



“ARTIFICIAL INTELLIGENCE STAKES IN THE SOCIETY” (AISS)

PRESENTATION DU PROJET AISS

L'intelligence artificielle (IA) est bien plus qu'une technologie : ses applications ont un impact sur tous les acteurs sociaux et pourraient modifier la place de l'humain dans la société, les normes, les pratiques, les croyances, les comportements, les questions d'éthique, de droit, de politique, d'interactions sociales et même les pratiques scientifiques. La complexité croissante de ce nouvel environnement le rend encore plus difficile à comprendre compte tenu de son contexte de déploiement mondial et de son impact potentiel sur l'ensemble des domaines de la connaissance.

La révolution de l'IA génère une quantité massive de signaux que les sciences humaines peuvent aider à comprendre. Laissée au seul domaine de l'informatique et des mathématiques, cette révolution pourrait renforcer les phénomènes de fracture sociale, de communautarisme, d'inégalités, d'exclusion, de discrimination, d'impact écologique, de manipulations de masse et de responsabilité faible ou nulle sur le comportement et le rôle des systèmes basés sur l'IA, tant au niveau local que mondial. Alors que nous sommes au tout début de la révolution de l'IA, le moment paraît unique : nous avons pour la première fois l'opportunité de comprendre les enjeux de l'IA dans la société, et donc, de prendre en main ces questions. Pour y parvenir, il paraît essentiel d'agir de manière proactive pour combler le fossé entre l'informatique et les mathématiques, et les sciences humaines et sociales.

Le projet de recherche AISS (Artificial Intelligence Stakes in the Society), financé dans le cadre de l'appel à projets 2020 *Paris 1 UNA EUROPA*, est organisé collaborativement par l'IRJS, le PIREH du LAMOP et le CRI avec le soutien de l'EDS et de l'IHPST de Paris 1 et d'UNA Europa. Il implique une équipe d'enseignants-chercheurs de l'Université Paris 1 Panthéon-Sorbonne et des partenaires UNA Europa (University of Edinburgh, Université Helsingin Yliopisto, Alma Mater Studiorum Università di Bologna, Freie Universität Berlin, Universidad Complutense de Madrid, Katholieke Universiteit Leuven ; Uniwersytet Jagielloński), respectivement choisis pour leur expertise en IA : en informatique et en mathématiques, ainsi qu'en sciences humaines et sociales (notamment en droit, philosophie, histoire, économie, sciences de la gestion et sciences politiques).

OBJECTIFS SCIENTIFIQUES DU PROJET AISS

Le projet AISS a pour objectif d'explorer l'impact social de l'intelligence artificielle (IA) en adoptant une approche réflexive et multidisciplinaire. Les objectifs scientifiques sont triples :

(1) mettre en perspective les principaux concepts qui structurent la recherche interdisciplinaire sur l'IA ;

(2) développer des outils de connaissance et de compréhension dans les disciplines concernées sous la forme de nouveaux concepts de recherche ;

(3) structurer des concepts pour favoriser le développement de l'IA dans la société, de la réglementation au développement des entreprises.

ORGANISATION DU PROJET AISS

Le projet AISS est organisé en 3 phases :

Phase 1 : Échange de connaissances. L'objectif est de permettre aux chercheurs des différentes disciplines et universités impliquées dans le projet de combler le fossé entre les disciplines et les cultures de l'IA. La principale attente de cette activité de découverte est qu'un noyau unique de compréhension partagée soit développé, établissant une pierre angulaire pour faciliter la recherche future entre les disciplines. Au cours de cette activité, un corpus de concepts partagés sera mis en évidence.

Phase 2 : Développement des connaissances. Le développement des connaissances se fera dans le cadre d'une série de séminaires « ateliers-discussion » où les chercheurs travailleront en coopération pour formaliser le corpus de concepts partagés. Pour se comprendre mutuellement, il s'agira alors de révéler les connaissances tacites, identifier les concepts qui semblent ambigus si l'on considère toutes les disciplines alors qu'ils semblent parfaitement bien définis dans chacune d'entre elles, et les formaliser de manière unique. Il en résultera un ensemble structuré de connaissances qui formalisera les concepts obtenus lors de la phase 1.

Phase 3 : Diffusion des connaissances. La diffusion se fera par le biais d'un symposium international, d'un livre et d'un site web. Le site web présentera : des articles de membres des projets, des rapports, des podcasts et des vidéos prises pendant les ateliers et le symposium international. Le symposium international clôturera le projet en présentant nos conclusions au public, aux autorités et aux chercheurs. Ces trois livrables seront le résultat d'un travail collectif - à l'image de l'organisation du projet -, plutôt qu'une somme de contributions individuelles.

Les ateliers-discussion se dérouleront en deux temps :

- Un premier séminaire « de rencontre » réunira les chercheurs de Paris 1 afin d'identifier les concepts communs et tracer les premières lignes d'un dialogue interdisciplinaire
- Un second séminaire « ouvert » réunira les collègues des Universités partenaires UNA EUROPA ainsi que des professionnels du secteur pour approfondir et diffuser auprès d'un plus large public ces connaissances partagées

PROGRAMME DES SEMINAIRES « PILOTES »

(i) Algorithmes. Décembre 2020-Janvier 2021

Les premières discussions entre experts révèlent que le terme "algorithme" est clairement utilisé avec des acceptions différentes dans des contextes différents. Nous devons clarifier le concept d'algorithme en comparant son utilisation dans divers domaines disciplinaires tels que la statistique et la calculabilité, l'informatique et la philosophie. L'idée de réintroduire le "mot" dans une perspective temporelle peut implicitement nous amener à susciter une polysémie qui se retrouve également dans la terminologie utilisée par le législateur et sa conformité à la réalité technique.

(ii) Discriminations. Mars-Avril 2021

Il s'agira d'identifier l'impact des biais algorithmiques sur la prise de décision afin de comprendre dans quelle mesure les effets discriminatoires potentiels (ethniques, de genre, culturels, etc.) peuvent être identifiés. Des solutions seront recherchées pour définir de nouvelles pratiques et améliorer la prise en compte de la diversité dans l'IA.

(iii) Explicabilité. Avril-Mai 2021

L'exigence émergente d'explicitation des décisions prises sur la conception de systèmes basés sur l'IA suppose de délimiter les limites des nouvelles réglementations et normes nationales et internationales. Le principal défi à relever ici est de savoir comment les algorithmes contemporains produisent ce que l'on attend réellement d'eux, et fournissent tout le matériel nécessaire pour interroger la responsabilité des divers acteurs interagissant avec ces systèmes.

(iv) Émotions. Mai-Juin 2021 (sous réserve)

L'objectif est de comprendre l'impact de l'IA émotionnelle et les défis de son exploitation, tant dans les processus décisionnels (par exemple dans le contexte du e-marketing) que sur la liberté individuelle.

(v) Colloque international de clôture. Juin 2021

Le Symposium international de clôture résumera les réflexions menées au cours des différents ateliers, rendra compte de leurs conclusions concernant l'impact social de l'AI et débattrà des propositions avec le public.

*The goal of the AISS (“**Artificial Intelligence Stakes in the Society**”) project is to explore the social impact of Artificial Intelligence (AI) with a reflexive and multidisciplinary approach reflecting the unique expertise of UNA Europa in AI - both Computer Sciences, and Mathematics, and Human and Social Sciences: Law, Philosophy, History, Economy, Management Sciences, Political Sciences, etc. The project will : (1) put in perspective the main concepts that structure interdisciplinary research about AISS (2) develop knowledge and understanding tools in the disciplines involved under the form of new concepts enactable for research; and (3) structuring concepts to foster the development of AI in the society, from regulation to business development.*