LLMOps — Déploiement, monitoring et industrialisation des modèles LLM

Lien : https://innov-systems.com/formation/llmops-deploiement-monitoring-et-industrialisation-des-modeles-llm

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI186

CATÉGORIE
Intelligence Artificielle Générative (IAG)

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre les principes du LLMOps et de la production des LLM
- Déployer des modèles de langage en environnement industriel
- Mettre en place des architectures RAG performantes
- Gérer le cycle de vie des prompts et modèles
- Superviser et monitorer les systèmes LLM en production
- Évaluer la qualité des réponses générées
- Optimiser les coûts et performances des LLM
- Intégrer les LLM dans des pipelines CI/CD
- Assurer la sécurité et la conformité des applications LLM
- Utiliser les bases vectorielles pour le retrieval
- Industrialiser des solutions IA générative
- Étudier des cas concrets d’applications LLM en entreprise

POUR QUI ?

- Machine learning engineers
- Data scientists
- AI engineers
- MLOps engineers
- LLM engineers
- Software engineers
- Backend developers
- DevOps engineers
- Data engineers
- Architectes IA
- Product engineers IA
- Consultants AI/ML
- Toute personne impliquée dans l’industrialisation des modèles LLM

Innov Systems

Programme détaillé

1 / Introduction au LLMOps

- Définition et enjeux du LLMOps
- Différences entre MLOps et LLMOps
- Cas d’usage des modèles de langage en production

2 / Architecture des systèmes LLM en production

- Composants d’un système LLM (API, prompt layer, retrieval)
- Architecture RAG (Retrieval Augmented Generation)
- Patterns d’intégration des LLM

3 / Déploiement des modèles LLM

- Déploiement via API (OpenAI, Anthropic, open-source)
- Containerisation et serving de modèles
- Gestion de la scalabilité et latence

4 / Gestion des prompts et prompt engineering industriel

- Versioning des prompts
- Prompt templates et chaînes de prompts
- Tests et validation des prompts

5 / Vector databases et systèmes de retrieval

- Utilisation de bases vectorielles (FAISS, Chroma, Pinecone)
- Indexation et recherche sémantique
- Optimisation des embeddings

Innov Systems

6 / Monitoring des LLM en production

- Suivi de la qualité des réponses
- Détection de dérives (hallucinations, drift)
- Logging et observabilité des requêtes

7 / Évaluation et métriques des LLM

- BLEU, ROUGE, embeddings similarity
- Evaluation humaine et automatisée
- A/B testing des modèles

8 / Sécurité et conformité des LLM

- Filtrage des contenus sensibles
- Protection des données utilisateurs
- Gestion des biais et conformité réglementaire

9 / Optimisation et performance

- Réduction des coûts d’inférence
- Caching et batching des requêtes
- Quantization et optimisation des modèles

10 / Industrialisation et cas pratiques

- Pipeline complet LLMOps en production
- Intégration CI/CD pour LLM
- Études de cas d’applications LLM en entreprise

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

08 au 12 Fév. 2027	Présentiel - Casablanca
05 au 09 Avr. 2027	Présentiel - Casablanca
31 Mai au 04 Juin 2027	Présentiel - Casablanca
26 au 30 Juil. 2027	Présentiel - Casablanca
20 au 24 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