



Offre n°2024-08428

## Post-Doctoral Research Visit F/M 2-years contract "Step-by-step guidance to clinical decision, by combining data- and knowledge-driven approaches"

*Le descriptif de l'offre ci-dessous est en Anglais*

Type de contrat : CDD

Niveau de diplôme exigé : Thèse ou équivalent

Fonction : Post-Doctorant

Niveau d'expérience souhaité : De 3 à 5 ans

### Contexte et atouts du poste

The 2-year position will take place in the HeKA team (Inria, Inserm, Université Paris Cité), physically located at ParisSanté Campus, 2-10 rue d'Oradour-sur-Glane, 75015 Paris. The work is a collaboration between the HeKA team, and the Pitié-Salpêtrière hospitals (HEGP) of the AP-HP (Assistance Publique - Hôpitaux de Paris).

This internship will take place in the context of the PEPR Digital Health project named ShareFAIR, funded by the French Research National Agency.

### Mission confiée

#### Motivation and background

In the last years, deep-learning-based clinical decision support has mostly been over-simplified to a single classification task, mostly for demonstrating the feasibility of reproducing a final medical decision on the basis of previously acquired healthcare data. This led to the development and publication of many models, with very little perfusion in clinics, due to their lack of explainability, fairness, usability, and need. For example, a single class prediction based on patient history appear to be of limited use to practitioners that first consider their last observations, and wait for decision support to integrate medical history to their own medical reasoning.

In previous works, we proposed to learn decision pathways from Electronic Health Records (EHRs) to propose a step-by-step guidance to the medical decision. These pathways are build on the basis of data collected along healthcare activity, (i.e., traces of real-word clinical practice) and for this reason follow to some extent medical reasoning.

This data-driven approach can be opposed to the classical decision processes in medicine that follow Clinical Practice Guidelines (CPGs), which compile in narratives the state-of-the-art knowledge authored by a college of specialists of the domain. We consider the latter as a knowledge-driven approach.

Both data- or knowledge- driven approaches have been shown to be associated to unfair decisions for various reasons, including sampling biases when constituting training datasets, or study cohorts used to build CPGs.

We think that Large Language Models (LLMs) are promising tools to support knowledge-driven decision support. But by default, they do not consider the amount of observational data acquired in EHRs. In addition, we think that hybrid systems that combines data- and knowledge-driven approaches offer the best promises for the design for explainable and fair decision supports in biomedicine. Consequently, the following research project proposes to study the combination of (i) LLM, as a provider to the state-of-the-art knowledge, and (ii) data-driven approaches such as the one developed by *Muyamæt al.* for decision support

### Principales activités

#### Objective

##### LLM as a provider of state-of-the-art knowledge --

Despite their undeniable role in clinical decision-making, CPGs have limitations. First, their development requires input from a diverse panel of experts, leading to a time-consuming and costly process. This implies that updates to CPG take several years, which makes them unsuitable for rapidly evolving medical practices, such as the introduction of new diagnostic tests or the emergence of a new disease. Second, due to the substantial resources needed to develop these guidelines, it is not practical to create

CPGs for every known medical condition. Third, guidelines are generally designed to address the needs of the majority of the population, potentially overlooking rare conditions or uncommon populations, potentially leading to unfair recommendations.

The first objective of this research project is to study how LLMs may embed state-of-the-art knowledge represented within CPGs, and thus support medical decision. Several challenges are associated with this objective such as the fact that CPGs include semi-structured information as decision trees or score tables, which are key to guide recommendations, or the fact that important element of knowledge may be spread within over several documents, including potential inconsistencies.

#### **Combine data-driven approach with LLM --**

Several methods can be leveraged to learn clinical pathways from data including sequence and process mining, to discover patterns within hospital processes. Reinforcement Learning (RL) methods are natural candidates to learning sequential decisions, Deep Reinforcement Learning (DRL) adopts deep neural networks to learn the optimal policy that is central to RL and facilitate handling cases where the number of states is potentially large. To identify optimal treatment regimens for patients, proposed a deep Q network-based model for EHR data. Patient records are used to model the state, action and transition probability of the model, used, in turn, to determine the individual optimal dose. Resulting constructs can in turn be compared with state-of-the-art CPGs, but both sources are not combined.

The second objective of the project is to enable LLM and a data-driven approach, such as a DRL one, to be combined. One potential direction is to find inspiration in the reinforcement learning (RL) from human feedback paradigm that is used to adapt LLM from human preferences. Here LLM model could be adapted through a number of RL iterations where the feedback is not provided by a human, but through healthcare data that trace historical clinical decision made by clinicians.

#### **Evaluate various approaches --**

A transversal objective of the project is to enable the evaluation of the various developed systems, with the following objectives in mind: correctness, optimality, fairness and utility. This relies, through others, on enabling the comparison of various clinical decision pathways in terms of precision, recall (for instance in regards to a ground truth diagnosis), number of steps (i.e., tests, questions), time before diagnosis, cost and others. To this matter, the candidate could propose metrics that will enable comparing pathways one-to-one and many-to-many.

One potential direction would be to rely on a common representation framework of clinical pathways by leveraging the experience on scientific workflows of the ShareFAIR PEPR project consortium. In particular, we could aim at designing a shareable, traceable, understandable format that supports reasoning mechanisms, to facilitate further comparison.

#### **Biomedical usecases --**

The project will be applied on three distinct biomedical use cases: the differential diagnosis of anemia, hypertension and the management of intracranial aneurysms

## **Compétences**

Technical skills and level required: Python, data science libraries

Languages: French and/or English

Relational skills: Good communication skills

Other valued appreciated: Interest for Biomedical data science

## **Avantages**

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

## **Informations générales**

- **Thème/Domaine** : Neurosciences et médecine numériques  
Statistiques (Big data) (BAP E)
- **Ville** : Paris
- **Centre Inria** : [Centre Inria de Paris](#)
- **Date de prise de fonction souhaitée** : 2025-02-01
- **Durée de contrat** : 2 ans
- **Date limite pour postuler** : 2025-01-04

## Contacts

- Équipe Inria : [HEKA](#)
- Recruteur :  
Coulet Adrien / [adrien.coulet@inria.fr](mailto:adrien.coulet@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

### 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.