LLMOps : gestion du cycle de vie des grands modèles de langage en production

Lien : https://innov-systems.com/formation/llmops-gestion-du-cycle-de-vie-des-grands-modeles-de-langage-en-production

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI208

CATÉGORIE
Intelligence Artificielle Générative (IAG)

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre les principes et les composants d'une démarche LLMOps
- Concevoir une architecture de gestion du cycle de vie des LLM
- Mettre en place des pipelines de données et de modèles automatisés
- Déployer des grands modèles de langage de manière sécurisée et scalable
- Superviser les performances et la qualité des modèles en production
- Détecter et gérer les dérives des modèles et des données
- Intégrer des pratiques CI/CD dans les projets d'IA générative
- Optimiser les performances, les ressources et les coûts d'exploitation
- Mettre en œuvre une gouvernance et une conformité adaptées aux LLM
- Assurer la maintenance et l'amélioration continue des modèles
- Industrialiser le déploiement de solutions d'IA générative
- Réaliser des cas pratiques de mise en production d'un LLM

POUR QUI ?

- AI Engineers
- Machine Learning Engineers
- MLOps Engineers
- LLMOps Engineers
- Data Scientists
- Data Engineers
- DevOps et DevSecOps Engineers
- Architectes Cloud
- Architectes IA
- Développeurs d'applications d'IA générative
- Consultants en intelligence artificielle
- Toute personne souhaitant industrialiser le déploiement, l'exploitation et la gouvernance des grands modèles de langage en production

Innov Systems

Programme détaillé

1 / Introduction au LLMOps

- Définition et principes du LLMOps
- Différences entre MLOps, AIOps et LLMOps
- Cycle de vie d'un grand modèle de langage (LLM)

2 / Architecture d'une plateforme LLMOps

- Composants d'une architecture LLMOps
- Environnements de développement, validation et production
- Intégration avec les plateformes cloud et hybrides

3 / Gestion des modèles et des versions

- Model Registry et versioning des LLM
- Gestion des artefacts et des dépendances
- Gouvernance du cycle de vie des modèles

4 / Préparation des données et pipelines LLM

- Collecte, nettoyage et transformation des données
- Gestion des jeux de données et des embeddings
- Automatisation des pipelines de données

5 / Déploiement et orchestration des LLM

- Déploiement avec Docker et Kubernetes
- Serving des modèles et API d'inférence
- Scalabilité, haute disponibilité et équilibrage de charge

Innov Systems

6 / Monitoring et observabilité des LLM

- Surveillance des performances et de la latence
- Détection de dérive (data drift, model drift et concept drift)
- Journalisation, traçabilité et métriques opérationnelles

7 / Sécurité, gouvernance et conformité

- Gestion des accès et des identités
- Protection des modèles, des données et des prompts
- Conformité réglementaire, auditabilité et IA responsable

8 / Optimisation des performances et des coûts

- Quantification, mise en cache (caching) et optimisation de l'inférence
- Gestion des ressources GPU/CPU
- Optimisation des coûts d'exploitation des LLM

9 / Automatisation et intégration CI/CD

- Pipelines CI/CD pour les LLM
- Validation, tests et déploiements automatisés
- Rollback, A/B testing et déploiements progressifs (canary)

10 / Cas pratiques et exploitation en production

- Mise en œuvre d'une plateforme LLMOps complète
- Déploiement d'un LLM en environnement de production
- Bonnes pratiques d'exploitation, maintenance et amélioration continue

Innov Systems

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

18 au 22 Jan. 2027	Présentiel - Casablanca
15 au 19 Mars 2027	Présentiel - Casablanca
10 au 14 Mai 2027	Présentiel - Casablanca
05 au 09 Juil. 2027	Présentiel - Casablanca
30 Août au 03 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-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés