



Université Claude Bernard



Lyon 1

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **27 octobre 2022**

Nom de famille et prénom de l'auteur : **Monsieur CHAPUT Rémy**

Titre de la thèse : « *Apprentissage de comportements alignés sur des valeurs morales dans un système multi-agent : guider l'apprentissage par renforcement avec des jugements symboliques* »



Résumé

Les précédentes décennies ont vu un immense progrès des techniques d'Intelligence Artificielle, dans de nombreux domaines, allant jusqu'à atteindre, voire dépasser, les performances humaines dans certains d'entre eux. Cela a mené des systèmes informatiques équipés de telles techniques d'IA à quitter les environnements contraints et artificiels des laboratoires, pour être déployés dans notre monde et notre société humaine, afin de résoudre des tâches ayant un impact bien réel. Ces systèmes ont une influence plus ou moins directe sur des humains, que ce soit leur vie pour les cas les plus extrêmes, ou de manière plus subtile, mais plus ubiquitaire, leur quotidien. Des questions se posent ainsi quant à leur capacité d'agir en accord avec les valeurs (morales) qui nous semblent importantes.

Divers champs de recherche se sont intéressés à des aspects de ce problème, tels que la capacité à fournir des décisions équitables et justes, ou encore la capacité à être intelligible, et ainsi fournir aux utilisateurs humains des raisons d'accorder leur confiance, et de savoir quand ne pas l'accorder. Dans cette thèse, nous nous concentrons particulièrement sur le domaine des *Machine Ethics*, qui consiste à produire des systèmes ayant les moyens d'intégrer des considérations éthiques, c'est-à-dire des systèmes ayant une prise de décision éthique, en accord avec les valeurs humaines qui sont importantes à la société.

Notre but est ainsi de proposer des systèmes, qui soient capables d'apprendre à exhiber des comportements jugés comme éthiques par les humains, à la fois dans des situations ayant des enjeux éthiques non en conflit, mais aussi dans les cas plus complexes de dilemmes entre les valeurs morales. Nous proposons 3 contributions, chacune ayant un objectif différent, pouvant être prises indépendamment les unes des autres, mais ayant été conçues pour s'associer afin de combiner leurs avantages, et de répondre à la problématique globale.

Premièrement, nous proposons un algorithme d'apprentissage par renforcement, capable d'apprendre à exhiber des comportements intégrant ces considérations éthiques à partir d'une fonction de récompense. Le but est ainsi d'apprendre ces enjeux éthiques dans de nombreuses situations, au fil du temps. Un cadre multi-agent est utilisé, ce qui augmente d'une part la richesse de l'environnement, et d'autre part offre une simulation plus réaliste, plus proche de notre société humaine, intrinsèquement multi-agent, et dans laquelle ces approches sont vouées à être déployées. Nous nous intéressons particulièrement à la question de l'adaptation des agents aux changements, à la fois aux dynamiques de l'environnement, tels que les changements saisonniers, mais aussi aux variations dans les mœurs éthiques couramment acceptées par la société.

Notre deuxième contribution se concentre sur la conception de la fonction de récompense, afin de guider l'apprentissage. Nous proposons l'intégration d'agents juges, se basant sur du raisonnement symbolique, chargés de juger les actions des agents apprenants et déterminer leur récompense, relativement à une valeur morale spécifique. L'introduction de multiples agents juges permet de rendre explicite l'existence de multiples valeurs morales. L'utilisation de jugement symbolique facilite la conception par des experts du domaine applicatif, et permet d'améliorer l'intelligibilité des récompenses ainsi produites, ce qui offre une fenêtre sur les motivations que reçoivent les agents apprenants.

Troisièmement, nous nous focalisons plus précisément sur la gestion des dilemmes. Nous profitons de l'existence de multiples valeurs morales afin de fournir plus d'informations aux agents apprenants, leur permettant ainsi d'identifier explicitement ces situations de dilemme, lorsque 2 valeurs morales (ou plus) sont en conflit et ne peuvent être satisfaites en même temps. L'objectif est d'apprendre, d'une part à les identifier, et d'autre part à les trancher, en fonction des préférences des utilisateurs et utilisatrices du système. Ces préférences sont contextualisées, elles dépendent ainsi à la fois de la situation dans laquelle se place le dilemme, et à la fois de l'utilisateur.

Nous nous inspirons des techniques d'apprentissage multi-objectif afin de proposer une approche capable de reconnaître ces conflits, d'apprendre à reconnaître les dilemmes qui sont similaires, autrement dit ceux qui peuvent être tranchés de la même manière, et finalement d'apprendre les préférences des utilisateurs et utilisatrices.

Nous évaluons ces contributions sur un cas d'application que nous proposons, définissons et implémentons : la répartition d'énergie au sein d'une *smart grid*.