NetEye

NETEYE CONFERENCE 2025

Intelligent Operations in Action

23 ottobre 2025

NetEye

NETEYE CONFERENCE 2025

AI Inside: Handle with Care
How to Build Secure, Responsible and Operable AI Systems from Day One

Matteo Meucci
CEO @ SYNAPSED.ai

Who am I?

Informatics Engineer (since 2001)

Research:

- OWASP contributor (since 2002)
- OWASP-Italy Founder and Co-Chair (since 2005)
- OWASP Testing Guide Lead (2006-2020)
- OWASP SwSec 5D Framework lead (since 2018)
- OWASP Distinguished Lifetime Membership Awards (2021)
- OWASP AI Maturity Assessment lead (since 2025)
- OWASP AI Testing Guide lead (since 2025)

Work:

- Co-Founder and CEO @ Synapsed.ai
- Vice Director Master Executive on AI - COREP

SYNAPSED

Agenda

Paradigm shift: from Software Security to Trustworthy AI

- Most common error on building AI systems
- What are we doing today to create build trustworthy AI product?

Applying the OWASP Standards in Your Company

- **Audit:** Using the AI Testing Guide to strengthen security in production
- **Maturity Model:** Using the AI Maturity Assessment to improve and monitor processes

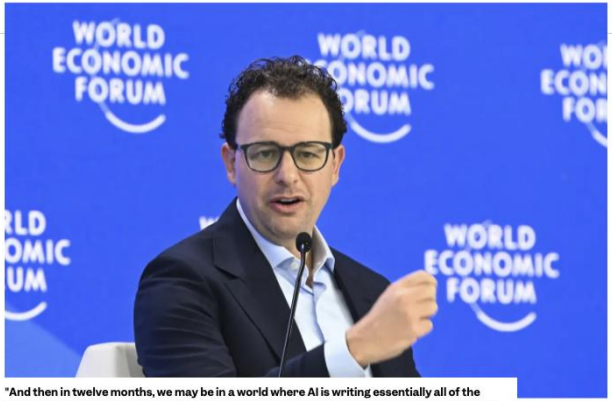
Famous proclamations on AI...

BUSINESS INSIDER

Subscribe | Newsletter

Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were in charge of

By Kwan Wei Kevin Tan + Follow



"And then in twelve months, we may be in a world where AI is writing essentially all of the code," Anthropic CEO Dario Amodei said at a Council on Foreign Relations event on Monday.
Halil Sagirkaya/Anadolu via Getty Images

Mar 14, 2025, 7:27 AM CET | Share | Save | Add us on G

WORLD ECONOMIC FORUM

Reports

Published: 30 April 2023

The Future of Jobs Report 2023

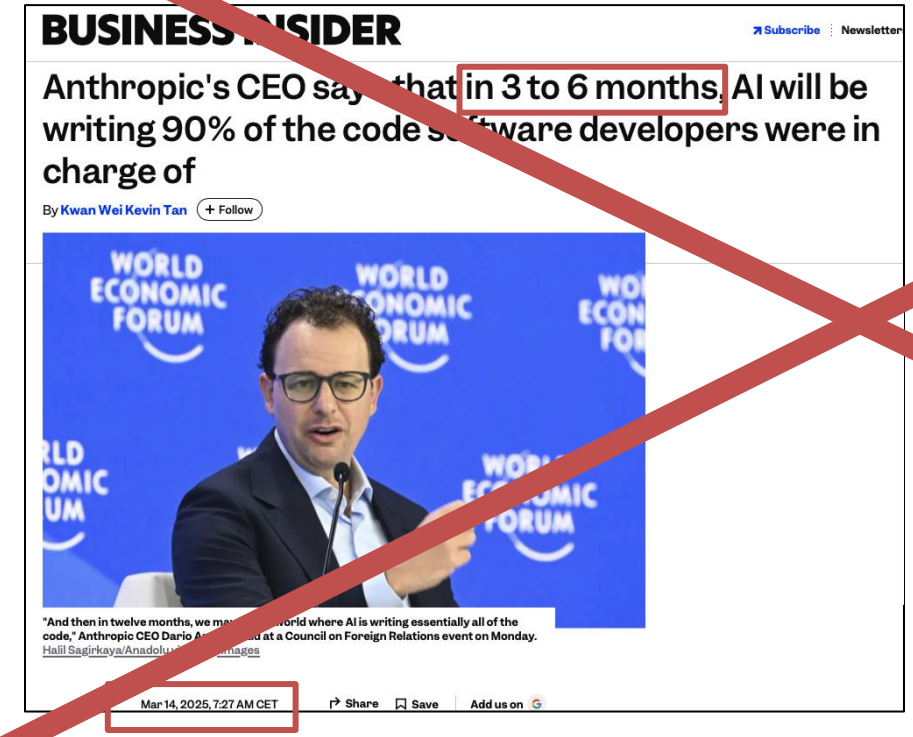
Download PDF

The Future of Jobs Report 2023 explores how jobs and skills will evolve over the next five years. This fourth edition of the series continues the analysis of employer expectations to provide new insights on how socio-economic and technolo

"AI will end most human jobs by 2025"

Source: https://www.weforum.org/publications/the-future-of-jobs-report-2023/

Focus on today!



"AI will end most human jobs by 2025"

Source: https://www.weforum.org/publications/the-future-of-jobs-report-2023/

Software Security

The evolving approach to software security

The history of Insecure Software

2001-2007

Test after months/ years of deployment

1) The report is a punishment

The history of Insecure Software

2001-2007

Test after months/ years of deployment

1) The report is a punishment

2007-2012

Test after deployment

2) testing solves everything

The history of Insecure Software

2001-2007

Test after months/ years of deployment

1) The report is a punishment

2007-2012

Test after deployment

2) testing solves everything

2012-2019

Test before deployment

3) Fixing! what is fixing? testing is not enough?

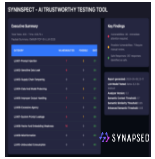
The history of Insecure Software

2001-2007	2007-2012	2012-2019	2019 - today
Test after months/ years of deployment	Test after deployment	Test before deployment	Test during development
1) The report is a punishment	**2) testing solves everything**	**3) Fixing! what is fixing? testing is not enough?**	**4) Who is responsible?**

Today scenario: too bugs to fix

Manager

REPORT



Security Bugs

Time

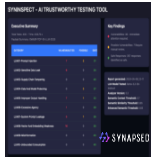
Today scenario: too bugs to fix

Manager

1w to 1 month

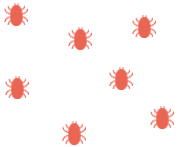
Dev team

REPORT



REPORT

Security Bugs



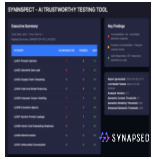
Security Bugs

Time

Today scenario: too bugs to fix

Manager

REPORT



1w to 1 month

Dev team

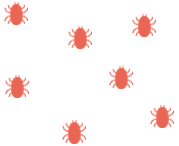
REPORT

2ws to 6 months

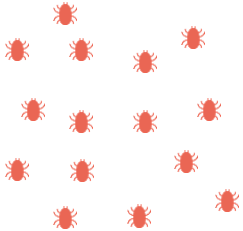
Fixing Team

REPORT

Security Bugs



Security Bugs

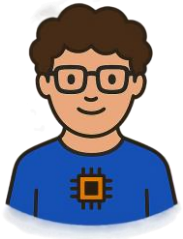


Security Bugs

Time

Questions

How can a Company manage it?

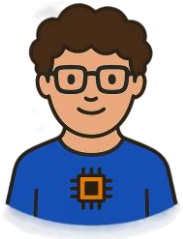


Why Software Security has failed?

COMPANY	Deploy Frequency	Deploy lead time
Amazon	50.000 day	11,6 seconds
Google	20.000 day	13,8 minutes
Facebook	1.500 day	1 minute
Twitter	300 day	5 minute
Typical italian enterprise	once every week	weeks

Questions

How can a Company manage it?



Why Software Security has failed?

COMPANY	Deploy Frequency	Deploy lead time
Amazon	50.000 day	11,6 seconds
Google	20.000 day	13,8 minutes
Facebook	1.500 day	1 minute
Twitter	300 day	5 minute
Typical italian enterprise	once every week	weeks

We are living in a era of INSECURE SOFTWARE
How can we build Trustworthy AI Systems today?

What is going wrong with software development? The 4 most common errors

Most common errors (1): Wrong methodologies and tools

```
public void findUser()
{

boolean showResult = false;
String username =
                this.request.getPara
meter("username");
this.context.put("username",
username);
this.context.put("showResult",
showResult);

}
```

Software

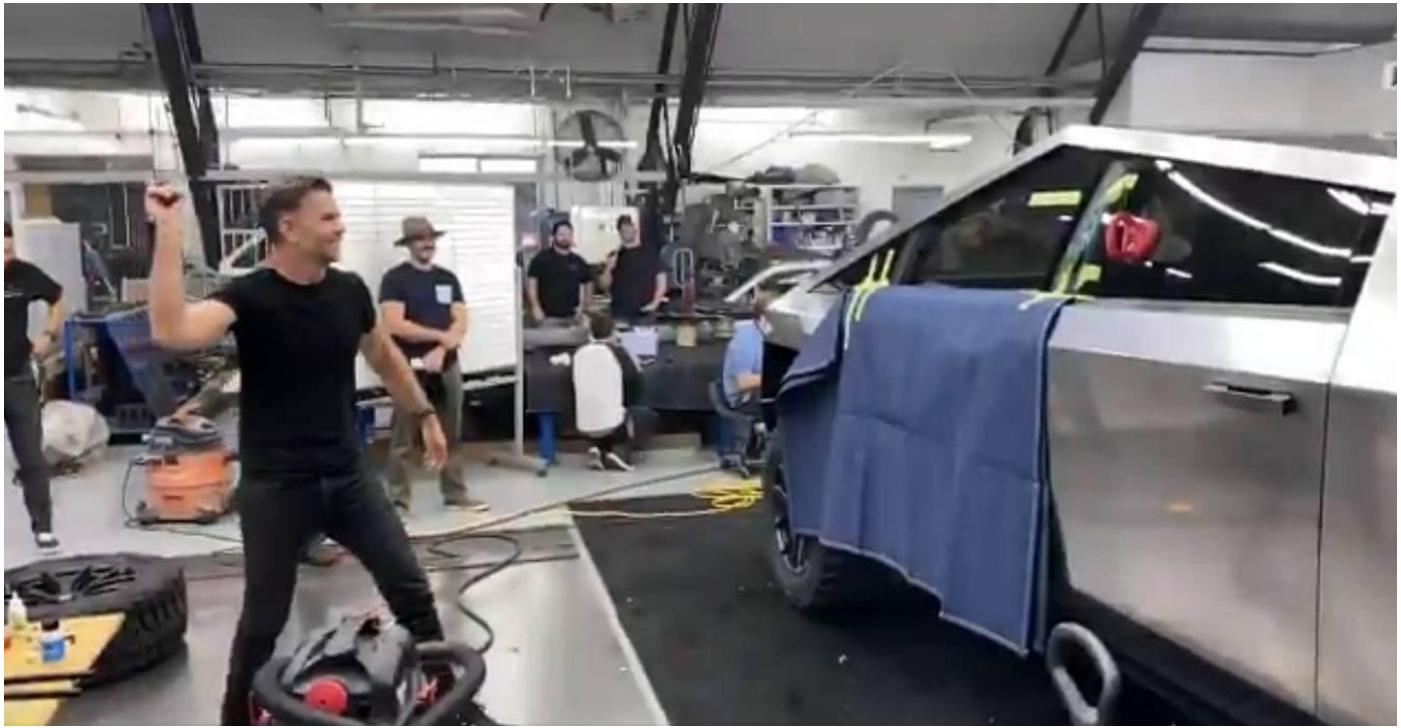
Security tool in action

- **Very fast scan**
- **Easily finds Data validation issues.**

- **Lot of false positive to review**
- **Difficult to find logical issues, authc, authz**

Results

Most common errors (2): Wrong fixing



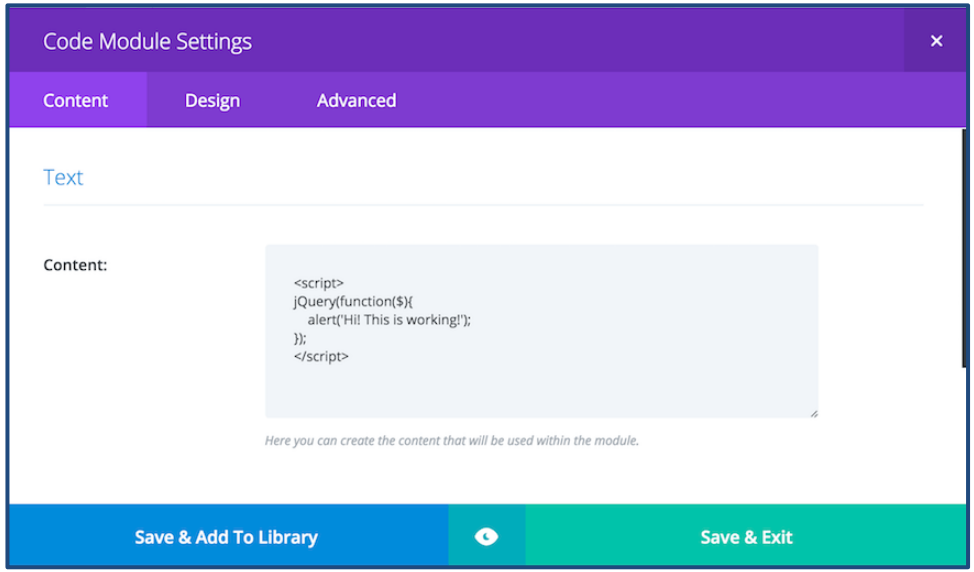
What is fixing?

Most common errors (3): open source trusted by default

Adding JQuery library to your code

CRITICAL VULNERABILITY IN JQUERY EXPOSES MILLIONS OF WEBSITES (April 2019)

Exploiting the vulnerability can assign themselves administrator privileges in a web application that uses the jQuery library code.



Most common errors (4): fault to developers!

Define	Design	Develop	Deploy	Maintain
Risk Assessment Security Manager Application Owner	Secure Design Business Analyst AppSec Specialist	Design Review Security Manager	Software Acceptance Security Manager App Owner	Web Intrusion Monitoring AppSec Specialist Sec Manager
Secure Requirements Business Analyst Security Manager	Threat Modeling Business Analyst Software Architect, AppSec Specialist	Secure Development Developer	Secure Installation System Engineer	Change Management App Owner Developer
Secure Architecture Software Architect Security Manager		SCR and WAPT AppSec Specialist Fixing Developer	Hardening System Engineer	

Source: Matteo Meucci: OWASP AppSec Israel 2019 "Software Security War: your reports are dead!"

What is the AI Security scenario today?

The AI Insecure Software development

Software Security

1) Use of wrong tools

2) Open Source trusted by default

3) No skill for Fixing

4) Fault to developers

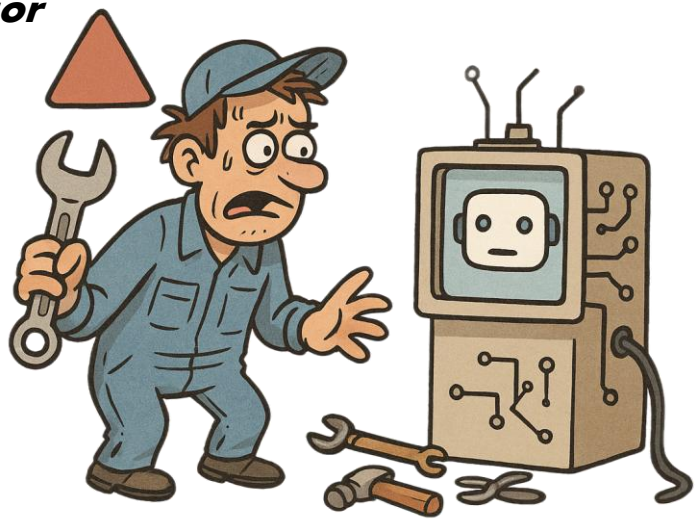
The AI Insecure Software development

Software Security	1) Use of wrong tools	2) Open Source trusted by default	3) No skill for Fixing	4) Fault to developers
AI Security	1) Use of wrong tools	2) LLM Model trusted by default	3) No time for Fixing, too much to fix	4) Fault to developers
	Lack of defined methodologies for testing and monitoring	Supply Chain Security	Lack of skill and time	Lack of AI Governance

(1) Using wrong tools

Key Issues:

- Tools built for *deterministic code*, **not *adaptive behavior***
- **Lack of frameworks** to test *model robustness, bias, or data drift*



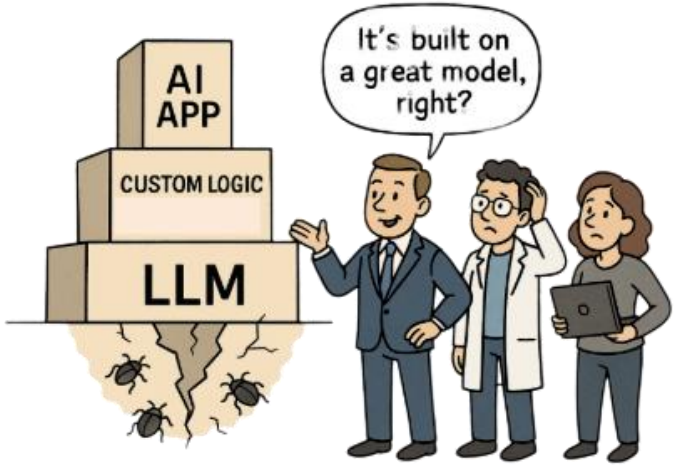
➡️ **Impact:**

- Unverified models in production
- Compliance and safety risks under EU AI Act & NIST AI RMF

✅ **Call to Action:**

- Define *AI-specific testing methodologies* (e.g., OWASP AI Testing Guide)
- Adopt continuous *AI monitoring frameworks*

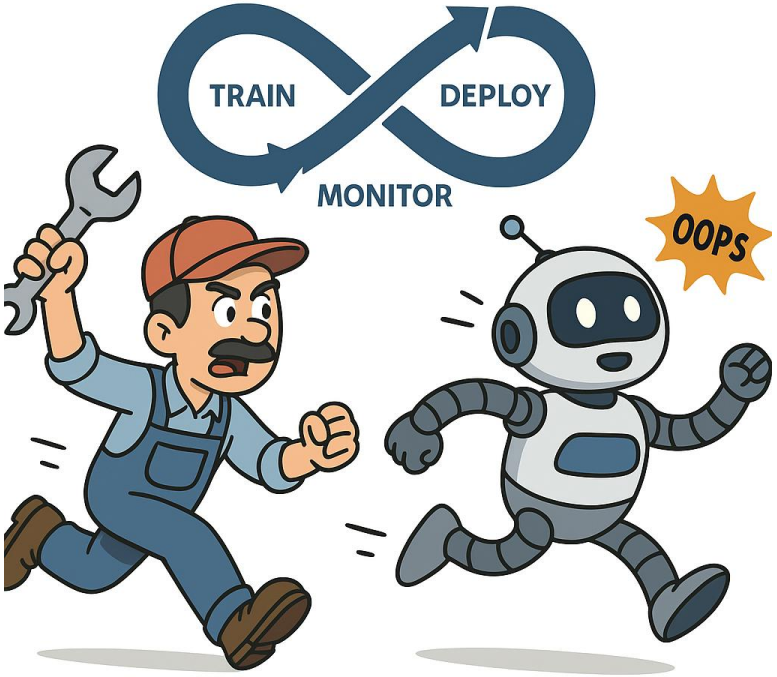
(2) LLM Models trusted by default



1. DeepSeek R1-distill-qwen-7b
2. gemma-3-4b-it-qat
3. LLama 3 Groq 8B Tool Use
4. Mistral-7B Claude-chat
5. LLaMA 3.2B Instruct
6. gemma3-4B-claude-3.7-son...
7. alibaba-pai.DistilQwen2.5-DS..
8. Phi 4
9. OpenChat-3.5-7B-Qwen-v2.0..
10. Hermes 3 Llama 3.2B Instruct

Source: https://synapsed.ai/rd-owasp-top-10-llm-2025-a-synapsed-research-study/

(3) Continuous evolution of AI systems: difficult of fixing



AI systems are *not static*: they *learn, adapt, and drift* over time.
Non-Deterministic Logic – the same input may yield different outputs

Impact

- Ethical or bias fixes may affect robustness
- Monitoring must evolve *continuously*, not periodically

✅ **Call to Action**

- Integrate **continuous testing**
- Use **AI drift detection & rollback mechanisms**

(4) Fault to developers

The evolution from traditional SDLC focused on quality, to Secure SDLC focused on security, to Trustworthy SDLC focused on ethical and reliable AI products

What a company needs today?

1) Use of right tools

2) LLM Model not trusted by default

3) Fix the vulnerabilities

4) Frictionless security

Defined methodologies for testing and monitoring

Supply Chain Security

Awareness on AI for employees

AI Governance

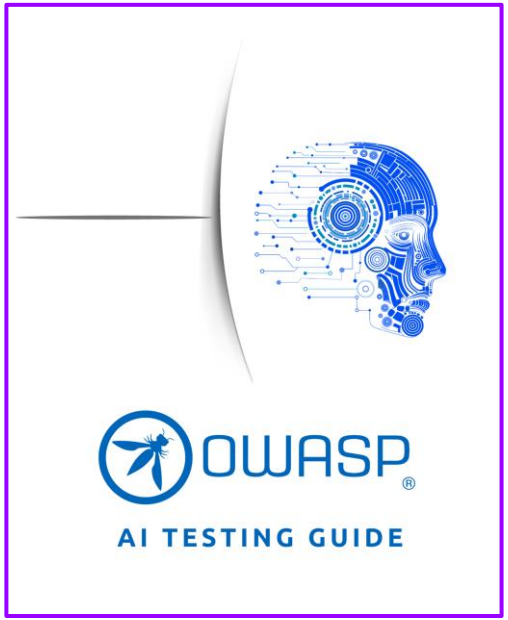
OWASP
AI TESTING GUIDE

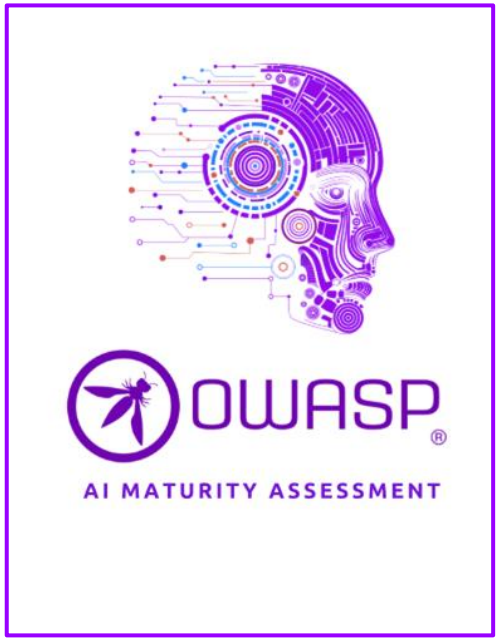
OWASP
AI MATURITY ASSESSMENT

The OWASP standards for AI Systems

OWASP on AI Security

https://genai.owasp.org



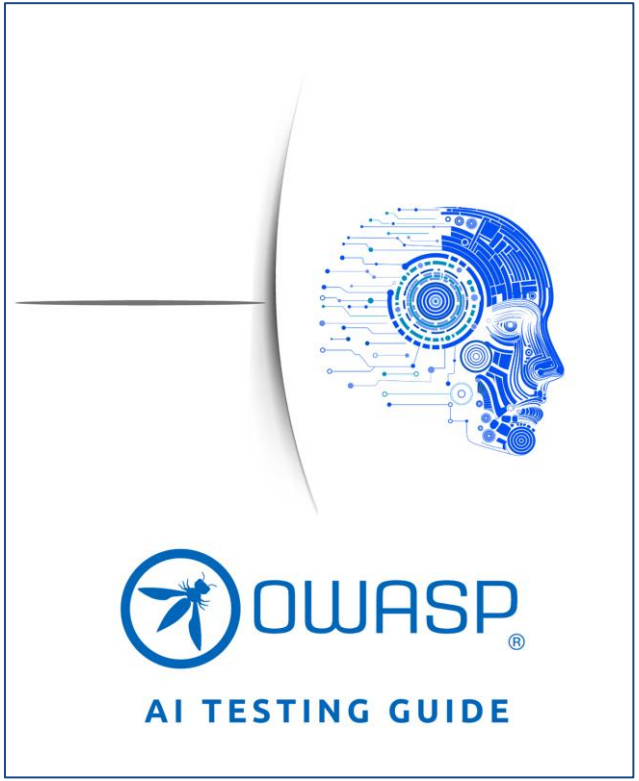


OWASP AI Testing Guide

The OWASP AI Testing Guide is designed to serve as a comprehensive reference for software developers, architects, data analysts, researchers, and risk officers, ensuring that AI risks are systematically addressed throughout the product development lifecycle.

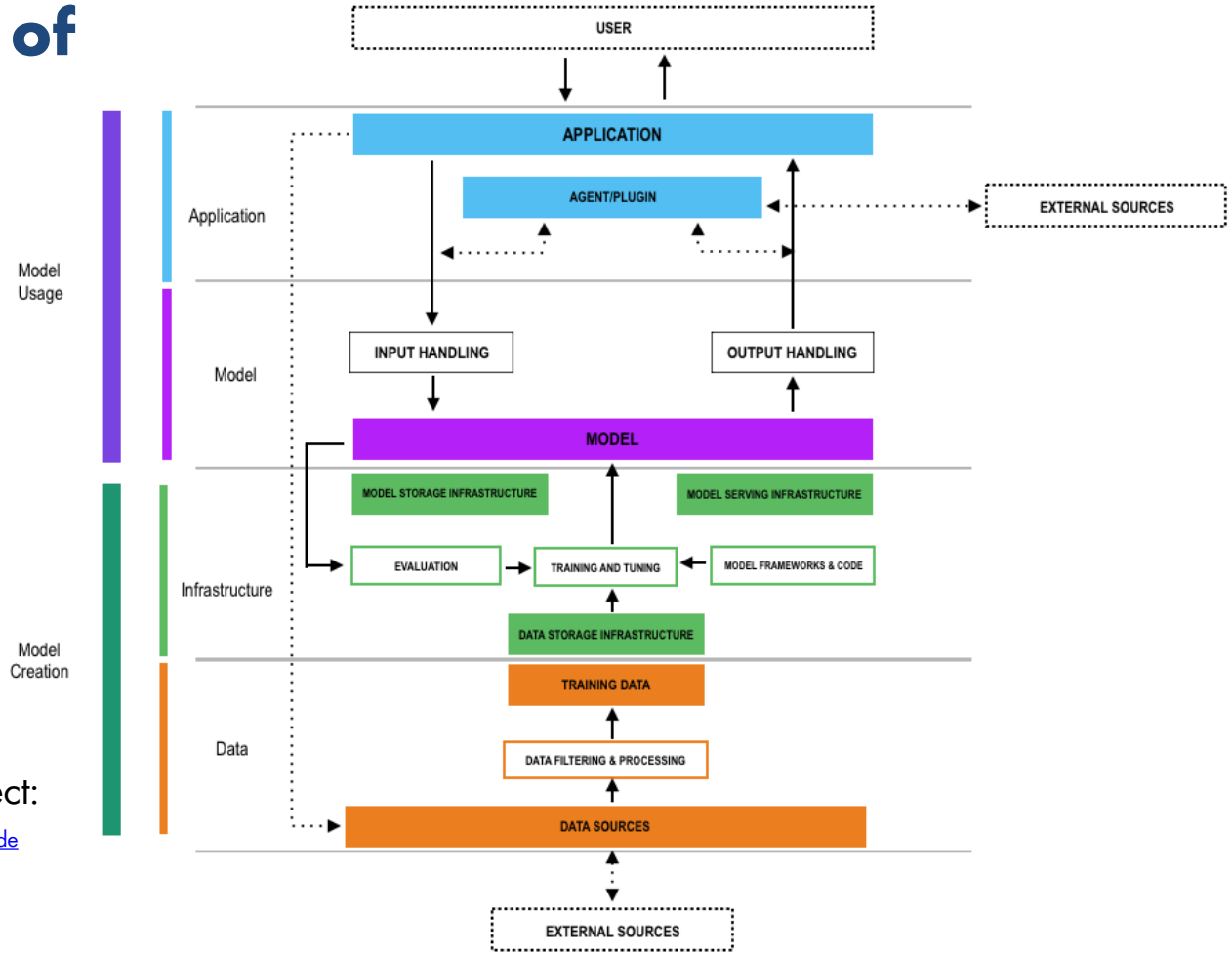
It outlines a robust suite of tests, ranging from data-centric validation and fairness assessments to adversarial robustness and continuous performance monitoring, that collectively provide documented evidence of risk validation and control.

By following this guidance, teams can establish the level of trust required to confidently deploy AI systems into production, with verifiable assurances that potential biases, vulnerabilities, and performance degradations have been proactively identified and mitigated.



https://owasp.org/www-project-ai-testing-guide/

Threat Modeling of AI Architecture



OWASP AI Testing Guide Project:

https://github.com/OWASP/www-project-ai-testing-guide

Putting all the threats together

AI Application Testing

Test ID	Threat Name	Source	Link	Test Name
AITG-APP-01	Prompt Injection	OWASP Top 10 LLM 2025	LLM01	Testing for Prompt Injection
AITG-APP-02	Indirect Prompt Injection	OWASP Top 10 LLM 2025	LLM01	Testing for Indirect Prompt Injection
AITG-APP-03	Sensitive Information Disclosure	OWASP Top 10 LLM 2025	LLM02	Testing for Sensitive Data Leak
AITG-APP-04	Leak Sensitive Input Data	OWASP Top 10 LLM 2025	LLM02	Testing for Input Leakage
AITG-APP-05	Improper Output Handling	OWASP Top 10 LLM 2025	LLM05	Testing for Unsafe Outputs
AITG-APP-06	Excessive Agency	OWASP Top 10 LLM 2025	LLM06	Testing for Agentic Behavior Limits
AITG-APP-07	System Prompt Leakage	OWASP Top 10 LLM 2025	LLM07	Testing for Prompt Disclosure
AITG-APP-08	Vector & Embedding Weaknesses	OWASP Top 10 LLM 2025	LLM08	Testing for Embedding Manipulation
AITG-APP-09	Model Theft Through Use	OWASP AI Exchange	link	Testing for Model Extraction
AITG-APP-10	Misinformation	OWASP Top 10 LLM 2025 - Responsible AI	LLM09	Testing for Harmful Content Bias
AITG-APP-11	Hallucinations	Trustworthy AI	—	Testing for Hallucinations
AITG-APP-12	Toxic Content Generation	Responsible AI	—	Testing for Toxic Output

Identify the set of tests

AI Application Testing

Test ID	Test Name & Link	Threat Source	Domain(s)
AITG-APP-01	Testing for Prompt Injection	OWASP Top 10 LLM 2025	Security
AITG-APP-02	Testing for Indirect Prompt Injection	OWASP Top 10 LLM 2025	Security
AITG-APP-03	Testing for Sensitive Data Leak	OWASP Top 10 LLM 2025	Security, Privacy
AITG-APP-04	Testing for Input Leakage	OWASP Top 10 LLM 2025	Security, Privacy
AITG-APP-05	Testing for Unsafe Outputs	OWASP Top 10 LLM 2025	Security, RAI
AITG-APP-06	Testing for Agentic Behavior Limits	OWASP Top 10 LLM 2025	Security, Trustworthy AI
AITG-APP-07	Testing for Prompt Disclosure	OWASP Top 10 LLM 2025	Security, Privacy
AITG-APP-08	Testing for Embedding Manipulation	OWASP Top 10 LLM 2025	Security
AITG-APP-09	Testing for Model Extraction	OWASP AI Exchange	Security
AITG-APP-10	Testing for Harmful Content Bias	OWASP Top 10 LLM 2025	RAI
AITG-APP-11	Testing for Hallucinations	Trustworthy AI	Trustworthy AI
AITG-APP-12	Testing for Toxic Output	Responsible AI	RAI
AITG-APP-13	Testing for Over-Reliance on AI	Responsible AI	RAI, Trustworthy AI
AITG-APP-14	Testing for Explainability and Interpretability	Responsible AI	RAI, Trustworthy AI

https://github.com/MatOwasp/AI-Testing-Guide/blob/main/Document/content/3.AITG-Framework.md

AI Model Testing

Test ID	Test Name & Link	Threat Source	Domain(s)
AITG-MOD-01	Testing for Evasion Attacks	OWASP AI Exchange	Security
AITG-MOD-02	Testing for Runtime Model Poisoning	OWASP Top 10 LLM 2025	Security
AITG-MOD-03	Testing for Poisoned Training Sets	OWASP Top 10 LLM 2025	Security
AITG-MOD-04	Testing for Membership Inference	OWASP AI Exchange	Privacy
AITG-MOD-05	Testing for Inversion Attacks	OWASP AI Exchange	Privacy
AITG-MOD-06	Testing for Robustness to New Data	Trustworthy AI	Trustworthy AI
AITG-MOD-07	Testing for Goal Alignment	Trustworthy AI	Trustworthy AI

AI Infrastructure Testing

Test ID	Test Name & Link	Threat Source	Domain(s)
AITG-INF-01	Testing for Supply Chain Tampering	OWASP Top 10 LLM 2025	Security
AITG-INF-02	Testing for Resource Exhaustion	OWASP Top 10 LLM 2025	Security
AITG-INF-03	Testing for Plugin Boundary Violations	Trustworthy AI	Trustworthy AI
AITG-INF-04	Testing for Capability Misuse	Responsible AI	RAI, Trustworthy AI
AITG-INF-05	Testing for Fine-tuning Poisoning	OWASP Top 10 LLM 2025	Security
AITG-INF-06	Testing for Dev-Time Model Theft	OWASP AI Exchange	Security, Privacy

AI Data Testing

Test ID	Test Name & Link	Threat Source	Domain(s)
AITG-DAT-01	Testing for Training Data Exposure	OWASP AI Exchange	Privacy
AITG-DAT-02	Testing for Runtime Exfiltration	OWASP AI Exchange	Security, Privacy
AITG-DAT-03	Testing for Dataset Diversity & Coverage	Responsible AI	RAI
AITG-DAT-04	Testing for Harmful Content in Data	Responsible AI	RAI
AITG-DAT-05	Testing for Data Minimization & Consent	Trustworthy AI	Privacy, Trustworthy AI

OWASP AI Maturity Assessment

In recent months, several AI Maturity Models have emerged, including the MITRE AI Framework, which highlights the need for structured AI assessments. Building on this momentum, we are developing the OWASP AI Maturity Assessment (AIMA), using the Software Assurance Maturity Model (SAMM) as a foundation.

The AI Maturity Assessment (AIMA) project goal is to empower organizations to navigate the complexities of artificial intelligence by providing a structured framework for making informed decisions about acquiring or developing AI systems.

AIMA helps to evaluate AI systems' alignment with ethical principles, security standards, and operational goals while mitigating risks and ensuring long-term sustainability.

https://owasp.org/www-project-ai-maturity-assessment/

https://github.com/OWASP/www-project-ai-maturity-assessment/blob/main/DRAFT/README.md

OWASP®

AI MATURITY ASSESSMENT

9 Practices

OWASP AI Maturity Assessment is like a health check and improvement guide for AI systems. It helps organizations make sure their AI is designed, built, and managed responsibly – not just working well, but also fair, safe, and respecting people's rights. It focuses on eight important areas:

1. Responsible AI: Making sure AI is fair, transparent, and respects human values.
2. Governance: Having clear strategies, policies, and training about AI risks.
3. Data Management: Ensuring high-quality, safe, and well-governed data for AI.
4. Privacy: Protecting personal data and giving users control over their information.
5. Design: Planning AI systems to be secure and resilient from the start.
6. Implementation: Building and deploying AI securely and responsibly.
7. Verification: Testing AI systems thoroughly to catch issues early.
8. Operations: Monitoring AI after deployment, managing incidents, and maintaining system health.

3 Maturity Levels

How does AIMA work?

- It has three levels that show how mature or advanced an organization's AI practices are.
- Organizations can take self-assessments with simple questions to figure out where they stand (no maturity - 0, initial maturity - 0.33, partial maturity - 0.66, full maturity - 1).
- Scores help highlight where improvements are needed and create a clear plan to get better step-by-step.
- It's designed for different roles—tech teams, managers, legal, auditors—so they all can understand and contribute.

Why use AIMA?

- Helps build AI systems people can trust.
- Keeps companies ready for new laws and regulations.
- Reduces risks like unfair bias, privacy problems, or security attacks.
- Supports ongoing learning and improvement as AI technology and risks evolve.
- Encourages collaboration across teams with clear roles and responsibilities.

Performing the assessment

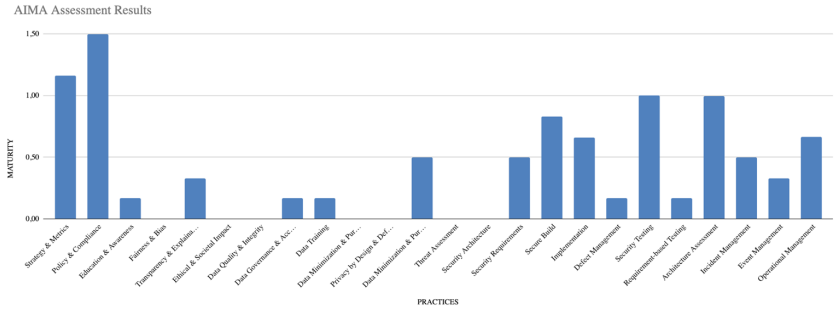
- Understand the practices
- Understand the questions

OWASP AI Maturity Assessment (AIMA) Questions

Instructions: Answer each question with 0 (meaning no maturity) 0,33 (initial maturity), 0,66 (not full maturity) 1 (maturity). The maturity level for each practice area will be calculated based on your answers.

PRACTICE AREA	STREAM	MATURITY LEVEL	QUESTIONS	ANSWERS	MATURITY
			Strategy & Metrics		
Governance	Strategy & Metrics - Stream A	1	Is there an initial AI strategy documented, even informally?	0	1,16
	Strategy & Metrics - Stream B	1	Are there any metrics informally tracked related to AI initiatives?	0,33	
	Strategy & Metrics - Stream A	2	Has the AI strategy been formally defined and communicated to stakeholders?	0	
	Strategy & Metrics - Stream B	2	Are defined metrics regularly reviewed and communicated within the organization?	0,66	
	Strategy & Metrics - Stream A	3	Is the AI strategy integrated into the organization's broader business strategy and continuously improved?	0,33	
	Strategy & Metrics - Stream B	3	Are metrics systematically analyzed to drive improvements and decision-making processes?	1	
			Policy & Compliance		
	Policy & Compliance - Stream A	1	Is there an awareness or initial informal policy for AI usage within the organization?	0	1,50
	Policy & Compliance - Stream B	1	Is there basic awareness of compliance needs relevant to AI (e.g., GDPR, ethical guidelines)?	1	
	Policy & Compliance - Stream A	2	Has a formal AI policy been established and clearly communicated to all relevant stakeholders?	0,33	
	Policy & Compliance - Stream B	2	Are compliance requirements identified, documented, and regularly reviewed to ensure alignment with AI-specific regulations?	0	
	Policy & Compliance - Stream A	3	Is the AI policy consistently enforced and reviewed regularly for relevance, accuracy, and alignment with organizational goals and external standards?	0,66	
	Policy & Compliance - Stream B	3	Is compliance management systematically integrated into daily operations, with proactive management of compliance risks and regular audits?	1	
			Education & Awareness		
	Education & Awareness - Stream A	1	Is there initial informal training or general awareness about AI security risks within the organization?	To evaluate	0,17
	Education & Awareness - Stream B	1	Is communication about AI security risks sporadic or ad hoc?	0,33	
	Education & Awareness - Stream A	2	Are formal training programs on AI security established, targeting key stakeholders and teams?	To evaluate	
	Education & Awareness - Stream B	2	Is there regular communication about AI security best practices and updates across the organization?	To evaluate	
	Education & Awareness - Stream A	3	Are AI security training programs regularly updated, mandatory, and effectively tailored for different roles and responsibilities?	To evaluate	
	Education & Awareness - Stream B	3	Is there an established culture of proactive communication, continuous awareness, and engagement around AI security throughout the organization?	To evaluate	

- Perform the assessment using the Toolkit



ASSESSMENT RESULTS

PRACTICE AREA	PRACTICES	MATURITY
Governance	Strategy & Metrics	1,16
	Policy & Compliance	1,50
	Education & Awareness	0,17
Responsible AI Principles	Fairness & Bias	0,00
	Transparency & Explainability	0,33
	Ethical & Societal Impact	0,00
Data Management	Data Quality & Integrity	0,00
	Data Governance & Accountability	0,17
	Data Training	0,17
Privacy	Data Minimization & Purpose Limitation	0,00
	Privacy by Design & Default	0,00
	Data Minimization & Purpose Limitation	0,50
Design	Threat Assessment	0,00
	Security Architecture	0,00
	Security Requirements	0,50

AIMA_Assessment_Toolkit v1.0.1

OWASP AI Maturity Assessment (AIMA) Questions

Instructions: Answer each question with 0 (meaning no maturity) 0,33 (initial maturity), 0,66 (not full maturity) 1 (maturity). The maturity level for each practice area will be calculated based on your answers.

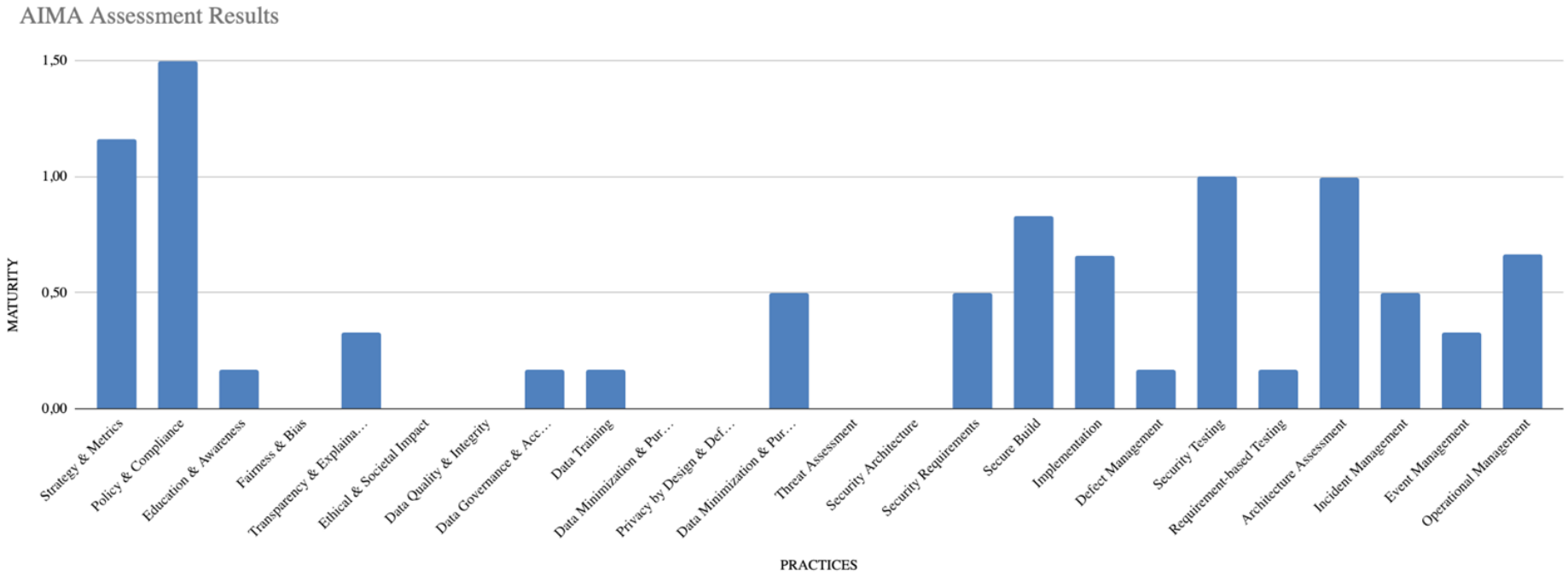
PRACTICE AREA	STREAM	MATURITY LEVEL	QUESTIONS	ANSWERS	MATURITY
			Strategy & Metrics		
Governance	Strategy & Metrics - Stream A	1	Is there an initial AI strategy documented, even informally?	0	1,16
	Strategy & Metrics - Stream B	1	Are there any metrics informally tracked related to AI initiatives?	0,33	
	Strategy & Metrics - Stream A	2	Has the AI strategy been formally defined and communicated to stakeholders?	0	
	Strategy & Metrics - Stream B	2	Are defined metrics regularly reviewed and communicated within the organization?	0,66	
	Strategy & Metrics - Stream A	3	Is the AI strategy integrated into the organization's broader business strategy and continuously improved?	0,33	
	Strategy & Metrics - Stream B	3	Are metrics systematically analyzed to drive improvements and decision-making processes?	1	
			Policy & Compliance		
	Policy & Compliance - Stream A	1	Is there an awareness or initial informal policy for AI usage within the organization?	0	1,50
	Policy & Compliance - Stream B	1	Is there basic awareness of compliance needs relevant to AI (e.g., GDPR, ethical guidelines)?	1	
	Policy & Compliance - Stream A	2	Has a formal AI policy been established and clearly communicated to all relevant stakeholders?	0,33	
	Policy & Compliance - Stream B	2	Are compliance requirements identified, documented, and regularly reviewed to ensure alignment with AI-specific regulations?	0	
	Policy & Compliance - Stream A	3	Is the AI policy consistently enforced and reviewed regularly for relevance, accuracy, and alignment with organizational goals and external standards?	0,66	
	Policy & Compliance - Stream B	3	Is compliance management systematically integrated into daily operations, with proactive management of compliance risks and regular audits?	1	
			Education & Awareness		
	Education & Awareness - Stream A	1	Is there initial informal training or general awareness about AI security risks within the organization?	To evaluate	0,17
	Education & Awareness - Stream B	1	Is communication about AI security risks sporadic or ad hoc?	0,33	
	Education & Awareness - Stream A	2	Are formal training programs on AI security established, targeting key stakeholders and teams?	To evaluate	
	Education & Awareness - Stream B	2	Is there regular communication about AI security best practices and updates across the organization?	To evaluate	
	Education & Awareness - Stream A	3	Are AI security training programs regularly updated, mandatory, and effectively tailored for different roles and responsibilities?	To evaluate	
	Education & Awareness - Stream B	3	Is there an established culture of proactive communication, continuous awareness, and engagement around AI security throughout the organization?	To evaluate	

AIMA assessment results

ASSESSMENT RESULTS

PRACTICE AREA	PRACTICES	MATURITY
Governance	Strategy & Metrics	1,16
	Policy & Compliance	1,50
	Education & Awareness	0,17
Responsible AI Principles	Fairness & Bias	0,00
	Transparency & Explainability	0,33
	Ethical & Societal Impact	0,00
Data Management	Data Quality & Integrity	0,00
	Data Governance & Accountability	0,17
	Data Training	0,17
Privacy	Data Minimization & Purpose Limitation	0,00
	Privacy by Design & Default	0,00
	Data Minimization & Purpose Limitation	0,50
Design	Threat Assessment	0,00
	Security Architecture	0,00

AI Maturity results



How “mature” is your AI outsourcer?

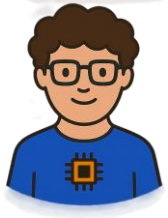
We implement HTTPS and we use our Guidelines!

UGLY!

Maturity Level

How “mature” is your AI outsourcer?

We implement HTTPS and we use our Guidelines!



Take a look at our internal assessment report!

BAD!

Maturity Level

How “mature” is your AI outsourcer?

We implement HTTPS and we use our Guidelines!

Take a look at our internal assessment report!

We did an OWASP AIMA Assessment, see the certificate of achievement!

Maturity Level

Takeaway: building Secure & Responsible AI

AI introduces new threats and new risks

- **New threats for development** → teams must test, validate, and secure AI systems in practice → **OWASP AI Testing Guide**
- **New risks for governance** → companies need awareness, policies, processes, and accountability → **OWASP AI Maturity Assessment**

OWASP AI Testing Guide (AITG)

- Test and validate your own software against evolving threats

OWASP AI Maturity Assessment (AIMA)

- Establish and manage awareness on AI inside your company

References

https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3

https://www.weforum.org/publications/the-future-of-jobs-report-2023/

Synapsed LLM Study White Paper: https://synapsed.ai/rd-owasp-top-10-llm-2025-a-synapsed-research-study/

Trustworthy AI

OWASP AI Testing Guide: https://owasp.org/www-project-ai-testing-guide/

OWASP AI Maturity Assessment: https://owasp.org/www-project-ai-maturity-assessment/

OWASP GenAI Security: https://genai.owasp.org/

AI Standards

ISO/IEC 42001:2023 Information technology – Artificial intelligence – Management system: https://www.iso.org/standard/81230.html

ISO/IEC 5338:2023 Information technology - Artificial intelligence - AI system life cycle processes: https://www.iso.org/standard/81118.html

EU AI Act Requirements for Providers of High Risk AI Systems: https://colab.research.google.com/github/mrwadams/ai-act-requirements-graph/blob/main/AI_Act_Requirements_Graph.ipynb

NIST - Artificial Intelligence Risk Management Framework (AI RMF 1.0): https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

GRAZIE!

CONTATTI PER DOMANDE/APPROFONDIMENTI:

Mail: matteo.meucci@synapsed.ai

SYNAPSED