STEMPRISE

Securing AI: A Playbook for Businesses from Startups to Utilities

AI Uncorked at Boca Raton, FL
October 2025

Marissa Morales-Rodriguez, PhD
Founder and Technology Security Strategist
STEMPRISE

DISCLAIMER

The views and opinions expressed in this presentation are those of the presenter and do not necessarily reflect the official policy or position of STEMPRISE or its partners. References to frameworks, organizations, or products are provided solely for illustrative purposes and do not constitute endorsement. The information presented is for educational and informational use only and should not be interpreted as legal, regulatory, or professional consulting advice.

MY EXPERIENCE

FOUNDER & TECHNOLOGY SECURITY STRATEGIST

STEMPRISE
(2025- Present)

- Principal Consultant
- Research and Development
- Focus on security and resilience of digital technologies.

CYBER PORTFOLIO LEAD AND TECHNOLOGY MANAGER

US Department of Energy
(2021 – 2025)

- Lead the development of technology driven cyber strategy for Energy Efficiency and Renewable Energy Office
- Technology Manager for projects related to operations, planning, resilience and cybersecurity for solar technologies

SCIENTIST

Oak Ridge National Laboratory
(2009 – 2021)

Managed a research portfolio implementing AI-driven predictive analytics and automation to enhance grid reliability and cybersecurity

AGENDA

- Introduction
- AI Systems & Security Considerations
- Scalable AI Security Playbook (SAISP)
- Takeaways

STEMPRISE

INTRODUCTION & MOTIVATION

MOTIVATION

AI Security Matters

AI systems are being rapidly deployed across organizations. In some cases, security practices and guidelines are not clear leaving end users vulnerable.

This presentation will:

- Increase awareness of security of AI systems
- Share information about open-source security guides
- Encourage adoption of security practices

Ensure Privacy & Reliability

Protect System Integrity

Build Trust & Compliance

AI SECURITY

STEMPRISE

What is AI Security?

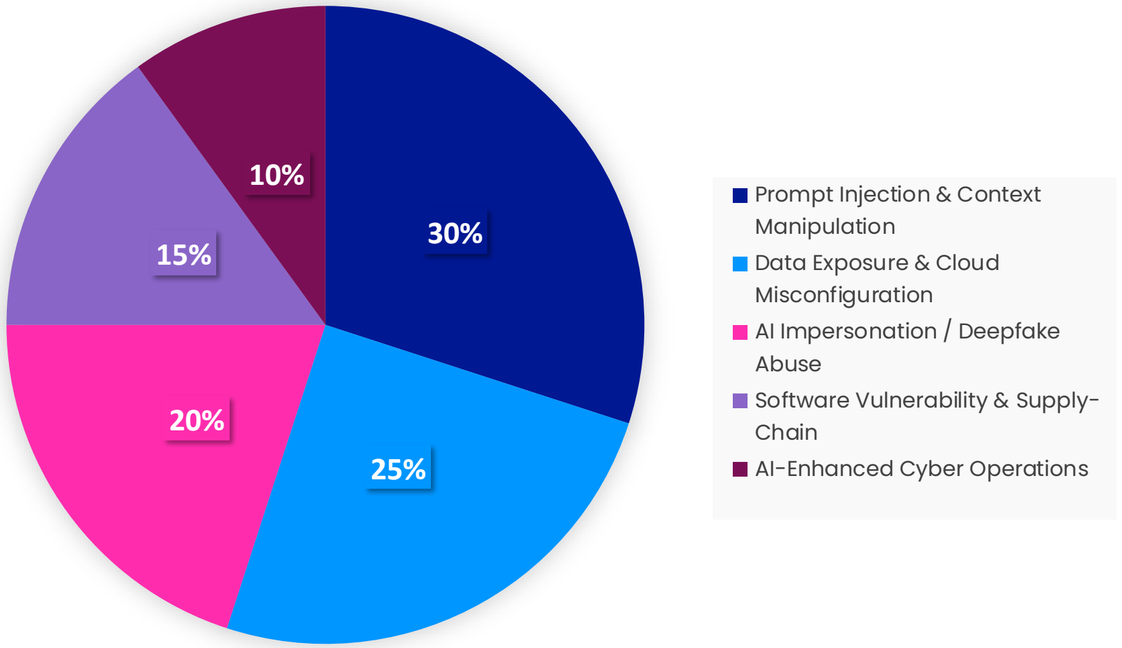
- Combines software security, machine learning security and cybersecurity
- Ability to minimize risks associated with each system component, lifecycle phase and use
- Resilient and trustworthy system
- Protection from threats while maintaining confidentiality, integrity, and availability

Security Challenges

- Rushed adoption and lack of governance
- Lack of specific system and user guidance
- Supply chain and vendor security posture
- System misconfiguration
- Black box problem – difficult to understand how models arrive at a decision
- Difficult to understand emerging threats and risks

REPORTED SECURITY INCIDENTS

2019-2025



Source: OWASP Gen Ai Incident & Exploit Round-up Q2 2025 & BBC

STEMPRISE

AI SYSTEMS & SECURITY CONSIDERATIONS

AI SYSTEMS

AI and Algorithms

- At its core, AI uses algorithms — step-by-step mathematical rules — to recognize patterns and make decisions from data. Traditional AI focuses on classification, optimization, and control.

Generative AI (GenAI)

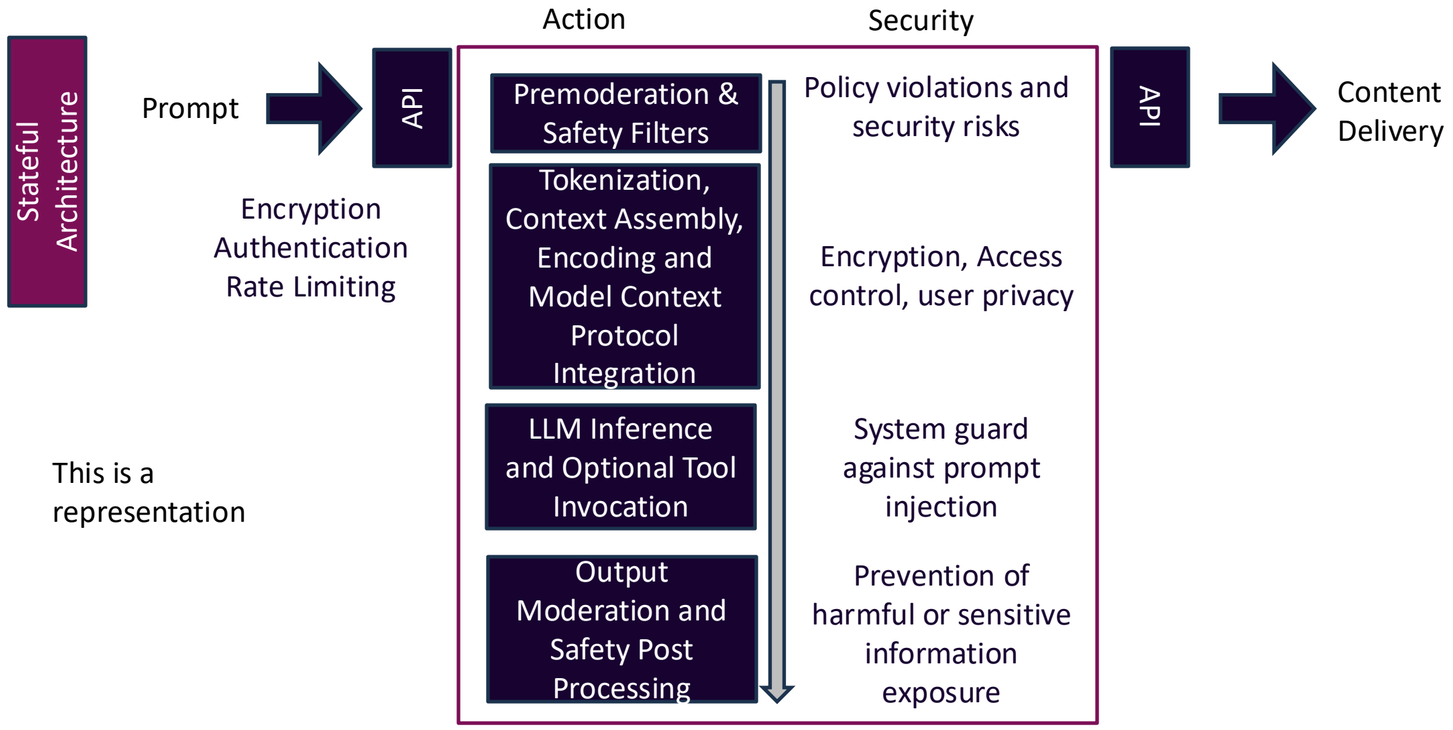
- A modern branch of AI that not only analyzes data but can also create new content (text, images, code) by learning underlying patterns. It’s like moving from just detecting a system fault to simulating realistic scenarios never explicitly seen before.

Agentic AI

- Builds on generative AI and large language models by adding planning, memory, and tool use. These agents don’t just respond — they can autonomously break down tasks, call external data sources, and act toward goals.

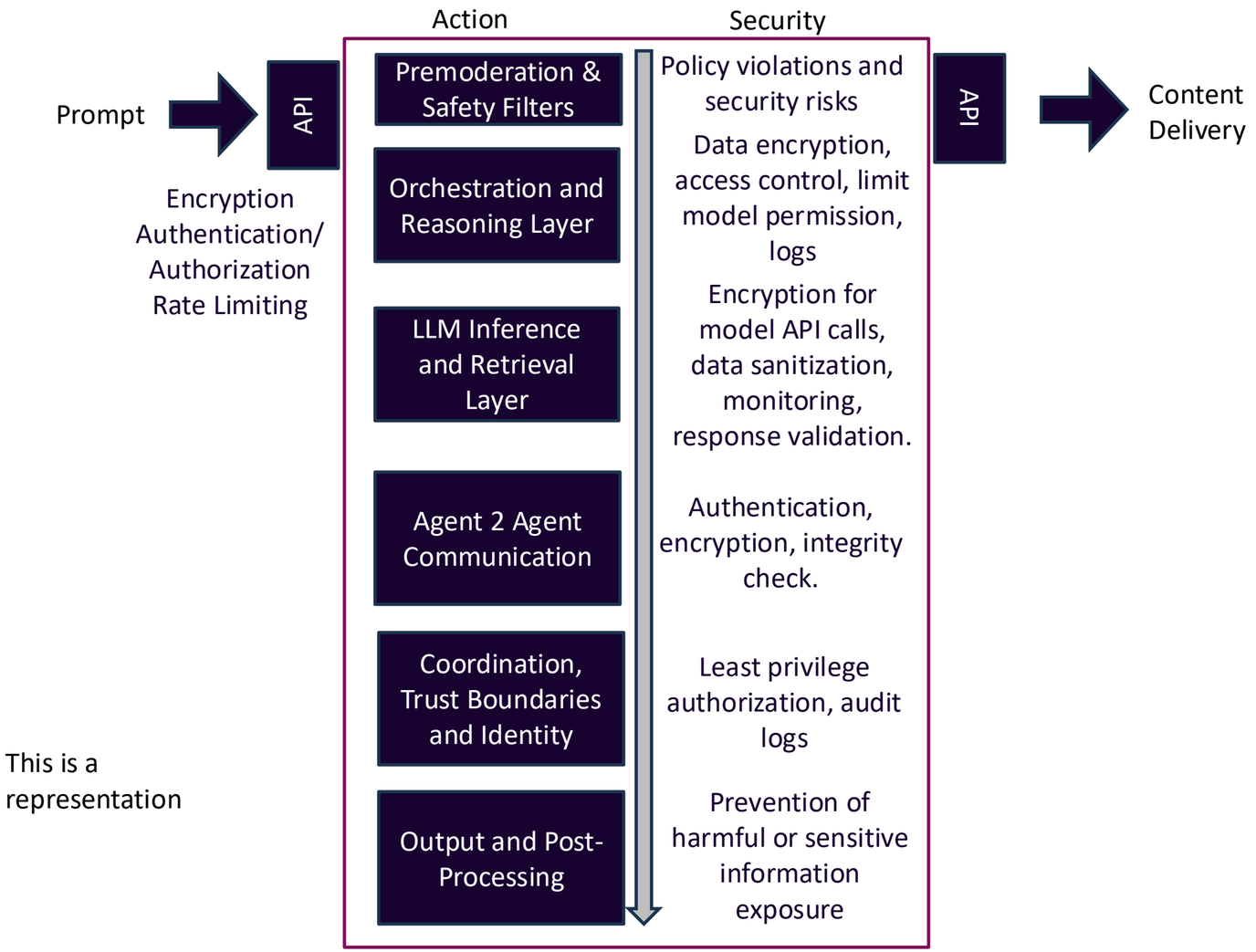
SECURITY CONSIDERATIONS FOR GenAI

Representative GenAI Architecture: From Prompt to Output



SECURITY CONSIDERATIONS FOR Agentic AI

Representative Agentic AI Architecture: From Prompt to Output



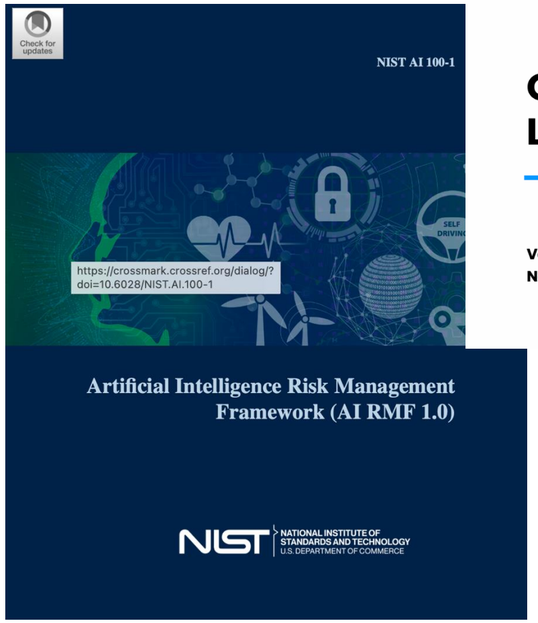
This is a representation

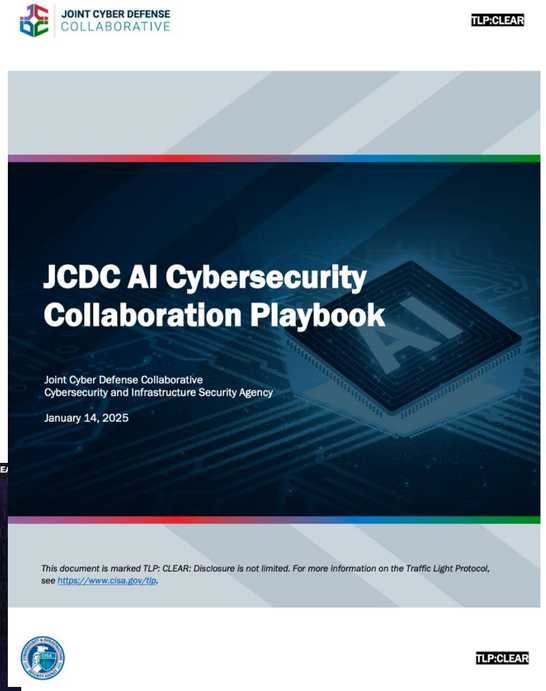
SCALABLE AI SECURITY PLAYBOOK (SAISP) 1.0

STEMPRISE

FOUNDATIONAL GUIDES

FOUNDATIONAL GUIDES



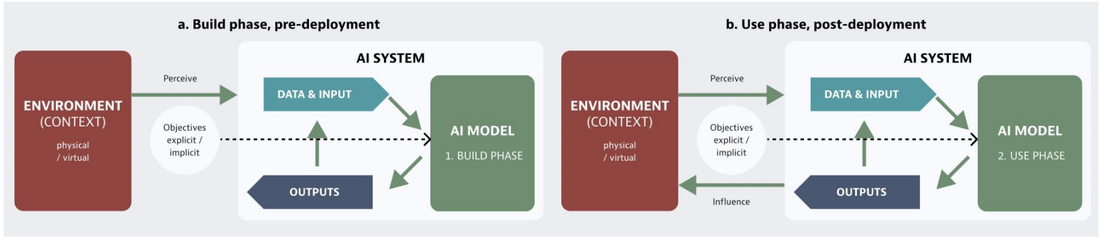


OECD AI PRINCIPLES

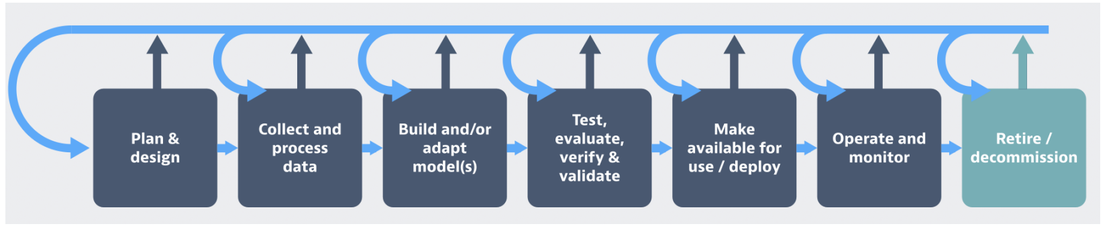
Organisation for Economic Co-operation and Development (OECD)

- Forum and knowledge hub for data, analysis and best practices in public policy. OECD work with over 100 countries across the world.
- Countries use the OECD AI Principles and related tools to shape policies and create AI risk frameworks, building a foundation for global interoperability between jurisdictions.
- Countries use the OECD's definition of an AI system and lifecycle in their legislative and regulatory frameworks and guidance.
- The principles, definition and lifecycle are all part of the OECD Recommendation on AI.

AI SYSTEM



AI SYSTEM LIFECYCLE

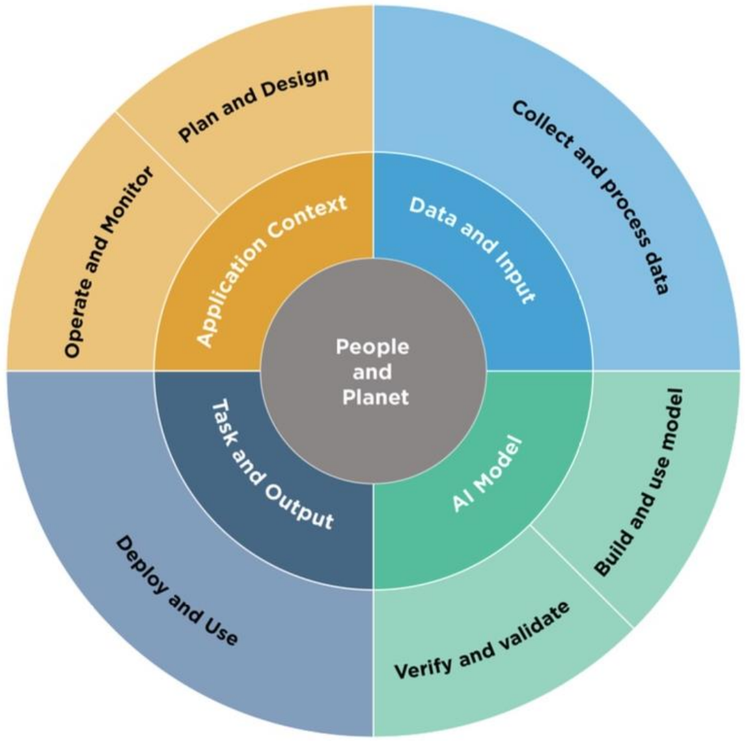


Source: OECD AI Principles Overview

NIST AI RISK MANAGEMENT FRAMEWORK (AI RMF) AI Lifecycle, Key Dimensions and Actors

National Institute of Standards and Technology (NIST)

The Framework is designed to equip organizations and individuals (AI actors) with approaches that increase the trustworthiness of AI systems, and to help foster the responsible design, development, deployment, and use of AI systems over time.



Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

In the context of the AI RMF *RISK* refers to the composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event

Source: Artificial Intelligence Risk Management Framework (AI RMF 1.0)

NIST AI RMF

National Institute of Standards and Technology (NIST)

Govern

- Policies, processes, procedures and practices across the organization
- Accountability
- Workforce diversity and Commitment to risk culture

Map

- Establish and understand your context: Categorize AI systems and actors
- AI capabilities, goals, benefits and costs
- Map risks, benefits and impacts

Measure

- Identify and apply methods and metrics of AI risks
- Evaluation of AI systems for trustworthy characteristics
- Mechanism for tracking identified risks and efficacy of measurements

Manage

- AI risk-based assessments from the Manage and Measure functions
- Strategies to maximize AI benefits minimizing negative impacts
- Plans for response, recovery and communication

Source: Artificial Intelligence Risk Management Framework (AI RMF 1.0)

OWASP: *Summary of Top 10 for LLM Applications*

Open Worldwide Application Security Project (OWASP)

Developers, Integrators, and End Users

LLM 01

PROMPT INJECTION

Alteration of LLM's behavior or output in unintended ways

LLM 02

SENSITIVE INFORMATION DISCLOSURE

Exposure of sensitive or proprietary data

LLM 03

SUPPLY CHAIN

Affect the integrity of training data, models and deployment platforms

LLM 04

DATA MODEL POISONONG

Manipulation of pre-training, fine-tuning, or embedding data to introduce vulnerabilities, backdoors, or biases.

LLM 05

IMPROPER OUTPUT HANDLING

Insufficient validation, sanitization, and handling of the outputs generated by LLMs

LLM 06

EXCESSIVE AGENCY

Enables damaging actions to be performed in response to unexpected, ambiguous or manipulated outputs from an LLM

LLM 07

SYSTEM PROMPT LEAKAGE

Systems prompts or instructions are used to steer the behavior of the model and may contain system sensitive information

LLM 08

VECTOR AND EMBEDDING WEAKNESSES

Injection of harmful content, manipulation of model outputs, or access to sensitive information

LLM 09

MISINFORMATION

LLMs produce false or misleading information that appears credible

LLM 10

UNBOUNDED CONSUMPTION

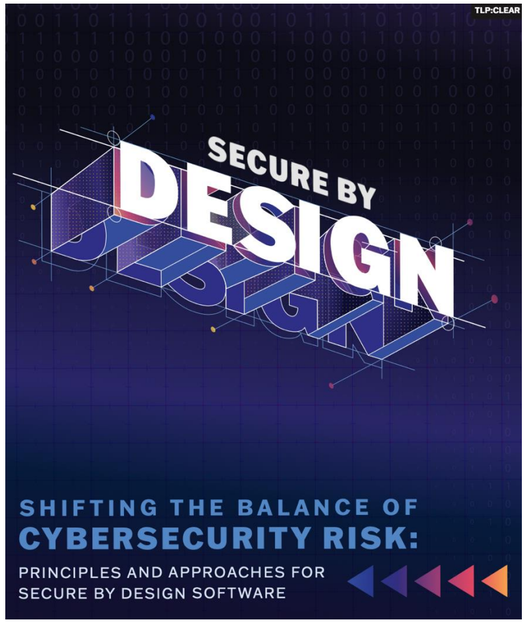
LLMs generates outputs based on input queries or prompts. Users are allowed to conduct excessive and uncontrolled inferences

Source: OWASP Top 10 LLM Applications 2025

CISA

Critical Infrastructure and Security Agency (CISA)

Secure by Design
Target Audience: Developers and Manufacturers



SOFTWARE PRODUCT SECURITY PRINCIPLES

Principle 1: Take ownership of customer security outcomes

Security must be baked not bolted on.

Principle 2: Embrace radical transparency and accountability

Share lessons learned from customers' deployments. Commitment to complete and accurate records of common vulnerability and exposure (CVE) records and advisories.

Principle 3: Build organizational structure and leadership to achieve these goals

Executives need to prioritize security as a critical element of product development across the organization, and in partnership with customers

- **Secure by design:** *products are built to protect against malicious cyber actors*
- **Secure by default:** *products are resilient against prevalent exploitation techniques without added charge.*

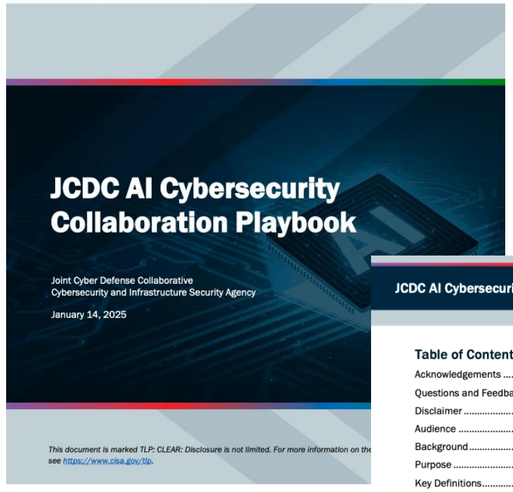
AI Cybersecurity Collaboration Playbook
Target Audience: All AI Actors

JOINT CYBER DEFENSE COLLABORATIVE

TLP:CLEAR

America's Cyber Defense Agency
NATIONAL COORDINATOR FOR CRITICAL INFRASTRUCTURE SECURITY AND RESILIENCE

JOINT CYBER DEFENSE COLLABORATIVE



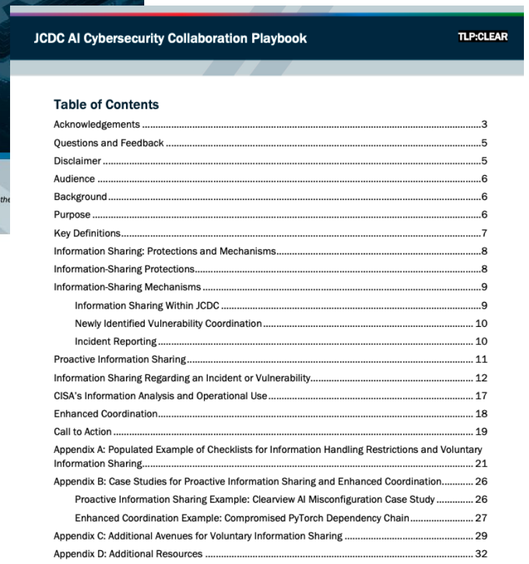

JCDC AI Cybersecurity Collaboration Playbook TLP:CLEAR

Table of Contents

Source: JCDC AI Cybersecurity Collaboration Playbook, 2025

SCALABLE AI SECURITY PLAYBOOK (SAISP) 1.0

SCALABLE AI SECURITY PLAYBOOK (SAISP)

STEMPRISE

Purpose & Scope

Provide a structured, scalable, and operational framework for implementing AI security practices across organizations.

Framework Alignment

- OECD
- NIST AI RMF
- OWASP Top 10 for LLM Applications
- CISA: Secure by Design & JCDC AI Cybersecurity Collaboration Playbook

AI Security Lifecycle

Ensure that protection, privacy, and governance are integrated into every phase of system design, development and operations

Operational Checklists

Life cycle phase specific checklist adapted for scalable AI security

Tier-Based Application

The proposed playbook is scaled across tiers to ensure relevance from small business to critical infrastructure operators.

Governance & Continuous Improvement

Ensure AI security remains a living discipline. Organizations may establish a repeatable review process.

STEMPRISE

AI SECURITY LIFECYCLE PHASES

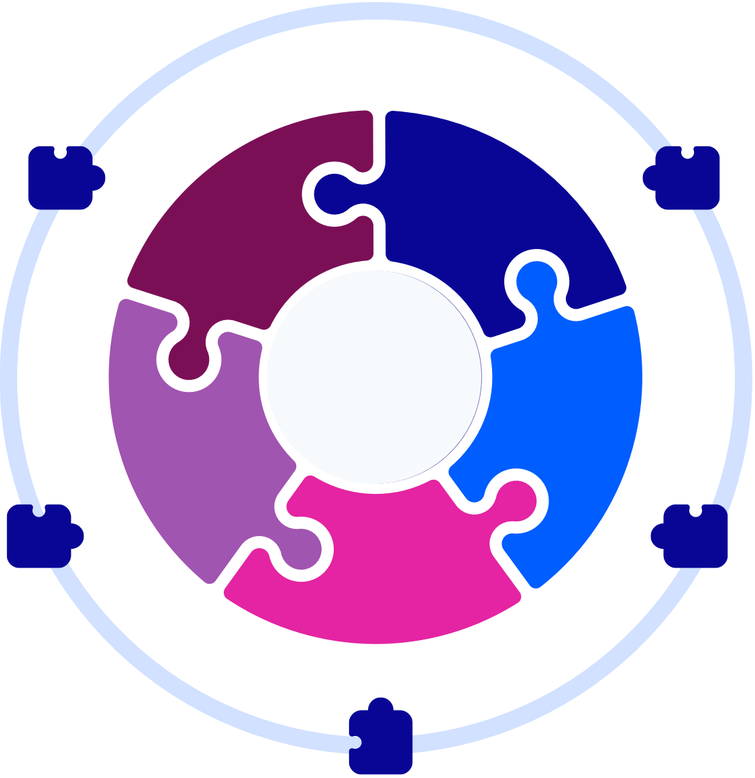
Ensure that protection, privacy, and governance are integrated into every phase of system design, development and operations

Govern Phase

Mantain accountability, oversight, and continuous improvement through specific policies, audits, and external collaboration.

Operate Phase

Continuously monitor for anomalies, apply human-in-the-loop oversight, and coordinate incident response



Design Phase

Define maturity readiness, system purpose, security requirements per AI deployment type and intended use. Apply secure architecture and data minimization principles

Build Phase

Develop and configure models securely, mantaining SBOMs, enforcing access control, and validating data integrity

Deploy Phase

Verify configurations, conduct security testing, and communicate AI use transparently to stakeholders

SAISP TIER-BASED APPLICATIONS

Based on Maturity Levels and Purpose

Tier	AI Adoption Stage & Maturity Assessment	Governance Focus	Risk Posture
Tier 1	Exploration & Enablement	Awareness, Responsible Use, Foundational Security	Low-impact, basic safeguards
Tier 2	Integration & Optimization	Formalized Controls, Measurement, Continuous Improvement	Moderate impact, proactive management
Tier 3	Full Adoption & Integration	Full Lifecycle Assurance, Transparency, Collaboration	High impact, resilient and auditable systems

STEMPRISE

SIASP OPERATIONAL CHECKLIST

Alignment with foundational guides

Lifecycle Phase	Objectives	General Controls
Design	❑ **Define AI purpose, risk context, and accountable owner.** ❑ **Data Classification** ❑ **Use Trusted Tools**	❑ Define AI purpose, risk context, and accountable owner. ❑ Perform threat modeling (prompt-injection, data leakage). ❑ Use trusted suppliers and document design assumptions.
Build	❑ **Access control** ❑ **Prompt Library Management** ❑ **Secure Credentials**	❑ Maintain SBOM and validate dependencies. ❑ Apply OWASP LLM Top-10 mitigations during development &testing. ❑ Secure secrets and implement reproducible builds.
Deploy	❑ **Configuration** ❑ **Model Verification** ❑ **Transparency**	❑ Require risk & security review before going live ❑ Enforce encryption, RBAC, and network segmentation ❑ Conduct red team / adversarial testing and enable logging.
Operate	❑ **Monitoring** ❑ **Incident Reporting** ❑ **Logging**	❑ Continuously monitor model drift and anomalies ❑ Follow JCDC incident response checklist (classify → share → mitigate). ❑ Apply sensitive markings and update models securely.
Govern	❑ **Policy Awareness** ❑ **Review & Audit** ❑ **External Collaboration**	❑ Maintain AI risk register and governance board. ❑ Conduct annual audits and update policies ❑ Share lessons learned through sector or partner collaboration.

*subject to change

USE CASES

STEMPRISE

SAISP Use Case #1: Considerations for Small Company Using ChatGPT for Content Development

Description

Tier 1 Exploration and Enablement

- A small company (e.g., marketing, consulting, or design) uses ChatGPT or similar hosted LLM to draft reports, articles, posts or web content.
- The AI system is externally hosted (off-premise) and accessed through a browser or API.
- The company must ensure confidentiality, accuracy, and ethical use

Lifecycle Phase	Objective	Action/Control	Framework Alignment
Design	❑ **Define AI purpose, risk context, and accountable owner.** ❑ **Data Classification** ❑ **Use Trusted Tools**	❑ Document the purpose, and intended users ❑ Identify data and sensitivity ❑ Use official tools/ vendor verification including third-party wrappers.	NIST AI RMF (Map), CISA SBD (protection by default and supplier trust)
Build	❑ **Access control** ❑ **Prompt Library Management** ❑ **Secure Credentials**	❑ AI tool access to approved employees ❑ Maintain a shared repository of approved prompt templates. Free form prompts may disclose internal information ❑ Use company SSO or multifactor authentication for AI accounts	CISA SBD (least privilege/secure authentication), NIST AI RMF (Measure/ Manage)
Deploy	❑ **Configuration** ❑ **Model Verification** ❑ **Transparency**	❑ Verify all browsers or API setting disable "share conversations" or "train on my data" ❑ Review model output and bias before publishing ❑ Include AI disclosure if require by company policy.	CISA SBD; OWASP LLM03 (data leakage), NIST AI RMF (Manage & Govern)
Operate	❑ **Monitoring** ❑ **Incident Reporting** ❑ **Logging**	❑ Check outputs for hallucination, misinformation, or copyright issues ❑ Define what is a data or output incident ❑ Keep monthly records of AI use (purpose, users, other)	OWASP LLM06 (Overreliance), JDC AI Playbook, NIST AI RMF (Measure)
Govern	❑ **Policy Awareness** ❑ **Review & Audit** ❑ **External Collaboration**	❑ AI acceptable user guidelines in employee handbooks. AI employee training. ❑ Reassess AI use periodically, update controls if new capabilities are added. ❑ Subscribe to official vendor or CISA advisories for AI platform updates	NIST AI RMF (Govern); CISA SBD (Continues Improvement) and JCDC AI Playbook Information Sharing

Disclaimer: The content is for educational and informational purposes. Does not imply specific recommendations.

STEMPRISE

SAISP Use Case #2:

Considerations for Municipal Distribution Utility Using AI-based Workflow Assistant for Solar + Storage Interconnection Studies

Description

Tier 2 Integration & Optimization

- GOAL: AI- based workflow assistant to accelerate interconnection studies for solar + storage.
- The tools requires access to sensitive data (e.g. CEII, GIS, SCADA, customer applications)
- Generates impact analysis and locations that require further review.
- The system uses LLM + RAG integrated with utility planning tools.
- Deployment Type: Private-cloud LLM

Disclaimer: The content is for educational and informational purposes. Does not imply specific recommendations.

Lifecycle Phase	Objective	Action/Control	Framework Alignment
Design	❑ **Define AI purpose, risk context, and accountable owner.** ❑ **Data Classification** ❑ **Use Trusted Tools and Sources**	❑ Define AI workflow boundaries (human validation), ❑ Identify data sources (e.g. CEII, GIS, SCADA, solar/storage system) ❑ Confirm data provenance, conduct threat modeling for data exposure and model misuse	NIST AI RMF (Map/Govern), CISA SBD (protection by default and supplier trust), OWASP LLM 03 (Sensitive Information Disclosure)
Build	❑ **Access control** ❑ **Prompt Library Management** ❑ **Secure Credentials**	❑ Implement MCP-based context isolation for scoped data access. ❑ Encrypt data in transit/at rest. Separate environment for development, test and production ❑ SBOM and validate dependencies ❑ Data provenance logs	CISA SBD (least privilege/secure development), NIST AI RMF (Measure/ Manage), OWASP LLM 05 (Supply Chain)
Deploy	❑ **Configuration** ❑ **Model Verification** ❑ **Transparency**	❑ Host in utility managed cloud ❑ Enforce SSO+ MFA and role-based access ❑ Enable default logging for AI-data interactions and sensitive dataset access.	NIST AI RMF (Manage & Govern), CISA SBD; OWASP LLM09 (Insecure Configuration)
Operate	❑ **Monitoring** ❑ **Incident Reporting** ❑ **Logging & Verification**	❑ Monitor for abnormal data access or prompt injection attempts ❑ Establish incident workflows ❑ Compare AI – generated study outputs with field data for validation	OWASP LLM06 (Overreliance), JDC AI Playbook, NIST AI RMF (Measure)
Govern	❑ **Policy Awareness** ❑ **Review & Audit** ❑ **External Collaboration**	❑ Maintain AI System Owner & Governance Board Oversight ❑ Review sensitive data access logs and audits periodically ❑ Incorporate validation results into model retraining and process updates.	NIST AI RMF (Govern); CISA SBD (Continues Improvement) and JCDC AI Playbook Information Sharing

STEMPRISE

SAISP Use Case #3:

Considerations for Health Care Facility for Off-Premise Multi-Agent System for Patient Data Processing

Description

Tier 3 High Assurance/Regulated Environment

- GOAL: Streamline documentation and billing while maintaining HIPAA compliance and prevent data leakage.
- AI clinical agents hosted in secure cloud environment (off-premise)
- Agents are used for
  - Summarize physician – patient conversation
  - Labs results from EHR
  - Treatment-coding and reimbursement alignment

Disclaimer: The content is for educational and informational purposes. Does not imply specific recommendations.

Lifecycle Phase	Objective	Action/Control	Framework Alignment
Design	❑ **Define AI purpose, risk context, and accountable owner.** ❑ **Data Classification** ❑ **Use Trusted Tools**	❑ Define AI purpose, agent roles, data boundaries, and HIPAA minimum necessary rules. ❑ Map PHI data flows, require explicit patient consent for ambient capture. ❑ Prohibit third-party wrappers and disable “train on my data”	NIST AI RMF (Map/Govern), CISA SBD (protection by default and supplier trust)
Build	❑ **Access control** ❑ **Prompt Library Management** ❑ **Secure Credentials**	❑ Implement MCP-based context isolation, access only to scoped data. ❑ Encrypt all data in transit/at rest, tokenize PHI for internal testing. ❑ Maintain SBOM and validate dependencies.	CISA SBD (least privilege/secure authentication), NIST AI RMF (Measure/ Manage), OWASP LLM03, LLM05
Deploy	❑ **Configuration** ❑ **Model Verification** ❑ **Transparency**	❑ Host in HIPAA – compliant private cloud ❑ Enforce SSO + MFA and role-based access ❑ Enable immutable audit logs for every agent interaction and HER call	CISA SBD; OWASP LLM09 and LLM 10NIST AI RMF (Manage & Govern)
Operate	❑ **Monitoring** ❑ **Incident Reporting** ❑ **Logging**	❑ Continuously monitor for abnormal data access or prompt-injection attempts. Establish incident logs. ❑ Use incident response guidelines ❑ Retrain or update models with sanitized data	NIST AI RMF (Manage), JDC AI Playbook
Govern	❑ **Policy Awareness** ❑ **Review & Audit** ❑ **External Collaboration**	❑ Maintain AI Governance Board ❑ Perform annual HIPAA audits and quarterly access reviews ❑ Continuous improvement and post-incident lessons learned	NIST AI RMF (Govern); CISA SBD (Continues Improvement) and JCDC AI Playbook Information Sharing

Sources:
Oracle Agents in Health Care: Benefits and Use Cases, 2025 &
Operant Networks Trust Layers for AI: Enabling Secure, Auditable Agent Ecosystem with Named Data Networking (NDN), 2025

TAKEAWAYS

Security is a critical component of the AI technology lifecycle. Systems security is the responsibility of multiple actors. Every stakeholder should aim for the highest level of security to increase trust and reliability of the system.

Security First

Ensure security is embedded into every phase of the AI Lifecycle

Strong Governance

Adoption and continuous assessments of policies and guidelines will increase trust and reliability of AI systems

Promote Collaboration

Information sharing across AI sectors strengthens the AI ecosystem across AI actors.

Education and Training

Stay up to date with new technology development and security requirements.

STEMPRISE

THANK YOU!

www.stemprise.com

marissa.morales@stemprise.com

Marissa Morales-Rodriguez, Ph.D
Founder and Technology Security Strategist
STEMPRISE