



<https://latribunelibre.com/emploi/ingenieur-ia-f-h-9>

Ingénieur IA F/H

Description

Au sein de l'IA Factory, nous sommes à la recherche d'un Ingénieur Intelligence Artificielle F/H pour renforcer nos équipes.

Vos missions seront les suivantes :

Prototypage rapide : Bâtir rapidement des preuves de concept (PoC) fonctionnelles pour valider les hypothèses produites.

Orchestration de LLM : Maîtriser les frameworks d'orchestration (ex: LangChain) pour construire des flux complexes (chaînes, agents, RAG).

Intégration de données (RAG) : Concevoir et implémenter des systèmes de Retrieval-Augmented Génération (RAG) en connectant les LLM à nos sources de savoir (bases SQL, bases vectorielles).

Prompt engineering & stratégie : Aller au-delà du prompt de base pour développer des stratégies d'interaction complexes afin de maximiser la fiabilité et la précision des modèles.

Développer et déployer des API d'IA sur le cloud privé du groupe ou sur la plateforme Google Cloud Platform (GCP) pour résoudre des problèmes spécifiques à l'entreprise, en se concentrant sur la modularité, la performance et la scalabilité, dans le respect des pratiques de développement du groupe.

Définir et maintenir l'architecture IA hybride (on-prem et GCP) et portable pour toutes les phases du MLOps (développement, inférence, suivi de modèle)

Veille technologique active : Évaluer et intégrer en continu les nouveaux modèles, outils et techniques de l'écosystème IA

Travailler en étroite collaboration avec les équipes de développement logiciel pour intégrer les modèles dans les applications existantes.

Qualifications

De formation bac+5 en Informatique ou Intelligence Artificielle, vous disposez d'au moins deux ans d'expérience dans le domaine d'ingénierie en IA dont au moins un an d'expérience dans la mise en production et l'orchestration des modèles génératifs.

Vos compétences techniques :

Maîtrise avancée de Python pour le développement applicatif. Vous écrivez du code propre, testable et performant.

Maîtrise des frameworks GenAI : Expérience solide et démontrée avec LangChain.

Compréhension des LLM : Connaissance approfondie du fonctionnement des modèles de langage.

Développement d'API : Solide expérience en création d'API REST.

Bases de données vectorielles : Compréhension et expérience pratique avec les VDB (ex: pgvector) pour les cas d'usage sémantiques.

Expérience en DevOps.

Maîtrise de Google Cloud Platform et de ses services (Kubernetes).

Capacité à appréhender les contraintes liées au développement de produits d'IA dans le secteur bancaire et assurantiel.

Organisme employeur
CREDIT MUTUEL ARKEA

Type de poste
Temps plein

Secteur
AUTRES INTERMÉDIATIONS
MONÉTAIRES

Lieu de poste
29235, LE RELECQ KERHUON, LE
RELECQ KERHUON, France

Date de publication
22 décembre 2025 à 16:03

Valable jusqu'au
21.01.2026

Si possible, connaissance de gestion de workflows complexes (Outil utilisé : Temporal).

Vos qualités personnelles :

Capacité à travailler en équipe et à collaborer avec les autres services,

Autonomie et curiosité : Capacité à naviguer dans un écosystème qui change chaque semaine et à apprendre par vous-même.

Vous savez vous organiser et bien gérer votre temps.

Vous êtes doté d'un esprit critique et faites preuve de curiosité intellectuelle.

Vous êtes reconnu pour votre sens de la communication et votre grande adaptabilité.

Le processus de recrutement :

1. Je postule à une offre et je reçois un mail de confirmation,
2. Après un premier échange téléphonique, je reçois des codes de connexion pour passer des tests en ligne,
3. Si les tests sont positifs, je passe un entretien avec un responsable RH et un manager opérationnel,
4. Enfin, si ma candidature est validée, un responsable RH me fait une proposition.