Université Lille Nord de France
Pôle de Recherche et d'Enseignement Supérieur

Ecole doctorale régionale Sciences Pour l'Ingénieur
Lille Nord-de-France - 072

Titre : Apport des Connaissances Sémantiques dans la mise en œuvre l'Explicabilité des processus d'Intelligence Artificielle
Financement prévu : 100 % ULCO
Cofinancement éventuel :
(Co)-Directeur de thèse : Mourad Bouneffa
E-mail : **mourad.bouneffa@univ-littoral.fr**
Encadrants : **Grégory Bourguin**
E-mail : **gregory.bourguin@univ-littoral.fr**

Laboratoire : LISIC (Laboratoire d'Informatique Signal et Image de la Côte d'Opale, EA 4491)
Equipe : Model

Description

Cette thèse est consacrée à la problématique de l'explicabilité des processus de raisonnement des systèmes d'Intelligence Artificielle. L'importance de cette problématique a été soulignée dans de nombreux travaux comme le rapport Villani sur l'IA où il a été clairement établi que le manque d ' "explicabilité des IA" constitue un frein sérieux au développement de l'IA dans des domaines critiques et/ou sensibles comme la médecine, la conduite automatique de véhicules ou les systèmes d'aide à la décision dans des domaines sociaux comme l'attribution des bourses d'étude et/ou de logements sociaux, etc.

De nombreux travaux ont été menés ces dernières années pour adjoindre des explications aux systèmes d'IA, en particulier ceux à base de réseaux de neurones profonds, qui sont à la fois très performants, mais aussi par principe conçus et mis en œuvre comme des boites noires.
Les travaux de recherche les plus récents explorent diverses pistes pour réaliser l'explicabilité. Par exemple, certains outils réalisent des traitements sur les modèles d'IA dans le but de réifier les liens entre des données et un modèle ayant conduit à une décision spécifique. D'autres travaux s'emploient à produire des versions simplifiées de modèles d'IA existants. Enfin, certains travaux tentent d'altérer le processus même de génération de modèles dans le but de directement fournir des modèles plus explicables. Si les approches divergent, leur but commun est de produire des systèmes d'IA permettant de fournir des explications qui soient compréhensibles par les acteurs humains concernés (médecins, ingénieur de production ou de maintenance, etc.). De notre point de vue, il s'agit de produire des explications à un niveau d'abstraction sémantique en adéquation avec leurs connaissances. L'étude de l'état de l'art permet constater que si ces notions de connaissances et de sémantique adaptée sont souvent implicites, elles ne sont encore elles-mêmes que rarement réifiées.

Les travaux de cette thèse rentrent dans ce cadre et ont pour but de montrer que les approches pour l'explicabilité gagneraient fortement à mieux considérer la connaissance et sa sémantique comme des

objets de premier ordre en les liant explicitement aux systèmes d'IA grâce aux outils qui permettent de les expliciter et de les manipuler, comme les graphes de connaissances et les ontologies. Elles concerneront aussi bien le domaine métier considéré que les profils des utilisateurs, ainsi que la nature des algorithmes d'IA utilisées et leurs spécificités. Ces travaux développés sont au cœur de la thématique du groupe de travail SysReIC (Systèmes Réflexifs et Ingénierie des Connaissances).

Références bibliographiques

1) Adeel Ahmad, Henri Basson, Mourad Bouneffa, For a better Assessment of Business Process Quality, International Conference on Computer, Software, Modeling and Simulation (CSMS2019), DEStech Transactions on Computer Science and Engineering (ISSN: 2475-8841). DOI: 10.12783/dtcse/iteee2019/28825, (2019)

2) Barredo Arrieta A., et al., Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, arXiv:1910.10045, (2019)

3) Basson H., Bouneffa M., Matsuda M., Ahmad A., Chung D., Arai E., Manufacturing Software Units: ISO 16300-3 main Guidelines for Interoperability Verification and Validation In: Popplewell K., Thoben KD., Knothe T., Poler R. (eds) Enterprise Interoperability VIII. Proceedings of the I-ESA Conferences, vol 9. Springer, Cham (2019) pp 167-176 DOI: 10.1007/978-3-030-13693-2_14

4) Bourguin Grégory, Lewandowski Arnaud, ABCT : The Activity Based Contextual Tagging Ontology, IC3K 2018, 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Seville, Spain, 18-20 September, (2018).

5) Confalonieri, Roberto et al. "An Ontology-based Approach to Explaining Artificial Neural Networks." ArXiv abs/1906.08362 (2019)

6) Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should i trust you?: Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, (2016).

7) Rudin, Cynthia. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." Nature Machine Intelligence 1.5 (2019): 206-215.

8) Villani, Cédric, et al. *Donner un sens à l'intelligence artificielle: pour une stratégie nationale et européenne*. Conseil national du numérique, 2018.

9) Zhang, Quanshi et al. “Interpretable Convolutional Neural Networks.” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2017): 8827-8836.

10) Zhong, Ray Y., et al. "Intelligent manufacturing in the context of industry 4.0: a review." *Engineering* 3.5 (2017): 616-630.