



ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

Formation Conception d'applications IA générative - LLM, RAG et architectures agentiques

Développer de nouvelles formes de contenus grâce aux LLM

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

La formation Conception d'applications IA générative : LLM, RAG et architectures agentiques constitue un socle technique avancé pour maîtriser la conception de solutions basées sur les modèles de langage (LLM). Elle permet de comprendre en profondeur les mécanismes des Transformers, du prompt engineering avancé, des architectures RAG (Retrieval Augmented Generation) et des approches de personnalisation comme le fine-tuning, indispensables pour concevoir des applications d'IA générative performantes, fiables et adaptées aux besoins métier. Orientée pratique et innovation, cette formation aborde également les architectures agentiques, le raisonnement autonome (ReAct), les outils LLMOps et les frameworks modernes tels que LangChain et LangGraph, afin de préparer efficacement les participants à passer de l'expérimentation à l'industrialisation de projets IA.

Objectifs

- | Comprendre l'utilité des LLM
- | Maîtriser l'usage des LLM et de l'IA générative
- | Développer de nouveaux projets avec l'aide des LLM

Public

- | Data Scientists
- | Chefs de projets
- | Ingénieurs
- | Architectes IA

Prérequis

- | Une connaissance de base des principes de Machine Learning et de Deep Learning
- | La maîtrise d'une langage de script type Python est recommandé

Programme de la formation

Comprendre le moteur : architecture et fonctionnement des LLM

- | Changement de paradigme de l'IA et évolution des LLM (lois d'échelle)
- | Fondamentaux du traitement automatique du langage naturel (NLP)
- | Transition vers l'ère post Deep Learning
- | Architecture des Transformers et principes fondateurs
- | Rôle de l'encodeur et du décodeur
- | Tokenisation et algorithme Byte Pair Encoding (BPE)
- | Fonctionnement de l'encodeur : embeddings et positionnement
- | Mécanisme d'attention multi-têtes
- | Spécialisation des têtes d'attention
- | Fonctionnement du décodeur et auto-attention masquée
- | Étapes d'entraînement d'un modèle Transformer
- | Évolutions récentes : décodeur seul et encodage RoPE
- | Nouvelles architectures : FlashMLA et Mixture of Experts (MoE)

Référence	BI108
Durée	2 jours (14h)
Tarif	2 050 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 12 au 13 octobre 2026
du 5 au 6 novembre 2026*
du 10 au 11 décembre 2026

[VOIR TOUTES LES DATES](#)

(*) session confirmée

Stratégie de sélection et écosystèmes LLM

- | Critères de sélection des modèles selon les cas d'usage
- | Différences entre modèles encodeur (BERT) et décodeur (GPT)
- | Panorama des modèles avancés (DeepSeek, GPT-4)
- | Optimisation des modèles avec LLaMA
- | APIs de modèles propriétaires (OpenAI, Anthropic)
- | Utilisation du Hub Hugging Face
- | Introduction à LangChain
- | Outils LLMops (type MLflow)

Design d'interactions : Prompt Engineering

- | Notion de token et distribution probabiliste
- | Paramétrage avancé (Top-K, Top-P, température)
- | Structure d'un prompt efficace
- | Instruction tuning et délimiteurs
- | Décomposition des tâches complexes
- | Techniques Zero-shot, One-shot, Few-shot
- | Templates de prompts avec LangChain
- | Chain-of-Thought
- | Self-consistency
- | Tree of Thoughts

Conception de systèmes augmentés (RAG et agents)

- | Limites du prompt seul
- | Gestion de la mémoire (court / long terme)
- | Types de mémoire avec LangChain
- | Architecture RAG
- | Découpage des documents
- | Génération d'embeddings
- | Bases vectorielles
- | Implémentation avec FAISS
- | RAG avancé (Corrective / Adaptive)
- | Agentic RAG
- | Intégration d'outils externes
- | Méthode ReAct

Fine-Tuning et déploiement

- | Cas d'usage du Fine-Tuning
- | Principes du réglage fin
- | Supervised Fine-Tuning (SFT)
- | RLHF
- | Stratégies d'adaptation des modèles
- | PEFT
- | LoRA
- | QLoRA
- | Model Serving
- | Model Registry
- | Déploiement sur endpoint
- | Évaluation des modèles (qualité, toxicité, tokens)
- | Métriques de lisibilité et LLM-as-a-Judge

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

Accessibilité



Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.