



Offre n°2022-04928

CDD Ingénieur de recherche en extraction d'informations

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Ingénieur scientifique contractuel

Niveau d'expérience souhaité : Jeune diplômé

Contexte et atouts du poste

Cette offre d'emploi est proposée dans le cadre du partenariat entre la Direction Interministérielle du Numérique (DINUM) et l'Inria appelée Lab IA. Le Lab IA a pour but de développer et accompagner des projets d'intelligence artificielle (IA) au sein d'acteurs publics. Le Ministère de la Transition Écologique (MTE) a proposé au Lab IA de l'aider à traiter des dossiers de projets soumis à étude d'impact écologique (par exemple un projet d'extension d'une station d'épuration ou de déploiement d'un parc éolien).

Les dossiers d'étude d'impact sont conséquents (centaine de pages), très techniques et l'emploi de l'IA vise à nourrir un système d'information intelligent permettant d'accélérer le travail des auditeurs devant rédiger un avis sur un projet.

L'IA permettra une réduction du temps nécessaire à traiter chacun des dossiers. Les agents pourront ainsi se concentrer sur d'autres tâches ou en profiter pour approfondir l'analyse des dossiers.

L'objectif est de développer une preuve de concept permettant de faciliter l'instruction des dossiers reçus par le MTE. Cette preuve de concept devra permettre de détecter et classifier automatiquement les paragraphes d'intérêt dans les documents d'étude d'impact (FEI) soumis à l'approbation du MTE.

Mission confiée

Avec l'aide des chercheurs de l'équipe ALMANACH, la personne recrutée sera amenée à développer la preuve de concept en liaison avec les équipes du MTE.

Deux étapes seront explorées pour traiter la problématique de recherche :

- Une phase amont d'acquisition de connaissances à partir du traitement linguistique d'un corpus de dossiers d'étude d'impact ainsi que de documents connexes (comme les réponses des auditeurs) et éventuellement de documents plus généraux relatifs aux questions écologiques. L'objectif est d'identifier les concepts (terminologie) et formulation des relations sémantique liant ces concepts pertinents par rapport aux impacts écologiques. Cette phase pourra également exploiter les référentiels (ontologie, thesaurus, ...) disponibles.
- Une phase de développement d'un outil d'annotation des documents permettant de marquer et typer les occurrences des concepts et relations pertinentes. Diverses approches seront envisagées, avec néanmoins un focus sur l'apprentissage d'un modèle neuronal de type transformeur au dessus du modèle de langue CamemBERT. Ce modèle sera entraîné sur un corpus pré-annoté par des outils TAL plus généraux. Des formes de supervision indirectes seront envisagées tirant parti des avis rendus par les auditeurs.

Ces étapes seront complétées par une intégration de l'outil d'annotation au sein d'un système de recherche d'information.

Principales activités

Principales activités :

- Étudier les méthodes d'acquisition de connaissances, d'extraction d'informations et de classification (annotation de segments) qui pourraient être utilisées dans le cadre du projet
- Étudier les différentes approches
- Explorer les étapes de recherche proposées au travers du développement d'une preuve de

concepts et d'expériences menées sur les données disponibles

- interactions avec les experts du MTE pour déterminer les informations pertinentes à extraire dans les dossiers et pour conduire des évaluations des outils développés
- Supervision éventuelle d'une campagne manuelle d'annotations et/ou de validation de pré-annotations
- Rédiger un rapport de recherche pour documenter le projet.

Compétences

Compétences techniques et niveau requis :

- Programmation : Python (avancé), Perl (notions)
- Expérience bibliotèques d'apprentissage automatique et profond (Pytorch, TensorFlow, Keras, transformers huggingface, Scikit-Learn)
- Expérience dans les architectures de réseaux de neurones (dont transformeurs) et modèles de langues
- Expérience académique en algorithmes de traitement automatique du langage naturel, en particulier sur l'extraction d'information et/ou l'acquisition de connaissances
- Une expérience dans une équipe de recherche

Langues : Français, Anglais

Compétences relationnelles : savoir interagir avec diverses audiences (académiques et experts métier), ouvert à la discussion, capacité d'écoute

Compétences additionnelles appréciées : capacités de rédaction

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail et aménagement du temps de travail (après 1 an d'ancienneté)
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle

Rémunération

Grille des contractuels de la fonction publique : en fonction du diplôme et de l'expérience professionnelle

Informations générales

- **Thème/Domaine** : Langue, parole et audio Statistiques (Big data) (BAP E)
- **Ville** : Paris
- **Centre Inria** : [Centre Inria de Paris](#)
- **Date de prise de fonction souhaitée** :2022-09-01
- **Durée de contrat** :12 mois
- **Date limite pour postuler** :2022-06-30

Contacts

- **Équipe Inria** : [ALMANACH](#)
- **Recruteur** :
Villemonde De La Clergerie Eric / Eric.Villemonde_De_La_Clergerie@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de

répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

Se sentir à l'aise dans un environnement de dynamique scientifique, aimer apprendre, poser les bonnes questions, et écouter sont des qualités essentielles pour réussir cette mission. Une sensibilisation aux enjeux écologiques est aussi bienvenue !

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.