

Offre de stage (6 mois) – Ingénieur IA F/H

XENOCS est un groupe international spécialisé dans le développement et la commercialisation de solutions innovantes pour la caractérisation de matériaux à l'échelle nanométrique en utilisant la diffusion des Rayons X. Notre clientèle internationale est composée de chercheurs travaillant dans des universités ou des laboratoires de recherche ainsi que de fabricants d'équipements d'analyse. La société, créée en 2000, est issue d'un essaimage de l'Institut Laue Langevin. Son siège social est à Grenoble en Isère. Le groupe compte quatre filiales situées aux Etats-Unis, au Danemark et en Chine pour un effectif d'environ 90 salariés.

Nous recherchons un(e) :

Stagiaire Ingénieur IA F/H

Nous vous proposons de rejoindre, dans un groupe international en croissance, une équipe multiculturelle, dynamique et innovante travaillant en mode collaboratif dans de nouveaux locaux situés sur la Presqu'île de Grenoble, conçus dans un esprit Qualité de Vie au Travail favorisant l'efficacité et la performance de nos collaborateurs.

MISSION

Dans le cadre du développement du projet **JARVIS**, assistant IA de cybersécurité autonome de Xenocs, vous participerez à la conception et à la mise en œuvre d'une infrastructure d'intelligence artificielle souveraine.

L'objectif du projet est de permettre à cet assistant d'interroger en temps réel les systèmes de sécurité des différentes entités du groupe (France, États-Unis, Chine), de répondre aux utilisateurs en langage naturel et de déclencher automatiquement, ou sous conditions, des actions de remédiation.

Les modèles open source seront inférés localement sur une infrastructure équipée de GPU. Vous travaillerez en étroite collaboration avec l'équipe IT (administration systèmes et réseaux et DSI) dans un environnement open source et souverain, limitant l'utilisation de services cloud externes au strict nécessaire.

Ce stage offre l'opportunité de participer à la mise en production d'un projet d'intelligence artificielle (LLM, NL2Query) depuis la preuve de concept jusqu'au déploiement opérationnel.

Architecture d'inférence hybride :

- Déployer vLLM pour servir les modèles lourds (Llama, Mistral, Codestral)
- Mettre en place une architecture d'inférence hybride avec Ollama
- Déployer un pipeline de mise à jour des modèles sans exposition des données

Développement du module NL2Query :

- Créer un module traduisant les questions utilisateurs en requêtes exploitables par les systèmes internes
- Implémenter des garde-fous afin de limiter les erreurs et les hallucinations des modèles

Fiabilité et score de confiance :

- Mettre en place une méthode heuristique de scoring basée sur la cohérence syntaxique SQL, la validité sémantique et la pertinence de l'intention utilisateur

Sécurisation des accès :

- Mettre en place une gestion des accès basée sur RBAC (Role Based Access Control)

Stack technique :

Base de données	Elasticsearch, Clickhouse
IA-Inference	vLLM, Ollama, Llama, nomic-embed-text, claude code et codestral (local)
Orchestration	Langchain, Llamaindex
Sécurité	Nemo Guardrails
Interface	Chainlit, FastAPI (interaction avec N8N)
Ingestion	Vector, fortidragon, elastic integration (crowdstrike), Cato networks github
Infrastructure&OS	Docker compose, Debian, AD, AZUR AD

Livrables attendus :

- Architecture technique documentée
- Module NL2Query opérationnel (MVP robuste)
- Rapport d'évaluation du système (précision, hallucination, performance)
- Pipeline de mise à jour des modèles IA en environnement air-gapped
- Démonstration exploitable en interne

PROFIL DU TITULAIRE

Diplôme, formation initiale : Étudiant(e) en École d'Ingénieur ou 2^{ème} année de Master IA / Data Science

Connaissances spécifiques requises :

- Maîtriser Python et les frameworks LLM (LangChain ou LlamaIndex)
- Connaissance des outils d'inférence (vLLM, Ollama)
- Aisance avec Elasticsearch, SQL ou ClickHouse

- Familiarité avec les environnements Linux et Docker
- Capacité à communiquer à l'écrit en anglais

Qualités souhaitées :

Vous êtes reconnu(e) pour votre curiosité technique et votre capacité à vous adapter à des environnements innovants ?

L'étudiant(e) que nous recherchons est rigoureux(se), autonome et force de proposition.

Curieux(se) et ouvert(e) d'esprit, il/elle sait collaborer avec différents interlocuteurs techniques et souhaite contribuer au développement d'une infrastructure d'IA souveraine performante et sécurisée.

Type de contrat : Stage conventionné, gratification selon dispositions légales

Lieu de travail : 1-3 Allée du Nanomètre, 38000 Grenoble

Les candidatures (CV et lettre de motivation) sont à adresser par mail à recrutement@xenocs.com.