

Guia de preparação

Edição 202610

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.

Conteúdo

1. Visão geral 4
2. Requisitos do exame 8
3. Lista de conceitos básicos 11
4. Literatura 15
5. Career Path 17

1. Visão geral

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

Escopo

A certificação EXIN AI Security Professional based on the OWASP AI Exchange confirma que o profissional detém conhecimento sobre como manter os sistemas de IA seguros durante o projeto, o desenvolvimento, a implantação e o uso diário, em conformidade com as políticas da organização, normas e leis aplicáveis.

Essa certificação inclui os seguintes tópicos:

- Segurança da IA na organização
- Ameaças de segurança da IA
- Controles de segurança da IA
- Testes de segurança da IA
- Privacidade e conformidade na segurança da IA

Resumo

Atualmente, a Inteligência Artificial (IA) impulsiona decisões, conteúdo e automação em todos os setores, ampliando oportunidades, mas também aumentando a superfície de ataque. Como resultado, as organizações enfrentam riscos em rápida evolução, incluindo ataques que envolvem corrupção de sistemas de IA, uso indevido de modelos ou vulnerabilidades interligadas entre ferramentas e fornecedores, enquanto os requisitos legais e de privacidade se tornam cada vez mais rigorosos. Nesse cenário, a segurança não pode ser uma preocupação secundária; ela deve estar integrada à forma como a IA é projetada, implantada e monitorada.

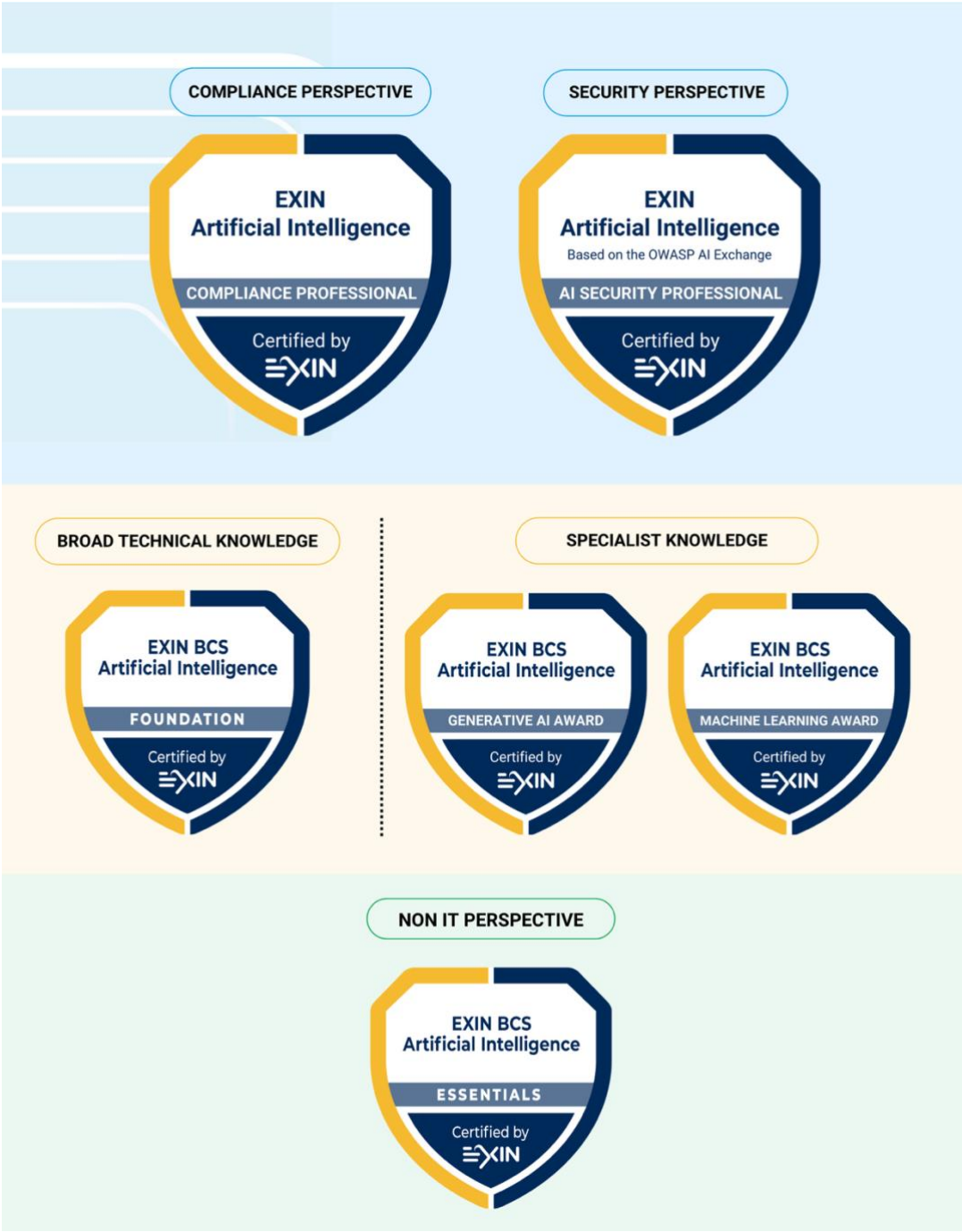
Nos últimos anos, a segurança da IA emergiu como uma prioridade crítica. O OWASP AI Exchange define segurança como a prevenção de acesso, uso, divulgação, interrupção, modificação ou destruição não autorizados, sendo que “modificação” inclui a manipulação do comportamento de um modelo de IA de maneiras indesejadas. À medida que a IA se torna cada vez mais integrada aos processos de tomada de decisão, garantir sua confiabilidade e integridade é essencial para promover a confiança e minimizar os riscos nesse complexo ambiente de segurança.

A certificação EXIN AI Security Professional based on the OWASP AI Exchange prepara profissionais para enfrentar esse desafio. Ela se baseia no OWASP AI Exchange, uma base de conhecimento aberta e orientada pela comunidade sobre riscos, ameaças, controles de segurança e estratégias de teste relacionados à IA. A certificação oferece um caminho para que os profissionais compreendam melhor a segurança da IA na organização, as principais categorias de ameaças de segurança, controles de segurança eficazes, testes de segurança de IA, bem como a privacidade e a conformidade ao longo de todo o ciclo de vida da IA.

A certificação AISP é relevante para profissionais das mais diversas formações e é particularmente atraente para aqueles que atuam ou têm interesse em funções relacionadas à segurança e privacidade, engenharia de IA, testes de IA, governança de segurança da IA e conformidade. Os detentores dessa certificação sabem como projetar, avaliar e aprimorar programas de segurança da IA que sejam resilientes, auditáveis e que levem em conta a privacidade, fortalecendo assim o trabalho do dia a dia.

Contexto

A certificação EXIN AI Security Professional based on the OWASP AI Exchange faz parte do programa de qualificação EXIN Artificial Intelligence.



Público-alvo

A certificação EXIN AI Security Professional based on the OWASP AI Exchange é adequada para um grupo diversificado de profissionais que trabalham com sistemas de IA, bem como para aqueles que se dedicam à segurança da IA e à privacidade desses sistemas. Essas funções exigem um nível de competência que permita às organizações conceber, desenvolver, adquirir, operar e manter aplicativos confiáveis.

Isso inclui, mas não se limita a:

- Profissionais de IA
- Profissionais de segurança (cibernética)
- Profissionais de proteção de dados e privacidade
- Engenheiros de aprendizado de máquina (segurança)
- Engenheiros de dados
- Especialistas em qualidade e avaliação
- Responsáveis por governança, risco e conformidade

Requisitos para a certificação

- Conclusão bem-sucedida do exame EXIN AI Security Professional based on the OWASP AI Exchange.

São altamente recomendáveis tanto a realização de treinamento para esta certificação quanto o conhecimento prévio da terminologia de Inteligência Artificial (IA), por exemplo, por meio da certificação EXIN BCS Artificial Intelligence Essentials ou da certificação EXIN BCS Artificial Intelligence Foundation.

Detalhes do exame

Tipo do exame:	Questões de múltipla escolha
Número de questões:	40
Mínimo para aprovação:	65% (26/40 questões)
Com consulta:	Não
Anotações:	Não
Equipamentos eletrônicos permitidos:	Não
Tempo designado para o exame:	90 minutos

As Regras e Regulamentos dos exames EXIN aplicam-se a este exame.

Nível Bloom

A certificação EXIN AI Security Professional based on the OWASP AI Exchange testa os candidatos nos Níveis Bloom 2 e 3 de acordo com a Taxonomia Revisada de Bloom:

- Nível Bloom 2: Compreensão – um passo além da lembrança. O entendimento mostra que os candidatos compreendem o que é apresentado e podem avaliar como o material de aprendizagem pode ser aplicado em seu próprio ambiente. Este tipo de pergunta pretende demonstrar que o candidato é capaz de organizar, comparar, interpretar e escolher a descrição correta de fatos e ideias.
- Nível Bloom 3: Aplicação – mostra que os candidatos têm a capacidade de utilizar as informações em um contexto diferente daquele em que elas foram aprendidas. Este tipo de pergunta pretende demonstrar que o candidato é capaz de resolver problemas em novas situações, aplicando o conhecimento adquirido, fatos, técnicas e regras de um modo novo ou diferente. A pergunta geralmente contém um breve cenário.

Treinamento

Horas de contato

A carga horária recomendada para este treinamento é de 14 horas. Isso inclui trabalhos em grupo, preparação para o exame e pausas curtas. Essa carga horária não inclui pausas para almoço, trabalhos extra aula e o exame.

Indicação de tempo de estudo

84 horas (3 ECTS), dependendo do conhecimento pré-existente.

Provedor de treinamento

Você encontrará uma lista de nossos provedores de treinamento credenciados em www.exin.com.

2. Requisitos do exame

Os requisitos do exame são definidos nas especificações do exame. A tabela a seguir lista os tópicos (requisitos do exame) e subtópicos (especificações do exame) do módulo.

Requisitos do exame	Especificações do exame	Peso
1. Segurança da IA na organização		**15%**
	1.1 Organização da segurança da IA	10%
	1.2 Modelagem de ameaças (threat modeling) e riscos da IA agêntica	5%
2. Ameaças de segurança da IA		**37,5%**
	2.1 Ameaças de segurança por meio de entrada	17,5%
	2.2 Ameaças de segurança em fase de desenvolvimento (development-time)	10%
	2.3 Ameaças convencionais de segurança em fase de execução (runtime)	10%
3. Controles de segurança da IA		**27,5%**
	3.1 Governança	12,5%
	3.2 Limitação de dados sensíveis	7,5%
	3.3 Limitação de comportamentos indesejados	7,5%
4. Testes de segurança da IA		**7,5%**
	4.1 Escopo das ameaças	5%
	4.2 Estratégias de testes de segurança da IA	2,5%
5. Privacidade e conformidade na segurança da IA		**12,5%**
	5.1 Privacidade e segurança da IA	5%
	5.2 Conformidade e regulamentação	7,5%
		100%

Especificações do exame

1 Segurança da IA na organização

1.1 Organização da segurança da IA

O candidato é capaz de...

1.1.1 aplicar as cinco etapas do modelo G.U.A.R.D para organizar a segurança da IA dentro de uma organização.

1.1.2 compreender como promover o uso responsável e confiável da IA na organização.

1.1.3 diferenciar a segurança da IA da segurança (cibernética) convencional.

1.1.4 descrever os ativos específicos de IA e suas principais ameaças.

1.2 Modelagem de ameaças (threat modeling) e riscos da IA agêntica

O candidato é capaz de...

1.2.1 descrever as etapas do gerenciamento de riscos para realizar a modelagem de ameaças com controles para a segurança da IA.

1.2.2 explicar os riscos típicos relacionados à IA agêntica.

2 Ameaças de segurança da IA

2.1 Ameaças de segurança por meio de entrada

O candidato é capaz de...

2.1.1 ilustrar os vários tipos de evasão com base no acesso dos invasores ao conhecimento: evasão de conhecimento zero, evasão de conhecimento perfeito, evasão de conhecimento parcial, ataque de transferência e evasão após envenenamento.

2.1.2 distinguir entre injeção direta de prompt e injeção indireta de prompt.

2.1.3 ilustrar as sete camadas de proteção contra a injeção de prompt.

2.1.4 explicar como a divulgação de dados sensíveis pode ocorrer em dados de saída dos modelos, inversão do modelo e inferência de associação (membership inference).

2.1.5 explicar como ocorre a exfiltração de modelo.

2.1.6 identificar ameaças de esgotamento de recursos por meio de entradas.

2.2 Ameaças de segurança em fase de desenvolvimento (development-time)

O candidato é capaz de...

2.2.1 identificar riscos de envenenamento de dados em fase de desenvolvimento.

2.2.2 explicar o envenenamento direto de modelo em fase de desenvolvimento.

2.2.3 explicar o envenenamento de modelos na cadeia de suprimentos.

2.2.4 interpretar diferentes tipos de vazamentos de dados sensíveis em fase de desenvolvimento: vazamento de dados em fase de desenvolvimento, vazamento direto de modelo em fase de desenvolvimento e vazamento de código-fonte/configuração.

2.3 Ameaças convencionais de segurança em fase de execução (runtime)

O candidato é capaz de...

2.3.1 explicar como as ameaças convencionais de segurança em fase de execução podem afetar os componentes de IA.

2.3.2 ilustrar o envenenamento direto de modelo em fase de execução e o vazamento direto de modelo em fase de execução.

2.3.3 descrever as ameaças de segurança associadas à injeção incorporada nas saídas de IA e ao vazamento de dados de entrada.

2.3.4 explicar como o vazamento e a manipulação de dados aumentados (augmented data) constituem ameaças de segurança.

3 Controles de segurança da IA

3.1 Governança

O candidato é capaz de...

3.1.1 aplicar controles gerais de governança para a supervisão da segurança da IA.

3.1.2 explicar a abrangência dos controles gerais de governança.

3.1.3 compreender como a implementação dos controles é dividida entre um fornecedor de modelos de IA (terceiro) e o desenvolvedor/responsável pela implantação do sistema de IA (modelo pronto para uso (ready-made model)).

3.2 Limitação de dados sensíveis

O candidato é capaz de...

3.2.1 implementar controles para limitação de dados sensíveis, a fim de reforçar a confidencialidade e a integridade.

3.2.2 compreender como a limitação de dados sensíveis reduz os riscos.

3.3 Limitação de comportamentos indesejados

O candidato é capaz de...

3.3.1 implementar controles para limitar o efeito de comportamento indesejado do modelo.

3.3.2 compreender como a limitação de comportamentos indesejados do modelo beneficia o desempenho e reduz riscos.

4 Testes de segurança da IA

4.1 Escopo das ameaças

O candidato é capaz de...

4.1.1 explicar por que os testes de segurança da IA são importantes e qual é o seu escopo.

4.1.2 compreender quais ameaças devem ser testadas além dos testes de segurança convencionais ao utilizar IA preditiva ou generativa.

4.2 Estratégias de testes de segurança da IA

O candidato é capaz de...

4.2.1 descrever estratégias de testes de segurança da IA utilizando a abordagem geral de testes.

5 Privacidade e conformidade na segurança da IA

5.1 Privacidade e segurança da IA

O candidato é capaz de...

5.1.1 explicar o que é privacidade e as questões relacionadas à privacidade associadas aos sistemas de IA.

5.1.2 aplicar princípios de privacidade da IA a um determinado cenário.

5.2 Conformidade e regulamentação

O candidato é capaz de...

5.2.1 explicar como as normas ISO/IEC 23894, ISO/IEC 27005, ISO/IEC 42001 e ISO/IEC 5338 auxiliam uma organização na implementação de controles e processos que estejam em conformidade com as regulamentações aplicáveis à IA.

5.2.2 compreender os desafios associados à conformidade dos sistemas de IA com as leis e regulamentações relevantes, incluindo a Lei da IA (AI Act) e o GDPR.

5.2.3 descrever estratégias para mitigar os riscos de violação de direitos autorais e apoiar a conformidade.

3. Lista de conceitos básicos

Este capítulo contém os termos e abreviaturas com que os candidatos devem se familiarizar.

Por favor, note que o conhecimento desses termos de maneira independente não é suficiente para o exame. O candidato deve compreender os conceitos e estar apto a fornecer exemplos.

Uma lista completa dos conceitos básicos e suas definições pode ser encontrada no Glossário (Literatura B).

Inglês	**Português**
accountability	responsabilização e prestação de contas
adversarial attack	ataque adversarial
adversarial testing	testes adversariais
agentic AI	IA agêntica
AI (artificial intelligence)	Inteligência Artificial (IA)
AI agent	agente de IA
AI assets	ativos de IA
AI jailbreak	jailbreak de IA
AI management system (AIMS)	Sistema de Gestão de IA (SGIA)
AI model	modelo de IA
AI model provider	fornecedor do modelo de IA
AI privacy principles • accuracy • consent • data minimization and storage limitation • data subject rights • explainability • fairness and lawfulness • transparency • use limitation and purpose specification	princípios de privacidade da IA • exatidão (dos dados) / acurácia (do modelo) • consentimento • minimização de dados e limitação de armazenamento • direitos do titular dos dados • explicabilidade • equidade e licitude • transparência • limitação de uso e especificação da finalidade
AI red teaming	red teaming de IA
AI security	segurança da IA
AI security governance	governança de segurança da IA
AI security testing approach	abordagem de testes de segurança de IA
AI system	sistema de IA
AI system provider	fornecedor do sistema de IA
anonymization	anonimização
API (application programming interface)	API (interface de programação de aplicações)
audit	auditoria
augmentation data	dados de aumento (augmentation data)
automated oversight	supervisão automatizada
automation	automação
bias	viés
blast radius	raio de impacto (blast radius)
chatbot	chatbot
chief AI officer (CAIO)	Chief AI Officer (CAIO)
Chief Information Security Officer (CISO)	Chief Information Security Officer (CISO)

code	código
compliance	conformidade
confidentiality	confidencialidade
configuration leak	vazamento de configuração
copyright	direitos autorais
data leak	vazamento de dados
data protection	proteção de dados
data protection impact assessment (DPIA)	avaliação de impacto sobre a proteção de dados (DPIA)
data security	segurança de dados
data subject	titular dos dados
denial-of-service (DoS)	negação de serviço (denial-of-service, DoS)
denial-of-wallet (DoW)	negação de carteira (denial-of-wallet, DoW)
deployer	responsável pela implantação
development-time	fase de desenvolvimento (development-time)
development-time security threats • data poisoning development-time • model poisoning development-time • sensitive data leak development-time	ameaças de segurança em fase de desenvolvimento (development-time) • envenenamento de dados em fase de desenvolvimento • envenenamento de modelo em fase de desenvolvimento • vazamento de dados sensíveis em fase de desenvolvimento
embedding	embedding (representação vetorial)
encryption	criptografia
ethical principles	princípios éticos
European Union (EU)	União Europeia (UE)
G.U.A.R.D (Govern, Understand, Adapt, Reduce, Demonstrate)	G.U.A.R.D (Governar, Entender, Adaptar, Reduzir, Demonstrar)
General Data Protection Regulation (GDPR)	Regulamento Geral de Proteção de Dados (GDPR)
generative AI	IA generativa
hijacking	sequestro
human oversight	supervisão humana
inference-time	fase de inferência (inference-time)
information security management system (ISMS)	Sistema de Gestão de Segurança da Informação (SGSI)
input data	dados de entrada
input security threats • evasion • model exfiltration • prompt injection • resource exhaustion • sensitive data disclosure through use	ameaças de segurança por meio de entrada • evasão • exfiltração de modelo • injeção de prompt • esgotamento de recursos • divulgação de dados sensíveis por meio do uso
integrity verification	verificação de integridade
intellectual property (IP)	propriedade intelectual (PI)
ISO/IEC 42001	ISO/IEC 42001
large language model (LLM)	modelo de linguagem de grande porte (LLM)
leakage	vazamento
machine learning	aprendizado de máquina
membership inference	inferência de associação (membership inference)
model architecture	arquitetura do modelo

model inversion	inversão do modelo
model leak	vazamento do modelo
open source	código aberto
output data	dados de saída
output sanitization	sanitização de saída
output validation	validação de saída
personal data	dados pessoais
predictive AI	IA preditiva
privacy	privacidade
privacy by default	privacidade por padrão (by default)
privacy by design	privacidade desde a concepção (by design)
privacy impact assessments (PIA)	avaliações de impacto sobre a privacidade (PIA)
privacy incident	incidente de privacidade
processing (of personal data)	tratamento (de dados pessoais)
prompt	prompt
pseudonymization	pseudonimização
publicly accessible	acessível ao público
ready-made model	modelo pronto para uso (ready-made model)
residual risk	risco residual
responsible AI	IA responsável
retrieval-augmented generation (RAG)	geração aumentada por recuperação (RAG)
risk	risco
risk analysis	análise de risco
risk levels for AI systems based on AI Act • minimal or no risk • limited risk • high risk • unacceptable risk	níveis de risco para sistemas de IA com base na Lei da IA (AI Act) • risco mínimo ou nulo • risco limitado • risco elevado • risco inaceitável
risk management	gerenciamento de riscos
risk mitigation	mitigação de riscos
risk rating	classificação de riscos
runtime	fase de execução (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	ameaças de segurança em fase de execução (runtime) • manipulação de dados de aumento (augmentation data) • vazamento direto de dados de aumento (augmentation data) • vazamento direto de modelo em fase de execução (runtime) • envenenamento direto de modelo em fase de execução (runtime) • vazamento de dados de entrada • saída contendo injeção convencional
safety of an AI system	segurança operacional de um sistema de IA
(security) attack	ataque (de segurança)
security breach	violação de segurança
security controls • controls for sensitive data limitation • controls to limit the effect of unwanted behavior	controles de segurança • controles para limitação de dados sensíveis • controles para limitar o efeito de comportamento indesejado

• governance controls	• controles de governança
security incident	incidente de segurança
security risks	riscos de segurança
security testing	testes de segurança
security threats	ameaças de segurança
sensitive data	dados sensíveis
service provider	prestador de serviços
source code	código-fonte
source code leak	vazamento de código-fonte
sponge attack	ataque esponja (sponge attack)
stakeholder	parte interessada
startup	startup
structured query language (SQL)	linguagem de consulta estruturada (SQL)
tampering	adulteração
third party	terceiro
threat modeling	modelagem de ameaças (threat modeling)
token	token
tokenization	tokenização
tool	ferramenta
training data	dados de treinamento
trustworthy AI	IA confiável
unwanted behavior	comportamento indesejado
vulnerability	vulnerabilidade

4. Literatura

Literatura do exame

O conhecimento necessário para o exame é coberto na seguinte literatura:

A. AI Security Community
OWASP AI Exchange
Publicação aberta ao público (abril 2026)
Esse material pode ser baixado gratuitamente em https://owaspai.org/.
Para baixar a versão em PDF usada na certificação (abril 2026), acesse www.exin.com. Clique em 'Profissionais' e, em seguida, em 'Certificações' para encontrar a certificação. O download gratuito pode ser encontrado sob 'Leitura obrigatória'.

Literatura adicional

B. EXIN
Glossário
EXIN (2026)
Esse material encontra-se disponível gratuitamente em www.exin.com.

C. União Europeia
Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial (Lei da Inteligência Artificial da UE)).
União Europeia (2024)
https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689

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

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

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

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

Comentário

A literatura adicional destina-se exclusivamente à referência e ao aprofundamento do conhecimento.

Matriz da literatura

Requisitos do exame	Especificações do exame	Referência
1. Segurança da IA na organização		
	1.1 Organização da segurança da IA	A, Capítulo 0
	1.2 Modelagem de ameaças (threat modeling) e riscos da IA agêntica	A, Capítulo 0
2. Ameaças de segurança da IA		
	2.1 Ameaças de segurança por meio de entrada	A, Capítulo 2
	2.2 Ameaças de segurança em fase de desenvolvimento (development-time)	A, Capítulo 3
	2.3 Ameaças convencionais de segurança em fase de execução (runtime)	A, Capítulo 4
3. Controles de segurança da IA		
	3.1 Governança	A, Capítulos 0, 1
	3.2 Limitação de dados sensíveis	A, Capítulo 1
	3.3 Limitação de comportamentos indesejados	A, Capítulo 1
4. Testes de segurança da IA		
	4.1 Escopo das ameaças	A, Capítulo 5
	4.2 Estratégias de testes de segurança da IA	A, Capítulo 5
5. Privacidade e conformidade na segurança da IA		
	5.1 Privacidade e segurança da IA	A, Capítulo 6
	5.2 Conformidade e regulamentação	A, Capítulos 0, 1, 2, 3, 6

5. Career Path

O EXIN acredita no valor do aprendizado contínuo e na importância de combinar diversas habilidades para prosperar no atual mundo dinâmico e em evolução. Com os EXIN Career Paths, os candidatos podem se preparar para funções específicas e continuar crescendo e avançando em sua jornada profissional. Para obter mais informações sobre os EXIN Career Paths, consulte https://www.exin.com/career-paths/.

A certificação EXIN AI Security Professional based on the OWASP AI Exchange faz parte do seguinte EXIN Career Path.

EXIN AI Security Specialist

O EXIN AI Security Specialist confirma que os profissionais são capazes de lidar com as complexidades das tecnologias de IA com segurança, manter a integridade dos dados e atender aos padrões globais de segurança da IA.





Certified for what's next

Contato EXIN

www.exin.com