

Stage Ingénieur Développement IA

Titre : Conception d'une pipeline de traitement de données multimodales pour une architecture LLM x KAG x RAG

Lieu : Bordeaux – Village by CA, 11 cours du 30 Juillet

Début : Janvier ou février 2026 – **Durée :** 5 à 6 mois

Organisation : Télétravail partiel possible

Rémunération : Selon profil

Candidature : CV + court mail de motivation à envoyer à rh@daryl-ai.com

Contexte de l'entreprise

DARYL révolutionne la vente de produits techniques dans l'industrie grâce à un assistant IA capable de générer, en quelques secondes, des diagnostics techniques et des devis commerciaux connectés aux outils CRM, CPQ et ERP.

Notre solution est déjà déployée chez plusieurs PME industrielles qui l'utilisent pour mieux qualifier leurs leads, accélérer leurs cycles de vente et réduire fortement les délais de réponse aux clients.

Basée au Village by CA Bordeaux, DARYL connaît une croissance soutenue portée par la demande d'outils d'IA appliqués au CPQ industriel.

Objectif du stage

L'architecture agentique de DARYL repose sur un processus robuste de préparation de données pour alimenter à la fois les bases vectorielles (RAG), le graphe de connaissance (KAG) et les tools agentiques (SmolAgent).

Les données à traiter sont multimodales (PDF, fichiers Excel avec formules, images, etc.). Vous aurez pour mission de concevoir une pipeline automatisée combinant modèles génératifs et traitements algorithmiques classiques, afin de limiter au maximum les aspects non déterministes induits par l'usage des LLM et de fiabiliser la qualité des données ingérées.

Missions principales

Vous proposerez une ou plusieurs architectures de pipeline de traitement de données visant à remplacer l'existant et à réduire au minimum les interventions humaines, notamment dans l'évaluation des outputs des modèles génératifs.

Vous définirez, mettrez en place et collecterez un ensemble d'indicateurs permettant de benchmarker les différentes architectures proposées, tant sur la qualité des données produites que sur les performances (coûts, latence, robustesse).

Vous ferez évoluer l'architecture retenue dans une approche itérative et incrémentale, en tenant compte du fait que, dans un contexte produit, le système est continuellement enrichi par de nouveaux documents et doit rester efficace et frugal dans le temps.

Profil recherché

Vous êtes étudiant(e) en école d'ingénieur ou en Master (IA, data, mathématiques appliquées) et vous souhaitez travailler au cœur d'une stack IA moderne.

Vous disposez de bonnes bases en data engineering, NLP, LLM, modèles d'embedding, modèles de reasoning, ETL et Python, et vous êtes à l'aise avec la manipulation de données à grande échelle.

Vous avez idéalement une expérience (ou un fort intérêt) pour des outils comme Docling, marker-pdf, DSA, LangChain, LangGraph et pour les infrastructures cloud.

Vous appréciez les systèmes complexes et industriels, et vous êtes sensible aux enjeux d'IA explicable et de fiabilité des pipelines de données.

Encadrement & environnement

Vous serez encadré(e) directement par le CTO de DARYL et intégré(e) à une équipe pluridisciplinaire mêlant IA, produit, tech et business.

Vous évoluerez dans un cadre stimulant au cœur de l'écosystème French Tech Bordeaux, avec la possibilité de travailler sur des cas d'usage concrets et rapidement déployés chez des clients industriels.