

Proposition de Stage Master 2 **Année 2023**

Titre du stage : Explicabilité dans les équipes humaines/informatiques

Filières recherchées : Informatique

Durée : 4-6 mois

Date de début : Février ou mars 2023

Lieu :

Laboratoire Connaissance et Intelligence Artificielle Distribuées (CIAD) EA 7533
Université de Bourgogne Franche-Comté, UTBM, F-90010 Belfort, FRANCE

Niveau : Master 2 (ou ingénieur projet de fin d'étude)

Indemnités de stage : environ 580 euros/mois

Encadrement :

Dr. Yazan MUALLA
Maître de Conférences
Contact : yazan.mualla@utbm.fr

Participants :

- Pr. Stéphane GALLAND, CIAD, France
- Pr. Franck GECHTER, CIAD, France
- Doctorant, à partir d'décembre 2022 (contrat doctoral établissement), CIAD, France
- Dr. Amro NAJJAR, Luxembourg Institute of Science and Technology, Luxembourg

Résumé :

L'interaction homme-machine (HCI) est un sujet important des prochaines années que ce soit dans le contexte des applications industrielles et pédagogiques. Au-delà des aspects techniques et technologiques, les verrous scientifiques portent également sur des problématiques liées à l'Humain que ce soit d'ordre cognitif, perceptif, voire biologique.

Comprendre le comportement des machines devient essentiel pour garantir une interaction fluide d'Humains/Machine. En effet, il n'est pas simple pour les humains de comprendre les processus qui ont amenés aux décisions des machines. De récentes études dans le domaine de l'intelligence artificielle explicable (XAI) ont confirmé qu'expliquer le comportement d'une machine à un humain favorise la compréhensibilité de la machine par ce dernier et augmente son acceptabilité.

Ce stage a pour objectif de traiter de la problématique de la explicabilité au sein d'équipes mixtes d'Humains/Machine devant accomplir collectivement une tâche donnée. Dans les projets existants, l'explicabilité est effectuée entre l'homme et les machines individuellement, où un robot explique son comportement et ses décisions à l'homme. Cependant, lorsque plusieurs machines ou ordinateurs fournissent des explications contradictoires, la compréhensibilité et la confiance de l'humain peuvent être affectées. De plus, il est possible

d'envisager des changements de rôle de l'humain en cours de fonctionnement de façon à mieux s'adapter au contexte. Il peut être humain sur la boucle (Human-on-the-loop) recevant des décisions et des plans, ou humain en boucle (Human-in-the-loop) qui modifie l'environnement.

Pour traiter ces problématiques, plusieurs aspects sont primordiaux pour spécifier les interactions entre les éléments du système Humain/Machine. Parmi ces aspects, nous pouvons citer la génération des explications par les machines/ordinateurs/robots, la communication des explications aux humains et la manière dont les humains les reçoivent, la communication entre les éléments du système, les actions de contrôle des entités autonomes, l'observation du comportement de l'être humain, la définition du comportement autonome de ces entités. Ceci implique donc des capacités d'analyse de grande quantité de données afin de pouvoir fournir à l'Humain une représentation synthétique de la situation.

Objectif

Sur les bases des simulateurs existants au sein du CIAD, il faudra mettre en place un simulateur intégrant des modèles et algorithmes d'explicabilité et devant permettre la simulation d'équipes mixtes Humains/Machines selon des scénarii liés aux tâches envisagées. Ces scénarii seront définis en collaboration avec le/la doctorant en fonction des besoins liés aux projets du laboratoire.

Résultat attendu

Une simulation logicielle permettant d'exécuter des algorithmes de explicabilité et générer des explications.

Références

Reuves internationales avec comité de lecture :

Yazan Mualla, Igor Tchappi Haman, Timotheus Kampik, Amro Najjar, Davide Calvaresi, Abdeljalil Abbas-Turki, Stéphane Galland, and Christophe Nicolle. *The Quest of Parsimonious XAI: a Human-Agent Architecture for Explanation Formulation*. Artificial Intelligence (SJR Quartile: **Q1**, Clarivate Analytics Impact Factor 2021: **9.088**) [Long article]. DOI: 10.1016/j.artint.2021.103573

Yazan Mualla, Amro Najjar, Alaa Daoud, Stéphane Galland, Christophe Nicolle, Ansar-UI-Haque Yasar, and Elhadi Shakshuki. *Agent-Based Simulation of Unmanned Aerial Vehicles in Civilian Applications: A Systematic Literature Review and Research Directions*. International Journal of Future Generation Computer Systems (SJR Quartile: **Q1**, Clarivate Analytics Impact Factor 2019: **7.187**), Elsevier, vol. 100, pp. 344-364 (2019) [Long article]. DOI: 10.1016/j.future.2019.04.051

Conférences internationales avec comité de lecture :

Yazan Mualla, Igor Tchappi Haman, Timotheus Kampik, Amro Najjar, Davide Calvaresi, Abdeljalil Abbas-Turki, Stéphane Galland, and Christophe Nicolle. *A Human-Agent Architecture for Explanation Formulation (An extended abstract)*. 33rd Benelux Conference on Artificial Intelligence, Belval, Esch-sur-Alzette (Luxembourg), 2021.

Yazan Mualla, Stéphane Galland, and Christophe Nicolle. *Explaining the Behavior of Remote Robots to Humans (Extended abstract)*. 33rd Benelux Conference on Artificial Intelligence, Belval, Esch-sur-Alzette (Luxembourg), 2021.

Yazan Mualla, Amro Najjar, Stéphane Galland, Christophe Nicolle, Igor Tchappi Haman, Ansar-UI-Haque Yasar, and Kary Främling. *Between the Megalopolis and the Deep Blue Sky: Challenges*

of Transport with UAVs in Future Smart Cities. International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Montreal, Canada, pp. 1649-1653 (2019) [*Regular Blue Sky paper, AAMAS is the most important international conference in the domain of Multi-agent systems, CORE Rank: A**]. [Link](#)

Yazan Mualla, Igor Tchappi Haman, Amro Najjar, Timotheus Kampik, Stéphane Galland, and Christophe Nicolle. *Human-Agent Explainability: An Experimental Case Study on the Filtering of Explanations*. 12th International Conference on Agents and Artificial Intelligence (ICAART 2020), Volume 1: HAMT, ISBN 978-989-758-395-7, ISSN 2184-433X, pages 378-385 [*Long paper, HAMT session, CORE Rank: C*]. DOI: 10.5220/0009382903780385

Workshops/ateliers internationaux avec comité de lecture :

Yazan Mualla, Timotheus Kampik, Amro Najjar, Stéphane Galland, and Christophe Nicolle. *Explainable Agents as Static Web Pages: UAV Simulation Example*. 2nd International Workshop on EXplainable TRansparent Autonomous Agents and Multi-Agent Systems, AAMAS, Auckland, New Zealand, pp. 149-154 (2020) [*Demo paper*]. [Link](#)

Davide Calvaresi, **Yazan Mualla**, Amro Najjar, Stéphane Galland, and Michael Schumacher. *Explainable Multi-Agent Systems through Blockchain Technology*. 1st International workshop on EXplainable TRansparent Autonomous Agents and Multi-Agent Systems, AAMAS, Montreal, Canada, pp. 41-58, Springer, Cham (2019) [*Long paper*]. [Link](#)

Egberdien Van der Torre, **Yazan Mualla**, Thiago Jorge Bourscheid, Sana Nouzri, Yolanda Spinola Elias and Amro Najjar. *XAI & Human Synergies to Explain the History of Art*. 3rd International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems [*Long paper*].

Alaa Daoud, Hiba Alqasir, **Yazan Mualla**, Amro Najjar, Gauthier Picard and Flavien Balbo. *Towards Explainable Recommendations of Resource Allocation Mechanisms in On-Demand Transport Fleets*. 3rd International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems [*Long paper*].

Communications orales sans actes dans un congrès international ou national :

Yazan Mualla, Amro Najjar, Timotheus Kampik, Igor Tchappi Haman, Stéphane Galland, and Christophe Nicolle. *Towards Explainability for a Civilian UAV Fleet Management Using an Agent-Based Approach*. 1st Workshop on Explainable AI in Automated Driving: A User-Centered Interaction Approach, Utrecht, Netherland, preprint arXiv:1909.10090 (2019) [*Short paper, 7 pages*]. [Link](#)