

Prompt Injection, jailbreak & sécurisation des LLM en entreprise

Lien :
https://innov-maroc.com/formation/prompt-injection-jailbreak-securisation-des-llm-en-entreprise

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI202

CATÉGORIE
Sécurité de l'Intelligence Artificielle

INNOV MAROC

🎯 OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- ✔ Comprendre les mécanismes de prompt injection et jailbreak
- ✔ Identifier les vulnérabilités des systèmes LLM en entreprise
- ✔ Sécuriser les applications basées sur l’IA générative
- ✔ Protéger les pipelines RAG et les bases vectorielles
- ✔ Mettre en place des guardrails et politiques de sécurité IA
- ✔ Détecter et prévenir les attaques sur les LLM
- ✔ Sécuriser les agents IA et les intégrations API
- ✔ Assurer la conformité et la gouvernance des systèmes IA
- ✔ Réduire les risques liés à l’exploitation des modèles
- ✔ Implémenter des stratégies de défense robustes
- ✔ Améliorer la résilience des applications IA
- ✔ Étudier des cas concrets d’attaques et de défenses LLM

👥 POUR QUI ?

- ✔ AI Engineers
- ✔ Machine Learning Engineers
- ✔ MLOps Engineers
- ✔ Data Scientists
- ✔ Cybersecurity Engineers
- ✔ DevSecOps Engineers
- ✔ Application Security Engineers
- ✔ Cloud Architects
- ✔ RSSI / CISO
- ✔ Consultants en cybersécurité
- ✔ Product Managers IA
- ✔ Toute personne impliquée dans le développement et la sécurisation des applications d’intelligence artificielle générative

INNOV MAROC



Programme détaillé

1 / Introduction aux risques des LLM en entreprise

- Fonctionnement des grands modèles de langage (LLM)
- Surface d’attaque des systèmes basés sur l’IA générative
- Enjeux de sécurité en contexte entreprise

2 / Comprendre le prompt injection

- Définition et mécanismes d’attaque
- Injection directe et indirecte
- Exploitation des données contextuelles et RAG

3 / Techniques de jailbreak des LLM

- Contournement des garde-fous des modèles
- Manipulation des instructions système
- Techniques d’évasion et d’ingénierie sociale

4 / Vulnérabilités des applications LLM

- Chaînes RAG et fuites de données
- Plugins, tools et agents IA
- Risques liés aux API et intégrations

INNOV MAROC

5 / Attaques sur les systèmes augmentés (RAG)

- Poisoning de bases vectorielles
- Injection via documents et sources externes
- Manipulation des embeddings

6 / Sécurisation des prompts et des workflows IA

- Prompt hardening et filtrage des entrées
- Séparation des instructions et des données
- Sanitization et validation des entrées utilisateur

7 / Protection des architectures LLM

- Guardrails et policy engines
- Sandboxing des outils et agents
- Contrôle des appels API externes

8 / Monitoring et détection des attaques

- Détection d’anomalies dans les interactions LLM
- Logging et audit des prompts
- Analyse comportementale des utilisateurs

9 / Gouvernance et conformité des LLM

- Politiques d’utilisation de l’IA générative
- Gestion des risques et conformité réglementaire
- Responsabilité et traçabilité des outputs

10 / Cas pratiques et défense en environnement réel

- Simulation d’attaques prompt injection
- Mise en place de protections sur un système RAG

INNOV MAROC

- Bonnes pratiques de sécurisation des LLM en entreprise

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

25 au 29 Jan. 2027	Présentiel - Casablanca
22 au 26 Mars 2027	Présentiel - Casablanca
17 au 21 Mai 2027	Présentiel - Casablanca
12 au 16 Juil. 2027	Présentiel - Casablanca
06 au 10 Sep. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-maroc.com
Web : https//www.innov-maroc.com

INNOV MAROC — Tous droits réservés