



OFFRE DE STAGE

MEng - MSc
Architecture Neuro-Symbolique

1. STAGE

Période de stage :

Localisation : Villeurbanne (déplacements à prévoir sur Clermont-Ferrand)

Nous proposons à nos stagiaires de participer à la mise en place d'une plateforme d'aide à la conformité réglementaire en y développant de nouvelles fonctionnalités, des algorithmes d'analyses et d'explorer de nouvelles pistes pour rendre la conformité plus accessible.

Cependant l'application sur des textes réglementaires représentent des limites techniques avec l'utilisation de modèles de NLP classiques basés sur les réseaux de neurones : ils sont adaptés pour lire des milliers de pages de normes, comprendre le langage naturel et résumer des concepts. Mais par leur caractère probabiliste, ils peuvent "halluciner" ou faire des erreurs de logique subtiles. Dans le nucléaire ou l'aéronautique, ces erreurs sont inacceptables. Par ailleurs, l'extraction automatique de graphes de connaissances à partir de textes normatifs pose des difficultés spécifiques, liées au caractère prescriptif du langage, à la complexité des structures conditionnelles et à l'exigence d'une interprétation juridiquement rigoureuse et cohérente des obligations formulées.

Pour tenter de remédier à ce défaut, nous souhaitons implémenter de l'IA "Symbolique" basée sur les graphes de connaissances en quoi peuvent être décomposées les textes normatifs et réglementaires basés sur des règles strictes (SI A ALORS B, SI A>x ALORS CONFORME...)

Ce projet vise à marier les deux : utiliser la puissance de lecture des LLMs pour nourrir une structure logique rigoureuse (le Graphe) et automatiser avec un haut degré de fiabilité le transfert du document initial en un format adapté à notre plateforme dynamique.

Ce stage sera mené en collaboration scientifique avec le laboratoire LIMOS, UMR CNRS, Université Clermont Auvergne, localisé à Clermont-Ferrand.

Le travail effectué pourra être poursuivi dans le cadre d'une thèse CIFRE.



Missions :

- Effectuer l'état de l'art des techniques et outils d'extraction de document, de construction de graphes de connaissances (graphes de connaissances ou ontologies ?). Analyser plus spécifiquement les méthodes d'extraction adaptées aux normes ainsi que l'apport et limites des LLMs dans ce contexte et identifier les bonnes pratiques.
- Etudier la modélisation cible : que représenter dans le graphe de connaissances ? dans quelle mesure peut-on exploiter des ontologies juridiques existantes ?
- A partir de l'ontologie actuelle, évaluer la faisabilité d'un pipeline d'extraction du contenu réglementaire et de mise en forme dans une base de données graphe.
- Mettre en place des indicateurs d'évaluation de la qualité

Merci de transmettre votre CV et votre lettre de motivation à l'adresse mail ci-dessous :

Personne à contacter : Jean GILBERT

Mail : jean.gilbert@seals.co

2. ENTREPRISE

2.1. À propos de SEALS

SEALS est une entreprise deeptech qui développe des solutions de conformité réglementaire à destination des industriels. À travers sa plateforme HIRI, SEALS ambitionne de permettre aux ingénieurs du monde entier d'identifier, comprendre et résoudre efficacement des problématiques techniques complexes, en s'appuyant sur une méthodologie d'ingénierie pilotée par les normes. HIRI permet de déterminer les textes réglementaires applicables, d'en extraire les exigences techniques essentielles à évaluer, et d'assurer la traçabilité de ces évaluations tout au long du cycle de vie des équipements et des installations.

SEALS s'appuie également sur un Bureau d'Études intégré, composé d'experts techniques intervenant directement sur le terrain auprès d'industriels. Cette proximité opérationnelle permet à l'entreprise de rester en prise directe avec les réalités technico-réglementaires, d'alimenter en continu la plateforme, et de garantir l'adéquation des solutions développées aux besoins réels du terrain.

Nous sommes 20 dans l'équipe dont 3 associés avec des expériences professionnelles dans différents secteurs, qui donne à l'entreprise une forte culture technique.

L'entreprise bénéficie de l'agrément CII/CIR du ministère de la Recherche et du label Deeptech de Bpifrance, gages de son expertise technique.

2.2. Notre approche

Notre entreprise se distingue par une organisation où chaque employé est au contact direct des différents services de l'entreprise. Cette approche vise à renforcer la compréhension globale des projets, offrant un défi intellectuel, une montée en compétences rapide et une efficacité accrue de chaque expert.

Nous accordons une grande importance aux méthodes de travail, car elles nous permettent de traiter des sujets techniques et exigeants tout en garantissant une qualité constante dans nos résultats et nos livrables.

Nous encourageons une démarche d'amélioration continue, en permettant à chacun de proposer des idées pour faire évoluer nos méthodes de travail et en consacrant chaque semaine du temps à l'amélioration et à la création d'outils internes. Nous mettons donc un fort accent sur la formation des stagiaires et apprentis.

3. PROFIL RECHERCHÉ

- **Formation** : En Master en Intelligence artificielle, Informatique, mathématiques appliquées ou Data Science ou équivalent.
- **Compétences techniques** :
 - Une bonne connaissance des outils et/ou langages de programmation dédiés à l'IA (Python, TensorFlow, PyTorch, etc.) sera appréciée.
 - Connaissances approfondies en machine learning, ou analyse de données.
 - Connaissances de base en architecture logicielle et environnements de développement,
 - Appétence pour le milieu industriel et des Bureaux d'Études Techniques, qui représente la composante « métiers » du projet.
- **Compétences analytiques** : Capacité à comprendre et analyser des jeux de données complexes.
- **Qualités personnelles** : Autonomie, rigueur scientifique, capacité à travailler en équipe, et intérêt pour la recherche appliquée.
- **STACK TECHNIQUE**
 - Back : Python Django, Django REST Framework (DRF) et Pytest
 - Front : Vue 3, TypeScript et Vitest/Cypress
 - MariaDB
 - API REST
 - Docker & Docker Compose
 - Sentry
 - YouTrack
 - GitLab & GitLab CI