Scalable AI Security Playbook (SAISP) 1.0

By Marissa E. Morales-Rodriguez, Ph.D. STEMPRISE, LLC
Date: November 20th, 2025

1. Introduction and Motivation

Artificial intelligence (AI) is rapidly reshaping critical sectors, from healthcare and energy to finance and public services. As organizations adopt large language models (LLMs) and agentic systems to automate workflows and accelerate decision-making, the security landscape has evolved beyond traditional Information technology (IT) or operational technology (OT) boundaries. Yet, most existing guidance either focuses narrowly on model performance or treats AI security as an abstract concern. Practitioners and end users are often left without clear, actionable methods to secure AI systems holistically.

To address this gap, **the Scalable AI Security Playbook (SAISP) 1.0** was developed as a practical framework to help organizations integrate secure-by-design and trustworthy AI principles into real-world operations. This short report shows how the playbook translates the objectives of frameworks such as the Organization for Economic Co-operation and Development (OECD) AI principles, *National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF 1.0)*, *Critical Infrastructure Security Agency (CISA) Secure-by-Design Guidelines*, *Joint Cyber Defense Collaborative (JCDC)AI Cybersecurity Collaboration Playbook*, and the *Open Worldwide Application Security Project (OWASP) Large Language Model (LLM) Top 10* into a set of operational checklists that scale by maturity and intended use. It guides adopters, regardless their maturity level, through a lifecycle approach that emphasizes protection, accountability, and transparency across every phase of AI deployment.

Unlike traditional cybersecurity programs that focus solely on protecting infrastructure, SAISP 1.0 focuses on the **integrated system** of AI: the models, data pipelines, APIs, and human interactions that define trustworthy outcomes. Its goal is to promote secure adoption by aligning risk management, governance, and user awareness from design to operations. End of life and system retirement will be described in the next iteration.

2. Defining AI Security

The term *AI security* is often used loosely to describe a wide spectrum of risks, from adversarial model attacks to data leakage and disinformation. In the context of this playbook, AI security encompasses software and machine learning security, cybersecurity, data security and the security of outcomes viewed as an integrated system from prompt to output. It protects not only the underlying code and infrastructure but also the accuracy, reliability, and ethical integrity of the results produced by AI systems.

This integrated view recognizes that the security of an AI application depends on how its components interact. A prompt may trigger a model to access sensitive data through an API; that output may influence a human or automated decision; and each stage introduces new risks. For end users, AI security therefore means understanding the complete chain of trust, from input to decision, so they can adopt and manage AI technologies responsibly.

Within SAISP 1.0, AI security is operationalized through lifecycle phases, **Design, Build, Deploy, Operate, and Govern,** that reflect how organizations plan, develop, and sustain AI systems over time. Each phase contains clear objectives, minimum security actions, and framework anchors to ensure that every user, whether a small business, integrator, or critical-infrastructure operator, can implement protections that match their maturity and regulatory context.

3. Foundational Frameworks

SAISP 1.0 does not create new compliance obligations or security models, it **operationalizes existing, trusted open-source frameworks** so organizations can apply them in a coherent and measurable way. Five open standards and best-practice sources form its foundation:

- **Organization for Economic Co-operation and Development (OECD) AI Principles**[1]
  The OECD AI Principles are used by over 100 countries to shape policies and create AI risk frameworks, building a foundation for global interoperability. SAISP uses the foundation and definition of AI systems and lifecycle.

[1] OCED Legal Instruments Recommendation of the Counsil on Artificial Intelligence, AI Principles, 2025

- **NIST AI Risk Management Framework (AI RMF 1.0)**[2]
  Provides a lifecycle-based structure for managing AI risks across *Map, Measure, Manage,* and *Govern* functions. SAISP builds on this to define practical security and assurance objectives within each phase of the AI system lifecycle.

- **CISA Secure-by-Design (SBD) and Secure-by-Default Principles**[3]
  Reinforce the responsibility of software developers and system owners to build security in from the start. SAISP incorporates these principles to ensure that AI workflows, data connectors, and APIs are deployed with minimal exposure and secure configurations by default.

- **JCDC AI Cybersecurity Collaboration Playbook**[4]
  Introduces an incident-response model for AI systems, promoting transparent information sharing through the Traffic Light Protocol (TLP). SAISP aligns its *Operate* and *Govern* phases with JCDC's coordinated response steps: *Classify → Contain → Report → Recover*.

- **OWASP Top 10 for LLM Application 2025**[5]
  Identifies the most common vulnerabilities in generative and agentic AI applications, such as prompt injection (LLM01), data leakage (LLM03), and insecure plugin use (LLM09) among others. SAISP directly maps controls to these vulnerabilities to provide actionable mitigations for developers and users alike.

Together, these frameworks ensure that the playbook is **framework-anchored but implementation-focused**, bridging the gap between strategic guidance and operational practice.

4. The Scalable AI Security Playbook (SAISP) 1.0

Purpose and Structure

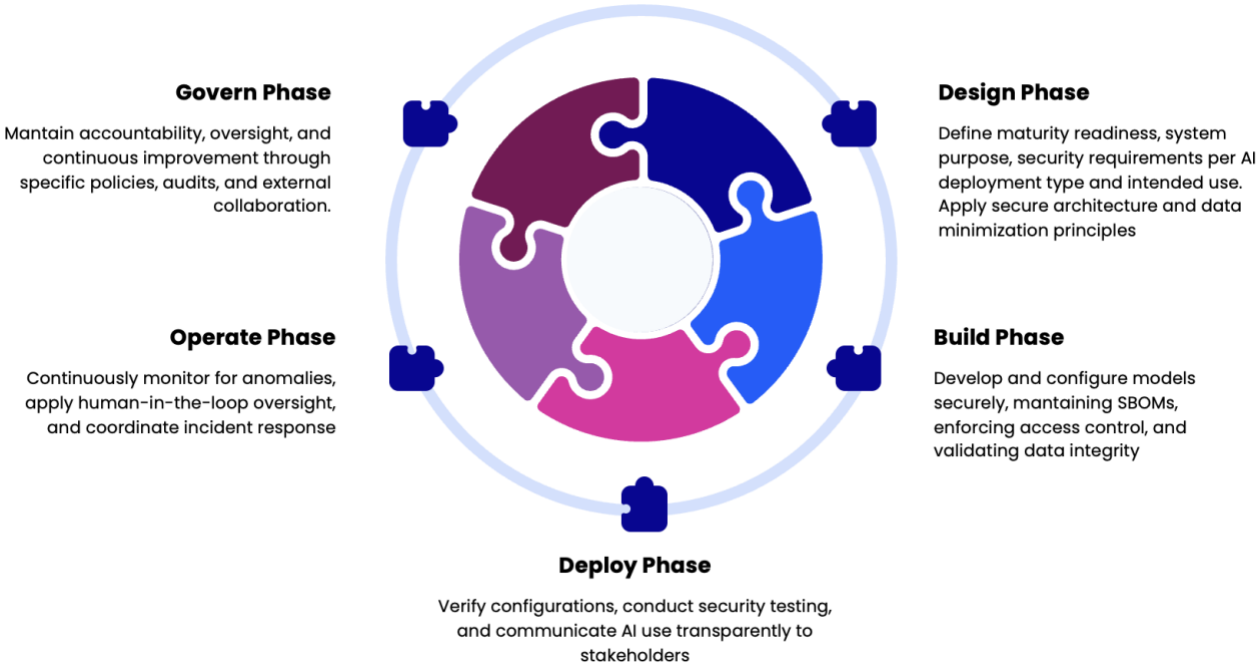
[2] NIST AI Risk Management Framework AI RMF 1.0, 2023
[3] Secure by Desing, Shifting the Balance of Ciber Risk: Principles and Approaches for Secure by Desing Software, 2023
[4] JCDC AI Cybersecurity Collaboration Playbook, 2025
[5] OWASP Top 10 for LLM Applications, 2025

SAISP provides a **lifecycle-based model** to help organizations embed protection, privacy, and governance into the design and operation of AI systems. It emphasizes that AI adoption is not a one-time event but an ongoing process of design, integration validation, monitoring, and improvement. Each phase **Design, Build, Deploy, Operate, Govern,** defines objectives, minimum security actions, and corresponding framework references. The structure can be scaled to any maturity level, allowing both small businesses and critical infrastructure operators to implement the same foundational controls at different depths of rigor. An assessment of maturity level frameworks will be completed in the next SAIS iteration.

Framework Alignment

The playbook translates the objectives of frameworks such as the *NIST AI Risk Management Framework (AI RMF)*, *CISA Secure-by-Design Guidelines*, *JCDC AI Cybersecurity Collaboration Playbook*, and the *OWASP LLM Top 10* into a set of operational checklists that scale by maturity and intended use

AI Security Lifecycle Phases



Operational Checklist

To guide the implementation of the playbook an operational checklist was developed specific to each lifecycle phase. This approach will guide the users to adopt security measures from governance to design to operations.

Lifecycle Phase	General Controls	Framework Alignment
Design	❑ Define AI purpose, risk context, and accountable owner. ❑ Perform threat modeling (prompt-injection, data leakage). ❑ Use trusted suppliers and document design assumptions.	NIST AI RMF (Map / Govern); CISA SBD; OWASP LLM01 & LLM03
Build	❑ Maintain SBOM and validate dependencies. ❑ Apply OWASP LLM Top-10 mitigations during dev/testing. ❑ Secure secrets and implement reproducible builds.	CISA SBD; OWASP LLM05 & LLM09; NIST AI RMF (Measure)
Deploy	❑ Require risk & security review before going live ❑ Enforce encryption, RBAC, and network segmentation ❑ Conduct red team / adversarial testing and enable logging.	NIST AI RMF (Manage); CISA SBD; OWASP LLM08 & LLM10
Operate	❑ Continuously monitor model drift and anomalies ❑ Follow JCDC incident response checklist (classify → share → mitigate). ❑ Apply TLP markings and update models securely.	JCDC AI Playbook; NIST AI RMF (Manage); CISA SBD
Govern	❑ Maintain AI risk register and governance board. ❑ Conduct annual audits and update policies ❑ Share lessons learned through sector or partner collaboration.	NIST AI RMF (Govern); CISA SBD; JCDC Playbook

Tiered Application

SAISP introduces **three maturity tiers** to match the capabilities and regulatory environments of different organizations. By combining lifecycle consistency with tiered scalability, SAISP ensures that *AI security is achievable for any organization*, not only those with large budgets or specialized teams.

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

5. Representative Use Cases

The Scalable AI Security Playbook (SAISP) can be applied across multiple sectors to demonstrate how lifecycle-based controls can scale with organizational maturity. The following examples illustrate how the same framework applies from small enterprises to critical infrastructure environments.

Tier 1 – Small Business Content Creation

Description:

- 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

By following SAISP's *Design* and *Operate* controls, the company limits data inputs to non-confidential content, enables multi-factor authentication (MFA), and disables data retention settings. A brief acceptable-use policy and quarterly review complete

its compliance at the Foundational tier. The result: faster workflows without exposing client information or violating data-handling policies.

Tier 2 – Municipal Utility Interconnection Studies for Solar plus Storage

Description:

- The goal is to develop an AI- based workflow assistant to accelerate interconnection studies for solar plus storage using a Private-cloud LLM.
- The tools require access to sensitive data (e.g. CEII, GIS, SCADA, customer applications)
- The outcome is the generation of impact analysis and locations for interconnection. Final reviews will be approved by humans.
- The system uses LLM + RAG integrated with utility planning tools.

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

SAISP guides the utility through model provenance tracking, encryption, role-based access, and continuous validation of AI-generated results against real-world performance. This structured maturity tier achieves measurable efficiency gains while maintaining full data-governance alignment.

Tier 3 – Healthcare Multi-Agent System

Description

- The objective is to 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

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

Operating under HIPAA, the organization applies SAISP's full lifecycle: secure design, model-context protocol (MCP) isolation, encryption, immutable audit logging, and JCDC-style incident-response coordination. The system reduces documentation burden and improves reimbursement accuracy while preserving privacy and regulatory compliance.

These cases demonstrate that AI security is not a single technology but a consistent process that adapts to organizational scale, data sensitivity, and regulatory expectations.

6. Key Takeaways and Path Forward

SAISP 1.0 reframes *AI security* as an **integrated, lifecycle discipline**—encompassing software and machine learning security, cybersecurity, and the security of AI outcomes. It provides a structured pathway for organizations to implement basic protection and assurance from *prompt to output*, bridging technical, governance, and human factors.

Key lessons include:

- **Security must be built in, not bolted on.** Embedding controls early in design and development prevents costly retrofits and compliance gaps.
- **Maturity, not size, defines readiness.** Even small entities can reach higher assurance levels through policy discipline, configuration management, and user awareness.
- **Visibility and accountability enable trust.** Continuous monitoring, provenance tracking, and auditable logs make AI systems explainable and defensible.
- **Collaboration drives resilience.** Open frameworks such as NIST AI RMF, CISA Secure-by-Design, and JCDC guidance foster a shared security language across industries.

As AI adoption accelerates, the need for **scalable, evidence-based governance** becomes urgent. SAISP offers a practical foundation for organizations seeking to integrate trustworthy AI securely helping end users understand how every interaction, from data input to model decision, contributes to overall trust and assurance.

At STEMPRISE we are helping organizations integrate AI into their workflows by developing robust and implementable secure AI strategy for your organization.

SAISP continues to develop. The next playbook version will include maturity level frameworks, end-of-life and retirement phases for AI systems, and an in-depth look at each lifecycle phase.

Disclaimer

The author's views do not represent STEMPRISE or its partners. Framework, organization, or product references are for illustration only and not endorsements. This information is educational and not legal or professional advice.

STEMPRISE

marissa.morales@stemprise.com
https://stemprise.com