Title: A hybrid approach to the explainability of artificial intelligence algorithms for personalised health care

Advisors:
Nédra Mellouli: n.mellouli@iut.univ-paris8.fr
&
Ivan Varzinczak: ivarzinczak@icloud.com

Internship location:
EID Team, LIASD, Université Paris 8. 140, rue de la nouvelle France, 93100 Montreuil.

Topic description:
Over the past decade, there has been active research into healthcare services and their technological advancements. In particular, the Internet of Things (IoT) has demonstrated its potential to connect numerous medical devices, sensors, and healthcare professionals to provide high-quality medical services in remote locations. This trend was greatly enhanced during the COVID-19 outbreak. The result is an increase in patient safety, a decrease in healthcare spending, an increase in accessibility of healthcare services, and an increase in the operational efficiency of the healthcare sector. However, all these benefits are not without negative consequences for patients and even for healthcare workers. Indeed, artificial intelligence is increasingly being integrated into diagnostic systems, taking advantage of the availability of big data. Deep Learning (DL) applied to medical images for the diagnosis of cancer, and other diseases has led to black-box diagnostics systems with astounding results in terms of accuracy that often surpass those by expert clinicians. However, to be used for effective decision support in a perhaps stressed situation, a black-box oracle answer positive/negative is not enough; some explanation is needed. Abduction and Argumentation are two forms of inference where conclusions are drawn according to an underlying theory. Typically, abduction aims to draw an explanation for a set of observations, while argumentation aims to give reasons, or arguments, that support a conclusion against other conflicting conclusions. Abduction is sometimes described as "deduction in reverse", whereby given a rule "A follows from B" and the observed result "A", we infer that the condition "B" of the rule (may) hold. More generally, in the context of a logic-based setting, given a set of sentences representing a theory T that models a medical diagnosis domain of interest, and a sentence representing an observation O, abduction returns a set of sentences representing an abductive explanation H for O. The distinguishing feature of this project is to design and develop such tools in a collaborative design (CD) process together with medical staff experienced in the diagnosis and who represents the final users of this technology.

The main research question of this internship is how to link abduction and formal argumentation theory with learning-based approaches to address the aforementioned problem. Indeed, reasoning and learning play a complementary role in decision-making: learning produces the knowledge taken for granted when reasoning, whereas systematic reasoning draws inferences that provide the inductive bias that is assumed as given when learning. Hence, the main goal of the internship is to exploit the synergy between learning and reasoning, especially abduction and argumentation, to enhance learning-based processes.

Duration and candidate profile:

The internship duration is six months ("stage fin d'études"). The starting date must be before the end of March 2023, preferably at the beginning of the month. We are looking for a candidate interested in this topic with a background in artificial intelligence, knowledge representation and reasoning, formal logic, deep learning, and human-computer interaction.

References:

1-Ben-Ari, M. Mathematical Logic for Computer Science. 3rd edition. Springer-Verlag, 2012.

2-Mooney, R.J. Integrating Abduction and Induction in Machine Learning. In Abduction and Induction; Flach, P.A., Kakas, A.C., Eds.; Applied Logic Series 18; Springer: Kluwer Academic Publishers, pp.181-191, 2000.

3-Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. Inf. Fusion **2020**, 58, 82–115. [Google Scholar] [CrossRef][Green Version]

5-von Wright, G.H. Explanation and Understanding; Cornell University Press: Cornell, NY, USA, 1971. [Google Scholar]

6-Stepin, I.; Alonso, J.M.; Catala, A.; Pereira-Fariña, M. A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence. IEEE Access **2021**, 9, 11974–12001. [Google Scholar] [CrossRef]

7-Chiffi, D.; Pietarinen, A.-V. Abductive Inference within a Pragmatic Framework. Synthese **2020**, 197, 2507–2523. [Google Scholar] [CrossRef]

8-Wang-Zhou Dai, Qiuling Xu, Yang Yu, and Zhi-Hua Zhou. Bridging Machine Learning and Logical Reasoning by Abductive Learning. In 33rd Conference on Neural Information Processing Systems (NeurIPS), pages 2815–2826, 2019.