

Guide de préparation

Édition 202611

Copyright © EXIN Holding B.V. 2026. All rights reserved.
EXIN® is a registered trademark of EXIN.

No part of this publication may be reproduced, stored, utilized or transmitted in any form or by any means, electronic, mechanical, or otherwise, without the prior written permission from EXIN.

Table des matières

1. Résumé 4
2. Conditions de l'examen 8
3. Liste des concepts de base 11
4. Bibliographie 15
5. Parcours de carrière 17

1. Résumé

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

Portée

La certification EXIN AI Security Professional based on the OWASP AI Exchange atteste que le professionnel sait comment assurer la sécurité des systèmes d'IA lors de leur conception, de leur développement, de leur déploiement et de leur utilisation quotidienne, conformément aux politiques de l'organisation, aux normes et à la législation en vigueur.

Cette certification couvre les thèmes suivants :

- Sécurité de l'IA au sein de l'organisation
- Menaces pour la sécurité de l'IA
- Contrôles de sécurité de l'IA
- Tests de sécurité de l'IA
- Protection de la vie privée et conformité en matière de sécurité de l'IA

Sommaire

L'intelligence artificielle (IA) est aujourd'hui au cœur des décisions, des contenus et de l'automatisation dans tous les secteurs, ce qui élargit les opportunités mais augmente également la surface d'attaque. En conséquence, les organisations sont confrontées à des risques en constante évolution, notamment des attaques impliquant la corruption de systèmes d'IA, l'utilisation abusive de modèles ou des vulnérabilités liées aux outils et aux fournisseurs, tandis que les exigences en matière de protection de la vie privée et les obligations légales se renforcent. Dans ce contexte, la sécurité ne peut être une considération secondaire ; elle doit être intégrée à la conception, au déploiement et à la surveillance de l'IA.

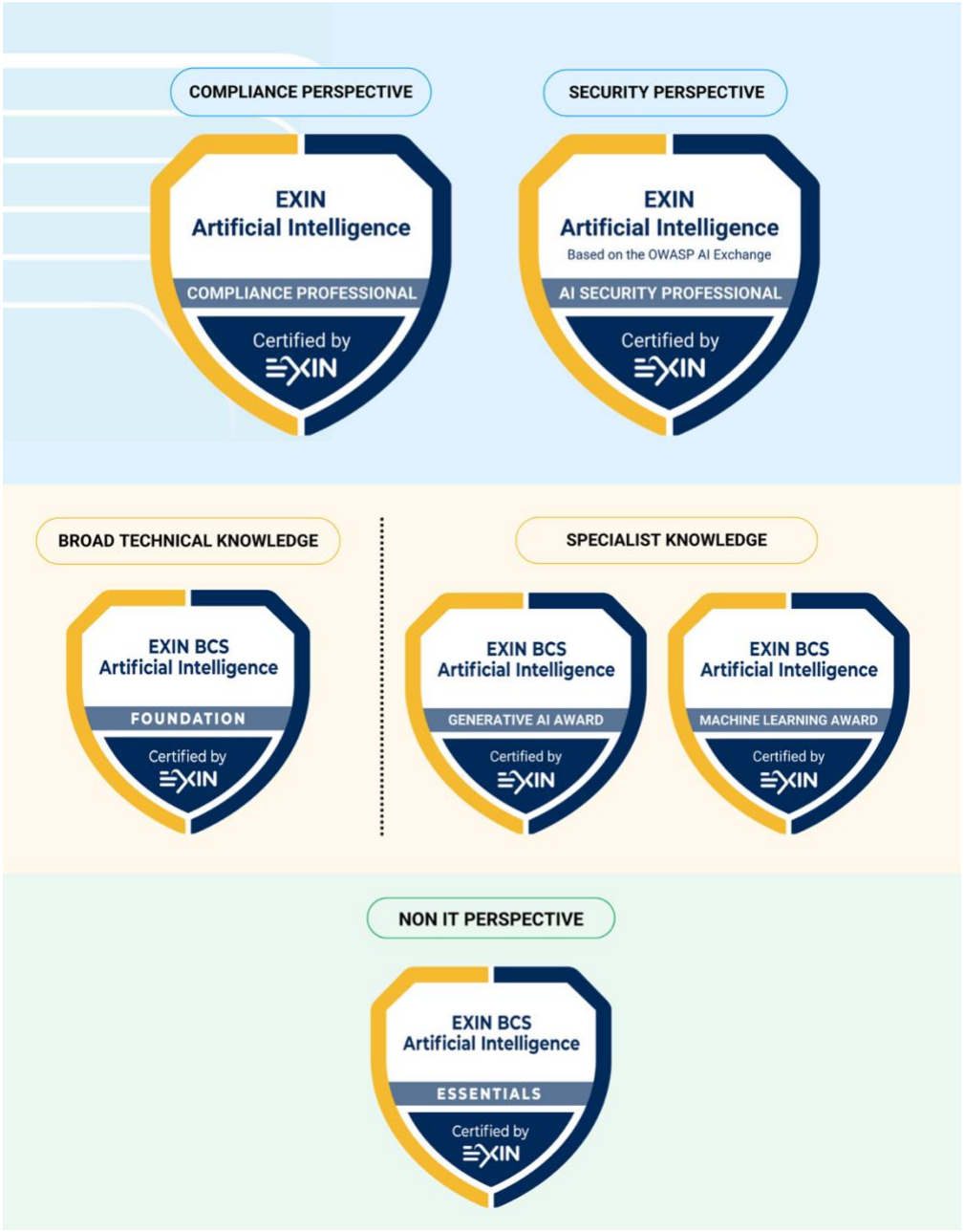
Au cours des dernières années, la sécurité de l'IA est devenue une priorité absolue. L'OWASP AI Exchange définit la sécurité comme la prévention de tout accès, utilisation, divulgation, perturbation, modification ou destruction non autorisés, la « modification » incluant la manipulation du comportement d'un modèle d'IA de manière indésirable. À mesure que l'IA s'intègre de plus en plus aux processus décisionnels, il est essentiel de garantir sa fiabilité et son intégrité pour renforcer la confiance et minimiser les risques dans cet environnement de sécurité complexe.

La certification EXIN AI Security Professional based on the OWASP AI Exchange prépare les professionnels à relever ce défi. Elle s'appuie sur l'OWASP AI Exchange, une base de connaissances ouverte et animée par la communauté, consacrée aux risques, menaces, contrôles de sécurité et stratégies de test liés à l'IA. Cette certification offre aux professionnels un parcours leur permettant de mieux comprendre la sécurité de l'IA au sein de l'organisation, les principales catégories de menaces, les contrôles de sécurité efficaces, les tests de sécurité de l'IA, ainsi que la protection de la vie privée et la conformité tout au long du cycle de vie de l'IA.

La certification AISP s'adresse à des professionnels issus d'horizons divers et concerne tout particulièrement ceux qui occupent des postes liés à la sécurité et à la protection de la vie privée, à l'ingénierie (de l'IA), aux tests d'IA, à la gouvernance et à la conformité, ou qui s'y intéressent. Les titulaires de cette certification savent comment concevoir, évaluer et améliorer des programmes de sécurité de l'IA qui soient résilients, vérifiables et respectueux de la vie privée, renforçant ainsi leur travail au quotidien.

Contexte

La certification EXIN AI Security Professional based on the OWASP AI Exchange fait partie du programme de qualification EXIN Artificial Intelligence.



Groupe ciblé

La certification EXIN AI Security Professional based on the OWASP AI Exchange s'adresse à un large éventail de professionnels travaillant avec des systèmes d'IA, ainsi qu'à ceux qui se concentrent sur la sécurité de ces systèmes et la protection de la vie privée. Ces fonctions exigent un niveau de compétence permettant aux organisations de concevoir, développer, acquérir, exploiter et maintenir des applications fiables.

Cela inclut, sans s'y limiter :

- les professionnels de l'IA
- les professionnels de la (cyber)sécurité
- les professionnels de la protection des données et de la vie privée
- les ingénieurs en apprentissage automatique (sécurité)
- les ingénieurs data (data engineers)
- les spécialistes de la qualité et de l'évaluation
- les responsables de la gouvernance, des risques et de la conformité

Exigences pour être certifié

- Réussite à l'examen EXIN AI Security Professional based on the OWASP AI Exchange.

Il est fortement recommandé de suivre une formation. La maîtrise de la terminologie de l'IA, acquise par exemple grâce à la certification EXIN BCS Artificial Intelligence Essentials ou à la certification EXIN BCS Artificial Intelligence Foundation, est vivement recommandée.

Précisions sur l'examen

Type d'examen :	Questions à choix multiples
Nombre de questions :	40
Note minimale pour réussir :	65% (26/40 questions)
Livre ouvert :	Non
Notes :	Non
Outils / aides électroniques autorisés :	Non
Durée de l'examen :	90 minutes

Les règles et règlements de l'EXIN en matière d'examens s'appliquent à cet examen.

Niveau Bloom

La certification d'EXIN AI Security Professional based on the OWASP AI Exchange teste les candidats aux niveaux Bloom 2 et Bloom 3 de la taxonomie révisée de Bloom :

- Niveau 2 : Compréhension - une étape au-delà de la mémorisation. La compréhension indique que les candidats peuvent saisir ce qui est présenté et sont en mesure d'évaluer l'application de l'apprentissage à leur propre environnement. Ce type de questions vise à dé montrer que le candidat est en mesure d'organiser, de comparer, d'interpréter et de choisir la description adéquate de faits et d'idées.
- Niveau 3 : Application - indique que les candidats ont la capacité d'utiliser des informations dans un contexte différent de celui dans lequel elles ont été apprises. Ce type de questions vise à démontrer que le candidat est en mesure de résoudre des problèmes dans de nouvelles situations, en utilisant des connaissances acquises, des faits, des techniques et des règles d'une manière nouvelle ou différente. Les questions comportent généralement un bref scénario.

Formation

Heures de contact

Le nombre recommandé d'heures de contact pour cette formation est de 14 heures. Cela comprend les exercices pratiques ou les exercices de groupe, la préparation aux examens ainsi que les brèves pauses. Ce nombre d'heures n'inclut pas les pauses déjeuners, les devoirs, et l'examen.

Charge de travail estimée

84 heures (3 ECTS), en fonction de connaissances existantes.

Organisme de formation

Une liste d'organismes de formation accrédités est disponible sur le site internet d'EXIN : www.exin.com.

2. Conditions de l'examen

Les conditions de l'examen sont détaillées dans les spécifications de l'examen. Le tableau ci-dessous énumère les sujets du module (conditions de l'examen) et les sous-sujets (spécifications de l'examen).

Conditions de l'examen	Spécifications de l'examen	Pondération
1. Sécurité de l'IA au sein de l'organisation		**15%**
	1.1 Organisation de la sécurité de l'IA	10%
	1.2 Modélisation des menaces et risques liés à l'IA agentique	5%
2. Menaces pour la sécurité de l'IA		**37,5%**
	2.1 Menaces liées aux données d'entrée	17,5%
	2.2 Menaces liées à la phase de développement	10%
	2.3 Menaces de sécurité conventionnelles en phase d'exécution	10%
3. Contrôles de sécurité de l'IA		**27,5%**
	3.1 Gouvernance	12,5%
	3.2 Limitation des données sensibles	7,5%
	3.3 Limitation des comportements indésirables	7,5%
4. Tests de sécurité de l'IA		**7,5%**
	4.1 Portée des menaces	5%
	4.2 Stratégies de test de la sécurité de l'IA	2,5%
5. Protection de la vie privée et conformité en matière de sécurité de l'IA		**12,5%**
	5.1 Protection de la vie privée et sécurité de l'IA	5%
	5.2 Conformité et réglementation	7,5%
		100%

Spécifications de l'examen

1 Sécurité de l'IA au sein de l'organisation

1.1 Organisation de la sécurité de l'IA

Le candidat est capable de…

1.1.1 appliquer les cinq étapes du modèle G.U.A.R.D pour organiser la sécurité de l'IA au sein d'une organisation.

1.1.2 comprendre comment favoriser une utilisation responsable et digne de confiance de l'IA au sein de l'organisation.

1.1.3 distinguer la sécurité de l'IA de la (cyber)sécurité classique.

1.1.4 décrire les actifs liés à l'IA et les principales menaces qui pèsent sur eux.

1.2 Modélisation des menaces et risques liés à l'IA agentique

Le candidat est capable de…

1.2.1 décrire les étapes de la gestion des risques pour réaliser une modélisation des menaces avec des contrôles de sécurité visant à assurer la sécurité de l'IA.

1.2.2 expliquer les risques typiques liés à l'IA agentique.

2 Menaces pour la sécurité de l'IA

2.1 Menaces liées aux données d'entrée

Le candidat est capable de…

2.1.1 illustrer les différents types d'évasion en fonction du niveau de connaissance dont disposent les attaquants : évasion à connaissance nulle, évasion à connaissance parfaite, évasion à connaissance partielle, attaque par transfert et évasion après empoisonnement.

2.1.2 distinguer l'injection directe de prompt de l'injection indirecte de prompt.

2.1.3 illustrer les sept niveaux de protection contre l'injection de prompt.

2.1.4 expliquer comment la divulgation de données sensibles peut se produire dans les sorties des modèles, l'inversion de modèle et l'inférence d'appartenance.

2.1.5 expliquer comment se produit l'exfiltration du modèle.

2.1.6 identifier les menaces d'épuisement des ressources liées aux données d'entrée.

2.2 Menaces liées à la phase de développement

Le candidat est capable de…

2.2.1 identifier les risques d'empoisonnement des données pendant la phase de développement.

2.2.2 expliquer l'empoisonnement direct du modèle en phase de développement

2.2.3 expliquer l'empoisonnement de modèle dans la chaîne d'approvisionnement.

2.2.4 interpréter les différentes fuites de données sensibles (phase de développement) : fuite de données (phase de développement), fuite directe de modèle (phase de développement) et fuite du code source ou de la configuration.

2.3 Menaces de sécurité conventionnelles en phase d'exécution

Le candidat est capable de…

2.3.1 expliquer comment les menaces de sécurité conventionnelles en phase d'exécution peuvent affecter les composants d'IA.

2.3.2 illustrer l'empoisonnement direct du modèle en phase d'exécution et la fuite directe du modèle en phase d'exécution.

2.3.3 décrire les menaces de sécurité associées à l'injection intégrée dans les données de sortie de l'IA et à la fuite des données d'entrée.

2.3.4 expliquer en quoi les fuites de données d'augmentation et leur manipulation constituent des menaces de sécurité.

3 Contrôles de sécurité de l'IA

3.1 Gouvernance

Le candidat est capable de…

3.1.1 appliquer des contrôles de gouvernance généraux pour la supervision de la sécurité de l'IA.

3.1.2 expliquer la portée des contrôles de gouvernance généraux.

3.1.3 comprendre comment la mise en œuvre des contrôles est répartie entre un fournisseur tiers de modèle d'IA et le développeur/déployeur du système d'IA (modèle prêt à l'emploi).

3.2 Limitation des données sensibles

Le candidat est capable de…

3.2.1 mettre en œuvre des contrôles visant à limiter les données sensibles afin de renforcer la confidentialité et l'intégrité.

3.2.2 comprendre en quoi la limitation des données sensibles réduit les risques.

3.3 Limitation des comportements indésirables

Le candidat est capable de…

3.3.1 mettre en œuvre des contrôles visant à limiter les effets des comportements indésirables du modèle.

3.3.2 comprendre en quoi la limitation des comportements indésirables du modèle améliore les performances et réduit les risques.

4 Tests de sécurité de l'IA

4.1 Portée des menaces

Le candidat est capable de…

4.1.1 expliquer pourquoi les tests de sécurité de l'IA sont importants et en décrire la portée.

4.1.2 comprendre quelles menaces doivent être testées au-delà des tests de sécurité classiques lors de l'utilisation d'une IA prédictive ou générative.

4.2 Stratégies de test de la sécurité de l'IA

Le candidat est capable de…

4.2.1 décrire les stratégies de test de la sécurité de l'IA en s'appuyant sur l'approche générale des tests de sécurité.

5 Protection de la vie privée et conformité en matière de sécurité de l'IA

5.1 Protection de la vie privée et sécurité de l'IA

Le candidat est capable de…

5.1.1 expliquer la notion de protection de la vie privée et les préoccupations qui y sont liées dans les systèmes d'IA.

5.1.2 appliquer les appliquer les principes de protection de la vie privée relatifs aux systèmes d'IA à un scénario donné.

5.2 Conformité et réglementation

Le candidat est capable de…

5.2.1 expliquer en quoi les normes ISO/IEC 23894, ISO/IEC 27005, ISO/IEC 42001 et ISO/IEC 5338 aident une organisation à mettre en œuvre des contrôles et des processus conformes à la réglementation applicable en matière d'IA.

5.2.2 comprendre les défis liés à la conformité des systèmes d'IA aux lois et réglementations pertinentes, notamment le règlement sur l'IA et le RGPD.

5.2.3 décrire des stratégies visant à atténuer les risques de violation du droit d'auteur et à favoriser la conformité.

3. Liste des concepts de base

Ce chapitre dresse une liste des termes que les candidats sont censés maîtriser à l'issue de la formation.

Prière de noter que la connaissance seule des termes ci-après ne suffit pas lors de l'examen. Le candidat doit impérativement comprendre les concepts que recouvre chacun de ces termes et être en mesure de citer quelques exemples.

La liste complète des concepts de base et de leurs définitions figure dans le Glossaire (bibliographie B).

Anglais	**Français**
accountability	responsabilité
adversarial attack	attaque adversariale
adversarial testing	test adversarial
agentic AI	IA agentique
AI (artificial intelligence)	IA (intelligence artificielle)
AI agent	agent IA
AI assets	actifs liés à l'IA
AI jailbreak	contournement des restrictions de l'IA (jailbreak)
AI management system (AIMS)	système de management de l'IA (SMIA)
AI model	modèle d'IA
AI model provider	fournisseur de modèle d'IA
AI privacy principles • accuracy • consent • data minimization and storage limitation • data subject rights • explainability • fairness and lawfulness • transparency • use limitation and purpose specification	principes de protection de la vie privée pour l'IA • exactitude • consentement • minimisation des données et limitation de la conservation • droits de la personne concernée • explicabilité • équité et licéité • transparence • limitation de l'utilisation et spécification de la finalité
AI red teaming	red teaming de l'IA (exercice de test adversarial)
AI security	sécurité de l'IA
AI security governance	gouvernance de la sécurité de l'IA
AI security testing approach	approche des tests de sécurité de l'IA
AI system	système d'IA
AI system provider	fournisseur de système d'IA
anonymization	anonymisation
API (application programming interface)	API (interface de programmation d'application)
audit	audit
augmentation data	données d'augmentation
automated oversight	contrôle automatisé
automation	automatisation
bias	biais
blast radius	rayon d'impact

chatbot	agent conversationnel (chatbot)
chief AI officer (CAIO)	directeur/directrice de l'IA (CAIO)
Chief Information Security Officer (CISO)	responsable de la sécurité des systèmes d'information (RSSI) / Chief Information Security Officer (CISO)
code	code
compliance	conformité
confidentiality	confidentialité
configuration leak	fuite de configuration
copyright	droit d'auteur
data leak	fuite de données
data protection	protection des données
data protection impact assessment (DPIA)	analyse d'impact relative à la protection des données (DPIA)
data security	sécurité des données
data subject	personne concernée
denial-of-service (DoS)	déni de service (DoS)
denial-of-wallet (DoW)	déni de portefeuille (DoW)
deployer	déployeur
development-time	phase de développement
development-time security threats • data poisoning development-time • model poisoning development-time • sensitive data leak development-time	menaces de sécurité en phase de développement • empoisonnement des données en phase de développement • empoisonnement du modèle en phase de développement • fuite de données sensibles en phase de développement
embedding	plongement vectoriel (embedding)
encryption	chiffrement
ethical principles	principes éthiques
European Union (EU)	Union Européenne (UE)
G.U.A.R.D (Govern, Understand, Adapt, Reduce, Demonstrate)	G.U.A.R.D (Gouverner, Comprendre, Adapter, Réduire, Démontrer)
General Data Protection Regulation (GDPR)	Règlement Général sur la Protection des Données (RGPD)
generative AI	IA générative
hijacking	détournement
human oversight	contrôle humain
inference-time	phase d'inférence
information security management system (ISMS)	système de management de la sécurité de l'information (SMSI)
input data	données d'entrée
input security threats • evasion • model exfiltration • prompt injection • resource exhaustion • sensitive data disclosure through use	menaces de sécurité liées aux données d'entrée • évasion • exfiltration du modèle • injection de prompt • épuisement des ressources • divulgation de données sensibles par l'utilisation
integrity verification	vérification de l'intégrité
intellectual property (IP)	propriété intellectuelle (PI)
ISO/IEC 42001	ISO/IEC 42001
large language model (LLM)	grand modèle de langage (LLM)

leakage	fuite
machine learning	apprentissage automatique
membership inference	inférence d'appartenance
model architecture	architecture du modèle
model inversion	inversion de modèle
model leak	fuite de modèle
open source	open source
output data	données de sortie
output sanitization	assainissement des sorties
output validation	validation des sorties
personal data	données à caractère personnel
predictive AI	IA prédictive
privacy	protection de la vie privée
privacy by default	protection de la vie privée par défaut
privacy by design	protection de la vie privée par conception
privacy impact assessments (PIA)	analyses d'impact relatives à la vie privée (PIA)
privacy incident	incident lié à la vie privée
processing (of personal data)	traitement (des données à caractère personnel)
prompt	prompt
pseudonymization	pseudonymisation
publicly accessible	accessible au public
ready-made model	modèle prêt à l'emploi
residual risk	risque résiduel
responsible AI	IA responsable
retrieval-augmented generation (RAG)	génération augmentée par récupération (RAG)
risk	risque
risk analysis	analyse des risques
risk levels for AI systems based on AI Act • minimal or no risk • limited risk • high risk • unacceptable risk	niveaux de risque des systèmes d'IA selon le règlement sur l'IA • risque minimal ou nul • risque limité • haut risque • risque inacceptable
risk management	gestion des risques
risk mitigation	atténuation des risques
risk rating	cotation du risque
runtime	phase d'exécution
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	menaces de sécurité en phase d'exécution • manipulation des données d'augmentation • fuite directe des données d'augmentation • fuite directe du modèle en phase d'exécution • empoisonnement direct du modèle en phase d'exécution • fuite des données d'entrée • sortie contenant une injection conventionnelle

safety of an AI system	sûreté d'un système d'IA
(security) attack	attaque (de sécurité)
security breach security controls • controls for sensitive data limitation • controls to limit the effect of unwanted behavior • governance controls	violation de sécurité contrôles de sécurité • contrôles de limitation des données sensibles • contrôles visant à limiter les effets des comportements indésirables • contrôles de gouvernance
security incident	incident de sécurité
security risks	risques de sécurité
security testing	tests de sécurité
security threats	menaces de sécurité
sensitive data	données sensibles
service provider	fournisseur de service
source code	code source
source code leak	fuite du code source
sponge attack	attaque éponge
stakeholder	partie prenante
startup	jeune pousse
structured query language (SQL)	langage de requête structuré (SQL)
tampering	altération
third party	tiers
threat modeling	modélisation des menaces
token	jeton
tokenization	tokenisation
tool	outil
training data	données d'entraînement
trustworthy AI	IA digne de confiance
unwanted behavior	comportement indésirable
vulnerability	vulnérabilité

4. Bibliographie

Bibliographie pour l'examen

Les connaissances requises pour passer l'examen sont couvertes par la bibliographie suivante :

A. AI Security Community
OWASP AI Exchange
Publication open source (avril 2026)
Ce document est disponible gratuitement sur https://owaspai.org/.
Pour télécharger la version PDF utilisée pour la certification (avril 2026), rendez-vous sur www.exin.com. Cliquez sur « Professionals », puis sur « Certifications » pour trouver la certification. Le téléchargement gratuit se trouve sous la rubrique « Required reading ».

Bibliographie additionnelle

B. EXIN
Glossaire
EXIN (2026)
Ce document est disponible gratuitement sur www.exin.com.

C. Union européenne
Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle (règlement sur l'intelligence artificielle)
Union européenne (2024)
https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32024R1689

D. ISO/IEC 23894:2023 (FR)
Technologies de l'information — Intelligence artificielle — Recommandations relatives au management du risque
Suisse, ISO/IEC (2023)
https://www.iso.org/home.html

E. ISO/IEC 27005:2022 (FR)
Sécurité de l'information, cybersécurité et protection de la vie privée — Préconisations pour la gestion des risques liés à la sécurité de l'information
Suisse, ISO/IEC (2022)
https://www.iso.org/home.html

F. ISO/IEC 42001:2023 (FR)
Technologies de l'information — Intelligence artificielle — Système de management
Suisse, ISO/IEC (2023)
https://www.iso.org/home.html

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

Commentaire

La bibliographie additionnelle est citée uniquement en référence et pour approfondir les connaissances.

Matrice de la bibliographie

Conditions de l'examen	Spécifications de l'examen	Référence
1. Sécurité de l'IA au sein de l'organisation		
	1.1 Organisation de la sécurité de l'IA	A, Chapitre 0
	1.2 Modélisation des menaces et risques liés à l'IA agentique	A, Chapitre 0
2. Menaces pour la sécurité de l'IA		
	2.1 Menaces liées aux données d'entrée	A, Chapitre 2
	2.2 Menaces liées à la phase de développement	A, Chapitre 3
	2.3 Menaces de sécurité conventionnelles en phase d'exécution	A, Chapitre 4
3. Contrôles de sécurité de l'IA		
	3.1 Gouvernance	A, Chapitres 0, 1
	3.2 Limitation des données sensibles	A, Chapitre 1
	3.3 Limitation des comportements indésirables	A, Chapitre 1
4. Tests de sécurité de l'IA		
	4.1 Portée des menaces	A, Chapitre 5
	4.2 Stratégies de test de la sécurité de l'IA	A, Chapitre 5
5. Protection de la vie privée et conformité en matière de sécurité de l'IA		
	5.1 Protection de la vie privée et sécurité de l'IA	A, Chapitre 6
	5.2 Conformité et réglementation	A, Chapitres 0, 1, 2, 3, 6

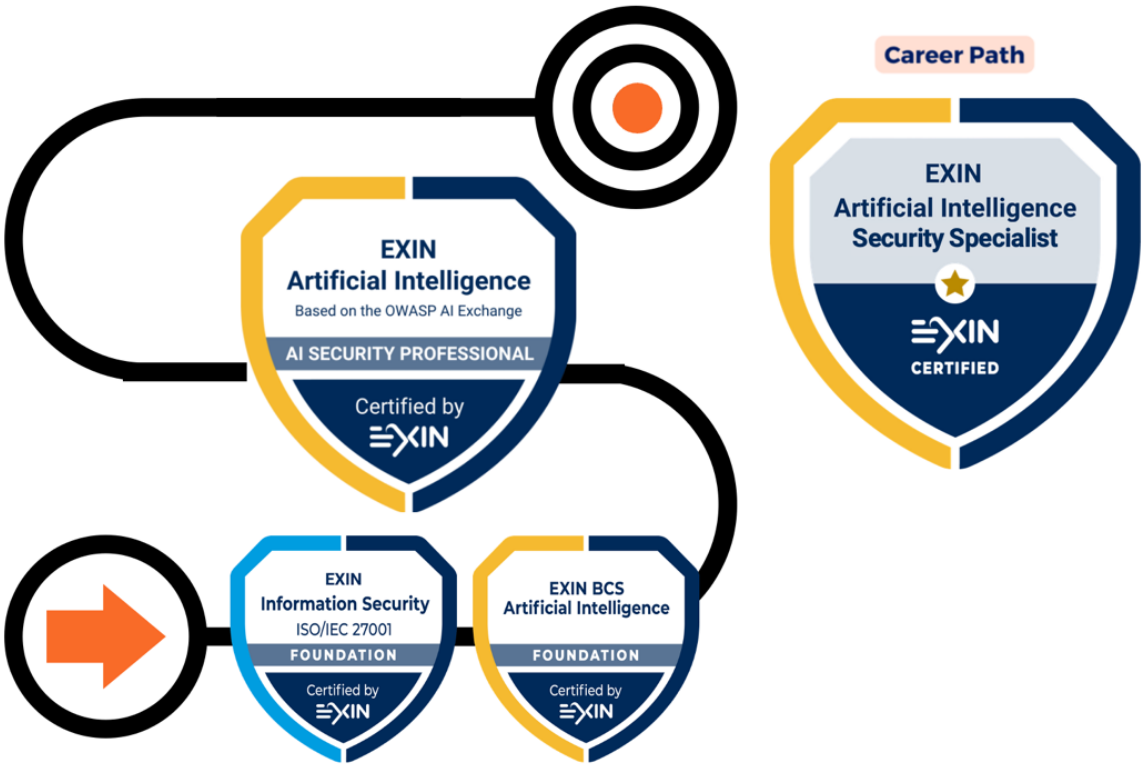
5. Parcours de carrière

Chez EXIN, nous croyons en la valeur de l'apprentissage tout au long de la vie et en l'importance de combiner des compétences variées pour réussir dans un monde dynamique et en constante évolution. Avec nos EXIN Career Paths, les professionnels peuvent se préparer à des rôles spécifiques et continuer à progresser dans leur parcours professionnel. Pour plus d'informations sur les EXIN Career Paths, veuillez consulter : https://www.exin.com/career-paths/.

La certification EXIN AI Security Professional based on the OWASP AI Exchange fait partie du parcours de carrière EXIN suivant.

EXIN AI Security Specialist

EXIN AI Security Specialist atteste que les professionnels sont capables de maîtriser en toute sécurité les complexités des technologies d'IA, de préserver l'intégrité des données et de respecter les normes de sécurité internationales.



Certified for what's next

Contacter EXIN

www.exin.com