OWASP AI Summit

Where to Start - A CISO Checklist

Scott Clinton
Project Core Team Lead

Scott Clinton

President, SCVentures, Ltd
Advisory Board, Zafran Security

Project Core Team Lead, OWASP LLM

Formerly: Trend Micro, VMWare, MobileIron, Qualys, Red Hat, Sun Microsystems

15+ Years Leading Cyber Security Product and Solution Portfolios, Multiple Security Acquisitions, 12+ Years in Big Data and AI Products.

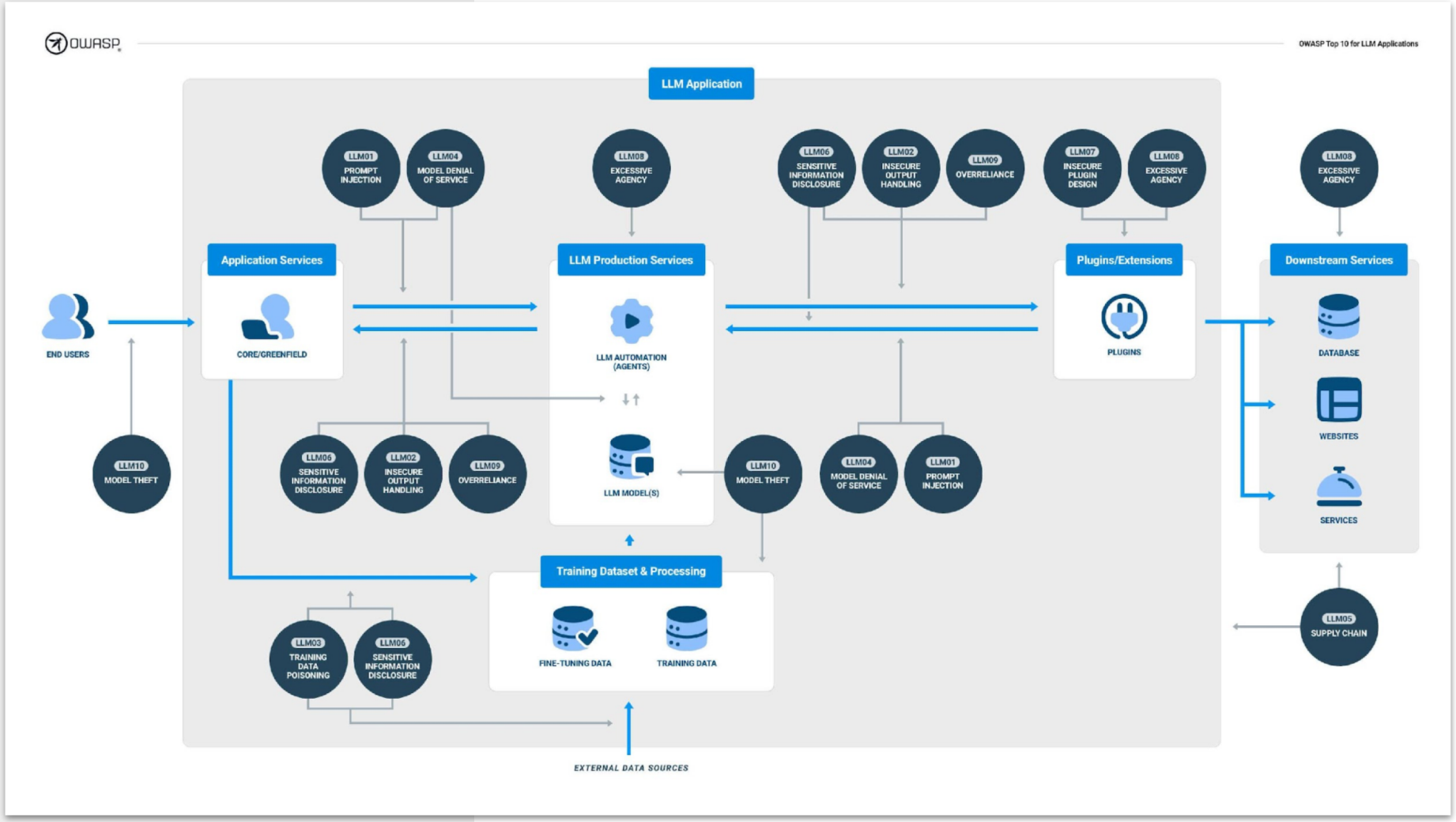
Co-lead Founding of the Project Liberty Alliance for Federated Identity (SAML)

Contributing Technical Author


Navigating the Gen AI Security Jungle

Expanded Risks Generated by AI Systems

- Dramatically Expanded Attack Surfaces
- New Insider and External Threat Vectors
- AI Accelerated Exploits

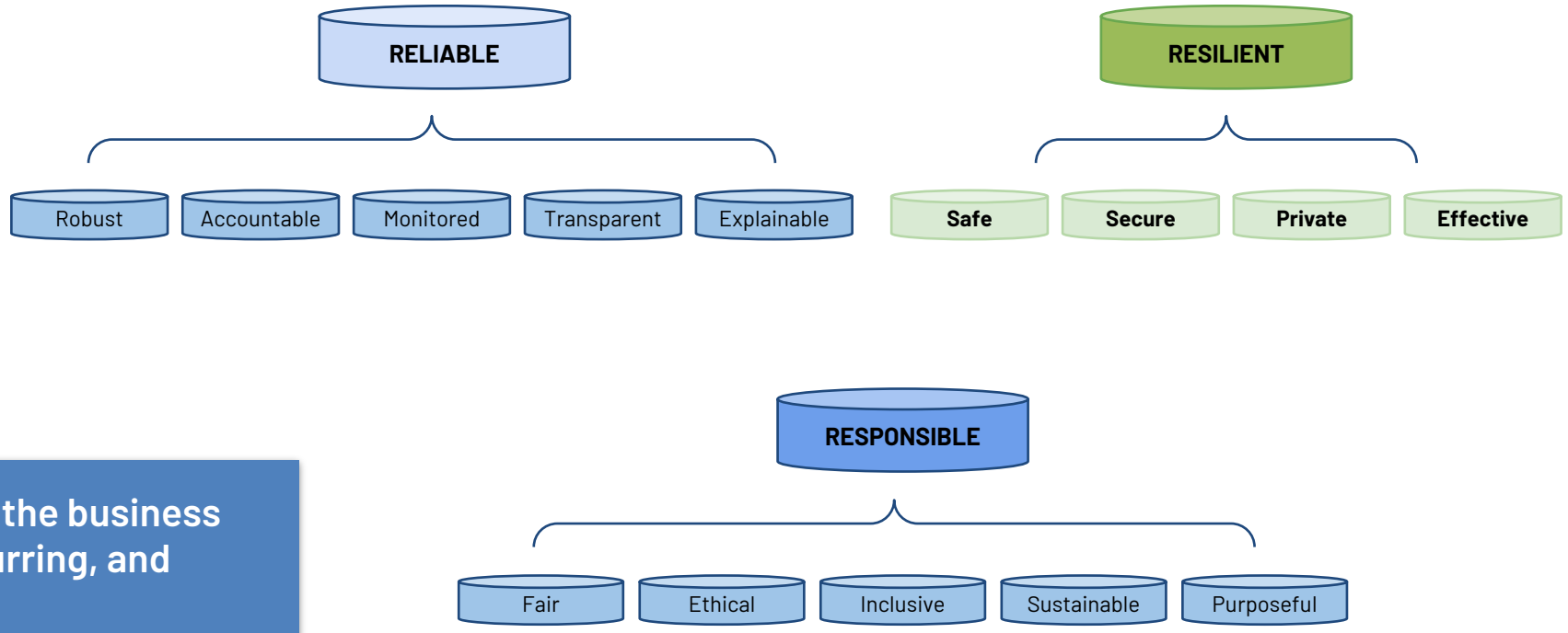


Rapid Adoption and Evolution of Model Technology

Lack of Understanding of Model Behavior

Playing Catch-up to Secure AI Apps

Tennants for Building Responsible and Trustworthy AI



Organizations need to understand the business and security risks they may be incurring, and what best practices look like.

Tom Patterson / Managing Director – Emerging Technology and Security, Accenture

CISO Challenges - Where to Start

AI security breaches average around $2.36 million for disruption of operations, with total annual costs exceeding $5.34 million

PEOPLE

Skill Gaps:
CISOs must ensure that their teams are equipped with advanced skills not only in cybersecurity but also in understanding AI technologies. Finding and training talent with these dual capabilities is a significant challenge.

Insider Threats:
The sophisticated nature of AI systems increases the risk of insider threats. CISOs need to implement strong security protocols and monitor for unusual activities that might suggest misuse of AI technologies.

Training and Awareness:
It is crucial for a CISO to maintain ongoing training programs focused on the evolving landscape of AI security threats and mitigation strategies. This includes ethical considerations in AI usage and awareness of AI-specific attack vectors.

PROCESS

Governance and Compliance:
Developing and enforcing governance frameworks that include AI risks is essential. CISOs must navigate existing and emerging regulations affecting AI, ensuring compliance while also advocating for practical standards in AI security.

Incident Response:
Traditional incident response strategies may be inadequate for AI-specific incidents, such as data poisoning or manipulative outputs. CISOs need to develop specialized protocols to effectively manage these unique challenges.

Continuous Monitoring and Assessment:
As AI systems are dynamic and continuously learning, CISOs must implement processes for ongoing evaluation and adjustment of security postures to protect against emerging threats.

TECHNOLOGY

Secure Development Lifecycle:
Integrating security from the outset in the development of AI applications is critical. This involves secure coding practices, robust testing for AI vulnerabilities, and the secure deployment of AI models.

Data Security:
Protecting the integrity and confidentiality of data used by AI systems is paramount. CISOs face the challenge of implementing effective controls to prevent data breaches and ensure data privacy.

Adapting to AI-Specific Threats:
AI technologies can be susceptible to unique forms of exploitation, such as adversarial attacks. CISOs need to stay ahead with cutting-edge technological defenses that can identify and mitigate these novel threats.

Building a Task Force

Establish a Center of Excellence (CoE) for Generative AI Security

The pace of AI technology is incredibly fast. We're moving into implementations quickly to capture value and stay relevant to developments in AI technology.. At the same time, governance and control have to be at the forefront of our strategy and consider and respect responsible AI tenets in everything we do.

- David Finney, director of IT Service Management, Microsoft

Assemble a Multi-disciplinary Team

Ethics and Governance

Data Scientists

Legal and Compliance

Security Team

AI/ML Developers

Risk Management

Human Resources

IT Infrastructure and Operations

USER/CONSUMERS

Marketing and Communications

Customer Support

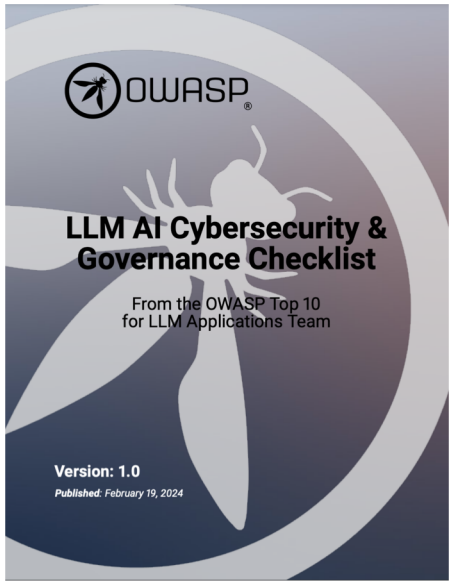
Operations Management

LoB Users

Diving Deeper Into The Checklist

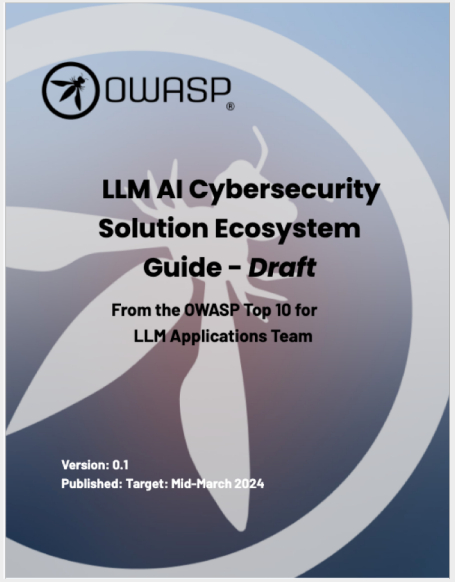
Top 10 List:
- Developers
- AppSec Teams

Leads: Steve Wilson & Ads Dawson



Checklist:
- CISOs
- Compliance Officers

Lead: Sandy Dunn

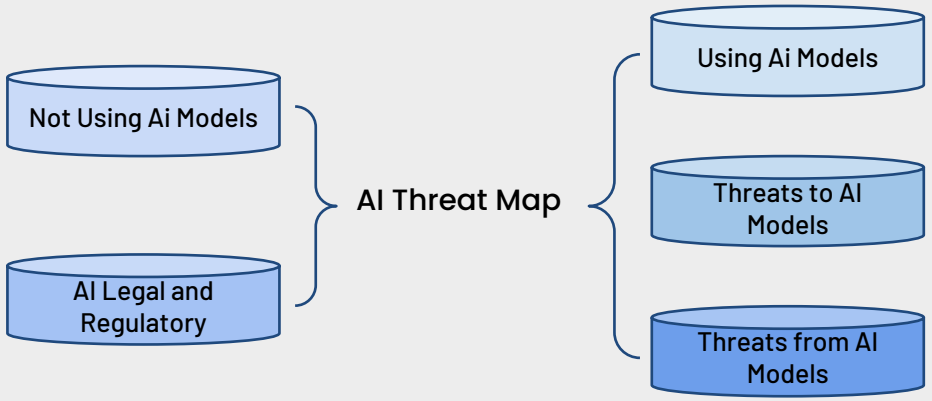


Solutions Guide:
- Development Leaders
- Security Operations

Lead: Scott Clinton

OWASP LLM, AI Cybersecurity Checklist

The checklist is intended to help technology and business leaders quickly understand the risks and benefits of using LLM, allowing them to focus on developing a comprehensive list of critical areas and tasks needed to defend and protect the organization as they develop a Large Language Model strategy.



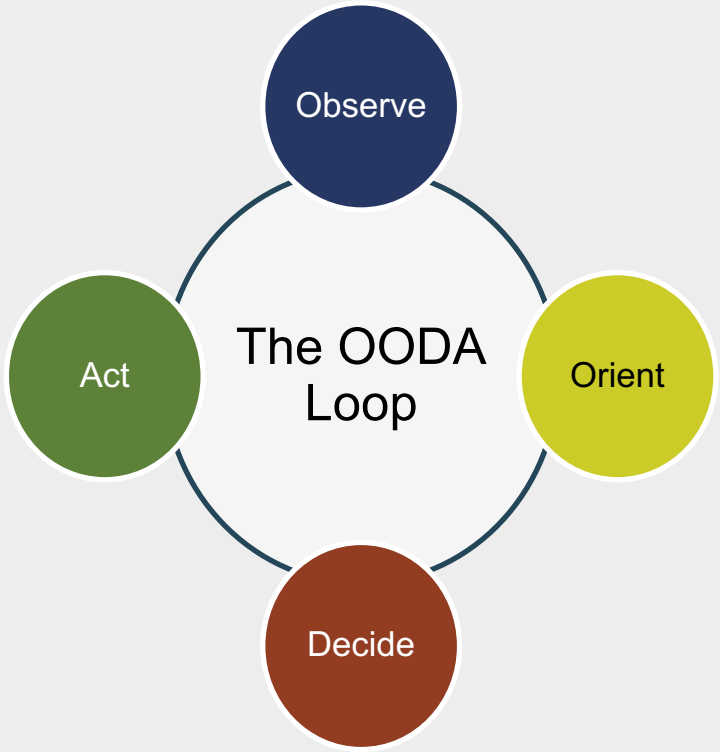
01 Currently Asymmetrical Warfare

02 Critical to Analyze the WHOLE Attack Surface

03 GenAI Curse AND Cure

04 Jagged Frontier

The checklist follows the OODA Loop decision making framework which provides leaders a method for prioritizing threats and making decisions quickly.



OWASP LLM AI Cybersecurity & Governance Checklist

Adversarial Risk

This manipulates a large language model (LLM) through crafty inputs,

Threat Modeling

Threat modeling is highly recommended to identify threats, examine processes and defenses.

AI Asset Inventory

An AI Asset inventory, as with any IT assets are essential to tracking and mitigating threats

AI Security and Privacy Training

This manipulates a large language model (LLM) through crafty inputs,

Governance

This manipulates a large language model (LLM) through crafty inputs,

Legal

This manipulates a large language model (LLM) through crafty inputs,

Regulatory

This manipulates a large language model (LLM) through crafty inputs,

Using or Implementing

This manipulates a large language model (LLM) through crafty inputs,

Testing, Evaluation, Verification, and Validation

This manipulates a large language model (LLM) through crafty inputs,

Establish Business Cases

Solid business cases are essential to determining the business value of any proposed AI solution.

Model and Risk Cards

This manipulates a large language model (LLM) through crafty inputs,

RAG: Model Optimization

This manipulates a large language model (LLM) through crafty inputs,

AI Red Teaming

AI Red Teaming is an adversarial attack test simulation of the AI System to validate there aren't vulnerabilities which can be exploited.

Adversarial Risk

Adversarial Risk includes competitors and attackers.

Highlighted Checklist Items

- **Scrutinize** how competitors are investing in artificial intelligence. Although there are risks in AI adoption, there are also business benefits that may impact future market positions.
- **Investigate** the impact of current controls, such as password resets, which use voice recognition which may no longer provide the appropriate defensive security from new GenAI enhanced attacks.
- **Update the Incident Response Plan** and playbooks for GenAI enhanced attacks and AIML specific incidents.

Read the full checklist . .

Threat Modeling

Threat modeling for GenAI accelerated attacks and before deploying LLMs is the most cost effective way to Identify and mitigate risks

Highlighted Checklist Items

- **How will attackers accelerate exploit attacks** against the organization, employees, executives, or users? Organizations should anticipate "hyper-personalized" attacks.
- **How could GenAI be used for attacks** on the business's customers or clients through spoofing or GenAI generated content?
- **Can the business detect and neutralize** harmful or malicious inputs or queries to LLM solutions?
- **Does the business have insider threat mitigation** to prevent misuse by authorized users?
- **Can the business prevent unauthorized access** to proprietary models or data to protect Intellectual Property?

Read the full checklist . .

AI Asset Inventory

An AI asset inventory should apply to both internally developed and external or third-party solutions

Highlighted Checklist Items

- **Catalog existing AI services,** tools, and owners. Designate a tag in asset management for specific inventory.
- **Include AI components in the Software Bill of Material (SBOM),** a comprehensive list of all the software components, dependencies, and metadata associated with applications.
- **Catalog AI data sources** and the sensitivity of the data (protected, confidential, public).
- **Establish if pen testing or red teaming** of deployed AI solutions is required to determine the current attack surface risk.
- **Ensure skilled IT admin staff is available** either internally or externally, following SBoM requirements.

Read the full checklist . .

AI Security and Privacy Training

ADD QUOTE, OR USE CASE

Highlighted Checklist Items

- **Actively engage with employees** to understand and address concerns with planned LLM initiatives
- **Establish a culture of open, and transparent communication** on the organization's use of predictive or generative AI within the organization
- **Train all users on ethics, responsibility, and legal issues** such as warranty, license, and copyright.
- **Update security awareness training** to include GenAI related threats.
- **Any adopted GenAI solutions should include training** for both DevOps and cybersecurity for the deployment pipeline to ensure AI safety and security assurances

Read the full checklist . .

Clear Business Cases

Solid business cases are essential to determining the business value of any proposed AI solution, balancing risk and benefits.

Example Use Cases

- Enhance customer experience.
- Better operational efficiency.
- Better knowledge management.
- Enhanced innovation.
- Market Research and Competitor Analysis.
- Document creation, translation, summarization, and analysis.

Read the full checklist . .

Governance

Corporate governance in LLM is needed to provide organizations with transparency and accountability.

Highlighted Checklist Items

- **Establish the organizationś AI RACI chart** (who is responsible, who is accountable, who should be consulted, and who should be informed)
- **Document and assign AI risk,** risk assessments, and governance responsibility within the organization.
- **Establish data management policies,** including technical enforcement, regarding data classification and usage limitations. Models should only leverage data classified for the minimum access level of any user of the system.
- **Create an AI Policy** supported by established policy (e.g., standard of good conduct, data protection, software use)
- **Document the sources and management of any data** that the organization uses from the generative LLM models

Read the full checklist . .

Legal

Many of the legal implications of AI are undefined and potentially very costly. An IT, security, and legal partnership is critical to identifying gaps and addressing obscure decisions

Highlighted Checklist Items

- **Review any risks to intellectual property.** Intellectual property generated by a chatbot could be in jeopardy if improperly obtained data was used during the generative process.
- **Restrict or prohibit the use of generative AI tools** for employees or contractors where enforceable rights may be an issue or where there are IP infringement concerns.
- **Review AI EULA agreements.** End-user license agreements for GenAI platforms are very different in how they handle user prompts, output rights and ownership.
- **Review Customer EULAs** , modify end-user agreements to prevent the organization from incurring liabilities related to plagiarism, bias propagation, or intellectual property infringement.
- **Review liability for potential injury** and property damage caused by AI systems.

Read the full checklist . .

Regulatory

The EU AI Act is anticipated to be the first comprehensive AI law but will apply in 2025 at the earliest.

Highlighted Checklist Items

- **Determine Country, State, or other Government specific AI compliance requirements.**
  - Examples:
    - Restricting electronic monitoring of employees and employment-related automated decision systems (Vermont, California, Maryland, New York,New Jersey)
    - Consent for facial recognition and the AI analysis of video required (Illinois, Maryland, Washington, Vermont)

- **Review any AI tools in use or being considered for employee hiring or management.**

- **Document any products using AI during the buying process.** Ask how the model was trained, and how it is monitored, and track any corrections made to avoid discrimination and bias.

Read the full checklist . .

Usage and Implementation

Highlighted Checklist Items

- **Threat Model**, define trust boundaries for LLM components and architecture.
- **Data Security**, verify how data is classified and protected based on sensitivity, including personal and proprietary business data. (How are user permissions managed, and what safeguards are in place?)
- **Access Control**, implement least privilege access controls and implement defense-in-depth measures.
- **Training Pipeline Security**, require rigorous control around training data governance, pipelines,models, and algorithms.
- **Input and Output Security**, evaluate input validation methods, as well as how outputs are filtered, sanitized, and approved.

Read the full checklist . .

Testing, Evaluation, Verification & Validation

NIST AI Framework recommends a continuous TEVV process throughout the AI lifecycle

Highlighted Checklist Items

- **Establish continuous testing,** evaluation, verification, and validation throughout the AI model lifecycle.
- **Provide regular executive metrics** and updates on AI Model functionality, security, reliability, and robustness.
- **Includes a range of tasks** such as system validation, integration, testing, recalibration, and ongoing monitoring for periodic updates to navigate the risks and changes of the AI system.

Read the full checklist . .

OWASP LLM, AI Cybersecurity Checklist

Leverage Model & Risk Cards

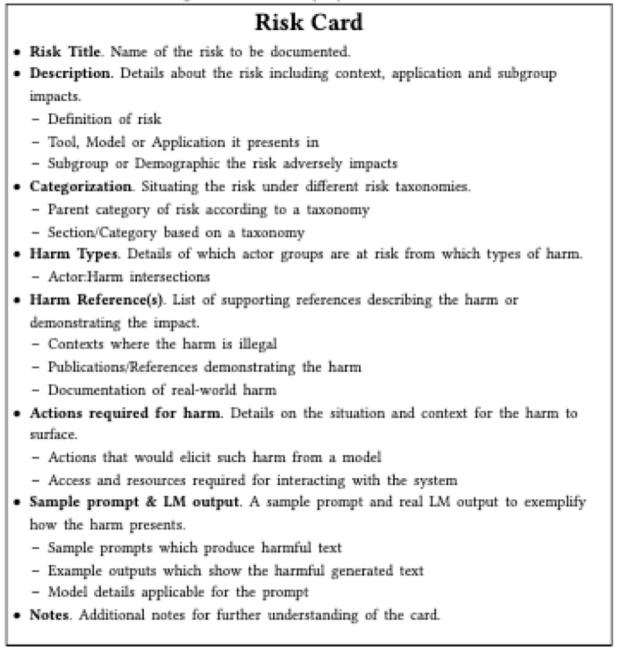
Model cards help users understand and trust AI systems by providing standardized documentation on their design, capabilities, and constraints, leading them to make educated and safe applications.

RiskCards, provide a framework for structured assessment and documentation of risks associated with an application of language models.

Highlighted Checklist Items

- **Apply ModelCard documentation as a standard practice** and requirement for both developed and consumed AI Models and applications.

- **Implement RiskCards along with ModelCards** documentation to define and document the riska a model or application may have to and organization



Risk Card

- **Risk Title**. Name of the risk to be documented.
- **Description**. Details about the risk including context, application and subgroup impacts.
  - Definition of risk
  - Tool, Model or Application it presents in
  - Subgroup or Demographic the risk adversely impacts
- **Categorization**. Situating the risk under different risk taxonomies.
  - Parent category of risk according to a taxonomy
  - Section/Category based on a taxonomy
- **Harm Types**. Details of which actor groups are at risk from which types of harm.
  - Actor:Harm intersections
- **Harm Reference(s)**. List of supporting references describing the harm or demonstrating the impact.
  - Contexts where the harm is illegal
  - Publications/References demonstrating the harm
  - Documentation of real-world harm
- **Actions required for harm**. Details on the situation and context for the harm to surface.
  - Actions that would elicit such harm from a model
  - Access and resources required for interacting with the system
- **Sample prompt & LM output**. A sample prompt and real LM output to exemplify how the harm presents.
  - Sample prompts which produce harmful text
  - Example outputs which show the harmful generated text
  - Model details applicable for the prompt
- **Notes**. Additional notes for further understanding of the card.

Example RiskCard Contents

- Risk Title
- Description
- Definition of risk
- Risk Categorization
- Harm Types.
- Harm Reference(s).
- Actions required for harm.
- Sample prompt & LM output.
- Example harmful outputs

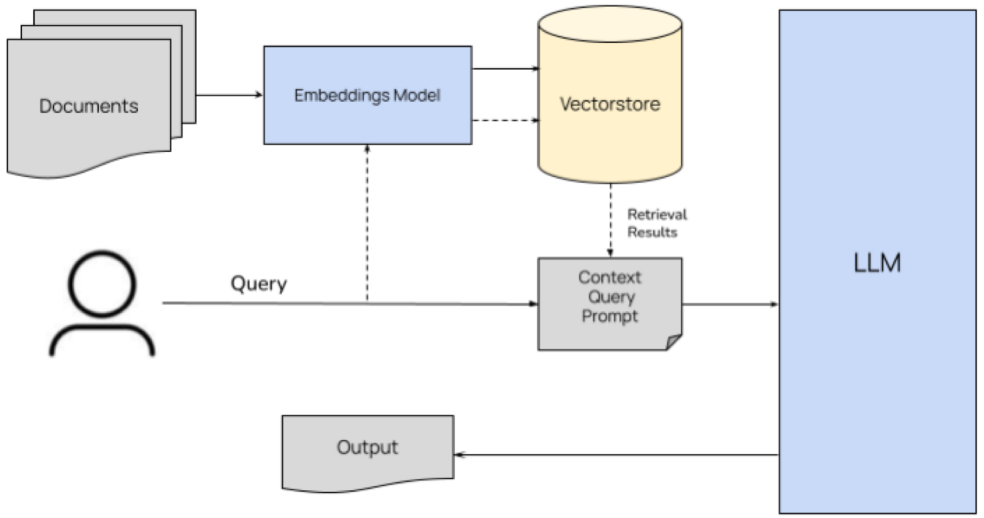
Read the full checklist . .

OWASP LLM, AI Cybersecurity Checklist

RAG: Model Optimization

This manipulates a large language model (LLM) through crafty inputs.

Highlighted Checklist Items



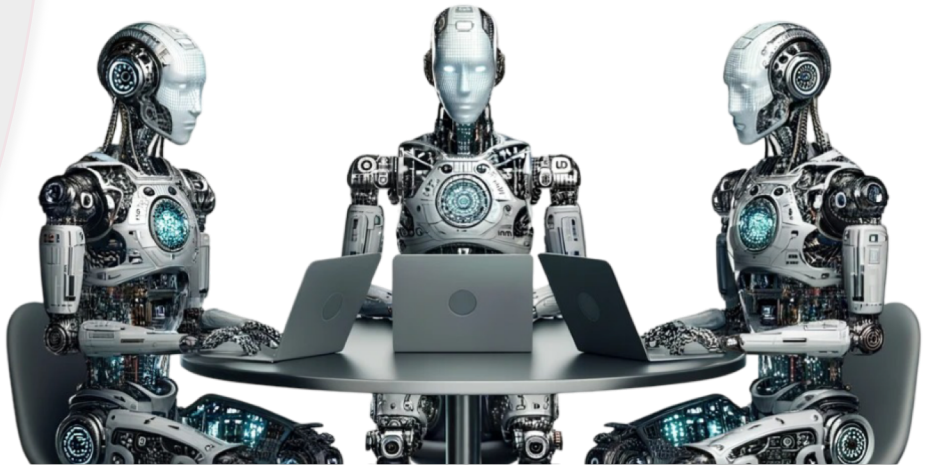
- **Create Feedback Loops,** to continuously gather input on the system's performance and security from users and security audits.
- **Implement Quality Control of Retrieved Data,** mechanisms to check the relevance and accuracy of retrieved data before it is used by the model.

Read the full checklist . .

Red Teaming

Highlighted Checklist Items

- **Incorporate Red Team testing** as a standard practice for AI Models and applications.
- **Validate there aren't any existing vulnerabilities** which can be exploited by an attacker.
- **Leverage AI systems themselves** to simulate complex attack scenarios both against AI systems and as a tool for hardening an organization's defenses.



Read the full checklist . .

The Next Steps

- Understand the Risks and Opportunities
- Assemble a Multiple-disciplinary COE
- Build Your Checklist
- Use the OWASP Checklist to Start
- Take a OODA Approach to Operationalize your Strategy

Subscribe to the Newsletter

Follow Us on LinkedIn


Thank You