



Offer #2025-08602

PhD Position F/M Designing for Explainability in Sustainable AI (M/F)

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

Created in 2008, the Inria center at the University of Lille employs 360 people, including 305 scientists in 15 research teams. Recognized for its strong involvement in the socio-economic development of the Hauts-De-France region, the Inria center at the University of Lille maintains a close relationship with large companies and SMEs. By fostering synergies between researchers and industry, Inria contributes to the transfer of skills and expertise in the field of digital technologies, and provides access to the best of European and international research for the benefit of innovation and businesses, particularly in the region.

For over 10 years, the Inria center at the University of Lille has been at the heart of Lille's university and scientific ecosystem, as well as at the heart of Frenchtech, with a technology showroom based on avenue de Bretagne in Lille, on the EuraTechnologies site of economic excellence dedicated to information and communication technologies (ICT).

Context

Background. Recent generational leaps in the complexity and capabilities of Machine Learning (ML) models have made Artificial Intelligence (AI) able to tackle challenges ranging from vision and graphics to natural language, and even creative tasks. These improvements, along with the growing availability and maturity of AI

technologies, also helped democratizing AI as a tool for a broad audience of researchers, industries, artists, and more. However this expansion also revealed the environmental and economic impacts of AI technologies when used at very large scales [3, 7]. The adoption of greener, less energy-consuming models by ML practitioners is a significant aspect in successfully improving AI impact in the future. However, there can exist hundreds of candidate algorithms to address a single category of problems, and the choice of a ML model for a given task is often driven by previous experience, domain understanding, or expertise availability. Adopting new technologies and approaches typically requires additional learning efforts in order to fully understand their purpose, strength, features, and adequacy to a task.

Explainability plays a crucial role in this process. Increasing awareness and adaption of new models goes beyond the accessibility of models, but further requires professionals to contextualize and balance potential trade-offs. In order assist ML practitioners with these processes, **this PhD will use human-centered design to develop tools that explain how a particular algorithm affects AI-waste and help them find more environmentally friendly models for their projects.** It will contribute to a larger *Sustainable ML* project which aims to develop a design framework and an associated toolkit to foster energy efficiency throughout the whole life cycle of ML applications: from the training and testing iterations of the design and exploration phases, to the final training of the production systems, and the continuous online re-training during and after deployment.

[1] M. Beaudouin-Lafon and W. E. Mackay. Reification, polymorphism and reuse: Three principles for designing visual interfaces. In Proc. of AVI '00, ACM.

[2] G. Casiez, E. Mackamul and S. Malacria. Clarifying and differentiating discoverability. Human Computer Interaction, 2024.

[3] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres. Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700, 2019.

[4] I. Lobo, J. Koch, J. Renoux, I. Batina, and R. Prada. When should I lead or follow: understanding initiative levels in human-ai collaborative gameplay. In Proc. of DIS '24, ACM.

[5] W. Mackay. Responding to cognitive overload: Co-adaptation between users and technology. Intellectica, 2000.

[6] X. Peng, J. Koch, and W. E Mackay. DesignPrompt: Using multimodal interaction for design exploration with generative AI. In Proc. of DIS '24, ACM.

[7] R. Schwartz, J. Dodge, N. A Smith, and O. Etzioni. Green AI. Communications of the ACM, 2020.

Administration. The candidate will be part of the Loki research team, based at Inria Lille in France and supervised by Prof. Géry Casiez and Dr. Janin Koch. We build on the principles such as instrumental interaction [1] and co-adaptation [5] to create interactive systems that are discoverable [2], appropriate [4], and expressive [6], that grow with the user to enhance rather than replace the users skills.

The 3-year doctoral position is funded by a European Union's Horizon 2020 grant for *SustainML: Application Aware, Life-Cycle Oriented Model-Hardware Co-Design Framework for Sustainable, Energy Efficient ML Systems*. The work will be in close collaboration with the DFKI (German Institute of Artificial Intelligence) and other partners.

Assignment

The goal of this Ph.D. position is to explore new directions of Human-AI interactions. This includes new interaction and explainability approaches that will allow users to interactively explore the trade-offs of competing ML models with the help of intelligent agents. Exploring ML model alternatives during the development process, before the models enter their full training cycles, requires users to express potentially ambiguous project objectives and to understand the trade-offs of ML model alternatives, e.g. time, computing hardware, or estimated CO2 footprint for a particular task. This requires the development of new design and evaluation methods to ensure effective interaction with intelligent systems, and specifically to:

- Develop new interactive methods for users to express and refine ML model needs and goals using a human-computer partnership approach.
- Develop new explainability approaches that intelligent systems can use to suggest and expose the trade-offs of alternative ML models in a context-dependent manner.
- Design new interactive visualizations to explore the design space of ML models with multiple competing objectives, including AI-waste minimization.

Main activities

The doctoral candidate will be expected to:

- Conduct empirical studies and workshops, e.g. participatory design workshops.
- Prototype, design and develop novel interactive systems.
- Design, run, and analyze controlled and field experiments to evaluate interaction and explainability techniques.
- Write research reports and scientific papers.

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

- Social security coverage

Remuneration

2200 € per month

General Information

- **Theme/Domain** : Interaction and visualization
Information system (BAP E)
- **Town/city** : Villeneuve d'Ascq
- **Inria Center** : [Centre Inria de l'Université de Lille](#)
- **Starting date** : 2025-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2025-03-30

Contacts

- **Inria Team** : [LOKI](#)
- **PhD Supervisor** :
Koch Janin / janin.koch@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.

The keys to success

We are looking for motivated students who are excited about creating human-centered exploratory tools and are interested in applying Human-Computer Interaction research methods to Machine Learning problems.

Suitable candidates should have experience in Human-Computer Interaction methods and strong programming skills, preferably Python, are required.

Background knowledge in Machine Learning and experience in web technologies is a plus.

The doctoral position will be in English.

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

CV + cover letter

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.