LLMs Et Applications Génératives Avancées

Lien : https://innov-systems.com/formation/llms-et-applications-generatives-avancees

DURÉE	RÉFÉRENCE	CATÉGORIE
4 jours (28h)	**BSI461**	**Intelligence Artificielle**

OBJECTIFS DE LA FORMATION

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

- Comprendre en profondeur les concepts, architectures et usages des LLMs
- Maîtriser les techniques avancées de prompt engineering pour optimiser les réponses
- Mettre en œuvre des stratégies RAG robustes et performantes
- Intégrer et utiliser efficacement les bibliothèques LangChain et LlamaIndex
- Structurer, vectoriser et stocker des données pour un usage optimal avec LanceDB
- Personnaliser un modèle via fine-tuning et optimisation du retrieval
- Évaluer et améliorer les performances d’un système LLM en conditions réelles
- Déployer une application LLM sécurisée, performante et évolutive

POUR QUI ?

- Développeurs, Data Scientists, Data Engineers, ML Engineers
- Architectes logiciels et chefs de projets IA
- Toute personne impliquée dans la conception, l’intégration ou le déploiement de solutions basées sur les LLMs

Innov Systems

Programme détaillé

1 / COMPRENDRE LES PRINCIPES ET L'ARCHITECTURE DES LLMS

- Fonctionnement des transformers et attention mechanism
- Différences entre modèles de langage, modèles multimodaux et modèles spécialisés
- Impact de la taille du modèle et des jeux de données sur les performances

2 / PANORAMA ACTUEL DES TECHNOLOGIES ET OUTILS

- Modèles propriétaires : GPT-4, Gemini, Claude
- Modèles open-source : LLaMA 3, Mistral, Falcon
- Écosystème logiciel : LangChain, LlamaIndex, API REST, SDKs

3 / LES BASES DU PROMPT ENGINEERING

- Types de prompts : instruction, conversationnel, contexte enrichi
- Chaînage de prompts et gestion des erreurs
- Exemples et bonnes pratiques

4 / INTRODUCTION AU RAG (RETRIEVAL AUGMENTED GENERATION)

- Principe, avantages et cas d'usage
- Intégration de données externes pour enrichir les réponses
- Mises en pratique : créer un "Hello World" RAG avec LlamaIndex

5 / ARCHITECTURE D'UNE PIPELINE RAG

- Récupération, nettoyage et prétraitement des données
- Vectorisation et stockage avec LanceDB
- Intégration d'API LLM (locale et cloud)

Innov Systems

6 / ACQUISITION ET PRÉPARATION DES DONNÉES

- Extraction depuis des sites web et documents divers
- Normalisation et formats optimisés pour LLMs
- Utilisation des embeddings pour la recherche contextuelle

7 / MISE EN ŒUVRE D'UN RAG COMPLET

- Indexation et requêtes en langage naturel
- Création d'une pipeline de bout en bout avec LlamaInde
- Mises en pratique : indexer la documentation Hugging Face et interroger en langage naturel

8 / INTRODUCTION AU FINE-TUNING

- Distinction entre fine-tuning complet, LoRA et prompt-tuning
- Utilisation de l'écosystème Hugging Face pour l'entraînement

9 / OPTIMISATION DU RETRIEVAL ET RE-RANKING

- Améliorer la pertinence des résultats avec des modèles de re-ranking
- Combinaison RAG + fine-tuning pour des cas d'usage spécifiques
- Mises en pratique : améliorer la recherche documentaire avec re-ranking

10 / GESTION DES PERFORMANCES ET DE LA QUALITÉ

- Métriques : MMLU, BLEU, ROUGE, LMSYS
- Tests de robustesse et de scalabilité
- Optimisation des coûts d'inférence

11 / SÉCURITÉ ET GOUVERNANCE DES LLMS

- Gestion des données sensibles et confidentialité
- Filtrage et modération de contenu généré

Inov Systems

12 / MODES DE DÉPLOIEMENT

- Déploiement cloud (AWS, Azure, GCP)
- Déploiement on-premise et hybride
- Conteneurisation avec Docker et orchestration avec Kubernetes

13 / CRÉATION ET DÉPLOIEMENT D'UNE APPLICATION FINALE

- Intégration avec une interface web ou mobile
- API REST pour consommation externe
- Déploiement et monitoring en production

14 / ATELIER FINAL : RÉALISATION D'UNE APPLICATION RAG PERSONNALISÉE

- De l'acquisition de données au déploiement final
- Présentation et démonstration par les participants

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

Inov Systems

Prochaines dates programmées

12 au 15 Jan. 2027	Présentiel - Casablanca
09 au 12 Mars 2027	Présentiel - Casablanca
04 au 07 Mai 2027	Présentiel - Casablanca
29 Juin au 02 Juil. 2027	Présentiel - Casablanca
31 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

Innov Systems