



HABILITATION A DIRIGER DES RECHERCHES

Date de la soutenance : **28 mai 2026**

Nom de famille et prénom de l'auteur : **Monsieur MAURI Andrea**

Titre des travaux : *Dimensions humaines des applications data-intensives*

Résumé



Les applications intensives en données jouent un rôle croissant dans des processus décisionnels critiques dans des domaines tels que la santé, l'élaboration des politiques publiques et les services numériques. Si les communautés du data management et du machine learning ont réalisé des progrès substantiels en matière d'efficacité, de passage à l'échelle et de précision de ces systèmes, le rôle des humains — à la fois comme producteurs de données et comme personnes affectées par les décisions fondées sur les données — demeure insuffisamment intégré dans leur conception et leur déploiement. Dans mes travaux, je défends une perspective globale et centrée sur l'humain des applications intensives en données, dans laquelle les facteurs humains sont considérés comme des préoccupations de premier plan tout au long du cycle de vie des données.

Ce travail est structuré autour d'un modèle conceptuel du pipeline de données comprenant la collecte, le prétraitement et l'analyse des données, et examine comment l'implication humaine peut être intégrée de manière systématique à chaque étape. Concernant la collecte de données, le manuscrit présente des méthodes participatives ainsi qu'une approche de « policy sandboxing » mobilisant l'empathie et l'exposition à des perspectives diversifiées afin de réduire les biais et de favoriser des décisions plus inclusives. Pour le prétraitement des données, il introduit des techniques interactives et centrées sur l'utilisateur visant à améliorer la qualité des données, avec un accent particulier sur les données de graphes. Ces travaux incluent de nouveaux cadres méthodologiques et des études empiriques sur la réparation de graphes centrée sur l'humain, ainsi que des explorations préliminaires de l'usage des grands modèles de langage pour assister ces processus. À l'étape d'analyse, le manuscrit étudie la manière dont les utilisateurs — en particulier les non-experts — interagissent avec des systèmes de données complexes, en présentant des études quantitatives et qualitatives sur l'apprentissage et l'usage des langages de requête pour graphes, et en formulant des recommandations concrètes pour la conception d'outils analytiques plus accessibles.

Au-delà des différentes étapes du pipeline, le manuscrit aborde le défi plus large de la conception d'applications intensives en données prenant en compte les impacts sociétaux, les tensions de valeurs et les considérations liées au care, en mobilisant des méthodes issues de l'interaction humain-machine telles que le design participatif et spéculatif. Dans son ensemble, ce travail propose un programme de recherche cohérent, des fondements

méthodologiques solides ainsi que des systèmes concrets et des résultats empiriques contribuant à faire progresser le champ du Human-Centered Data Management, en montrant comment l'intégration des facteurs humains peut conduire à des applications intensives en données plus fiables, inclusives et efficaces.