![Inria logo]

Offer #2023-05775

Post-Doctoral Research Visit F/M Transformer for unnatural language: scientific information extraction and generation of new fuel molecules

Contract type : Fixed-term contract

Renewable contract : Yes

Level of qualifications required : PhD or equivalent

Fonction : Post-Doctoral Research Visit

Level of experience : Recently graduated

Context

Inria is the French national research institute for digital science and technology. World-class research, technological innovation and entrepreneurial risk are its DNA. In 215 project teams, most of which are shared with major research universities, more than 3,900 researchers and engineers explore new paths, often in an interdisciplinary manner and in collaboration with industrial partners to meet ambitious challenges.
As a technological institute, Inria supports the diversity of innovation pathways: from open source software publishing to the creation of technological startups (Deeptech).

Strengthening partnerships with the State's Security-Defense sphere is a strategic priority for Inria. In this context, Inria is in the process of creating a Defense and Security Department whose mission is to federate, in the most readable and operational way possible, the various Inria actions that can meet the digital needs of the Defense and Security sphere.

Assignment

This post-doctorate is part of the CLEE (Carburants Liquides à Énergie Élevées) project, set up in partnership by the start-up Alysophil, MBDA and the Defense & Security department of Inria.

The objective of the CLEE project is to develop new fuels offering better performance, for example in terms of their viscosity, density, calorific value, etc., thus allowing greater autonomy with reduced volume, or to reduce the environmental footprint of production units. In order to identify new candidate molecules, the project explores their automatic generation using artificial intelligence.

To describe a molecule, different encodings allow to represent it as a string of characters (e.g. SMILES, SELFIES languages...). The hypothesis that motivates this post-doctoral fellowship is that approaches from natural language processing can be generalized to the discovery of new molecules associated with their properties in order to support generation of new molecules.

The post-doctoral fellow will work under the supervision of Lauriane Aufrant (researcher in charge of language activities within Inria Defense & Security), in close collaboration with industrial partners.

Main activities

The post-doctoral fellow will initially focus on the analysis of existing molecules (prediction of properties: viscosity, density, etc.), in order to identify the optimal architecture for processing SMILES or SELFIES encodings. The first avenue to be explored is Transformer-type architectures, but other approaches may be considered depending on the obtained results. Scientific challenges include the choice of the input representation of the model (e.g. experimentation with CharacterBERT architectures) and the small volume of existing datasets (e.g. experimentation with data augmentation methods, transfer, semi-supervision, etc.).

In order to overcome the lack of data, and depending on the results obtained on the pre-existing data, it is planned to use in parallel more exploratory approaches to collect new data (molecules and/or properties), such as the extraction of information from scientific publications.

In a second step, the work done on property prediction will be used to generate new molecules with the desired properties. Other algorithmic approaches will then be implemented in coupling with the architecture initially chosen for the analysis. Various approaches could be explored, including GANs, VAEs, graph grammars, reinforcement learning, genetic algorithms, etc.

Throughout the work, the post-doctoral fellow will be able to benefit from the expertise in fuel chemistry provided by the partner companies, in order to focus on the algorithmic aspects of the project. The final validation of the proposed new molecules will be carried out manually by chemical experts.

Skills

- PhD in natural language processing or deep learning, or about to obtain one,
- Theoretical and practical knowledge of Transformer models, comfortable with training models,
- Experience with at least one of the following topics: semi-supervised learning, data augmentation, information extraction from scientific texts, reinforcement learning,
- Willingness to diversify his/her skills by applying known algorithms to new domains,
- Strong interest in collaborative and multidisciplinary work.

Benefits package

- 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

General Information

- **Town/city :** Le Chesnay
- **Inria Center :** Siège
- **Starting date :** 2023-12-01
- **Duration of contract :** 2 years
- **Deadline to apply :** 2023-12-08

Contacts

- **Inria Team :** MIS-DEFENSE
- **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.

> **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.

Instruction to apply

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.