Loki
technology & knowledge for interaction

Designing for Explainability in Sustainable AI

Duration 36 months
Level Ph.D.
Team Loki
Advisor(s) Janin Koch (Janin.Koch@inria.fr), Géry Casiez (gery.casiez@univ-lille.fr)

The goal of this Ph.D. is to explore new directions of explainability approaches that allow users to interactively explore the trade-offs of competing ML models with the help of intelligent agents.

Context

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.

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 user's 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.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
Inria
Université de Lille
CRIStAL
Centre de Recherche en Informatique, Signal et Automatique de Lille

Objectives

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.

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.

References

[1] M. Beaudouin-Lafon and W. E. Mackay. Reification, polymorphism and reuse: three principles for designing visual interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '00, pages 102–109, Palermo, Italy. Association for Computing Machinery, 2000. DOI: 10.1145/345513.345267.

[2] G. C. Eva Mackamul and S. Malacria. Clarifying and differentiating discoverability. *Human–Computer Interaction*, 0(0):1–26, 2024. DOI: 10.1080/07370024.2024.2364606.

[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 *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, pages 2037–2056, Copenhagen, Denmark. Association for Computing Machinery, 2024. DOI: 10.1145/3643834.3661583.

[5] W. Mackay. Responding to cognitive overload: co-adaptation between users and technology. *Intellectica*, 30, July 2000. DOI: 10.3406/intel.2000.1597.

[6] X. Peng, J. Koch, and W. E. Mackay. Designprompt: using multimodal interaction for design exploration with generative ai. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, pages 804–818, 2024.

[7] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni. Green ai. *Communications of the ACM*, 63(12):54–63, 2020.

Candidate

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.

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Inria

Université
de Lille

CRIStAL
Centre de Recherche en Informatique,
Signal et Automatique de Lille