RAG (Retrieval-Augmented Generation) : architectures et déploiement en entreprise

Lien : https://innov-systems.com/formation/rag-retrieval-augmented-generation-architectures-et-deploiement-en-entreprise

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI216

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 architectures RAG
- Concevoir une solution RAG adaptée aux besoins de l’entreprise
- Préparer, gouverner et indexer les données pour la recherche sémantique
- Mettre en œuvre des bases de données vectorielles performantes
- Développer des pipelines RAG avec les principaux frameworks
- Déployer des LLM dans des environnements cloud, hybrides ou on-premise
- Sécuriser les données et les applications d’IA générative
- Évaluer et optimiser les performances des systèmes RAG
- Industrialiser une solution RAG avec les pratiques MLOps et LLMOps
- Intégrer le RAG aux applications métiers et aux systèmes d’information
- Réduire les hallucinations et améliorer la pertinence des réponses
- Réaliser des projets concrets de déploiement RAG en entreprise

POUR QUI ?

- Architectes IA
- Architectes solutions
- Data engineers
- Machine Learning engineers
- Ingénieurs IA générative
- Développeurs Python
- MLOps / LLMOps engineers
- DevOps engineers
- Architectes cloud
- Consultants en intelligence artificielle
- Chefs de projet innovation et transformation digitale
- Toute personne impliquée dans la conception, le déploiement ou l’exploitation de solutions d’IA générative en entreprise

Innov Systems

Programme détaillé

1 / Introduction au RAG et aux architectures d’IA générative

- Principes du Retrieval-Augmented Generation (RAG)
- Différences entre LLM seuls, RAG et Fine-Tuning
- Cas d’usage en entreprise

2 / Architecture d’une solution RAG

- Composants d’une architecture RAG
- Flux d’ingestion, d’indexation et de génération
- Intégration avec les systèmes d’information

3 / Collecte, préparation et gouvernance des données

- Sources de données structurées et non structurées
- Nettoyage, segmentation (chunking) et enrichissement
- Gouvernance et qualité des données

4 / Bases de données vectorielles et recherche sémantique

- Embeddings et représentation vectorielle
- ChromaDB, Milvus, Pinecone, Weaviate, FAISS
- Indexation, recherche hybride et optimisation des performances

5 / Orchestration des pipelines RAG

- Frameworks LangChain, LlamaIndex et Semantic Kernel
- Construction des pipelines de récupération et de génération
- Gestion du contexte et optimisation des prompts

Innov Systems

6 / Déploiement des modèles de langage

- LLM propriétaires et open source
- Déploiement sur cloud, on-premise et hybride
- Optimisation des performances, coûts et latence

7 / Sécurité, conformité et gouvernance IA

- Contrôle des accès aux données
- Protection des informations sensibles et conformité réglementaire
- Prévention des fuites de données et sécurité des modèles

8 / Évaluation et optimisation des systèmes RAG

- Mesure de la pertinence des réponses
- Réduction des hallucinations et amélioration de la précision
- Monitoring, observabilité et amélioration continue

9 / Industrialisation et exploitation en production

- Architecture scalable et haute disponibilité
- CI/CD, MLOps et LLMOps
- Supervision, journalisation et gestion des versions

10 / Cas pratiques et déploiement d’une solution RAG

- Mise en œuvre d’un assistant documentaire d’entreprise
- Intégration avec Microsoft 365, SharePoint, CRM ou ERP
- Bonnes pratiques et retours d’expérience

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

15 au 19 Fév. 2027	Présentiel - Casablanca
12 au 16 Avr. 2027	Présentiel - Casablanca
07 au 11 Juin 2027	Présentiel - Casablanca
02 au 06 Août 2027	Présentiel - Casablanca
27 Sep. au 01 Oct. 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