

Partenariat NSF-Amazon sur IA et équité : trois projets retenus portés par des chercheurs de l'université Carnegie Mellon



Le 10 février 2020, la *National Science Foundation* (NSF) et l'entreprise Amazon ont annoncé [11 nouveaux projets](#) portés par 37 chercheurs de différentes universités américaines qui s'intéressent à l'équité et l'intelligence artificielle (IA). Les quatre thèmes couverts dans cette seconde phase d'une coopération initiée en 2019 sont : (1) fondations théoriques et algorithmiques, (2) principes pour l'interaction humain-systèmes d'IA, (3) technologies permettant la compréhension du langage naturel (*Natural Language Understanding* ou NLU) et la vision par ordinateur (*Computer Vision*), et (4) applications de l'IA dans le cadre de décisions d'embauche, pour l'éducation, la justice pénale ou les services sociaux.

1. Historique sur la coopération NSF-Amazon

Pour rappel, dès mars 2019, la NSF et Amazon [avaient annoncé ce partenariat](#) pour la recherche sur l'équité de l'intelligence artificielle (*fairness in AI*) doté d'un budget initial de \$20 millions pour financer des projets sur les trois prochaines années.

Lors de cette première phase, [10 projets avaient été primés](#) et s'intéressaient déjà principalement à quatre domaines : (1) assurer l'équité dans les algorithmes et

les systèmes qui les intègrent, en commençant par définir et quantifier le concept d'équité, (2) responsabilité et transparence dans les algorithmes d'IA, (3) utilisation de l'IA pour promouvoir l'équité dans la société, et (4) assurer que les bénéfices dérivés de l'IA soient accessibles à tous.

Parmi ces dix premiers projets, le montant des dotations variait entre \$300k et \$1 million. Le [projet de l'université de l'Illinois à Chicago](#) qui s'intéresse aux systèmes d'aide à la prise de décisions, en partenariat avec la ville de Chicago et l'association [Wild Me](#) (qui utilise l'IA pour la recherche sur la conservation des espèces animales en danger) avait reçu le plus grand financement. Porté par l'université Johns Hopkins, un autre projet s'intéressait à la [quantification des conséquences directes et indirectes des inégalités raciales sur les résultats médicaux des patients ayant subi une opération de chirurgie cardiaque](#).

Sur les onze nouveaux projets dévoilés en 2021, Carnegie Mellon University (CMU) à Pittsburgh (Pennsylvanie) tire son épingle du jeu puisque trois de ces projets sont portés par des chercheurs de l'université.

2. Organisation d'audits pour détecter les biais dans les systèmes de *Machine Learning* (ML)

Porté par le [Pr. Jason Hong](#), chercheur principal, ce projet a pour but de [d'aider les équipes de développement en IA à détecter et réduire les biais et stéréotypes négatifs](#) qu'ils peuvent parfois introduire de manière inconsciente. Ces biais sont particulièrement difficiles à identifier lorsque les systèmes développés sont déployés à l'échelle mondiale, dans des contextes socio-culturels et historiques variés, pour lesquels les indicateurs actuels sont insuffisants et des connaissances spécifiques liées aux contextes sont requises.

Le projet cherche à développer un système d'audit des biais « par la foule » (*crowd audit service*) qui tireraient parti du grand nombre de participants (des bénévoles mais aussi des « travailleurs de masse », *crowd workers*) dont la tâche est de repérer les cas de discrimination ou d'iniquité (*unfairness*) dans les systèmes ML. Cette évaluation *par la foule* serait synthétisée pour que les équipes de développement puissent intervenir et corriger leurs algorithmes.

3. Une IA juste (*fair AI*) dans les politiques publiques

Le second projet est mené par [Hoda Heidari](#), assistante professeure et affiliée au [département de Machine Learning](#) ainsi qu'à l'[Institute for Software Research](#).

Partant du constat qu'un nombre grandissant de décisions prises par les agences gouvernementales mais également par les organismes à but non lucratif sont assistées par des algorithmes et que ces décisions peuvent avoir des conséquences importantes sur le long terme, notamment pour les populations historiquement désavantagées, [le projet se focalise sur trois axes qui lient politiques publiques et usage d'algorithmes ML](#).

Le premier axe consiste en une cartographie reliant les objectifs d'une politique publique (*policy goals*) avec leur formulation algorithmique. Ce travail permet d'engager des discussions avec les parties prenantes (décideurs politiques, personnes ou organismes chargés de la mise en oeuvre des politiques publiques et communautés affectées par ces politiques) sur les données à prendre en compte pour juger de l'équité des décisions prises. Parmi les domaines étudiés, le projet s'intéressera notamment aux politiques de santé, de protection des enfants et de justice pénale.

Le second axe étudie plus spécifiquement l'équité à toutes les étapes du processus d'aide algorithmique à la décision en analysant les conséquences qu'impliquent différents choix dans les méthodes de collecte des données, d'imputation, de sélection des modèles, d'implémentation et d'évaluation des politiques.

Enfin, le troisième axe de recherche porte sur la modélisation des conséquences pour la société que peuvent avoir sur le long terme les décisions en matière de politique publique assistées par les algorithmes.

En termes de méthodologie, le projet s'appuie sur des activités innovantes d'enseignement et de formation qui permettent d'accroître la participation des populations minoritaires dans la conception des recherches et de la technologie.

4. Quantifier et atténuer les disparités dans les technologies du langage

Le [troisième projet](#) mené par des chercheurs de CMU permettra des avancées dans le traitement automatique des langues (*Natural Language Processing* ou NLP) en développant des méthodes permettant aux systèmes de mieux comprendre la façon de parler d'un plus grand nombre de personnes. Ces méthodes permettront notamment de prendre en compte les variations dans les dialectes, le vocabulaire et la mécanique du langage de différentes populations.

[Graham Neubig](#), professeur associé au [Language Technologies Institute](#) est le chercheur principal et collaborera avec des collègues du [Human-Computer Interaction Institute](#) de CMU mais aussi avec [Antonios Anastasopoulos](#) de l'[université George Mason](#), basée à Fairfax en Virginie.

Pour retrouver la liste complète des projets :

- [NSF-Amazon projets 2020](#)
- [NSF-Amazon projets 2021](#)

Rédacteur :

Kévin KOK HEANG, Attaché adjoint pour la Science et la Technologie, deputy-ntics@ambascience-usa.org