

**PROGRAMME INTITUTS ET INITIATIVES**  
**Appel à projet – campagne 2021**  
**Proposition de projet de recherche doctoral (PRD)**  
**SCAI - Sorbonne Center of Artificial Intelligence**

**Intitulé du projet de recherche doctoral (PRD): Analyse multi-critères de l'équité d'algorithmes de fouille de données éducatives**

**Directeur.rice de thèse porteur.euse du projet (titulaire d'une HDR) :**

NOM : **Luengo** Prénom : **Vanda**  
Titre : Professeur des Universités ou  
e-mail : vanda.luengo@lip6.fr  
Adresse professionnelle : pierre et Marie Curie,  
(site, adresse, bât., bureau) Couloir 26-00, Étage 3, Bureau 306  
4 place Jussieu

**Unité de Recherche :**

Intitulé : LIP6  
Code (ex. UMR xxxx) : UMR 7606

**École Doctorale de rattachement de l'équipe (future école ED130-EDITE  
doctorale du.de la doctorant.e) :**

**Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s,  
leur année de 1<sup>e</sup> inscription et la quotité d'encadrement) :**

**Eleonore ferrier-barbut 2018 30%**

**Morais Canellas Camila 2018 30%**

**Olivier Allegre 2020 50%**

-----

**Co-encadrant.e :**

NOM : **Bouchet** Prénom : **François**  
Titre : Maître de Conférences des Universités ou HDR   
e-mail : françois.bouchet@lip6.fr

**Unité de Recherche :**

**Unité de Recherche :**

Intitulé : LIP6  
Code (ex. UMR xxxx) : UMR 7606

**École Doctorale de rattachement de l'équipe (future école doctorale du.de la doctorant.e) :** ED130-EDITE

**Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1<sup>e</sup> inscription et la quotité d'encadrement) :**

MORAIS CANELLAS Camila 40 %  
QUELENNEC Katia 50%  
SERGENT Thomas 25%

**Selon vous, ce projet est-il susceptible d'intéresser une autre Initiative ou un autre Institut ?**

Non  Oui, précisez Choisissez l'institut ou l'initiative :

## **Description du projet de recherche doctoral (*en français ou en anglais*) :**

**Mots-clé :** équité ; éducation ; multi-critère ; combinaison d'algorithmes ; fouille de données

### **1 Contexte**

De plus en plus d'algorithmes d'IA utilisés en éducation intègrent une part statistique et dépendent donc de la disponibilité de données existantes pour être entraînés et calibrés, et ils sont utilisés pour prendre des décisions pouvant affecter la réussite ou l'orientation d'élèves ou d'étudiants.

Cependant, l'équité de traitement est un élément essentiel en éducation, au coeur aussi bien des politiques éducatives à large échelle (e.g. bourses pour compenser certains facteurs sociaux) [4], que des choix pédagogiques au sein d'une unité d'enseignement (e.g. évaluation anonyme, harmonisation des notes). Il est donc crucial que (1) les jeux de données utilisés pour l'entraînement des algorithmes, et (2) les algorithmes de décision automatique (dans un système de type tuteur intelligent) ou d'aide à la décision (dans une approche de type learning analytics) prennent en compte dans (1) leur collecte et (2) leur conception, l'équité pour garantir l'absence de biais. Ces biais peuvent être relatifs à différents facteurs comme le genre, le milieu socio-culturel ou même l'âge alors que la formation tout au long de la vie est désormais indispensable. Cette prise en compte est d'autant plus critique que des études montrent que ce sont les étudiants les plus privilégiés qui bénéficient le plus des nouvelles technologies éducatives [20], d'où un vrai risque d'accroissement des inégalités.

Au-delà de l'éducation, des exemples récents ont montré comment le déploiement de modèles opaques peut involontairement ancrer des pratiques discriminatoires mal identifiées au préalable [1]. Cette prise de conscience a amené à des ajustements de biais d'algorithmes, par exemple par rapport au genre dans les systèmes de traitement automatique de la langue [5], ou de reconnaissance faciale [7]. Parallèlement, certaines questions transverses relatives à l'utilisation

de l'IA sont étudiées par la communauté de recherche FAT\* (Fairness, Accountability, Transparency) [12], notamment pour le développement de métriques de mesure de biais ou de méthodes de corrections de ceux-ci. Ces techniques visent à permettre une utilisation équitable du data mining et machine learning [3] car des modèles avec une même précision globale peuvent avoir des distributions d'erreurs très différentes au sein d'une population donnée (e.g. entre garçons et filles). Mais ce n'est que très récemment que la communauté EDM (Educational Data Mining) a commencé à s'intéresser aux spécificités de ces problématiques appliquées à l'éducation [13, 14]. Les travaux actuels proposent par exemple des techniques d'analyse de données permettant de quantifier le biais d'algorithmes de prédiction de performance ou d'attrition par rapport au genre des étudiants [10], de mesurer l'équité de modèles de notation [15], l'équité d'algorithmes de prédiction de compétences acquises dans des tuteurs intelligents [9], ou les effets perçus de choix (non-)équitables effectués par des algorithmes dans les processus d'admission en université [17]. Par ailleurs, au-delà du besoin scientifique, il y a également une demande sociétale forte. En effet, alors qu'au niveau international la question se pose encore de la propriété des données éducatives [16], dans le contexte européen, le RGPD établit le besoin de garantir l'utilisation faite des données collectées. Sa mise en place a également permis une prise de conscience de la part des utilisateurs du besoin de comprendre la manière dont leurs données sont traitées, renforçant les attentes en termes de respect de la vie privée et d'équité de traitement [8].

## 2 Problématique

L'objectif principal de cette thèse est de proposer une méthode d'analyse multi-critères de l'équité pour différents types d'algorithmes communément employés dans la communauté EDM en fonction de différents jeux de données. En effet, les travaux récents se focalisent tous sur l'évaluation d'un critère unique (le genre, l'ethnicité [en milieu anglo-saxon], l'établissement d'origine...) mais la diversité des critères étudiés montrent bien l'aspect multidimensionnel requis pour prendre des décisions qui soient équitables de manière globale. Cette problématique générale se décline en plusieurs questions de recherche complémentaires qui pourront être investiguées par le candidat retenu : 1. Est-il possible de combiner différents algorithmes équitables selon des critères différents pour obtenir une décision globalement plus équitable ? 2. Peut-on déterminer automatiquement le risque de biais d'équité dans un jeu de données en fonction de divers critères afin de recommander la collecte de nouvelles données auprès d'une population particulière ? 3. Peut-on neutraliser les biais de différents algorithmes dans le temps, en alternant les méthodes de choix pour équilibrer les biais selon différents critères ?

## 3 Pistes

Pour répondre aux différentes questions, nous nous appuyerons notamment sur des travaux antérieurs de l'équipe MOCAH [18] concernant l'utilisation de données multi-aspects (par exemple un profil de joueur et un profil d'apprenant) pour le choix d'une activité, à l'aide des Q-matrices [2]. Nous explorerons ainsi si cette approche de décision peut-être appliquée pour maintenir une meilleure équité globale en fonction de différents aspects. Une autre approche pourra consister à s'appuyer sur les méthodes ensemblistes, précédemment employée notamment dans la thèse de Fatima Harrak [11], en s'intéressant à leur potentiel pour réduire les biais des algorithmes qu'elles combinent.

Pour les aspects expérimentaux, nous nous appuyerons également sur l'accès de l'équipe à de multiples jeux de données éducatives issues : (1) de MOOC locaux (SorbonneX [accès aux données en tant que data czar local]) ou d'autres universités (MOOC Gestion de Projet de Centrale Lille [collaboration depuis 2015 [6]]), (2) de LMS comme Moodle (en collaboration avec le projet LAPAD de Capsule porté par Yves Noël, membre associé MOCAH depuis 2020 [19]) (3) des plateformes des données des entreprises EDTECH avec lesquels on collabore (EvidenceB, Tralalère,..). Des implémentations seront également possibles dans ces différentes plateformes pour tester des variantes des algorithmes développés dans la thèse et permettre de mesurer l'impact réel des corrections de biais d'équité.

## 4 Bibliographie

- [1] Bolukbasi, T., Chang, K.W., Zou, J.Y., Saligrama, V., Kalai, A.T. : Man is to computer programmer as woman is to homemaker ? debiasing word embeddings. In : Advances in neural information processing systems. pp. 349–4357 (2016)
- [2] Bouchet, F., Labarthe, H., Yacef, K., Bachelet, R. : Comparing Peer Recommendation Strategies in a MOOC. In : ACM Extended Proceedings of UMAP 2017. Bratislava, Slovakia (Jul 2017).
- [3] Buolamwini, J., Gebru, T. : Gender shades : Intersectional accuracy disparities in commercial gender classification. In : Conference on fairness, accountability and transparency. pp. 77–91 (2018)
- [4] Cerulus, L., Scott, M. : Europe’s new privacy rules : 1 month in, 7 takeaways. In : Politico (Jun 2018), <https://www.politico.eu/article/gdpr-europe-new-privacy-rules-7-takeaways/>
- [5] Doroudi, S., Brunskill, E. : Fairer but not fair enough on the equitability of knowledge tracing. In : 9th International Conference on Learning Analytics & Knowledge. pp. 335–339 (2019)
- [6] Gardner, J., Brooks, C., Baker, R. : Evaluating the fairness of predictive student models through slicing analysis. In : Proceedings of the 9th International Conference on Learning Analytics & Knowledge. pp. 225–234 (2019)
- [7] Harrak, F., Bouchet, F., Luengo, V. : Categorizing students’ questions using an ensemble hybrid approach. In : Lynch, C.F., Merceron, A., Desmarais, M., Nkambou, R. (eds.) Proc. of the 12th International Conference on Educational Data Mining. pp. 312–317. Montréal, Canada (Jul 2019).
- [8] Hildebrandt, M., Castillo, C., Celis, E., Ruggieri, S., Taylor, L., Zanfir Fortuna, G. : FAT\* ’20 : Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In : Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. ACM 2020).
- [9] Holmes, W., Bektik, D., Woolf, B., Luckin, R. : Ethics in AIED : Who cares ? Chicago (2019), <http://oro.open.ac.uk/60361/>
- [10] Holstein, K., Doroudi, S. : Fairness and equity in learning analytics systems. In : Companion Proceedings of the Ninth International Learning Analytics & Knowledge Conference (2019)
- [11] Hutt, S., Gardner, M., Duckworth, A.L., D’Mello, S.K. : Evaluating Fairness and Generalizability in Models Predicting On-Time Graduation from College Applications. International Educational Data Mining Society (2019)
- [12] Lynch, C.F. : Who prophets from big data in education ? New insights and new challenges. Theory and Research in Education 15(3), 249–271 (Nov 2017).
- [13] Marcinkowski, F., Kieslich, K., Starke, C., Lünich, M. : Implications of AI (un-) fairness in higher education admissions : the effects of perceived AI (un-) fairness on exit, voice and organizational reputation. In : Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 122–130 (2020)
- [14] Monterrat B., Yessad A, Bouchet F, Lavoué E., Luengo V. : “MAGAM: A Multi-Aspect Generic Adaptation Model for Learning Environments”, Data Driven Approaches in Digital Education, vol. 10474, Lecture Notes in Computer Science, Tallinn, Estonia, pp. 139-152, (Springer) (2017)
- [15] Noël, Y., Mergoil, R., Luengo, V., Bouchet, F. : Towards a modular and flexible Learning Analytics framework. In : Proc. of the 10th International Conference on Learning Analytics and Knowledge (LAK20). Frankfurt, Germany (Mar 2020)
- [16] Reich, J., Ito, M. : From good intentions to real outcomes : Equity by design in learning technologies. Digital Media and Learning Research Hub (2017)