity? <input type="checkbox"/> Is this information related to a vulnerability? Include the Common Vulnerabilities and Exposures (CVE) assignment, if available. <input type="checkbox"/> Was this information obtained directly or indirectly (via another organization)? <input type="checkbox"/> Was this information obtained from a privileged or non-public source? <input type="checkbox"/> What is the confidence level of this information? Is this information confirmed to be related to malicious activity or is it unconfirmed (i.e., suspicious activity)?
---	--

This document is marked TLP: CLEAR: Disclosure is not limited. For more information on the Traffic Light Protocol, see <https://www.cisa.gov/tlp>.







Checklist for Voluntary Information Sharing

<p> How the incident or vulnerability was first detected</p>	<ul style="list-style-type: none"><input type="checkbox"/> Initial access vector.<input type="checkbox"/> Detection method (e.g., Structured Threat Information Expression (STIX) indicators).<input type="checkbox"/> Indicators of Compromise (IOCs)<input type="checkbox"/> Indicators of attack.<input type="checkbox"/> Sample attack information or screenshots.<input type="checkbox"/> IP (Internet Protocol) addresses, domains, and hashes.<input type="checkbox"/> Timestamps to include dates/times related to when the information was active or observed.<input type="checkbox"/> What are the IOCs being used for (e.g., initial access, command and control [C2] infrastructure)?
<p> System and network vulnerabilities</p>	<ul style="list-style-type: none"><input type="checkbox"/> Known and previously disclosed vulnerabilities being maliciously exploited in the wild.<input type="checkbox"/> Vulnerabilities of critical concern (from a JCDC partner’s perspective), even if exploitation evidence has not been found yet.<input type="checkbox"/> Publicly known proofs of concept in open-source platforms (i.e., news reporting, social media).<input type="checkbox"/> Note: Due to sensitivity concerns, non-public or lesser-known proofs of concept should be shared with CISA through the “Report a Vulnerability” link on CISA’s Coordinated Vulnerability Disclosure Process page, which includes a section to report exploitation information. See also the “Newly Identified Vulnerability Coordination” section.







Checklist for Voluntary Information Sharing

<p> Affected AI artifact(s) and systems</p>	<ul style="list-style-type: none"><input type="checkbox"/> Any known model information about the training dataset: model name, model version, model task, model architecture, model source (author or location), and lifecycle phase.<input type="checkbox"/> Any known information about the AI model developer.<input type="checkbox"/> Any agentic, copilot, or third-party platforms in use.<input type="checkbox"/> Any known information about Application Programming Interface (API) and libraries.<input type="checkbox"/> Software/hardware configuration and access specific to the AI model.<input type="checkbox"/> The software underpinning the affected system(s).<input type="checkbox"/> AI application information (i.e., author information, AI application accesses).
<p> Affected users or victims</p>	<ul style="list-style-type: none"><input type="checkbox"/> If known, specific or type (i.e., sector) of victims targeted based on JCDC partner's interactions and/or campaign attributes.<input type="checkbox"/> Geographic location of affected users, if relevant.<input type="checkbox"/> Types and scope of information that was lost or exploited.<input type="checkbox"/> Category (e.g., financial, reputational) and severity of harm (i.e., negligible, minor, moderate, severe).<input type="checkbox"/> List of systems or products whose users might be impacted by the incident.<input type="checkbox"/> Estimated number of directly impacted users.<input type="checkbox"/> List of external systems to which the AI model possibly had direct access.






Checklist for Voluntary Information Sharing

<p> Broader impacts of the attack</p>	<ul style="list-style-type: none"><input type="checkbox"/> Lateral movement identified and impact.<input type="checkbox"/> Suspected exploitation vector.<input type="checkbox"/> Exfiltration impact.<input type="checkbox"/> Impact to business operations.<input type="checkbox"/> Supply chain impacts (e.g., information on trusted vendors, third-party considerations, data provenance).<input type="checkbox"/> Known or suspected impact to specific critical infrastructure sectors or the U.S. federal government.<input type="checkbox"/> Impact of vulnerability found and level of access required to exploit.
<p> Mitigations</p>	<ul style="list-style-type: none"><input type="checkbox"/> Mitigation status.<input type="checkbox"/> Category of implemented mitigation (i.e., risk acceptance, risk avoidance, and risk transfer).<input type="checkbox"/> Remediation technique (e.g., rollback or updating of specific components, including models).





Checklist for Voluntary Information Sharing

 **Attribution and malicious actor profile**

- Attacker identity, if known, or similarities observed between attack details (IOCs/tactics, techniques, and procedures (TTPs)) and a known threat actor.
- Level of confidence (i.e., unverified, speculative, confirmed).
- Specific techniques demonstrated, citing (if possible) MITRE ATT&CK® framework and the MITRE ATLAS framework.
- Specific cyber defense controls targeted, subverted, or evaded by attacker (including technique, if observed).
- Patterns or themes the attacker relied on in targeted attacks.
- Control and access obtained by the malicious actor.
- Type of adversarial AI attack and attack procedure used.
- Underlying system component.
- Adversary tooling used.
- Anti-forensics or actor cleanup efforts witnessed.
- Whether the specific threat actor is known or suspected.





Checklist for Voluntary Information Sharing

Technical data and analysis

- How a threat actor uses certain TTPs or IOCs.
- Include adversarial prompt along with identified response content that illustrates the attack's success and overall structure.
- Is the information novel or has it been previously observed or publicly reported?
- "Abnormal" registry behavior and activity.
- Code overlap from other known/historical malware or attack samples.
- Known overlap with historical attack on C2 infrastructure and APIs or third parties.
- File extension modification.
- Campaign artifacts (i.e., recycle bin or other file removal/app modification).

