SCIENCE & CREATIVITY TO INVENT A SUSTAINABLE WORLD

We're recruiting

Post-Doctoral Researcher in Machine Learning

Establishment	IMT Mines Alès (Ecole Nationale Supérieure des Mines d'Alès)
Main assignment	Centre for Education and Research in Informatics and Systems (CERIS)
Administrative residence	Alès (Department Gard - Region Occitanie)
Type of contract	18-month contract - Public law contract - Full-time
Starting date	01/07/2024

This study is part of the **European ENFIELD project**. ENFIELD is set to establish a distinctive **European Center of Excellence focused on advancing fundamental research in Adaptive, Green, Human-Centric, and Trustworthy AI**. These pillars represent novel, strategic elements crucial for the successful development, deployment, and acceptance of AI in Europe. The Institut Mines-Télécom (IMT), a partner in the project through its Data Analytics and AI scientific community, is in charge of the workpackage that brings together the project's research on these four scientific pillars. This study falls under the human-centric AI pillar. More specifically, the post-doc will be involved in the "HC-AI.3 Interpretable Data Driven Decision Support Systems" sub-theme of the pillar.

Presentation of our establishment and the CERIS Centre

Institut Mines-Télécom

The Institut Mines-Télécom (IMT) is a public scientific, cultural and professional establishment (EPSCP) under the supervision of the ministers for industry and digital technology. It is the largest group of engineering schools in France, with 11 public engineering schools throughout the country, training 13,500 engineers and PhDs. The ITM employs 4,500 people and has an annual budget of €400 million, 40% of which comes from its own resources. The ITM has 2 Carnot institutes, 35 industrial chairs, produces 2100 A-rank publications annually, 60 patents and carries out €110M of contract research.

IMT Mines Alès

Founded in 1843, IMT Mines Alès currently has 1,400 students (including 250 foreign students) and 380 staff. The school has 3 research and teaching centres of a high scientific and technological level, working in the fields of materials and civil engineering (C2MA), the environment and risks (CREER), artificial

6 avenue de Clavières
30319 Alès Cedex - France
☎ : +33 (0)4 66 78 50 00
www.imt-mines-ales.fr

intelligence and industrial and digital engineering (CERIS). It has 12 technology platforms and 1,600 partner companies. The person recruited will be assigned to the Centre d'Enseignement et de Recherche en Informatique et Systèmes (CERIS).

CERIS Centre

CERIS comprises two research teams, ISOAR for *Ingénierie des Systèmes et des Organisations pour les Activités à Risque* and I3A for *Informatique, Image et Intelligence Artificielle. CERIS* also runs two teaching departments: 2IA for *Computer Science and Artificial Intelligence* and PRISM for *PeRformance Industrielle et Systèmes Mécatroniques*, as well as 2 technology platforms: AIHM for *Alès Imaging and Human Metrology* and PFM for *Plateforme Mécatronique*.

Background to the study

Nowadays, there are machine learning techniques that enable us to develop high-performance predictive systems for solving a wide range of tasks. Some of these systems are integrated into decision-making processes involving human operators, and in which the predictor will be considered a decision aid, e.g. making delicate decisions in complex contexts that are only partially appreciated by the predictive system (partial visibility of the decision-making context), or because of legal constraints on the final decision that will be taken. In this case, the human operator takes advantage of the predictions provided by the system to make their decision.

To improve this type of human-machine collaboration, it may be desirable to provide the human operator with information on the overall behaviour of the predictor, or even to enrich the predictions provided with information intended to explain them. For example, the prediction of a trained deep learning model for a given input, e.g. why does this Vision Transformer model provide this classification result for a given image? XAI (eXplainable AI) techniques have been proposed and studied in the literature. They can be used to evaluate the overall behaviour of trained models, or even to explain specific predictions as a function of the input (local interpretability methods). These XAI techniques are based on assumptions about the way in which predictive models work, and the importance given to the information supplied to them.

What is the impact of XAI techniques on Human-Machine collaboration? May the type of XAI technique we use influence our evaluation of trained models? In many cases, AI models are complex systems that process huge amounts of data and perform computations that are difficult to interpret intuitively. The aim of explainability is to provide users with information about the decision-making logic of the model so that they can understand and trust the results of the AI. We want to contribute questioning if XAI techniques may bias the trust we place on AI in specific contexts. We also want to evaluate if specific XAI techniques influence Human-Machine collaborations evolving towards a relationship of mediation, delegation, or substitution.

Job description

You will contribute to :

- ▶ Evaluation of variations in the results produced by XAI methods in specific study contexts (for example, image classification tasks and XAI techniques based on local interpretability using attribution methods).
- ▶ Assessing the impact of XAI methods on human-machine collaboration in simplified decision-making contexts:

    o evaluation of the human operator's performance in carrying out a task, in different contexts: alone, with the help of a predictive model for which decisions will be explained/not explained, using an XAI technique.

    o assessment of man-machine collaboration: delegation, substitution or mediation.

    o assessment of potential biases induced by CAE techniques.

6 avenue de Clavières
30319 Alès Cedex - France
☎ : +33 (0)4 66 78 50 00
www.imt-mines-ales.fr

Page **2** on **4**

You will be involved in defining :

- ▶ Study contexts (e.g. games, image classification) and test protocols to be taken into account.
- ▶ The selection and implementation of predictive models and XAI techniques.
- ▶ Setting up the tools needed to carry out the experiments covered by the study protocols, for example the development of simple games and decision-making interfaces.
- ▶ The implementation of the above-mentioned protocols on cohorts of human operators.
- ▶ Assessing and promoting the results obtained.

Depending on your profile, certain aspects will be explored in greater depth than others, and complementary aspects will be considered concerning the identification of XAI techniques likely to respond to the limitations of existing techniques identified in the literature or by the tests carried out. We remain open to suggestions for contributions from candidates based on their areas of interest.

Profile sought and general assessment criteria

Skills, knowledge and experience required:

- ▶ Deep learning models, and their implementation via PyTorch (ability to train and refine pre-trained models on specific datasets on dedicated GPU computing resources), evaluation of trained models according to standard protocols.
- ▶ XAI techniques; knowledge of the main XAI methods (e.g. local and attribution methods) and tools in the field. Skills can be improved during the assignment, but knowledge of these aspects is desirable.

Skills, knowledge and experience appreciated :

- ▶ An important part of the assignment is the evaluation of human-machine collaborations, i.e. assessing the impact of AI and XIA models on human decision-making; initial experience of working with human cohorts would be a plus.

Minimum level of training and/or experience required :

- ▶ Doctorate in Computer Science on a topic related to machine learning or deep learning

Application

📑 Administrative conditions for applying

The position offered by IMT Mines Alès is a 18-month, full-time, public law contract governed by the management framework of the Institut Mines-Télécom, profession P, Post Doctorant, category II. The project will be carried out in close collaboration with IMT Business School (BS), in particular with Nicolas Soulié. The position will be based at IMT Mines Ales, although a location at IMT BS (Evry, Ile-de-France) may be considered.

Salary: to be defined according to profile and experience.

📄 How to apply

Applications (CV and covering letter) should be sent exclusively to :
https://institutminestelecom.recruitee.com/o/post-doctorant-ou-post-doctorante-en-apprentissage-automatique

6 avenue de Clavières
30319 Alès Cedex - France
☎ : +33 (0)4 66 78 50 00
www.imt-mines-ales.fr

Page **3** on **4**

📅 **Recruitment schedule**

Closing date for applications : 12/05/2024

Approximate date of the jury : Snd fortnight in May 2024

Desired starting date : 01/07/2024

📇 **Contacts**

▶ Job content:

Sébastien HARISPE, Lecturer and researcher
⌨ : sebastien.harispe@mines-ales.fr

Nicolas SOULIE, Lecturer and researcher
⌨ : nicolas.SOULIE@imt-bs.eu

▶ Administrative aspects:

Géraldine BRUNEL, Head of Human Relations Department
⌨ : geraldine.brunel@mines-ales.fr

6 avenue de Clavières   Page **4** on **4**
30319 Alès Cedex - France
☎ : +33 (0)4 66 78 50 00
www.imt-mines-ales.fr