



HAL
open science

EISAI: Ethical Information System based on Artificial Intelligence

Saïd Assar, Christine Balagué, Loréa Baïada-Hirèche

► **To cite this version:**

Saïd Assar, Christine Balagué, Loréa Baïada-Hirèche. EISAI: Ethical Information System based on Artificial Intelligence. Séminaire de présentation des projets de recherche Good In Tech, Institut Mines-Télécom Business School (IMT-BS); Télécom SudParis (TSP); Fondation du Risque Louis Bachelier; SciencesPo Paris, Dec 2020, Paris (Séminaire de recherche en ligne), France. hal-03123998

HAL Id: hal-03123998

<https://hal.archives-ouvertes.fr/hal-03123998>

Submitted on 28 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RECHERCHE EISAI

Nature de la demande :

Projet de recherche

ACRONYME et titre du projet :

EISAI

Ethical Information System based on Artificial Intelligence

Auteurs :

Saïd Assar, Professeur en Systèmes d'Information

Christine Balagué, Professeur en Sciences de Gestion

Loréa Baiada-Hirèche, Maître de Conférence en Sciences de Gestion

Institut Mines-Télécom Business School, Laboratoire LITEM, Université Paris Saclay

said.assar@imt-bs.eu

christine.balague@imt-bs.eu

lorea.baiada-hireche@imt-bs.eu

Contexte du projet de recherche :

Durant ces dernières années, l'explosion des données massives issues de technologies numériques et les capacités accrues de l'intelligence artificielle ont généré de nombreuses applications et services au sein des entreprises et plus largement des organisations (systèmes de recommandations, traitement automatique d'images, analyse de réseaux sociaux, modèles prédictifs, détection de fraude, traitement automatique de données vidéos pour le recrutement, prix dynamique, etc.). Demain, le développement de la 5G dans de nombreux pays va accroître la collecte des données et leur stockage dans le cloud, connecter des millions d'objets entre eux quasiment en temps réel, et créer de nouveaux usages. Cependant, ces évolutions souvent présentées comme innovations positives ont des impacts négatifs à la fois sur les individus et la société, montrés par de nombreux rapports et publications scientifiques. Dès 2016, le rapport « Big Data : a report on algorithmic systems, opportunity and civil rights » du bureau exécutif du président Obama montrait les effets de discriminations des systèmes algorithmiques sur les minorités et de l'impact des technologies sur les libertés individuelles. La même année, la FTC publiait le rapport « Big data : a tool for inclusion or exclusion ? » montrant que les technologies utilisant des données massives entraînaient une exclusion d'une partie de la population. La sociologue Virginia Eubanks en 2017 dans son ouvrage « Automating inequalities » décrit une automatisation des inégalités entre les individus par la très large diffusion des services numériques et des systèmes informatiques aux Etats-Unis. Plusieurs articles de recherche ont montré les impacts sociétaux de l'intelligence artificielle et des algorithmes : Eipstein (2015) a montré les effets de manipulation des moteurs de recherche sur les intentions de vote ; Basky et al. (2015) ont révélé dans la revue Science le phénomène des bulles filtrantes sur Facebook et la limite d'accès à une diversité d'opinion ; Lambrecht et Tucker (2019) ont mis en avant des effets de discrimination selon le genre dans les algorithmes de recrutement ; Obermayer et al. (2019) dans Science montrent la discrimination des populations noires dans les algorithmes de santé largement utilisés aux Etats-Unis pour l'accès aux soins. Cette prise en compte des impacts sociétaux négatifs des algorithmes de l'intelligence artificielle et du big data a entraîné ces

dernières années plusieurs mobilisations et travaux internationaux sur la nécessité de développer des technologies responsables et éthiques. La déclaration de Montréal en décembre 2018, portée par l'Université de Montréal, porte dix principes pour une intelligence artificielle responsable et éthique. Le HLEG (High Level Expert Group) sur l'intelligence artificielle, mis en place par la commission européenne en 2018, a publié un rapport en 2019 sur les sept exigences pour une intelligence artificielle de confiance, et a mis en place des groupes pilotes dans les différents pays européens pour avancer sur ces sujets. Plus récemment, le 19 février, la commission européenne a publié le livre blanc « Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance », prônant un cadre de régulation européen, une plus grande transparence des systèmes algorithmiques, et leur certification pour des usages dits sensibles (santé, recrutement, etc...). Selon Virginia Dignum (2019), « *it is not an option to ignore our responsibility. AI systems are artefacts decided upon, designed, implemented and used by people. We, people, are responsible* ». Notre projet de recherche se positionne dans ce courant international de nécessité de développer des technologies plus responsables et éthiques, qui est aussi l'objectif de la chaire Good in Tech.

Etat de l'art et question de recherche :

Les publications sur l'éthique de l'intelligence artificielle sont majoritairement dans les champs disciplinaires de l'informatique, la philosophie, le droit et la sociologie. Paradoxalement, les recherches en systèmes d'information, qui pourtant implémentent et souvent pilotent ces technologies au sein des organisations, restent peu nombreuses sur le sujet. Nous proposons avec ce projet de recherche de combler cet écart et d'étudier comment un projet en systèmes d'informations, utilisant l'intelligence artificielle, peut être mené de manière éthique. Les enjeux éthiques dans les systèmes d'information plus généralement ont fait l'objet de plusieurs publications : Cordoba (2006) mobilise les théories de Foucault, Leclercq-Vandelannoitte et Bertin (2018) reprennent également le concept de gouvernementalité de Foucault pour l'appliquer à la gouvernance des systèmes d'information. Leclercq-Vandelannoitte (2017) présente une perspective éthique pour contrôler les systèmes d'information. Paraschakis (2017) propose un cadre éthique pour les systèmes de recommandations, mais plus globalement, peu de recherches portent sur l'éthique de systèmes d'information utilisant de l'intelligence artificielle. Or, l'un des challenges des systèmes d'information combinant big data, analytics et intelligence artificielle est de transformer ces données en connaissances et systèmes d'aide à la décision (Kersting et Meyer, 2018). Selon Zuboof (2015), nous vivons dans une civilisation de l'information, à la source d'un capitalisme de la surveillance fondé sur une logique d'accumulation de données visant à prédire et modifier le comportement humain comme un moyen de produire du revenu et de contrôler les marchés. Les enjeux éthiques des systèmes d'information utilisant le big data et l'intelligence artificielle sont donc majeurs.

Notre recherche vise d'abord à mieux comprendre les pratiques actuelles dans la prise en compte des questions éthiques dans les projets en systèmes d'information, et ensuite, à proposer un cadre conceptuel de ces pratiques qui servira à élaborer des recommandations sur les différentes dimensions dans un projet.

Méthodologie :

La méthodologie de cette recherche se fonde sur une revue de la littérature pour identifier des hypothèses qui serviront à la conduite d'une vingtaine d'entretiens qualitatifs avec des chefs de projets en systèmes d'information. Ces entretiens seront menés au sein de l'entreprise CGI, partenaire de la chaire Good in Tech. Nous avons l'accord de Christophe Maillet, VP conseil & services chez CGI, et avons déjà organisé une réunion téléphonique avec l'équipe des chefs de projets de SI basés à Brest.

L'objectif des entretiens qualitatifs est d'interroger les pratiques sur les enjeux éthiques des systèmes d'information et de faire réagir les chefs de projets SI participant à l'étude aux dimensions éthiques identifiées. Ces entretiens pourront faire émerger de nouvelles dimensions permettant d'élaborer un cadre éthique pour les systèmes d'information utilisant l'intelligence artificielle. Les entretiens seront ensuite retranscrits et soumis à une analyse de contenu.

Les recommandations issues de cette recherche permettront d'ouvrir des voies de recherche en systèmes d'information sur l'intégration de l'éthique dans les SI utilisant l'intelligence artificielle et le big data, tout en ayant une portée applicative plus large pour tous les chefs de projets en SI souhaitant mettre en place de tels systèmes.

Bibliographie

Bakshy E., Messing S., Adamic L.A. (2015): Exposure to ideologically diverse news and opinion on Facebook, *Science*, vol. 348, issue 23

Barn, B.S., Barn, R. (2018): Human and Value Sensitive Aspects of Mobile App Design: A Foucauldian Perspective. In: International Conference on Advanced Information Systems Engineering (CAiSE). pp. 103–118. Springer, Cham.

Córdoba, J.-R (2006).: Using Foucault to analyse ethics in the practice of problem structuring methods, *Journal of the Operational Research Society*, 57, 1027–1034.

Dignum V. (2019): Responsible Artificial Intelligence: how to develop and use AI in a responsible way, *Artificial Intelligence: Foundations, theory and algorithms*, Springer Nature Switzerland AG2019

Epstein R. and Ronald E. Robertson Ronald E. (2015): The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections, *PNAS* August 18, 2015 112 (33) E4512-E4521

Kersting, K., & Meyer, U. (2018). From Big Data to Big Artificial Intelligence? *Algorithmic Challenges and Opportunities of Big Data*, 32(1), 3-8.

Lambrecht A. et Tucker C. (2019): Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads, *Management Science*, Jul2019, Vol. 65 Issue 7, p2966-2981

Leclercq-Vandelannoitte, A., Bertin, E. (2018): From sovereign IT governance to liberal IT governmentality? A Foucauldian analogy. *European Journal of IS (EJIS)*. 27, 326–346.

Leclercq-Vandelannoitte, A. (2017): An Ethical Perspective on Emerging Forms of Ubiquitous IT-Based Control. *Journal of Business Ethics*. 142, 139–154.

Obermeyer Z., Powers B., Vogeli C ., Mullainathan S. (2019) : Dissecting racial bias in an algorithm used to manage the health of populations, *Science* 25, Vol. 366, Issue 6464, pp. 447-453

Paraschakis, D.(2017): Towards an ethical recommendation framework. In: 11th Int. Conf. on Research Challenges in Information Science (RCIS). pp. 211–220.

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), 626-639.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.