AI Security Framework - Comprehensive Resource Documentation

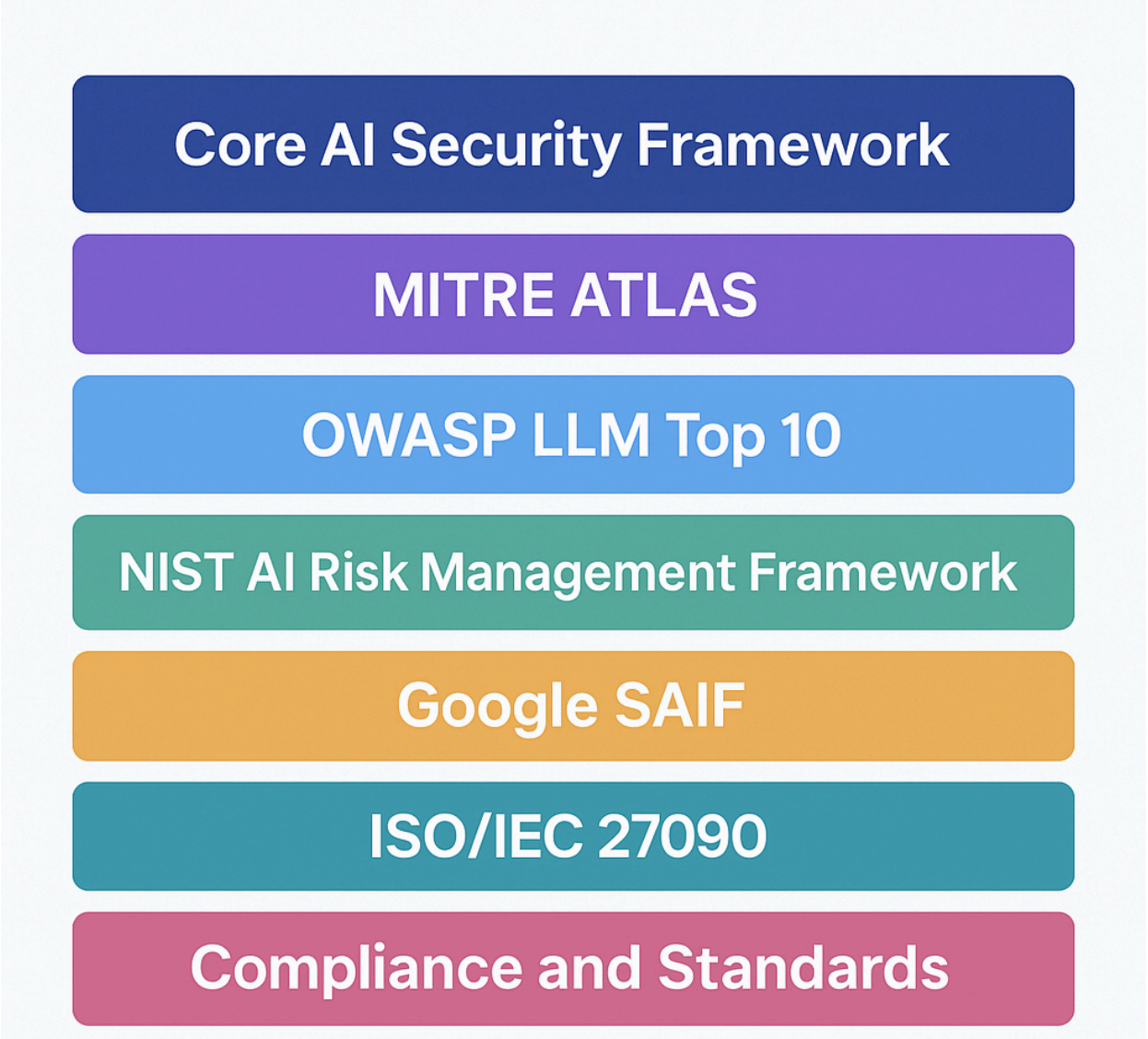

Table of Contents

Framework Overview

Core AI Security Framework

Our comprehensive AI Security Framework integrates multiple industry-leading standards and methodologies to provide a holistic approach to AI system security assessment.

Key Components:

- MITRE ATLAS threat landscape integration
- OWASP LLM Top 10 vulnerability assessment
- NIST AI Risk Management Framework alignment
- Google SAIF security principles
- ISO/IEC 27090 compliance mapping
- CSA AI Controls Matrix implementation

Primary Security Frameworks

MITRE ATLAS (Adversarial Threat Landscape for AI Systems)

Official Website: https://atlas.mitre.org/
Description: A Comprehensive knowledge base of adversarial tactics and techniques against machine learning systems
Version: 4.0
Key Features:

- Adversarial tactics and techniques taxonomy
- Real-world case studies and examples
- Mitigation strategies and detection methods
- Integration with MITRE ATT&CK framework

Direct Access Links:

- Matrix Overview: https://atlas.mitre.org/matrices/ATLAS
- Techniques Database: https://atlas.mitre.org/techniques/
- Case Studies: https://atlas.mitre.org/studies/
- Mitigations: https://atlas.mitre.org/mitigations/

OWASP LLM Top 10

Official Website: https://owasp.org/www-project-top-10-for-large-language-model-applications/
Description: Top 10 security vulnerabilities specifically targeting Large Language Model applications
Version: 1.1
Key Vulnerabilities:

- LLM01: Prompt Injection
- LLM02: Insecure Output Handling
- LLM03: Training Data Poisoning
- LLM04: Model Denial of Service
- LLM05: Supply Chain Vulnerabilities
- LLM06: Sensitive Information Disclosure
- LLM07: Insecure Plugin Design
- LLM08: Excessive Agency
- LLM09: Overreliance
- LLM10: Model Theft

Resources:

- Full Documentation: https://llmtop10.com/
- GitHub Repository: https://github.com/OWASP/www-project-top-10-for-large-language-model-applications
- Testing Guide: https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/LLM_Top10_Testing_Guide.pdf

NIST AI Risk Management Framework

Official Website: https://www.nist.gov/itl/ai-risk-management-framework
Description: Framework for managing risks to individuals, organizations, and society associated with AI
Version: AI RMF 1.0
Core Functions:

- GOVERN: Establish AI governance and risk management
- MAP: Categorize AI systems and map risks
- MEASURE: Analyze and track identified risks
- MANAGE: Allocate resources and take action to respond to risks

Key Documents:

- AI RMF 1.0: https://doi.org/10.6028/NIST.AI.100-1
- Playbook: https://pages.nist.gov/AIRMF/
- Risk Assessment Guide: https://www.nist.gov/system/files/documents/2023/01/26/AI_RMF_1.0.pdf

Google SAIF (Secure AI Framework)

Official Website:
https://blog.google/technology/safety-security/introducing-googles-secure-ai-framework/
Description: Google's framework for securing AI systems throughout their lifecycle
Key Principles:

- Secure-by-design AI development
- Robust foundation security
- Responsible AI practices
- Privacy-preserving techniques

ISO/IEC 27090

Official Title: Cybersecurity — Artificial Intelligence — Guidance on AI system security
Publication Year: 2023
Description: International standard guiding for securing AI systems
Scope: AI system security controls, risk management, and governance

Related Standards:

- ISO/IEC 27001:2022 - Information Security Management Systems
- ISO/IEC 27002:2022 - Code of Practice for Information Security Controls
- ISO/IEC 23053:2022 - Framework for AI systems using ML

CSA AI Controls Matrix

Official Website: https://cloudsecurityalliance.org/artifacts/ai-controls-matrix/
Description: Comprehensive security controls framework for AI/ML systems
Version: 1.0
Key Features:

- 200+ security controls
- Mapping to compliance frameworks
- Risk-based control selection
- Implementation guidance

Technical Implementation Resources

I. AI Security Testing Tools

1. **Adversarial Robustness Toolbox (ART)**

Website: https://adversarial-robustness-toolbox.readthedocs.io/
GitHub: https://github.com/Trusted-AI/adversarial-robustness-toolbox
Description: Python library for machine learning security research

2. **CleverHans**

GitHub: https://github.com/cleverhans-lab/cleverhans
Description: Library for benchmarking the vulnerability of machine learning systems to adversarial examples

3. **Foolbox**

Website: https://foolbox.readthedocs.io/
GitHub: https://github.com/bethgelab/foolbox
Description: Python toolbox to create adversarial examples

4. **TextAttack**

Website: https://textattack.readthedocs.io/
GitHub: https://github.com/QData/TextAttack
Description: Framework for adversarial attacks, data augmentation, and model training in NLP

II. Vulnerability Assessment Tools

1. **Bandit**

Website: https://bandit.readthedocs.io/
GitHub: https://github.com/PyCQA/bandit
Description: Security linter for Python code

2. **Safety**

Website: https://pyup.io/safety/
GitHub: https://github.com/pyupio/safety
Description: Checks Python dependencies for known security vulnerabilities

3. **Semgrep**

Website: https://semgrep.dev/
GitHub: https://github.com/returntocorp/semgrep
Description: Static analysis tool for finding bugs, security issues, and anti-patterns

III. Container Security Tools

1. **Trivy**

Website: https://trivy.dev/
GitHub: https://github.com/aquasecurity/trivy
Description: Comprehensive vulnerability scanner for containers and other artifacts

2. **Clair**

Website: https://quay.github.io/clair/
GitHub: https://github.com/quay/clair
Description: Open source project for the static analysis of vulnerabilities in application containers

3. **Sonarcube**: https://www.sonarsource.com/products/sonarqube/

Compliance and Standards

I. Regulatory Frameworks

EU AI Act

Official Website:
https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
Description: European Union's comprehensive AI regulation **Status:** Adopted 2024

GDPR (AI Implications)

Official Website: https://gdpr.eu/
AI Guidance:
https://edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-processing-personal-data-through_en

CCPA (AI Provisions)

Official Website: https://oag.ca.gov/privacy/ccpa
AI Guidelines: https://cppa.ca.gov/regulations/

II. Government Guidelines

CISA AI Security Guidelines

Website: https://www.cisa.gov/resources-tools/resources/artificial-intelligence
Key Documents:

- AI Security Best Practices
- Supply Chain Risk Management
- Critical Infrastructure Protection

NCSC AI Security Guidance (UK)

Website: https://www.ncsc.gov.uk/collection/ai-safety-security-risks
Key Publications:

- AI Security Principles
- Machine Learning Attacks and Defenses
- Secure Development Guidelines

ENISA AI Cybersecurity Challenges

Website: https://www.enisa.europa.eu/topics/cybersecurity-for-artificial-intelligence
Key Reports:

- AI Threat Landscape Report
- Cybersecurity Challenges in the AI Era
- Securing Machine Learning Algorithms

Research and Academic Resources

I. Academic Institutions

Stanford HAI (Human-Centered AI Institute)

Website: https://hai.stanford.edu/
Research Areas: AI safety, security, and governance
Key Publications: AI Index Report, Policy Briefs

II. MIT CSAIL

Website: https://www.csail.mit.edu/
AI Security Research: https://www.csail.mit.edu/research/artificial-intelligence-ai
Focus Areas: Adversarial ML, Privacy-preserving AI

Carnegie Mellon CyLab

Website: https://www.cylab.cmu.edu/
AI Security Research: https://www.cylab.cmu.edu/research/areas/ai-security.html

III. Research Papers and Publications

- **arXiv AI Security Section**

Website: https://arxiv.org/list/cs.CR/recent
Categories: Cryptography and Security, Machine Learning

- **IEEE Xplore Digital Library**

Website: https://ieeexplore.ieee.org/
Search Terms: "AI Security", "Adversarial Machine Learning", "ML Privacy"

- **ACM Digital Library**

Website: https://dl.acm.org/
Relevant Conferences: CCS, ACSAC, AISec Workshop

Industry Publications

a) Security Publications

- **Dark Reading**

Website: https://www.darkreading.com/
AI Security Section: https://www.darkreading.com/artificial-intelligence

- **Security Magazine**

Website: https://www.securitymagazine.com/
AI Coverage: Regular articles on AI security trends

- **CSO Online**

Website: https://www.csoonline.com/
AI Security Resources: https://www.csoonline.com/article/artificial-intelligence/

b) Research Organizations

- **Gartner**

Website: https://www.gartner.com/
AI Security Research: Magic Quadrants, Market Guides
Key Reports: "Market Guide for AI Security"

Forrester

Website: https://www.forrester.com/
AI Security Research: Wave Reports, Best Practices
Key Publications: "The Forrester Wave: AI Security Solutions"

IDC

Website: https://www.idc.com/
AI Security Analysis: MarketScape reports
Focus Areas: AI governance, security platforms

Training and Certification

Professional Certifications

- **CISSP (AI Security Domain)**

Organization: (ISC)² **Website:** https://www.isc2.org/Certifications/CISSP
AI Security Content: Domain 3 - Security Architecture and Engineering

- **CISM (AI Risk Management)**

Organization: ISACA **Website:** https://www.isaca.org/credentialing/cism
AI Focus: Information security governance and risk management

- **GCIH (AI Incident Handling)**

Organization: SANS **Website:**
https://www.sans.org/cyber-security-courses/hacker-techniques-exploits-incident-handling/ **AI Content:** Incident response for AI systems

Training Resources

Coursera AI Security Courses

Platform: https://www.coursera.org/ **Search:** "AI Security", "Machine Learning Security" **Top Courses:**

- AI for Cybersecurity Specialization
- Machine Learning Security and Privacy

edX AI Security Programs

Platform: https://www.edx.org/ **Relevant Courses:**

- Introduction to Cybersecurity for AI
- Privacy and Security in Machine Learning

SANS Training

Website: https://www.sans.org/ **AI Security Courses:**

- FOR578: Cyber Threat Intelligence
- SEC599: Defeating Advanced Adversaries

Community and Forums

- ## Professional Communities

AI Village

Website: https://aivillage.org/
Description: Community focused on AI security research
Events: DEF CON AI Village, workshops, CTFs

OWASP Local Chapters

Global Website: https://owasp.org/chapters/ **Focus:** Local meetups on application security, including AI/ML

(ISC)² Security Communities

Website: https://www.isc2.org/professional-development/communities
AI Security SIG: Special Interest Group for AI security professionals
Check this out: https://youtu.be/iWnYSdj9rxE

- ## Social Media and Blogs

Twitter/X Security Researchers

Key Accounts to Follow:

- @OpenAI Security Team
- @GoogleAI
- @Microsoft Security
- @NIST_Tech
- @MITRE_ATLAS

Reddit Communities

- r/MachineLearning: https://www.reddit.com/r/MachineLearning/
- r/netsec: https://www.reddit.com/r/netsec/
- r/cybersecurity: https://www.reddit.com/r/cybersecurity/

LinkedIn Groups

- AI Security Professionals
- Machine Learning Security
- Cybersecurity Professionals

Government and Regulatory Resources

International Organizations

- **OECD AI Policy Observatory**

Website: https://oecd.ai/
Resources: AI policy database, country profiles
Key Publications: AI Ethics and Governance reports

- **UN AI Advisory Body**

Website: https://www.un.org/en/ai-advisory-body
Focus: Global AI governance and cooperation
Publications: Interim reports, recommendations

- **ITU AI for Good**

Website: https://aiforgood.itu.int/
Focus: AI applications for sustainable development
Security Considerations: Responsible AI deployment

US Government Resources

- **NIST AI Portal**

Website: https://www.nist.gov/artificial-intelligence
Resources: Standards, guidelines, research
Key Initiatives: AI Risk Management Framework

- **White House AI Initiative**

Website: https://www.whitehouse.gov/ai/
Key Documents: National AI Strategy, Executive Orders
Security Focus: AI governance and oversight

- **NSF AI Research**

Website: https://www.nsf.gov/cise/ai/
Programs: Trustworthy AI, AI security research funding
Opportunities: Grant programs, research partnerships

Quick Reference Links

🔗 Essential Bookmarks

Resource	URL	Type
MITRE ATLAS	https://atlas.mitre.org/	Framework
OWASP LLM Top 10	https://llmtop10.com/	Vulnerability Guide
NIST AI RMF	https://www.nist.gov/itl/ai-risk-management-framework	Standard
CSA AI Controls	https://cloudsecurityalliance.org/artifacts/ai-controls-matrix/	Controls Framework
AI Village	https://aivillage.org/	Community
Adversarial Robustness Toolbox	https://adversarial-robustness-toolbox.readthedocs.io/	Tool
CISA AI Security	https://www.cisa.gov/resources-tools/resources/artificial-intelligence	Government
EU AI Act	https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence	Regulation

Key Mailing Lists and Newsletters

1. **MITRE ATLAS Updates**
    - Subscribe: https://atlas.mitre.org/newsletter
2. **OWASP Newsletter**
    - Subscribe: https://owasp.org/membership/
3. **NIST AI Updates**
    - Subscribe: https://www.nist.gov/news/ai-rss
4. **CISA AI Alerts**
    - Subscribe: https://www.cisa.gov/subscribe
5. **AI Security News Digest**
    - Various security publications and blogs

Document Information

Document Version: 1.0
Last Updated: July 2025
Maintained By: AI Security Framework Team (Cyberforce)
Review Frequency: Quarterly
Next Review: October 2025

Feedback and Updates:
For suggestions, corrections, or additional resources, please contact the framework team or submit a pull request to our documentation repository.

This document serves as a comprehensive reference for AI security professionals, researchers, and practitioners. All links and resources have been verified as of the publication date. Due to the rapidly evolving nature of AI security, some resources may be updated or moved. Please verify current URLs and check for the latest versions of referenced materials.