

Formation LLMs pour data science augmentée

Présentation

Cette formation avancée couvre la conception, l'optimisation et l'évaluation de solutions de data science augmentée par les LLMs. En 2 jours, les participants conçoivent des pipelines RAG/Multi-RAG, orchestrent des agents outillés (tools/Function Calling), optimisent prompts et hyperparamètres (temperature, top-p), et mettent en place des métriques d'évaluation (precision@k, faithfulness). Le cours alterne théorie, démonstrations et TP guidés sur des cas d'usage data et MLOps.

Durée : 14,00 heures (2 jours)

Tarif INTRA : Nous consulter

Objectifs de la formation

- Concevoir des architectures LLM-centric pour la data science (RAG, agents, toolformer) adaptées aux cas d'usage et aux contraintes métier
- Optimiser le prompt engineering avancé (prompt chaining, self-consistency, ReAct) et les paramètres d'inférence pour la robustesse et la performance
- Mettre en place un pipeline RAG (indexation, chunking, embeddings, retrieval) et l'intégrer à un workflow analytique
- Orchestrer des agents outillés avec gestion d'outils (API/SQL/Pandas) et contrôle de l'hallucination (guardrails, schema-guided)
- Évaluer et monitorer les systèmes LLM (métriques intrinsèques/extrinsèques, evals basées sur LLM, tests unitaires de prompts) et industrialiser (MLOps/LLMOps)

Prérequis

- Solides bases en data science (pandas, scikit-learn), manipulation de données et feature engineering
- Connaissance des API de LLM (appels, paramètres, coûts) et des embeddings
- Notions de MLOps (environnements, versioning données/modèles)

Public



Data scientists, ML engineers, data/AI engineers, lead data, architectes IA en contexte entreprise

Programme de la formation

Jour 1 – Architecture LLM-centric, RAG avancé et prompt engineering

Session du matin :

- Panorama des patterns LLM pour la data science : RAG, multi-RAG, agents, tools, function calling, guardrails
- Conception d'un pipeline RAG : préparation du corpus, stratégies de chunking (fixed, semantic), embeddings (choix, dimensions), index (FAISS/Chroma)
- Prompt engineering avancé : ReAct, chain-of-thought (sans déverbalisation sensible), self-consistency, few-shot sélection (semantic similarity)

TP / Exercice :

- Construire un RAG pour QA analytique

Session de l'après-midi :

- Implémentation d'un RAG de bout en bout : ingestion, indexation, retrieval (BM25 vs vectoriel), re-ranking, fusion
- Réglages d'inférence : température, top-p, top-k, fréquence/presence penalty, longueur de contexte et coût/performance
- Observabilité et sécurité : traces, guardrails (schémas, listes contrôlées), gestion de PII et conformité

TP / Exercice :

- Construire un RAG pour QA analytique : créer un corpus, ajuster le chunking et les embeddings, comparer les retrieveurs, intégrer un re-ranker et évaluer precision@k et exact match. Livrable : notebook avec pipeline RAG, métriques et justification des choix

Points clés & takeaways :

- Savoir concevoir et évaluer un pipeline RAG robuste
- Maîtriser les prompts et paramètres d'inférence pour la stabilité et les coûts

Jour 2 – Agents outillés, LLMOps et évaluation des systèmes**TP / Exercice :**

- Orchestrer un agent analytique outillé
- Agents et tools pour la data science : planification, exécution outillée (SQL, Pandas, API), function calling/schema-guided
- Contrôle d'hallucination et robustesse : verification steps, tool-use first, constrained decoding, règles/guardrails
- Évaluation : métriques extrinsèques (task success, latency, coût), métriques intrinsèques (faithfulness, groundedness), évaluations par LLM et tests de régression de prompts

Session de l'après-midi :

- LLMOps : versioning prompts/datasets, traces, monitoring, AB testing, canary, budget/cost guards
- Déploiement : API, conteneurisation, secrets, gestion des clés et journalisation
- Gouvernance et risques : biais, conformité, sécurité des données, policy enforcement

TP / Exercice :

- Orchestrer un agent analytique outillé : mise en œuvre d'un agent capable de répondre à des questions business via SQL/Pandas et vérification de réponses. Livrable : notebook et script d'évaluation (scénarios, métriques) avec rapport d'itérations

Points clés & takeaways :

- Concevoir des agents fiables avec outillage contrôlé

- Structurer un cadre d'évaluation et d'industrialisation (LLMOps)

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes. Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétences.

Moyens pédagogiques et techniques

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation :

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

Tout au long de la formation :

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

A la fin de la formation :

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Evaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.