



Offre n°2025-09095

Post-Doctorant F/H Postdoc Améliorer la frugalité et la robustesse d'un modèle de langage pour les textes médicaux en français

Type de contrat : CDD

Niveau de diplôme exigé : Thèse ou équivalent

Fonction : Post-Doctorant

A propos du centre ou de la direction fonctionnelle

Le centre de recherche Inria de Saclay a été créé en 2008. Sa dynamique s'inscrit dans le développement du plateau de Saclay, en partenariat étroit d'une part avec le pôle de l'**Université Paris-Saclay** et d'autre part avec le pôle de l'**Institut Polytechnique de Paris**. Afin de construire une politique de site ambitieuse, le centre Inria de Saclay a signé en 2021 des accords stratégiques avec ces deux partenaires territoriaux privilégiés.

Le centre compte **40 équipes-projets**, dont 32 sont communes avec l'Université Paris-Saclay ou l'Institut Polytechnique de Paris. Son action mobilise **plus de 600 personnes**, scientifiques et personnels d'appui à la recherche et à l'innovation, issues de 54 nationalités.

Le centre Inria Saclay - Île-de-France est un acteur essentiel de la recherche en sciences du numérique sur le plateau de Saclay. Il porte les valeurs et les projets qui font l'originalité d'Inria dans le paysage de la recherche : l'excellence scientifique, le transfert technologique, les partenariats pluridisciplinaires avec des établissements aux compétences complémentaires aux nôtres, afin de maximiser l'impact scientifique, économique et sociétal d'Inria.

Contexte et atouts du poste

Dans le cadre du projet Partages, coordonné par le Health Data Hub et financé par BPI France, l'objectif est de développer un modèle open source pour les données médicales françaises.

L'intelligence artificielle générative et les grands modèles de langage ont récemment ouvert des perspectives sur l'utilisation de l'IA dans divers domaines. Pour concrétiser ce potentiel, le projet national collaboratif "PARTAGES" répond à un besoin crucial : le développement d'un modèle de

langage open-source en français, spécialisé dans les données de santé. Un tel outil accélérera considérablement et démocratisera l'utilisation de l'IA pour la santé, apportant des bénéfices massifs à l'organisation du système de santé, aux conditions de travail des soignants et, en fin de compte, à la santé publique. Une préoccupation principale lors de l'analyse des données textuelles de santé est la confidentialité des patients, et c'est pourquoi il est stratégique de développer des modèles open source pouvant être utilisés dans les hôpitaux et autres acteurs de santé. Pour cela "Partages" se déroulera en deux étapes : (i) générer des comptes rendus médicaux synthétiques en combinaison avec la littérature scientifique biomédicale pour affiner un modèle open-source, (ii) raffiner le modèle sur des comptes rendus médicaux réels, au sein de chacun des 18 établissements de soins partenaires. L'équipe Soda est impliquée dans la réalisation de tâches de la première partie : travail sur des données uniquement publiques, et mise à disposition du modèle en open-source.

Mission confiée

En nous appuyant sur l'expertise de l'équipe Soda en matière d'IA robuste et frugale, nous participons à la création d'un modèle de base qui est robuste face à un nouveau vocabulaire (que ce soit de nouveaux concepts médicaux apparaissant à l'avenir, ou des fautes d'orthographe ou des abréviations utilisées dans les textes médicaux), et frugal. Le postdoc recruté, Joel Mba Kouhoue travaillera sur la distillation de modèles, afin d'obtenir un modèle (dit ``étudiant`) plus économe en ressources de calcul que le modèle originel (dit ``enseignant`). Plusieurs architectures et catégories de modèles seront envisagées pour le modèle étudiant : modèles "encoder-only" de type BERT, en particulier pour les tâches de détection d'entités et d'annotations de documents, avec une comparaison avec des modèles génératifs sur ces tâches, et modèles "encoder-decoder" pour des tâches nécessitant un modèle génératif, comme le résumé, la génération d'exercices pour la formation des personnels soignants. L'ensemble des résultats de ces expériences sera publié et partagé en open-source.

Le travail consistera d'une part à distiller des gros modèles ``enseignants" existants en utilisant pour la distillation des corpus proches de ceux applicatifs (et donc avec un vocabulaire médical), d'autre part à affiner et valider ces modèles sur des tâches proxys, telles que de l'extraction d'entités nomées. L'étude fera varier le choix du modèle enseignant (en faisant attention au choix de licence compatible avec nos objectifs d'open source), cherchera le meilleur corpus (y compris en utilisant de l'augmentation de données) et mesurera les compromis taille de modèle / performance sur les tâches proxys.

Principales activités

Le candidat travaillera à l'élaboration et l'évaluation de modèles frugaux et robustes.

Compétences

Compétences techniques et niveau requis :

Langues :

Compétences relationnelles :

Compétences additionnelles appréciées :

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
- 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
- Sécurité sociale

Rémunération

Salaire brut mensuel : 2 788 euros brut/mois

Informations générales

- **Thème/Domaine** : Langue, parole et audio
Ingénierie logicielle (BAP E)
- **Ville** : Palaiseau
- **Centre Inria** : [Centre Inria de Saclay](#)
- **Date de prise de fonction souhaitée** : 2025-09-01
- **Durée de contrat** : 1 an, 10 mois
- **Date limite pour postuler** : 2025-08-31

Contacts

- **Équipe Inria** : [SODA](#)
- **Recruteur** :
Abecassis Judith / judith.abecassis@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

Vous pouvez donner là, un portrait à "gros traits" du (de la) collaborateur(trice) attendu(e) : ce que vous voyez comme nécessaire et suffisant et qui peut associer :

- goûts et appétences,
- domaine d'excellence,
- éléments de personnalité ou de caractère,
- savoir et savoir faire transversaux...

Cette rubrique permet de compléter et alléger (réduire) la liste plus formelle des compétences :

- "Se sentir à l'aise dans un environnement de dynamique scientifique, aimer apprendre et écouter sont des qualités essentielles pour réussir cette mission."
- " Passionné(e) par l'innovation, avec une expertise dans le développement Ruby on Rail et une grande capacité de conviction. Une thèse dans le domaine *** constitue un réel atout."

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.