+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 13/11/2026

S'inscrire

Formation Certification Certified Artificial Intelligence Security Professional

5 jours (35 heures)

Présentation

La certification officielle Certified Artificial Intelligence Security Professional de PECB valorise les compétences nécessaires pour sécuriser les systèmes d'intelligence artificielle. Du machine learning aux LLM et aux agents IA autonomes, sa préparation couvre les menaces, les contrôles de protection et la gouvernance sur tout le cycle de vie.

Notre Formation Certification Certified Artificial Intelligence Security Professional vous permettra de préparer l'examen tout en développant une démarche opérationnelle de sécurité de l'IA adaptée aux environnements professionnels.

Vous apprendrez à expliquer les architectures, à identifier les risques liés aux données et aux modèles, puis à analyser les attaques avec MITRE ATLAS et le OWASP LLM Top 10 comme référentiels. Les exercices vous entraîneront à sélectionner et à évaluer les mesures de protection appropriées.

À l'issue de la formation, vous serez en mesure de définir une architecture sécurisée pour l'IA, d'organiser la surveillance, de conduire une investigation et de structurer la réponse aux incidents spécifiques aux modèles et aux applications génératives.

Cette préparation associe ateliers pratiques et entraînement aux questions d'examen. Elle aborde également la chaîne d'approvisionnement, la conformité et le red teaming des systèmes IA. L'obtention du titre professionnel reste soumise aux conditions d'expérience et de certification de PECB.

Comme toutes nos formations, celle-ci vous présentera **la dernière version stable** de la technologie et ses nouveautés.

Objectifs

- Expliquer le fonctionnement du machine learning et du deep learning ainsi que les architectures des LLM et des agents IA
- Identifier les risques liés à la conception, à l'entraînement et au déploiement des systèmes IA
- Analyser les menaces sur l'ensemble du cycle d'attaque avec MITRE ATLAS et le OWASP LLM Top 10 comme référentiels
- Appliquer des contrôles de sécurité, des stratégies de surveillance et des workflows de renseignement sur les menaces
- Concevoir et évaluer un programme de sécurité IA intégrant gouvernance, conformité, réponse aux incidents et red teaming
- Préparer l'examen Certified Artificial Intelligence Security Professional par des cas pratiques et des questions d'entraînement

Public visé

- Ingénieur cybersécurité
- Architecte sécurité
- Analyste SOC
- Ingénieur IA et machine learning
- Consultant en gouvernance

Pré-requis

- Connaissances des principes de cybersécurité, du contrôle d'accès et de la protection des données
- Familiarité avec la modélisation des menaces, la supervision de sécurité et la réponse aux incidents
- Compréhension élémentaire du machine learning et des étapes d'entraînement et d'inférence
- Capacité à lire des journaux techniques et des échanges d'API dans un environnement applicatif
- Compréhension de l'anglais technique pour consulter les ressources et préparer l'examen

Pré-requis techniques

- Ordinateur sous Windows 11 ou macOS compatible avec les outils pédagogiques, avec 16 Go de mémoire vive recommandés
- Navigateur Chrome ou Firefox à jour pour accéder aux laboratoires distants, sans carte graphique dédiée nécessaire
- Connexion Internet stable avec accès autorisé aux plateformes pédagogiques et aux environnements isolés des ateliers
- Compte personnel myPECB et application PECB Exams pour le passage de l'examen en ligne, avec vérification préalable de compatibilité
- Webcam, microphone et espace calme pour la formation à distance et la télésurveillance de l'examen

Programme de notre formation Certified Artificial Intelligence Security Professional

[Jour 1 - Matin]

Fondamentaux IA et périmètre de certification

- Les cinq domaines de compétences de la certification et les objectifs de préparation à l'examen
- Distinctions entre intelligence artificielle, machine learning, deep learning et IA générative
- Principes des apprentissages supervisé, non supervisé et par renforcement
- Cycle de vie des systèmes IA, actifs sensibles et responsabilités des parties prenantes
- Atelier pratique : cartographier les composants, les données et les actifs critiques d'un service IA à partir d'un cas de certification.

[Jour 1 - Après-midi]

Architectures ML et risques des modèles

- Fonctionnement des réseaux de neurones, entraînement, validation et inférence
- Pipelines de données, qualité, séparation des jeux et surapprentissage
- Risques de conception, biais, confidentialité et limites de l'explicabilité
- Attaques par empoisonnement, évasion, extraction de modèle et inférence d'appartenance
- Atelier pratique : comparer les résultats d'un modèle sur des données normales et altérées, puis justifier les risques et les protections retenues.

[Jour 2 - Matin]

LLM, RAG et agents IA

- Architecture Transformer, tokenisation, embeddings et mécanisme d'attention
- Préentraînement, fine-tuning, alignement et risques des données d'apprentissage
- Architecture RAG, bases vectorielles et frontières de confiance documentaire
- Fonctionnement des agents IA, mémoire, orchestration, outils et délégation de permissions
- Atelier pratique : analyser une application RAG avec agent et repérer les frontières de confiance ainsi que les chemins d'accès aux données sensibles.

[Jour 2 - Après-midi]

Analyse des attaques avec ATLAS et OWASP

- Lecture de MITRE ATLAS, tactiques, techniques et progression du cycle d'attaque
- Analyse du OWASP LLM Top 10 et correspondances avec les scénarios de certification
- Prompt injection directe et indirecte, jailbreaking et divulgation d'informations sensibles
- Compromission de la chaîne d'approvisionnement, faiblesses vectorielles, sorties non sécurisées et autonomie excessive
- Fuite du prompt système, désinformation et consommation non bornée des ressources

- Atelier pratique : tester des injections dans une application pédagogique isolée et cartographier le scénario dans MITRE ATLAS et le OWASP LLM Top 10.

[Jour 3 - Matin]

Contrôles de protection et architecture sécurisée

- Défense en profondeur, segmentation et application des principes Zero Trust aux composants IA
- Gestion des identités, moindre privilège, secrets et autorisations des outils d'agents
- Validation des entrées, traitement sécurisé des sorties et limites des garde-fous
- Protection des données, isolation des contextes, quotas et validation humaine des actions sensibles
- Atelier pratique : renforcer les contrôles d'une application IA vulnérable et comparer les résultats avant et après protection.

[Jour 3 - Après-midi]

Surveillance et détection des activités adverses

- Collecte de télémétrie IA, traces d'appels, versions des modèles et exécution des outils
- Journalisation des interactions et minimisation des données sensibles dans les traces
- Détection d'anomalies, dérive des données et distinction entre dégradation et activité hostile
- Intégration SIEM et SOAR, corrélation des événements et priorisation des alertes
- Mesure de l'efficacité des contrôles, couverture de détection et gestion des faux positifs
- Atelier pratique : construire une règle de détection sur des journaux IA et qualifier une séquence suspecte d'appels et d'accès aux données.

[Jour 4 - Matin]

Threat intelligence et investigation assistée par IA

- Cycle du renseignement sur les menaces, collecte, analyse, diffusion et retour opérationnel
- Exploitation des IoC, indicateurs d'attaque et tactiques, techniques et procédures adverses
- Corrélation, enrichissement et synthèse assistés par IA avec validation des sources
- Threat hunting, analyse de code et de rapports de logiciels malveillants avec assistance IA
- Maîtrise des hallucinations, de la confidentialité et des contenus hostiles dans les sources analysées
- Atelier pratique : produire une note de renseignement à partir d'un dossier d'incident fictif et vérifier les conclusions proposées par l'assistant IA.

[Jour 4 - Après-midi]

Réponse aux incidents et déploiement maîtrisé

- Préparation des playbooks IA, classification des incidents et critères d’escalade
- Confinement des services, révocation des accès et suspension des actions d’agents
- Préservation des preuves, chronologie et investigation des données, modèles et configurations
- Restauration depuis des artefacts fiables, validation de sécurité et retour en production
- Sécurisation des pipelines, intégrité des modèles et retour d’expérience après incident
- Atelier pratique : conduire un exercice de réponse à une compromission d’agent, du confinement à la décision de remise en service.

[Jour 5 - Matin]

Gouvernance, conformité et chaîne d’approvisionnement

- Construction d’un programme de sécurité IA, rôles, politiques, registre des risques et indicateurs
- Évaluation des fournisseurs, provenance des données et modèles, dépendances et chaîne d’approvisionnement
- Analyse des exigences applicables, articulation avec l’AI Act, ISO/IEC 42001 et le NIST AI RMF
- Préparation du red teaming, périmètre autorisé, critères de réussite et suivi des corrections
- Atelier pratique : concevoir puis évaluer un programme de sécurité IA comprenant architecture cible, exigences de conformité et campagne de tests adverses.

[Jour 5 - Après-midi]

Synthèse et entraînement à l’examen

- Révision transversale des cinq domaines de la certification Certified Artificial Intelligence Security Professional
- Méthode de résolution des questions à choix multiple et analyse des scénarios
- Gestion du temps, organisation des ressources autorisées et identification des points de vigilance
- Modalités de planification du passage officiel sur une session dédiée et distinction entre réussite à l’examen et attribution du titre
- Atelier pratique : réaliser un examen blanc ciblé et chronométré, analyser les réponses et construire votre plan de révision individuel.

Sociétés concernées

Cette formation s’adresse à la fois aux particuliers ainsi qu’aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d’auto-évaluation nous permettant d’apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d’anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être

problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l’acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.

Organisme de formation enregistré sous le numéro 11 75 54743 75. Cet enregistrement ne vaut pas agrément de l’État.
© Ambient IT 2015-2026. Tous droits réservés. Paris, France - Suisse - Belgique - Luxembourg