

# Missions de Stage AXA

## **1. Acquisition & prétraitement des données**

- Collecte et nettoyage de documents scannés (PDF, images)
- Extraction du texte et de la structure via des modèles OCR
- Enrichissement sémantique via NER
- Constitution de corpus annotés de haute qualité
- Technos : LayoutLMv3, TrOCR, Tesseract OCR 5, VLMs, spaCy

## **2. Modélisation & fine-tuning**

- Fine-tuning de LLMs via LoRA / QLoRA
- Adaptation de SLMs / TLMs pour des tâches NLP ciblées
- Évaluation des performances (BLEU, ROUGE, F1, OCR accuracy)
- Spécialisation de modèles sur des cas métiers AXA
- Technos : Hugging Face Transformers, PEFT, LoRA

## **3. Pipelines & orchestration**

- Construction de pipelines ETL & Feature Engineering
- Intégration des modèles dans des APIs REST
- Garantie de la reproductibilité et de la traçabilité
- Technos : Kedro, FastAPI, Docker

## **4. Entraînement & optimisation**

- Entraînement des modèles sur Azure ML / OpenShift AI
- Hyperparameter tuning (Optuna, Ray Tune)
- Arbitrage performance / coûts

## **5. Déploiement & MLOps**

- Containerisation des modèles

- Déploiement sur Kubernetes via Helm / Kustomize
- Mise en place de pipelines CI/CD
- Sécurisation, scalabilité et robustesse des services
- Technos : Docker, Kubernetes, Azure DevOps

## 6. Monitoring & maintenance

- Suivi de la performance en production
- Détection de dérives (concept drift, latence, erreurs)
- Automatisation du retraining
- Technos : OpenTelemetry, Dynatrace, Azure ML Pipelines

## 7. Collaboration & documentation

- Code reviews, pair programming
- Documentation technique et MLOps
- Outils : Markdown, MkDocs, Confluence

### Votre profil

Vous êtes étudiant en Master 2 ou dernière année de cycle d'ingénieur spécialisé en Data Science, Statistiques, Mathématiques appliquées, Informatique, IA. Vous recherchez un stage de fin d'études (PFE) de 6 mois à partir de février / avril 2026.

### Compétences techniques :

#### Programmation

- Python (pandas, numpy, PyTorch)

#### Machine Learning & NLP

- Classification, régression, clustering
- Transformers & LLMs (BERT, GPT, LLaMA, T5)
- OCR & Document AI (spaCy, Tesseract, LayoutLM, TrOCR)

#### MLOps & Cloud

- Git & bonnes pratiques de code

- Pipelines ML
- Microservices
- Docker, Kubernetes, Helm, Kustomize
- Expérience Cloud (Azure apprécié)