

2022-05247 - PhD Position F/M Exploring Human-AI Collaboration and Explainability for Sustainable ML

Type de contrat : CDD
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

Located at the heart of the main national research and higher education cluster, member of the Université Paris Saclay, a major actor in the French Investments for the Future Programme (Idex, LabEx, IRT, Equipex) and partner of the main establishments present on the plateau, the centre is particularly active in three major areas: data and knowledge; safety, security and reliability; modelling, simulation and optimisation (with priority given to energy).

The 450 researchers and engineers from Inria and its partners who work in the research centre's 28 teams, the 60 research support staff members, the high-level equipment at their disposal (image walls, high-performance computing clusters, sensor networks), and the privileged relationships with prestigious industrial partners, all make Inria Saclay Île-de-France a key research centre in the local landscape and one that is oriented towards Europe and the world.

Contexte et atouts du poste

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 chosen candidate will be supervised by Prof. Wendy Mackay and Dr. Janin Koch. The work will be in close collaboration with the DFKI (German Institute of Artificial Intelligence) and other partners.

Mission confiée

Contexte

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 [5, 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. "AI waste", or the environmental cost of using these tools in terms of hardware and energy resources, is seldom taken into consideration.

This **PhD in Human-Computer Interaction (HCI)** aims to help ML practitioners better understand and address these issues, using collaborative algorithmic agents able to work together with them and in particular to inquire and understand their needs and to suggest the best and "greener" algorithms to address them. It will contribute to a larger Sustainable AI 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 retraining during and after deployment. This PhD project will therefore address the HCI aspects of these objectives.

The candidate will be part of the ExSitu team, a France-based research group that investigates alternate approaches to traditional human-computer roles in order to establish true human-computer partnerships. We build on the principles of instrumental interaction [1, 2] and co-adaptation [6] to create interactive systems that are discoverable, appropriable [3], and expressive [4], growing with their users to augment rather than replace their skills.

Objective

The goal of this doctoral position is to explore new directions of Human-AI interactions. This includes new interaction and explainability approaches to allow ML practitioners 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 a better understanding of how users plan, search for, and select ML models.
- Identify the relevant characteristics of a ML project in terms of constraints and context.
- Develop new interactive methods for users to express and refine these characteristics 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.

Bibliography

[1] Michel Beaudouin-Lafon. Instrumental interaction: An interaction model for designing post-wimp user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00, page 446–453, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 1581132166. doi: 10.1145/332040.332473. URL <https://doi.org/10.1145/332040.332473>.

Informations générales

- **Thème/Domaine** : Interaction et visualisation
Systèmes d'information (BAP E)
- **Ville** : Palaiseau
- **Centre Inria** : CRI Saclay - Île-de-France
- **Date de prise de fonction souhaitée** : 2022-11-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2022-08-31

Contacts

- **Equipe Inria** : EX-SITU
- **Directeur de thèse** :
Koch Janin / janin.koch@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 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 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 180 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.

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.

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.

[2] Michel Beaudouin-Lafon and Wendy E. Mackay. Reification, polymorphism and reuse: Three principles for designing visual interfaces. In Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '00, page 102–109, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 1581132522. doi: 10.1145/345513.345267. URL <https://doi.org/10.1145/345513.345267>.

[3] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. May ai? design ideation with cooperative contextual bandits. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi: 10.1145/3290605.3300863. URL <https://doi.org/10.1145/3290605.3300863>.

[4] Janin Koch, Nicolas Taffin, Michel Beaudouin-Lafon, Markku Laine, Andrés Lucero, and Wendy Mackay. ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. Proceedings of the ACM on Human-Computer Interaction, 4 (CSCW1):1–27, May 2020. doi: 10.1145/3392850. URL <https://hal.archives-ouvertes.fr/hal-02867303>.

[5] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700, 2019.

[6] Wendy Mackay. Responding to cognitive overload: Co-adaptation between users and technology. Intellectica, 30, 07 2000. doi: 10.3406/intel.2000.1597.

[7] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. Green ai. Communications of the ACM, 63(12):54–63, 2020.

Principales activités

Specific 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.

Expected Results

- Advance knowledge about what drives ML practitioners to select models and what it takes to make them experiment with new ones.
- Novel methods to identify and suggest ecology-friendly ML models that do not diminish performance.
- Full interactive system(s) that demonstrate new communication methodology and explainability approaches for exploring ML-model alternatives in early stage projects.

Compétences

Required Skills

The primary domain of this PhD position is **Human-Computer Interaction**. 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 past experience and **skills 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.

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 (after 6 months of employment) 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