Preparation Guide

Edition 202606

Content

1. Overview

EXIN AI Security Professional based on the OWASP AI Exchange (AISP.EN)

Scope

The EXIN AI Security Professional based on the OWASP AI Exchange certification confirms that the professional knows how to keep AI systems secure during design, development, deployment, and daily use in line with organizational policies, standards, and applicable laws.

This certification includes the following topics:

- AI security in the organization
- AI security threats
- AI security controls
- AI security testing
- Privacy and compliance in AI security

Summary

Artificial Intelligence (AI) powers decisions, content, and automation across every sector nowadays, expanding opportunities but also increasing the attack surface. As a result, organizations deal with fast-evolving risks, including attacks that involve AI systems corruption, misuse of models, or vulnerabilities linked across tools and vendors, while privacy and legal requirements grow stricter. In this landscape, security cannot be an afterthought; it must be built into how AI is designed, deployed, and monitored.

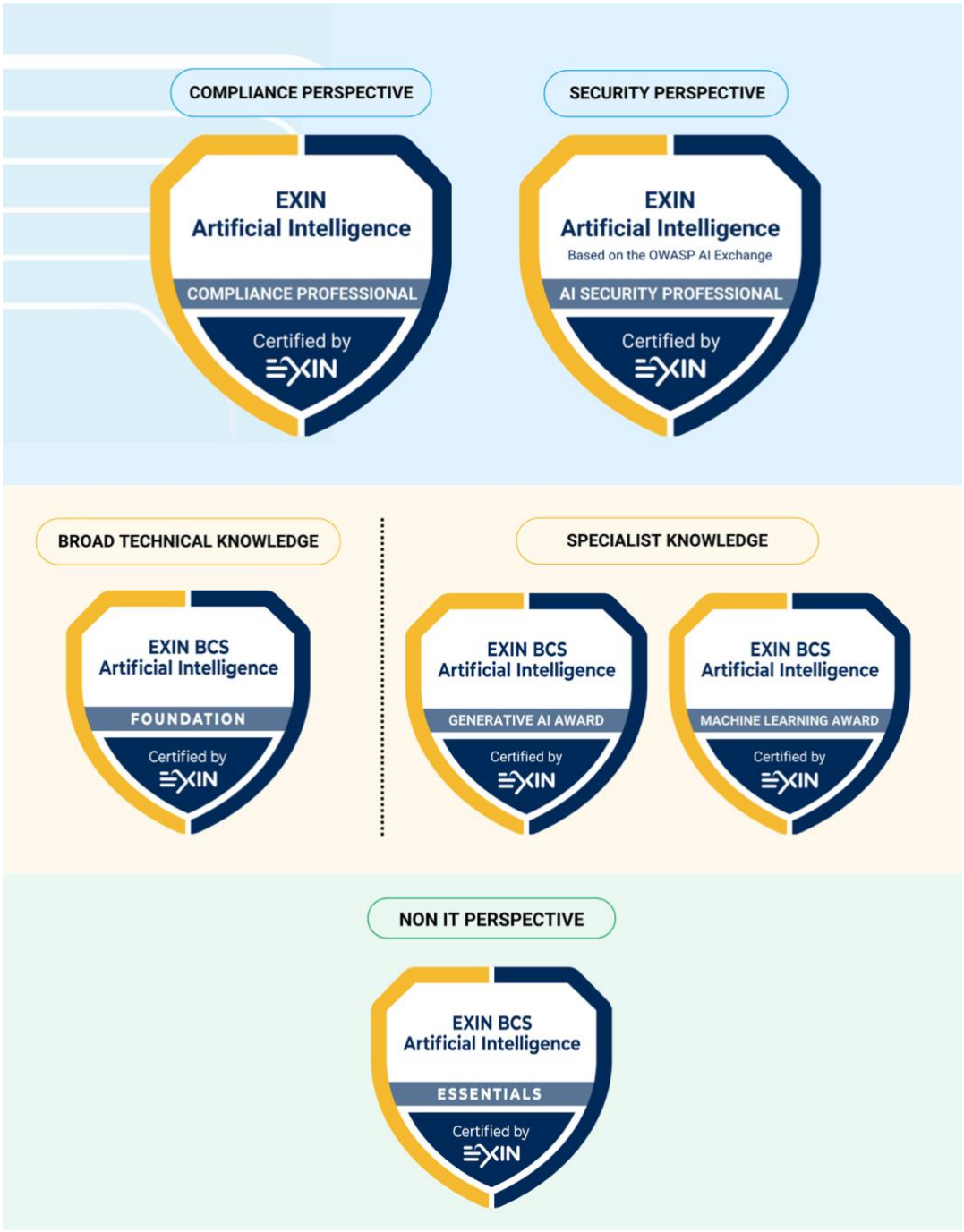
In the past few years, AI security has emerged as a critical priority. The OWASP AI Exchange defines security as preventing unauthorized access, use, disclosure, disruption, modification, or destruction, where 'modification' includes manipulating the behavior of an AI model in unwanted ways. As AI becomes increasingly integrated into decision-making processes, ensuring its reliability and integrity is essential for fostering trust and minimizing risks in this complex security environment.

The EXIN AI Security Professional based on the OWASP AI Exchange certification prepares professionals to meet this challenge. It is based on the OWASP AI Exchange, an open, community-driven knowledge base of AI risks, threats, controls, and testing strategies. The certification offers a path for professionals to better understand AI security in the organization, key threat categories, effective security controls, AI security testing, and privacy and compliance across the AI lifecycle.

The AISP certification is relevant for professionals from diverse backgrounds and is particularly appealing to those working in or interested in roles related to security and privacy, (AI) engineering, AI testing, governance, and compliance. Holders of this certification understand how to design, evaluate, and improve AI security programs that are resilient, auditable, and privacy-aware, thereby strengthening day-to-day work.

Context

The EXIN AI Security Professional based on the OWASP AI Exchange certification is part of the EXIN Artificial Intelligence qualification program.



Target group

The EXIN AI Security Professional based on the OWASP AI Exchange certification is suitable for a diverse group of professionals working with AI systems, as well as those focusing on its security and privacy. These roles require competent level to enable organizations to conceive, develop, acquire, operate, and maintain applications that can be trusted.

This includes, but is not limited to:

- AI (cyber)security professionals
- Data protection and privacy professionals
- AI developers
- AI architects
- AI DevOps engineers
- AI operations professionals
- AI auditors
- Machine learning (security) engineers
- AI software/infrastructure engineers
- (Cyber) security analysts
- Data engineers
- Quality & evaluation specialists
- Security testers
- QA engineers
- Red Team testers
- Model validation engineers
- Governance, risk and compliance officers

Requirements for certification

- Successful completion of the EXIN AI Security Professional based on the OWASP AI Exchange exam.

Training is highly recommended. Knowledge of AI terminology, for instance through the EXIN BCS Artificial Intelligence Essentials certification or the EXIN BCS Artificial Intelligence Foundation certification, is highly recommended.

Examination details

Examination type:	Multiple-choice questions
Number of questions:	40
Pass mark:	65% (26/40 questions)
Open book:	No
Notes:	No
Electronic equipment/aides permitted:	No
Exam duration:	90 minutes

The Rules and Regulations for EXIN's examinations apply to this exam.

Bloom level

The EXIN AI Security Professional based on the OWASP AI Exchange certification tests candidates at Bloom levels 2 and 3 according to Bloom's revised taxonomy:

- Bloom level 2: Understanding – a step beyond remembering. Understanding shows that candidates comprehend what is presented and can evaluate how the learning material may be applied in their own environment. This type of questions aims to demonstrate that the candidate is able to organize, compare, interpret and choose the correct description of facts and ideas.
- Bloom level 3: Application – shows that candidates have the ability to make use of information in a context different from the one in which it was learned. This type of questions aims to demonstrate that the candidate is able to solve problems in new situations by applying acquired knowledge, facts, techniques and rules in a different, or new way. These questions usually contain a short scenario.

Training

Contact hours

The recommended number of contact hours for this training course is 14. This includes practical group assignments, exam preparation and short breaks. This number of hours does not include lunch breaks, homework and the exam.

Indication study effort

84 hours (3 ECTS), depending on existing knowledge.

Training organization

You can find a list of our Accredited Training Organizations at www.exin.com.

2. Exam requirements

The exam requirements are specified in the exam specifications. The following table lists the topics of the module (exam requirements) and the subtopics (exam specifications).

Exam requirements	Exam specifications	Weight
1. AI security in the organization		**15%**
	1.1 Organizing AI security	10%
	1.2 Threat modeling and agentic AI risks	5%
2. AI security threats		**37.5%**
	2.1 Input threats	17.5%
	2.2 Development-time threats	10%
	2.3 Runtime conventional security threats	10%
3. AI security controls		**27.5%**
	3.1 Governance	12.5%
	3.2 Limiting sensitive data	7.5%
	3.3 Limiting unwanted behavior	7.5%
4. AI security testing		**7.5%**
	4.1 Threats scope	5%
	4.2 AI security testing strategies	2.5%
5. Privacy and compliance in AI security		**12.5%**
	5.1 Privacy and AI security	5%
	5.2 Compliance and regulation	7.5%
		100%

Exam specifications

1 AI security in the organization

1.1 Organizing AI security

The candidate can…

1.1.1 apply the five G.U.A.R.D steps to organize AI security within an organization.

1.1.2 understand how to facilitate the responsible and trustworthy use of AI in the organization.

1.1.3 differentiate AI security from conventional (cyber)security.

1.1.4 outline AI-specific assets and their key threats.

1.2 Threat modeling and agentic AI risks

The candidate can…

1.2.1 outline risk management steps to conduct threat modeling with controls for AI security.

1.2.2 explain typical risks related to agentic AI.

2 AI security threats

2.1 Input threats

The candidate can…

2.1.1 illustrate the various types of evasion based on the attackers' access to knowledge: zero-knowledge evasion, perfect-knowledge evasion, partial-knowledge evasion, transfer attack, and evasion after poisoning.

2.1.2 distinguish between direct prompt injection and indirect prompt injection.

2.1.3 illustrate the seven layers of prompt injection protection.

2.1.4 explain how the disclosure of sensitive data may occur in model outputs, model inversion and membership inference.

2.1.5 explain how model exfiltration occurs and applicable countermeasures.

2.1.6 identify input-driven resource exhaustion threats.

2.2 Development-time threats

The candidate can…

2.2.1 identify data poisoning risks during development-time.

2.2.2 explain direct development-time model poisoning.

2.2.3 explain supply-chain model poisoning.

2.2.4 interpret different sensitive data leaks development-time: development-time data leak, direct development-time model leak, and source code/configuration leak.

2.3 Runtime conventional security threats

The candidate can…

2.3.1 explain how conventional runtime security threats may affect AI components.

2.3.2 illustrate direct runtime model poisoning and direct runtime model leak.

2.3.3 describe the security threats associated with injection embedded in AI outputs and input data leak.

2.3.4 explain how augmentation data leak and manipulation constitute security threats.

3 AI security controls

3.1 Governance

The candidate can…

3.1.1 apply general governance controls for AI security oversight.

3.1.2 explain general governance controls coverage.

3.1.3 understand how the implementation of controls is divided between a third-party AI model provider and the AI system developer/deployer (ready-made model).

3.2 Limiting sensitive data
The candidate can…
3.2.1 implement controls to limit sensitive data in order to enhance confidentiality and integrity.
3.2.2 understand how limiting sensitive data reduces risks.
3.3 Limiting unwanted behavior
The candidate can…
3.3.1 implement controls to limit the effects of unwanted model behavior.
3.3.2 understand how limiting unwanted model behavior benefits performance and reduces risk.

4 AI security testing

4.1 Threats scope
The candidate can…
4.1.1 explain why AI security testing is important and its scope.
4.1.2 understand which threats to test for beyond conventional security testing when using predictive or generative AI.
4.2 AI security testing strategies
The candidate can…
4.2.1 describe AI security testing strategies by using the general testing approach.

5 Privacy and compliance in AI security

5.1 Privacy and AI security
The candidate can…
5.1.1 explain privacy and privacy concerns associated with AI systems.
5.1.2 apply privacy principles for AI systems to a given scenario.
5.2 Compliance and regulation
The candidate can…
5.2.1 explain how the ISO/IEC 23894, ISO/IEC 27005, ISO/IEC 42001 and ISO/IEC 5338 standards support an organization in implementing controls and processes that comply with applicable AI regulations.
5.2.2 understand the challenges associated with compliance of AI systems to relevant laws and regulations, including the AI Act and the GDPR.
5.2.3 describe strategies to mitigate the copyright-infringement risks and support compliance.

3. List of basic concepts

This chapter contains the terms and abbreviations with which candidates should be familiar.

Please note that knowledge of these terms alone does not suffice for the exam. The candidate must understand the concepts and be able to provide examples.

accountability
adversarial attack
adversarial testing
agentic AI
AI (artificial intelligence)
AI agent
AI assets
AI management system (AIMS)
AI model
AI model provider
AI privacy principles
- accuracy
- consent
- data minimization and storage limitation
- fairness and lawfulness
- privacy rights or privacy by design
- security and safeguards
- transparency and explainability
- use limitation and purpose specification

AI red teaming
AI security
AI security governance
AI security testing approach
AI system
AI system provider
anonymization
API (application programming interface)
audit
automated oversight
automation
bias
blast radius
chatbot
Chief AI Officer (CAIO)
Chief Information Security Officer (CISO)
code
compliance
confidentiality
configuration leak
consent
copyright
data accuracy
data leak
data protection
data protection impact assessment (DPIA)
data security
data subject
denial-of-service (DoS)
denial-of-wallet (DoW)
deployer
development-time
development-time security threats
- data poisoning development-time
- model poisoning development-time
- sensitive data leak development-time

embedding
encryption
ethical principles
European Union (EU)
fairness
G.U.A.R.D (Govern, Understand, Adapt, Reduce, Demonstrate)
General Data Protection Regulation (GDPR)
generative AI
hijacking
human oversight
inference-time
information security management system (ISMS)
input data
input security threats
- evasion
- model exfiltration
- prompt injection
- resource exhaustion
- sensitive data disclosure through use

integrity verification
intellectual property (IP)
large language model (LLM)
leakage
machine learning
membership inference
model architecture
model inversion
model leak
open source
output data
output sanitization
output validation
personal data

predictive AI
privacy
privacy by default
privacy by design
privacy impact assessments (PIA)
privacy incident
processing (or personal data)
prompt
pseudonymization
publicly accessible
ready-made model
residual risk
responsible AI
retrieval-augmented generation (RAG)
risk
risk analysis
risk levels for AI systems based on EU AI Act
- minimal or no risk
- limited risk
- high risk
- unacceptable risk

risk management
risk mitigation
risk rating
runtime
runtime security threats
- augmentation data manipulation
- direct augmentation data leak
- direct runtime model leak
- direct runtime model poisoning
- input data leak
- output containing conventional injection

safety of an AI system
security breach
security controls
- controls for sensitive data limitation
- controls to limit the effect of unwanted behavior
- governance controls

security incident
security risks
security testing
security threats
(security) attack
sensitive data
service provider
source code
source code leak
sponge attack
stakeholder
startup
structured query language (SQL)
tampering
third party
threat modeling
token
tokenization
tool
training data
trustworthy AI
unwanted behavior
vulnerability

4. Literature

Exam literature

The knowledge required for the exam is covered in the following literature:

A. AI Security Community
OWASP AI Exchange
Open-source publication (April 2026)
This material is freely available at https://owaspai.org/.
To download the PDF version used for the certification (April 2026), go to www.exin.com. Click on 'Professionals' and then on 'Certifications' to find the certification. The free download can be found under 'Required reading'.

Additional literature

B. European Union
Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonized rules on artificial intelligence (Artificial Intelligence Act).
European Union (2024)
https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

C. ISO/IEC 23894:2023 (EN)
Information technology — Artificial intelligence — Guidance on risk management
Switzerland, ISO/IEC (2023)
https://www.iso.org/home.html

D. ISO/IEC 27005:2022 (EN)
Information security, cybersecurity and privacy protection — Guidance on managing information security risks
Switzerland, ISO/IEC (2022)
https://www.iso.org/home.html

E. ISO/IEC 42001:2023 (EN)
Information technology — Artificial intelligence — Management system
Switzerland, ISO/IEC (2023)
https://www.iso.org/home.html

F. ISO/IEC 5338:2023 (EN)
Information technology — Artificial intelligence — AI system life cycle processes
Switzerland, ISO/IEC (2023)
https://www.iso.org/home.html

Comment

Additional literature is for reference and depth of knowledge only.

Literature matrix

Exam requirements	Exam specifications	Reference
1. AI security in the organization		
	1.1 Organizing AI security	A, Chapter 0
	1.2 Threat modeling and agentic AI risks	A, Chapter 0
2. AI security threats		
	2.1 Input threats	A, Chapter 2
	2.2 Development-time threats	A, Chapter 3
	2.3 Runtime conventional security threats	A, Chapter 4
3. AI security controls		
	3.1 Governance	A, Chapter 0, 1
	3.2 Limiting sensitive data	A, Chapter 1
	3.3 Limiting unwanted behavior	A, Chapter 1
4. AI security testing		
	4.1 Threats scope	A, Chapter 5
	4.2 AI security testing strategies	A, Chapter 5
5. Privacy and compliance in AI security		
	5.1 Privacy and AI security	A, Chapter 6
	5.2 Compliance and regulation	A, Chapter 0, 1, 2, 3, 6

Certified for what's next

Contact EXIN

www.exin.com