



**Offre n°2022-05506**

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

Type de contrat : CDD

Contrat renouvelable : Oui

Niveau de diplôme exigé : Thèse ou équivalent

### **Contexte et atouts du poste**

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.

### **Mission confiée**

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.

### **Principales activités**

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.

## Compétences

- Titulaire d'un doctorat en traitement automatique du langage naturel ou apprentissage profond, ou s'appêtant 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,

## Avantages

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

## Rémunération

Selon le cadrage prévu par le statut.

Localisation du poste : Rocquencourt ou Saclay

## Informations générales

- **Ville** : Le Chesnay
- **Centre Inria** : [Siège](#)
- **Date de prise de fonction souhaitée** : 2024-01-01
- **Durée de contrat** : 2 ans
- **Date limite pour postuler** : 2024-01-06

## Contacts

- **Équipe Inria** : MIS-DEFENSE (DIRECTION)
- **Recruteur** :  
Maillet Florence / [florence.maillet@inria.fr](mailto:florence.maillet@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.

**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

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.

**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.