

## Ingénieur d'études en traitement des données

### 1. Identification du poste

**Statut :** Agent contractuel

**Catégorie :** A

**Corps :** Ingénieur d'études (ITRF)

**Durée:** 1 an

**Composante de formation, unité de recherche, direction / service :** Centre d'études internationales de la propriété intellectuelle (CEIPI) - 7, rue de l'Écarlate FR-67082 STRASBOURG Cedex

#### Fonction exercée

Ingénieur d'études en traitement du langage naturel et apprentissage automatique

#### Situation du poste dans l'organigramme

Au sein du CEIPI, sous l'autorité hiérarchique du directeur général du CEIPI et en lien fonctionnel principal avec le responsable scientifique de la chaire « *Dynamique des normes en Europe face aux technologies émergentes* ».

### 2. Contexte

Ce poste entre dans le développement du CEIPI dans un contexte de transformation numérique de la composante répondant aux enjeux de la digitalisation de ses activités.

Ce poste s'inscrit en particulier dans le développement des activités prévues dans le cadre de la chaire « *Dynamique des normes en Europe face aux technologies émergentes* » (DENoTE) portée par le CEIPI.

### 3. Mission

L'agent interviendra dans le domaine de l'ingénierie informatique en assurant la conception et le développement d'outils de traitement automatique du langage et d'analyse de données appliqués aux données juridiques, notamment textuelles, à l'analyse du lien entre science, technologies et normes et aux applications des modèles au service de la description des processus créatif.

## 4. Activités

### ➤ **Activités principales :**

#### **1. Développement d'outils d'analyse textuelle de données juridiques et scientifique :**

Le/la titulaire du poste sera responsable de la conception et de la mise en place d'une méthode innovante de représentation des brevets et des articles académiques, optimisée pour le suivi des changements et de l'évolution des thématiques dans divers domaines scientifiques et technologiques. Cette représentation devra intégrer l'information temporelle et les contributions techniques majeures des documents tout en restant invariante face aux styles de rédaction.

- **Développer une méthode de représentation des documents :** concevoir et développer une méthode de représentation combinant l'information temporelle et les contributions techniques principales des brevets et articles académiques.
- **Générer une vue globale de l'évolution des concepts :** utiliser la nouvelle méthode de représentation des documents pour créer une vue d'ensemble de l'évolution des concepts clés dans des domaines spécifiques.
- **Détection de l'émergence de nouveaux concepts :** mettre en place des méthodes de détection des nouveaux concepts en mesurant les différences entre les nouveaux clusters de concepts et ceux déjà existants, afin d'évaluer la nouveauté et l'importance des nouveaux concepts dans un domaine.

#### **2. Développement d'outils de représentation des données de style :**

Le/la titulaire du poste sera responsable de la conception et du développement d'outils d'analyse de la notion de style artistique.

- **Représentation des styles :** développer des techniques basées sur des embeddings pour représenter et analyser l'évolution des "styles artistiques" à partir de données textuelles (critiques, descriptions, théories artistiques) ou d'images d'œuvres d'artistes.
- **Modélisation de l'évolution stylistique :** Identifier et modéliser les caractéristiques stylistiques des œuvres et auteurs pour créer une représentation évolutive de l'évolution des styles artistiques.
- **Visualisation :** Utiliser les embeddings pour visualiser les similarités et les différences entre les styles, et suivre l'émergence et la transformation des styles artistiques au fil du temps.

#### **3. Développement d'un système de RAG (Retrieval-Augmented Generation) dédié à l'exploitation de textes juridiques :**

Le/la titulaire du poste étudiera la faisabilité et proposera des solutions pour le développement d'outils de représentation et d'indexation de documents textuels juridiques (jurisprudences, textes, réglementaires, codes, etc.).

- **Représentation vectorielle des documents :** élaborer une méthode de représentation spécifique aux textes juridiques, prenant en compte les particularités du langage juridique, les structures des documents (ex. articles de lois, jugements, contrats) et les relations entre les différentes parties de texte.

- **Mise en œuvre d'une base de donnée vectorielle :** Concevoir une architecture de base de données vectorielle optimisée pour le stockage et la récupération efficace de textes juridiques représentés par des embeddings, permettant une recherche sémantique avancée et une analyse comparative.

- **Interface :** Développer une interface utilisateur interactive permettant d'exploiter ces textes juridiques, incluant des fonctionnalités de recherche, d'analyse et de visualisation des relations et de l'évolution des concepts juridiques, en prenant en charge la mise à jour dynamique des données et l'intégration des nouvelles sources de textes juridiques.

➤ **Activités associées :**

- Assurer et partager une veille sur les outils et technologies applicables à l'analyse de données juridiques.
- Participer aux activités transversales et aux événements organisés dans le domaine de tu TAL et de l'analyse de données (conférences, séminaires).

## 5. Compétences

### a) Savoir sur l'environnement professionnel

- Maîtrise des techniques de traitement et d'analyse des données textuelles et visuelles.
- Expérience dans l'utilisation d'embeddings et de modèles de représentation de données complexes.
- Compétences en développement de systèmes de RAG (Retrieval-Augmented Generation) et en conception d'architectures de bases de données vectorielles.
- Compétences avancées en visualisation de données et développement d'interfaces interactives.
- Expérience en science des données et en apprentissage automatique, particulièrement pour la détection et la modélisation des thématiques et des styles évolutifs.
- Connaissance des bases de données scientifiques, artistiques et brevets.
- Capacité à travailler de manière autonome et en équipe dans un environnement de recherche.

### b) Savoir-faire opérationnel

Le/la candidat(e) idéal(e) doit posséder un diplôme de niveau master en science des données, intelligence artificielle, linguistique computationnelle, avec une expérience préalable en traitement de données textuelles, notamment dans l'analyse du lien entre science et technologie à partir de plongements sémantiques. Une expérience démontrée dans le domaine de l'exploitation de données liées à l'innovation scientifique, technologique sera un atout majeur.

- Aptitude à dégager des actions prioritaires et à établir un plan d'action.
- Capacités de communication écrite et orale (en français et en anglais), notamment pour la rédaction de documentation de projets informatique.
- Capacité à rendre compte et à capitaliser les expériences.
- Compréhension des principes d'application des outils d'IA générative.
- Capacités organisationnelles et de planification ; gestion de projets.

Université

de Strasbourg

- Expérience préalable dans la conception et la gestion de projets liés au TAL et à l'analyse du lien science et technologie
- Maîtrise de l'anglais.

### c) **Savoir-faire comportemental**

- Initiative, autonomie, sens de l'organisation, reporting, respect des délais.
- Capacités d'observation, d'analyse et de synthèse.
- Disponibilité et flexibilité.
- Capacité à innover, curiosité intellectuelle.
- Capacité à travailler en équipe, à collaborer.
- Capacité à s'adapter dans différents contextes.
- Rigueur professionnelle.
- Capacité d'écoute et d'accompagnement.

## 6. **Environnement et contexte de travail**

### **Descriptif du service**

Nom du service : CEIPI

Nombre d'agents à encadrer (éventuellement) : /

Lieu d'exercice : CEIPI - Bâtiment LE CARDO - 7, rue de l'Ecarlate à Strasbourg

### **Conditions particulières d'exercice**

- Possibilité de déplacements occasionnels pour assister à des conférences, des événements ou des réunions liées au travail.
- Activité à rythme variable selon les actions à mener.
- Contraintes de délais à respecter.