CORSEN AI Academy

Sécurité IA et Red Teaming

Programme de formation

Version 2026-11

Présentation

Prompt injection, jailbreaking, guardrails, modération, red teaming et conformité RGPD pour l'IA.

Objectifs pédagogiques

- Cartographier les risques propres aux applications fondées sur des modèles de langage
- Identifier et prévenir les attaques par injection de prompt et les tentatives de jailbreak
- Mettre en place des garde-fous de validation en entrée et en sortie
- Exploiter les API de modération et le filtrage multi-couche
- Conduire un exercice de red teaming et documenter ses conclusions

Public visé et prérequis

Développeurs, ingénieurs et intégrateurs souhaitant exploiter ChatGPT par API dans leurs applications.

Expérience confirmée du développement d'applications et de leur mise en production.

Durée et modalités

Durée estimée : 1h45. Formation 100 % en ligne, en autonomie, accessible 24h/24 depuis un navigateur. Progression enregistrée, reprise possible à tout moment. Durée estimée : 1h45.

Contenu

- Menaces et surface d'attaque (4 leçons)
- Défenses (5 leçons)
- Red teaming (3 leçons)
- Conformité (2 leçons)

Modalités d'évaluation

Exercices d'application proposés au fil des leçons. Chaque leçon se termine par ses points clés, support de l'auto-évaluation. Un quiz final de 5 questions corrigées, en fin de dernière leçon, permet de valider les acquis.

Accessibilité

Formation accessible aux personnes en situation de handicap : contenus textuels compatibles avec les lecteurs d'écran, navigation au clavier et contrastes conformes. Pour tout besoin d'adaptation, écrivez à hello@corsen.ai.

Tarif et reconnaissance

Gratuit

Attestation de fin de formation délivrée par Corsen AI (non éligible au CPF à ce jour).

Délai d'accès

Immédiat après la création du compte et la validation de l'adresse email.

Corsen AI — Paris, France — SIRET 10297040700015
Organisme de formation en cours de déclaration d'activité.
academy@corsen.ai — https://academy.corsen.ai