Conception d’Applications RAG avec LLM et Agents IA

Lien : https://innov-systems.com/formation/conception-dapplications-rag-avec-llm-et-agents-ia

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI467

CATÉGORIE
Intelligence Artificielle

OBJECTIFS DE LA FORMATION

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

- Comprendre les concepts fondamentaux des LLM et leurs limites pour un usage raisonné
- Concevoir, développer et optimiser une application RAG complète de bout en bout
- Maîtriser LlamaIndex pour la gestion et l’indexation des données
- Construire des workflows complexes avec LangGraph
- Intégrer et orchestrer des agents IA au sein d’un pipeline RAG avec smolagents
- Appliquer les meilleures pratiques de performance, de sécurité et de déploiement en production

POUR QUI ?

- Développeurs, data scientists et ingénieurs en intelligence artificielle
- Chefs de projets techniques souhaitant intégrer le RAG dans leurs solutions
- Architectes de solutions IA cherchant à optimiser la recherche et l’intégration de données non structurées

Innov Systems

Programme détaillé

1 / FONDAMENTAUX DES LLM ET DU RAG

- Historique et évolution de l'intelligence artificielle
- Comprendre les LLM : structure, fonctionnement et limites
- Découverte du Transformer et de ses composants clés
- Concepts essentiels : tokens, embeddings, hyperparamètres, couches
- Panorama des modèles du marché : GPT, BERT, LLaMA et alternatives open-source
- Introduction à l'architecture RAG : principes, avantages et cas d'usage
- Les briques essentielles : base vectorielle, modèle sémantique, recherche par proximité

2 / ENVIRONNEMENT DE TRAVAIL ET OUTILS

- Présentation et prise en main de Google Colab pour le prototypage rapide
- Installation et configuration des bibliothèques RAG (LangChain, LlamaIndex, LangGraph, smolagents)
- Bonnes pratiques d'organisation de projet IA
- Gestion des dépendances et versions

3 / DÉVELOPPEMENT D'UN PIPELINE RAG DE A À Z

- Assemblage des briques pour construire un pipeline RAG fonctionnel
- Préparation et pré-traitement des données : nettoyage, normalisation, découpage
- Création et optimisation des embeddings : taille, précision, coût
- Gestion des erreurs et résolution des problèmes fréquents
- Évaluation de la pertinence des réponses avec des métriques spécifiques
- Optimisation des hyperparamètres pour de meilleures performances

4 / INDEXATION ET REQUÊTES AVANCÉES AVEC LLAMAINDEX

- Architecture et fonctionnement interne de LlamaIndex

Inov Systems

- Structurer et enrichir ses données pour l’indexation
- Génération et stockage des représentations vectorielles
- Construire et exécuter des requêtes sémantiques complexes
- Stratégies pour gérer de grands volumes de données

5 / CRÉATION DE WORKFLOWS INTELLIGENTS AVEC LANGGRAPH

- Introduction à LangGraph et à ses concepts : nœuds, arêtes, états persistants
- Conception de workflows modulaires pour le RAG
- Intégration de multiples sources et traitements dans un même flux
- Gestion de la persistance et reprise après interruption
- Mise en place de branches conditionnelles et de traitements parallèles

6 / INTÉGRATION D’AGENTS IA AVEC SMOLAGENTS

- Qu’est-ce qu’un agent IA ? Rôle et bénéfices dans un RAG
- Présentation de smolagents et de son approche modulaire
- Création d’agents spécialisés (recherche, extraction, analyse)
- Orchestration d’agents multiples dans un workflow LangGraph
- Cas pratiques : RAG multi-agents pour des scénarios complexes

7 / OPTIMISATION, SÉCURITÉ ET BONNES PRATIQUES

- Optimisation des performances : réduction de latence et gestion de mémoire
- Mise en cache des résultats et réutilisation des embeddings
- Sécurisation des données sensibles et filtrage des réponses
- Stratégies pour limiter les hallucinations des LLM
- Tests et validation des pipelines RAG

8 / DÉPLOIEMENT ET MISE EN PRODUCTION

- Choix des environnements de déploiement : cloud, serveur dédié, conteneurs
- Intégration avec des API et systèmes externes

Inov Systems

- Surveillance et maintenance d’une application RAG en production
- Retour d’expérience : exemples de déploiements réussis
- Perspectives d’évolution et veille technologique

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

19 au 22 Jan. 2027	Présentiel - Casablanca
16 au 19 Mars 2027	Présentiel - Casablanca
11 au 14 Mai 2027	Présentiel - Casablanca
06 au 09 Juil. 2027	Présentiel - Casablanca
07 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-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés