

2022-05506 - Post-Doctorant F/H Transformer pour le langage non-naturel : extraction d'information scientifique et génération de nouvelles molécules de carburants

Contract type : Fixed-term contract
Renewable contract : Oui
Level of qualifications required : PhD or equivalent
Fonction : Post-Doctoral Research Visit

Context

Inria, l'institut national (public) de recherche en sciences et technologies du numérique, s'assume au travers de son contrat d'objectifs et de performance 2019-2023 comme un outil de la souveraineté et de l'autonomie stratégique numérique de la Nation.

Le renforcement des partenariats avec la sphère Sécurité-Défense de l'État est une priorité stratégique. Dans ce contexte, Inria est en train de créer un Département Défense et Sécurité dont la mission est de fédérer, de façon la plus lisible et opérationnelle possible, les différentes actions d'Inria pouvant répondre aux besoins numériques de la sphère Défense et Sécurité.

Assignment

Ce post-doctorat s'inscrit dans le cadre du projet CLEE (Carburants Liquides à Énergie Élevée), monté en partenariat par la start-up Alysophil, l'entreprise MBDA et le département Défense & Sécurité d'Inria.

L'objectif du projet CLEE est de développer de nouveaux carburants offrant de meilleures performances, par exemple au niveau de leur viscosité, densité, pouvoir calorifique, etc., permettant ainsi une plus grande autonomie à volume réduit, ou de réduire l'empreinte environnementale des unités de production. Afin d'identifier de nouvelles molécules candidates à évaluer, l'approche explorée est leur génération par intelligence artificielle.

Pour décrire une molécule, différents encodages permettent de la représenter sous la forme d'une chaîne de caractères (ex : langages SMILES, SELFIES...). L'hypothèse qui motive ce post-doctorat est donc que des approches issues du traitement du langage naturel peuvent se généraliser à l'analyse et génération de molécules.

Le ou la post-doctorante travaillera sous la supervision de Lauriane Aufrant (chercheuse responsable des activités langage au sein d'Inria Défense & Sécurité), et en collaboration étroite avec les partenaires industriels.

Main activities

Le post-doctorat se concentrera dans un premier temps sur l'analyse de molécules existantes (prédiction de propriétés : viscosité, densité, etc.), afin d'identifier l'architecture optimale pour le traitement d'encodages SMILES ou SELFIES. La première piste à explorer porte sur les architectures de type Transformer, mais d'autres approches pourront être considérées en fonction des résultats obtenus. Les défis scientifiques à relever incluent notamment le choix de la représentation d'entrée du modèle (ex : expérimentation avec des architectures de type CharacterBERT) et le faible volume des jeux de données existants (ex : expérimentation avec des méthodes d'augmentation de données, transfert, semi-supervision, etc.).

Afin de pallier au manque de données, et en fonction des résultats obtenus sur les données pré-existantes, il est prévu de recourir en parallèle à des approches plus exploratoires pour collecter de nouvelles données (molécules et/ou propriétés), telles que l'extraction d'information dans les publications scientifiques.

Dans un deuxième temps, les travaux menés sur la prédiction de propriétés seront valorisés pour passer à la génération de nouvelles molécules sous contrainte de propriétés souhaitées. D'autres approches algorithmiques seront alors à mettre en œuvre en couplage avec l'architecture initialement retenue pour l'analyse. Diverses pistes pourront être explorées, incluant les GANs, VAEs, graph grammars, l'apprentissage par renforcement, algorithmes génétiques, etc.

Tout au long des travaux, le ou la post-doctorante pourra bénéficier de l'expertise en chimie des carburants apportée par les entreprises partenaires, afin de se concentrer sur les aspects algorithmiques du projet. La validation finale des nouvelles molécules proposées sera menée manuellement par des experts chimistes.

Skills

- Titulaire d'un doctorat en traitement automatique du langage naturel ou apprentissage profond, ou s'appropriant à soutenir,
- Connaissance théorique et pratique des modèles Transformer, aisance avec l'entraînement de modèles,
- Expérience sur au moins l'une des thématiques suivantes : apprentissage semi-supervisé, augmentation de données, extraction d'information dans les textes scientifiques, apprentissage par renforcement,
- Volonté de diversifier ses compétences en appliquant des algorithmes connus à des domaines nouveaux
- Intérêt marqué pour le travail collaboratif et pluridisciplinaire,

Benefits package

- Navette privée gratuite depuis Paris Place de l'Etoile ou Versailles-Chantiers / Rive Droite, le

General Information

- Town/city : Le Chesnay
- Inria Center : [Siège](#)
- Starting date : 2023-01-01
- Duration of contract : 2 years
- Deadline to apply : 2023-06-01

Contacts

- Inria Team : DRH-SRH (DIRECTION)
- Recruiter :
Maillet Florence / florence.maillet@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Instruction to apply

Nous vous remercions de déposer une lettre de motivation accompagnée de votre CV.

Des lettres de recommandation ou des indications de références seraient appréciées mais ne sont pas obligatoires.

Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

- matin et le soir.
- Remboursement partiel des frais de transport public (50 %)
 - Remboursement d'une partie de la mutuelle dans le cadre de la Protection Sociale Complémentaire
 - Parking sur place gratuit Restaurant d'entreprise subventionné et sur site
 - Installations sportives sur site (gymnase, sport collectif, salle de musculation, squash)
 - Congés = 45 jours soit 35 jours de Congés Annuels + 10 jours de RTT (sur la base d'un temps plein / année complète travaillée)
 - Possibilité de télétravail (2 jours / semaine)
 - Aménagement possible du temps de travail (5 jours ou 4.5 jours / semaine)
 - Possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
 - Prestations sociales, culturelles et sportives (AGOS - Association de Gestion des Œuvres Sociales d'Inria)
 - Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)

Remuneration

Selon le cadrage prévu par le statut.

Localisation du poste : Rocquencourt ou Saclay