

**Domaines d’action clés pour une IA qui intègre davantage la problématique genre**

Note d’information No. 3 **-** Alex Berryhill et Lucía Mesa Vélez (Ladysmith) **-** Décembre 2022

**Objectifs de cette ressource :** Présenter les principaux domaines d’action pour faire progresser une IA qui intègre davantage la problématique genre, afin de s’assurer que les conversations autour de l’égalité des sexes et l’inclusion passent de la théorie à l’opérationnel. *Cette synthèse de la littérature disponible est loin d’être exhaustive, et a pour but de servir de point de départ à des discussions et des actions autour de la quête de l’excellence en matière d’égalité des sexes et d’inclusion pour les laboratoires et les hubs d’AI4D.*

|  |
| --- |
| Ceci est la troisième ressource de la **série de Note d’information de l’équipe de soutien au projet d’intégration d’ESI dans AI4D** : * **[Note d’information No. 1](https://docs.google.com/document/d/1aJ5XdZ5OWkukFFXYZG0cK7Wp7waZMLCQzuMhkbyDoeI/edit)**résume les considérations prioritaires en matière d’ESI pour la conception des appels à propositions AI4D.
* **[Note d’information No. 2](https://docs.google.com/document/d/1scFDntd_eHwyLfahBRYblim74J4VcNpmPD0mOx3oqC0/edit)**fournit une définition opérationnelle des projets qui intègre la problématique genre dans le contexte de la recherche et de l’innovation axées sur l’IA, et synthétise les recherches disponibles qui illustrent pourquoi les considérations en matière d’égalité des sexes et inclusion sont importantes pour la conception et l’utilisation de la prise de décision algorithmique.
 |

L’intégration de la problématique genre semble être un défi particulier pour les projets de recherche et d’innovation en IA, comme l’illustre une base de données de plus en plus importante sur les nombreuses façons dont les applications de l’IA ont contribué et exacerbé les inégalités de genre, raciales, économiques et mondiales (voir la [Note d’information No 2)](https://docs.google.com/document/d/1scFDntd_eHwyLfahBRYblim74J4VcNpmPD0mOx3oqC0/edit).[[1]](#footnote-1) **Il n’existe pas de « listes de contrôle » totalement concluantes sur la façon de « résoudre » la problématique genre du secteur de l’IA** - bien que certaines considérations méthodologiques et organisationnelles puissent aider à « faire bouger l’aiguille » vers une recherche et une innovation en IA qui intègre davantage la problématique genre, dont certaines sont décrites ci-dessous. Pour un impact plus transformateur et durable, les équipes qui exploitent l’IA doivent continuellement réfléchir à *leurs* rôles et responsabilités dans le cadre d’un changement plus large des systèmes organisationnels, culturels et politiques.

En effet, il est important de noter que *l’égalité des sexes et l’inclusion aura une signification distincte pour chaque équipe, projet et contexte spécifique.* Dans cette optique, ces domaines d’action clés peuvent être considérés comme des points de départ essentiels, fondés sur des preuves.

# Considérations clés pour l’intégration de l’égalité des sexes et l’inclusion dans la recherche et l’innovation axées sur l’IA.

1. **Créer des équipes représentatives avec des structures décisionnelles équitables :** La prise en compte du genre commence par la façon dont les équipes sont formées et dont les décisions sont prises. Le secteur de l’IA est bien connu pour son manque de diversité sexuelle et raciale, en particulier dans les postes de leadership et de décision - tout comme les conséquences de cette sous-représentation dans la conception et l’utilisation de l’IA. Lorsque l’on considère la diversité, il est important de prendre en compte l’identité, les expériences, ainsi que la formation professionnelle ou les disciplines (c’est-à-dire qu’une diversité de perspectives interdisciplinaires contribue également à la sensibilité d’une équipe aux questions de genre).

Les premières étapes importantes pour assurer la représentativité d’une équipe de projet consistent à mener une évaluation et une analyse de la composition de l’équipe, ainsi qu’une analyse des obstacles que rencontrent les femmes et les autres groupes marginalisés pour entrer et progresser dans le secteur, dans le contexte spécifique de l’équipe de projet. Ces données peuvent ensuite servir à la conception d’interventions adaptées au contexte.

|  |
| --- |
| **Outils et ressources pertinents :**  |
| * To Build Less-Biased AI, Hire a More-Diverse Team ([Li 2020](https://hbr.org/2020/10/to-build-less-biased-ai-hire-a-more-diverse-team))
* Diverse Teams build better AI. Here’s why ([Shastri 2020](https://www.forbes.com/sites/arunshastri/2020/07/01/diverse-teams-build-better-ai-heres-why/?sh=3dd0710577b3))
* Women in Tech ([here](https://women-in-tech.org/))
* 2021 Women in Tech Report ([TrustRadius 2021](https://www.trustradius.com/buyer-blog/women-in-tech-report))
* How to keep human bias out of AI. ([TED Talk by Kriti Sharma](https://www.youtube.com/watch?v=BRRNeBKwvNM)).
 |

1. **Utiliser des données qualitatives et quantitatives pour élaborer des théories de changement adaptées à la situation :** L’IA peut être utilisée de nombreuses façons pour avoir un impact social positif, notamment pour faire progresser l’égalité des sexes. Cependant, l’un des principaux défis du secteur de l’IA est de supposer que les nouvelles technologies vont à elles seules « réparer » les inégalités et les injustices.

Au lieu de cela, les projets qui intègrent la dimension genre s’appuient sur des éléments probants concernant les causes profondes des défis identifiés, qui nécessitent le plus souvent (presque toujours) une variété d’interventions pour briser les cycles d’inégalité et d’injustice (plutôt que des interventions technologiques seules). Il est important que ces preuves comprennent des données qualitatives et quantitatives, ainsi que les points de vue de ceux qui sont les plus susceptibles de comprendre le défi identifié, comme les communautés touchées et les organisations de la société civile.[[2]](#footnote-2) Pour les défis liés à l’inégalité entre les sexes, il convient d’inclure spécifiquement les mouvements féministes locaux et les organisations de défense des droits des femmes, car ils sont les plus susceptibles de comprendre les points d’entrée clés pour obtenir des résultats positifs en matière d’égalité entre les sexes, ainsi que les risques potentiels pour les activités proposées.

|  |
| --- |
| **Outils et ressources pertinents :**  |
| * Stanford’s Gendered Innovations in Science and Engineering Project ([here](http://genderedinnovations.stanford.edu/index.html))
* Women at the Table’s <AI & Equality> Human Rights Toolbox ([here](https://www.womenatthetable.net/project/ai-equality-human-rights-toolbox/))
* AI4COVID Fiche Technique 1: Concevoir des projets de données qui intègrent la dimension de genre ([Ladysmith 2021](https://covidsouth.ai/research/download/English-Technical-Brief-1-Designing-gender-responsive-data-projects))
* Meeting the challenge of gender inequality through gender transformative research: lessons from research in Africa, Asia, and Latin America ([Njuki et al 2022](https://doi.org/10.1080/02255189.2022.2099356))
* Transformer les relations entre les hommes et les femmes : Enseignements de la recherche du CRDI ([IDRC 2019](https://issuu.com/idrc_crdi/docs/wd_13_000_gender_e-file_en?e=34655515/70235030)).
 |

1. **Identifier et traiter de manière appropriée les biais des données statistiques et sociales :** Il faut partir du principe que toutes les données comportent un certain degré de biais. Il peut s’agir de divers types de « biais statistiques (c’est-à-dire des préoccupations relatives à l’échantillonnage non représentatif et à l’erreur de mesure) » ainsi que de biais sociétaux « (c’est-à-dire des préoccupations relatives aux structures sociales répréhensibles et aux injustices passées qui sont représentées dans les données) ».[[3]](#footnote-3) Bien qu’il n’existe pas de définition concluante de la manière de définir ou de mesurer un ensemble de données « équitable » ou « inclusif », la reconnaissance des biais imbriqués dans les données utilisées pour informer les applications d’IA est un premier pas essentiel vers une plus grande sensibilité au genre. En identifiant les biais potentiels, les équipes peuvent ensuite déterminer les méthodes les plus appropriées pour diagnostiquer plus précisément les biais et y remédier de manière adéquate.

Notez qu’ici, une fois de plus, une variété de perspectives disciplinaires est essentielle pour identifier les biais potentiels qui pourraient autrement passer inaperçus. Par exemple, les sociologues et les mathématiciens peuvent être mieux équipés pour identifier et répondre à différents types de biais, d’où la nécessité d’une collaboration interdisciplinaire.

|  |
| --- |
| **Outils et ressources pertinents :**  |
| * Algorithmic Origins of Bias ([Mandal 2021](https://www.womenatthetable.net/research-advocacy/algorithmic-accountability/algorithmic-origins-of-bias/))
* A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle ([here](https://arxiv.org/pdf/1901.10002.pdf))
* Implicit Association Test (IAT) by Harvard University ([here](https://implicit.harvard.edu/implicit/takeatest.html))
* IBM’s open-source Library AI Fairness 360 to detect and mitigate biases ([here](https://github.com/Trusted-AI/AIF360))
 |

1. **Surveiller les performances et les impacts des modèles, y compris sur les normes sociales et les inégalités :** Une deuxième étape critique vers des ensembles de données plus inclusifs, et donc une recherche et une innovation en IA plus sensibles à la dimension genre, consiste à identifier une méthodologie appropriée pour surveiller de près la performance des modèles de prédiction et toute conséquence involontaire potentielle, y compris les impacts sociaux négatifs potentiels comme l’exacerbation des normes de genre discriminatoires. En plus d’une méthodologie de suivi solide et régulière, les équipes sensibles à la dimension genre s’assurent qu’il existe des moyens efficaces de relier les données de suivi à la prise de décision programmatique (par exemple, devons-nous itérer sur la conception ou l’utilisation de notre modèle, ou devons-nous mettre fin au projet ?)

|  |
| --- |
| **Outils et ressources pertinents :**  |
| * IBM’s Watson OpenScale performs bias checking and mitigation in real time when AI is making its decisions ([here](https://www.ibm.com/cloud/watson-openscale/model-risk-management))
* Google’s ‘What-If tool’ to test algorithms performance with different datasets ([here](https://pair-code.github.io/what-if-tool/index.html#about))
* Local Interpretable Model-Agnostic Explanations (LIME) ([here](https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/))
* FairML: a Python open-source toolbox that is used to audit machine learning predictive models ([here](https://github.com/adebayoj/fairml))
 |

1. **Faciliter la participation des communautés affectées à la conception, la mise en œuvre, le suivi et l’évaluation du projet :** La recherche et l’innovation en matière d’IA tenant compte de la dimension genre impliquent de manière significative les communautés affectées tout au long de la conception, de la mise en œuvre, du suivi et de l’évaluation du projet. Ceci est essentiel du point de vue des droits de l’homme (c’est-à-dire respecter le droit des individus et des communautés à participer aux décisions qui ont un impact sur leur vie), mais aussi du point de vue de l’impact et de la durabilité : La participation des personnes les plus touchées par les nouvelles interventions peut aider à identifier et à atténuer les risques potentiels, et à garantir la pertinence du projet dans les contextes locaux.

La formalisation des partenariats est une première étape importante pour engager avec succès les communautés dans des projets de recherche et d’innovation. Il peut s’agir d’un contrat officiel ou d’un protocole d’entente pour assurer une compréhension commune des détails de la relation. Les équipes doivent également prévoir du temps et de l’espace pour la participation de la communauté dans les plans et processus du projet, afin de s’assurer que ces activités sont privilégiées.

Pour faciliter la participation des communautés touchées tout au long du cycle de vie d’un projet, il faut instaurer un climat de confiance et de bonne volonté avec les organisations représentatives, ce qui peut être un défi dans les contextes où les technologies axées sur l’IA ont créé des dommages, ou lorsque les outils d’IA sont moins bien compris. Pour renforcer les partenariats et créer des opportunités significatives d’engagement et de retour d’information, les équipes doivent investir dans une communication continue, être prêtes à adapter les plans de projet en fonction du retour d’information, et envisager d’investir dans le renforcement des capacités mutuelles.

|  |
| --- |
| **Outils et ressources pertinents :**  |
| * AI4COVID Fiche technique 3 : Engagement des parties prenantes pour une recherche en santé tenant compte de la dimension genre ([Ladysmith 2021)](https://covidsouth.ai/research/download/English-Technical-Brief-3-Stakeholder-engagement-for-gender-responsive-health-research)
* AI4COVID Fiche technique 4 : Connecter les données sur le genre à l’action [(Ladysmith 2022)](https://covidsouth.ai/research?lang=en&page=3)
* L’IA responsable et ses parties prenantes ([Lima et al 2020)](https://deepai.org/publication/responsible-ai-and-its-stakeholders)
 |

# Annexe A : Termes clés

* **Genre** : Le genre fait référence aux rôles, responsabilités et relations socialement construits que la société considère comme appropriés pour les femmes et les hommes. Il a donc également des implications pour les individus et les groupes qui s’identifient comme non conformes au genre. Le genre est soutenu par des institutions politiques, économiques, sociales et culturelles. Le genre est spécifique au contexte et au temps, et donc également changeant.
* **Sexe :** La somme des caractéristiques biologiques et physiologiques qui définissent typiquement les hommes et les femmes, comme les organes reproducteurs, la constitution hormonale, les schémas chromosomiques, les schémas de croissance des cheveux, la répartition des muscles et des graisses, la forme du corps et la structure du squelette.
* **Intersectionnalité :** La manière cumulative dont les effets de multiples formes d’oppression, de discrimination et d’exclusion (y compris, mais sans s’y limiter, le racisme, le sexisme et le classisme) se combinent, se chevauchent ou se croisent.
* **Inclusion :** L’objectif de l’inclusion est d’englober toutes les personnes sans distinction de race, de sexe, de handicap, de besoins médicaux ou autres. Il s’agit de donner un accès et des chances égales et de se débarrasser de la discrimination et de l’intolérance (suppression des barrières). Elle touche tous les aspects de la vie publique.
1. [Mandal, A. (2021). «](https://www.womenatthetable.net/research-advocacy/algorithmic-accountability/algorithmic-origins-of-bias/)Algorithmic Origins of Bias ». Women at the Table.

 [↑](#footnote-ref-1)
2. [Fuentes, L. et Cookson, T.P. (2020).](https://doi.org/10.1080/0966369X.2019.1681371) "Compter l'(in)égalité des sexes ? une critique géographique féministe de la "révolution des données sur le genre"". *Gender, Place & Culture* 27(6) : 881-902. [↑](#footnote-ref-2)
3. [Mitchell, S., et al. (2021). "](https://doi.org/10.1146/annurev-statistics-042720-125902) Algorithmic Fairness : Choices, Assumptions, and Definitions". *Annual Review of Statistics and Its Application* 8:141-163. [↑](#footnote-ref-3)