**Qu’est-ce que l’IA qui intègre la problématique genre, et pourquoi est-ce important ?**

Note d’information No. 2 - Alex Berryhill et Lucía Mesa Vélez (Ladysmith) - Décembre 2022

**Objectifs de cette ressource :** Fournir 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 synthétiser les recherches disponibles qui illustrent pourquoi les considérations en matière d’égalité des sexes et d’inclusion (ESI) sont importantes pour la conception et l’utilisation de la prise de décision algorithmique. *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 deuxième 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. 3**,](https://docs.google.com/document/d/1njqcW3vZycnqSB9t2qAURA5SJWxdccl-gp_24TRhHS8/edit) dans la continuité de la note de synthèse n°2 (cette ressource), identifie les domaines d’action clés pour faire progresser la recherche et l’innovation en matière d’IA qui intègrent la problématique genre. |

**Qu’est-ce que l’IA qui intègre la problématique genre ?**

**Les projets qui intègre la problématique genre** font progresser l’égalité entre les sexes et la justice sociale par leurs pratiques organisationnelles, leurs méthodologies, leurs produits et leurs résultats. De tels projets sont au cœur de la mission du CRDI, comme l’illustre [l’**énoncé d’égalité du CRDI**](https://www.idrc.ca/en/about-idrc/equality-statement) : "Nous appuyons la production de connaissances - notamment par des personnes de sexe divers, de collectivités, d’histoires et d’expériences diverses - qui ***s’attaquent aux systèmes qui perpétuent les inégalités fondées sur l’identité*** *"* (mis en gras par nous).

**L’intégration de la problématique genre semble constituer un défi particulier pour les projets de recherche et d’innovation en matière d’IA**. Qu’il s’agisse des résultats de recherche d’images sexistes de Google, des algorithmes de recadrage d’images racistes de Twitter, des jeux vidéo basés sur l’IA qui encouragent la maltraitance et l’exploitation sexuelle des enfants, ou du diagnostic erroné de maladies potentiellement mortelles[[1]](#footnote-1) – un nombre croissant de preuves illustre les nombreuses façons dont les applications de l’IA ont contribué aux inégalités de genre, raciales, économiques et mondiales, ou les ont exacerbées.

**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 dans la [Note d’information n°3.](https://docs.google.com/document/d/1njqcW3vZycnqSB9t2qAURA5SJWxdccl-gp_24TRhHS8/edit)  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.

|  |  |
| --- | --- |
| **Caractéristiques principales des projets qui intègrent la problématique genre** | |
| **Basé sur les droits de l’homme** | Les droits de l’homme sont au cœur des projets qui intègrent la problématique genre. Cela signifie, entre autres, que les préoccupations de protection sont prioritaires par rapport aux désirs de données.[[2]](#footnote-2) Par exemple, alors que les données sur la violence basée sur le genre peuvent contribuer à éclairer les programmes et les politiques, certaines pratiques de collecte de données peuvent présenter des risques de créer un préjudice supplémentaire. Dans de tels cas, les droits à la protection de la personne concernée doivent être considérés comme prioritaires par rapport à toute autre priorité potentiellement concurrente (comme les intérêts bien intentionnés de surveiller les tendances en matière de violence basée sur le genre, ou de collecter des données qui pourraient ensuite contribuer à des modèles algorithmiques pertinents).[[3]](#footnote-3) |
| **Participatif** | Les projets qui intègrent la problématique genre *répondent* efficacement aux normes locales et obtiennent des résultats positifs en matière d’égalité des sexes, précisément en raison de leur nature profondément participative avec les communautés concernées. Ces projets vont au-delà des approches symboliques ou descendantes de la participation (c’est-à-dire qu’ils appellent certains groupes « partenaires » mais ne leur permettent pas d’influencer la prise de décision) et investissent plutôt dans des partenariats équitables (c’est-à-dire que les partenaires peuvent influencer la direction d’un projet et sont traités comme des collaborateurs équitables, plutôt que comme des « sujets » de recherche). |
| **Intersectionnel** | L’intersectionnalité est un terme qui fait référence à 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. Pour obtenir des résultats positifs en matière d’égalité des sexes, il faut procéder à une analyse sexospécifique intersectionnelle du sujet donné et de la population concernée. En d’autres termes, la recherche et l’innovation en matière d’IA tenant compte de la dimension au genre doivent commencer par comprendre *comment* et *quels* facteurs (par exemple, l’âge, la race, le handicap, le revenu, le lieu) se croisent pour accentuer l’oppression parmi les individus et les différentes sous-populations dans le contexte donné et à différents niveaux, y compris les systèmes interpersonnels, communautaires et institutionnels. |
| **Interdisciplinaire** | Pour atteindre bon nombre de ces caractéristiques, il faut réunir des disciplines, des formations et des expériences professionnelles diverses. Les équipes d’IA sont souvent composées de professionnels issus des sciences naturelles (par exemple, des mathématiciens, des ingénieurs, des biologistes). Cependant, les perspectives des sciences sociales (sociologues, économistes, psychologues, etc.) et humaines (histoire, arts visuels, droit, etc.) sont essentielles pour mener des analyses de genre intersectionnelles et pour relier ces analyses à une conception plus holistique et plus efficace de l’IA.[[4]](#footnote-4) Les perspectives interdisciplinaires sont également essentielles pour identifier et atténuer les risques de protection potentiels à l’aide de méthodes de collecte de données qualitatives et quantitatives[[5]](#footnote-5), pour élaborer des théories du changement fondées sur des preuves et pour faire participer les communautés touchées, entre autres éléments essentiels d’une IA responsable et tenant compte de la dimension genre. |
| **Réflexif** | Les projets tenant compte de la dimension genre reconnaissent que l’obtention de résultats positifs en matière d’égalité des sexes exige plus que des solutions technologiques à court terme - il faut plutôt un changement de système plus large.[[6]](#footnote-6) Ainsi, les projets sensibles à la dimension de genre réfléchissent régulièrement à leurs rôles individuels et collectifs dans ces systèmes, et à leurs responsabilités vis-à-vis d’un changement plus large des systèmes organisationnels, culturels et politiques. |
| **Axé sur l’action** | Dès le début de leur travail, les projets qui intègrent la problématique genre sont axés sur l’action et sont clairs quant aux résultats souhaités en matière d’égalité des sexes. Les relations intersectionnelles entre les sexes (c’est-à-dire la compréhension de la manière dont le sexe et les autres systèmes d’oppression façonnent les relations entre et au sein des sous-populations) sont prises en compte et intégrées dans les notes conceptuelles, les propositions, les grilles d’évaluation et les autres documents du projet afin de garantir que les résultats du projet soutiennent des résultats durables et positifs en matière d’égalité des sexes.[[7]](#footnote-7)    Dès le début, les projets qui intègrent la problématique genre impliquent les principales parties prenantes, y compris les détenteurs du pouvoir au sein du système particulier concerné. |

|  |
| --- |
| **Les projets tenant compte de la dimension de genre concernent-ils « uniquement les femmes » ?** |
| Les projets tenant compte de la dimension genre sont clairs quant aux résultats souhaités en matière d’égalité des sexes, dès le début et jusqu’à la fin du cycle de vie de leur projet, et traitent l’’égalité des sexes et l’inclusion (ESI) comme une composante intégrale de *leur* approche aux projets, plutôt que comme une activité individuelle ou un « ajout ». Cela signifie que les projets tenant compte de la dimension s au genre doivent aller au-delà d’une approche[[8]](#footnote-8) du type « ajoutez des femmes et remuez » et analyser plutôt *pourquoi* les femmes, parmi d’autres groupes opprimés (voir le principe d’intersectionnalité ci-dessus), sont systématiquement exclues de la conception et des avantages des applications de l’IA.    Considérez le projet théorique suivant : Une équipe développe de nouveaux outils de prédiction basés sur l’IA qui cherchent à renforcer la résilience des communautés face aux épidémies. Ce projet pourrait être considéré comme intégrant la dimension genre si (1) l’avancement de l’égalité des sexes et l’inclusion est un objectif clair, et (2) s’il investit dans des pratiques, des méthodologies et des activités organisationnelles qui s’alignent sur cet objectif, par exemple :   * La direction de l’équipe est **représentative des communautés qu**’elle souhaite toucher, présente une diversité de perspectives disciplinaires et a créé des partenariats avec des organisations de la société civile locale et des défenseurs de la communauté ; * Ils comprennent que les crises de santé publique ont des **impacts différents** sur les groupes en fonction d’identités multiples et croisées, et s’assurent que les données qui alimentent leurs modèles prédictifs sont inclusives et tiennent compte des identités locales et croisées (par exemple, en incluant les données générées par la **recherche-action participative[[9]](#footnote-9))** * Parce que l’équipe dispose de **pratiques de suivi et de réflexion** solides et bien intégrées, elle est en mesure d’identifier efficacement les nouveaux risques, les changements dans les contextes locaux (y compris les dynamiques qui influencent les systèmes d’oppression ou créent de nouvelles formes de discrimination), ou d’autres apprentissages clés, et sur cette base, d’**itérer, d’adapter et de renforcer** le projet en permanence, si nécessaire ; * Enfin, l’équipe du projet souhaite aller au-delà de la « collecte de données pour le plaisir de collecter des données »,[[10]](#footnote-10) . Elle investit donc dans des activités d’**utilisation des données** et travaille en étroite collaboration avec les principales parties prenantes (y compris les détenteurs du pouvoir) dès le début du projet, afin de créer des espaces de dialogue délibératif, de tirer parti de la recherche pour le plaidoyer et, par conséquent, d’informer des services, des programmes et des politiques de santé publique plus fondés sur les faits et plus inclusifs. |

**Pourquoi les considérations relatives à l’égalité des sexes et l’inclusion (ESI) sont-elles importantes pour l’IA ?**

En plus de garantir que les activités ne causent aucun tort aux communautés touchées, l’intégration efficace des considérations d’ESI peut améliorer la qualité, la nature innovante et l’impact durable des projets. Voici comment :

* 1. **Ne pas nuire** : Compte tenu de l’échelle, de la vitesse et de l’impact uniques des applications de l’IA dans de nombreux aspects de notre vie quotidienne (y compris des impacts souvent inconnus[[11]](#footnote-11) ), les technologies basées sur l’IA sont incontestablement des outils puissants. L’IA a le potentiel d’apporter de grands avantages à la société, mais elle a aussi le pouvoir de causer (intentionnellement ou non) des dommages importants. Par exemple, les professionnels de la santé s’appuient de plus en plus sur l’IA pour diagnostiquer les maladies. Cependant, les données utilisées pour informer ces algorithmes comportent souvent des préjugés sexistes ou raciaux, ce qui entraîne des diagnostics erronés qui touchent de manière disproportionnée les femmes, les personnes transgenres et les personnes à la peau plus foncée - exacerbant ainsi les inégalités en matière de santé et exposant ces groupes à un risque accru de maladies potentiellement mortelles.[[12]](#footnote-12) En d’autres termes, le principe de « ne pas nuire » a été gravement négligé dans cette utilisation particulière de l’IA.

Les pratiques qui intègrent la problématique genre permettent d’identifier et d’atténuer plus efficacement (et de manière responsable) ces préjudices : Les équipes qui impliquent de manière significative les communautés concernées tout au long des phases de conception, de mise en œuvre et de suivi de leur projet sont plus à même d’identifier et d’atténuer les risques ou les préjugés potentiels.[[13]](#footnote-13) Cela est particulièrement vrai lorsque l’équipe du projet *n’*est *pas* entièrement représentative des communautés concernées, et donc moins susceptible de comprendre intuitivement les nombreuses façons dont la recherche proposée ou les applications de l’IA peuvent avoir un impact sur divers groupes et individus. Les partenariats avec des organisations de défense des droits des femmes et d’autres parties prenantes en matière d’égalité des sexes, en particulier, peuvent aider à comprendre comment certains projets de recherche et d’innovation fondés sur l’IA peuvent renforcer les normes de genre et les inégalités discriminatoires, *ou* peut-être même remettre en question ces normes - et quelles peuvent en être les conséquences pour les femmes, les filles, les personnes LGBTIQ+ et les autres groupes marginalisés.

* 1. [**Qualité**](https://www.nature.com/articles/d41586-018-05581-4?WT.ec_id=NATURE-20180705&utm_source=nature_etoc&utm_medium=email&utm_campaign=20180705&spMailingID=56937555&spUserID=ODUwOTgwNjU3MDQS1&spJobID=1440528511&spReportId=MTQ0MDUyODUxMQS2) **:** La recherche inclusive est une recherche de meilleure qualité - que la qualité soit définie du point de vue de la rigueur scientifique, de la légitimité, de la pertinence ou de l’applicabilité, comme l’illustre l’approche [**Research Quality Plus (RQ+)**](https://www.idrc.ca/en/rqplus) du CRDI. La relation entre la qualité et l’inclusivité est particulièrement vraie pour les applications de l’IA, étant donné l’importance fondamentale des données pour façonner la performance et l’impact des modèles de prise de décision algorithmique. L’exclusion disproportionnée (ou la représentation stéréotypée) de groupes marginalisés (tels que les femmes, les filles et les personnes LGBTIQ+) des ensembles de données entraîne des biais statistiques et sociaux qui nuisent à la qualité globale des applications d’IA formées sur ces ensembles de données.[[14]](#footnote-14) Les ensembles de données visuelles,

par exemple, sont souvent biaisés en faveur des visages d’hommes blancs. Par conséquent, lorsque les technologies de reconnaissance faciale sont formées sur ces données, leurs algorithmes sont moins susceptibles de reconnaître les femmes et les visages non blancs. Il en résulte des technologies de moindre qualité ou moins efficaces, tout en contribuant aux micro- et macro-agressions sexistes et raciales.[[15]](#footnote-15) Des approches plus participatives et fondées sur les droits (c’est-à-dire impliquant *toutes les* parties prenantes clés dans le système ou le contexte ciblé) ainsi que des équipes plus inclusives sont essentielles pour réduire ces risques de préjudice et améliorer les possibilités d’une utilisation plus positive, réussie et efficace de l’application d’IA donnée.

* 1. **Innovation :** Les équipes diversifiées, inclusives et plus collaboratives sont aussi des équipes plus innovantes. Ces conclusions sont étayées par un nombre croissant d’études,[[16]](#footnote-16) , qui montrent que des équipes plus *diversifiées* (c’est-à-dire dont la composition présente une diversité d’identités, d’origines et de disciplines - et cette diversité se reflète également dans les organes et les processus décisionnels) sont plus susceptibles d’aborder les défis de conception sous un angle nouveau, de susciter de nouvelles idées, de collaborer et donc d’innover. Par conséquent, pour les organisations et les chercheurs qui cherchent à utiliser les applications de l’IA pour relever les défis les plus pressants de la société, il est très pratique et précieux d’inclure un éventail diversifié de perspectives - en particulier celles qui sont le plus souvent marginalisées du secteur et/ou celles qui sont le plus susceptibles d’être touchées par les applications proposées de l’IA.

* 1. **Impact durable** : Enfin, la recherche tenant compte de la dimension genre est également plus susceptible d’avoir un impact durable, à long terme. Et ce, pour plusieurs raisons : Les partenariats et l’engagement étroit avec les communautés touchées et les organisations qui les représentent, ainsi que l’inclusion d’une expertise plus interdisciplinaire dans les équipes d’IA, sont essentiels pour comprendre en profondeur les contextes locaux et, sur cette base, évaluer et atténuer les risques. Par conséquent, la recherche tenant compte de la dimension de genre est *moins* susceptible d’avoir des impacts négatifs involontaires, qui pourraient (et parfois, devraient) entraîner l’arrêt prématuré d’un projet. [[17]](#footnote-17)

Deuxièmement, la recherche sensible au genre est axée sur l’action, ce qui signifie que l’impact social positif est une motivation centrale de la recherche sensible au genre. Grâce à des théories du changement fondées sur des données probantes et à des partenariats approfondis avec les communautés touchées, les organisations de la société civile et les dirigeants communautaires, les projets sensibles au genre sont plus susceptibles de comprendre les facteurs fondamentaux des défis sociaux et les principaux points d’entrée pour relever ces défis. Par conséquent, les pratiques fondamentales qui définissent la sensibilité à la dimension genre sont également des pratiques essentielles pour développer des projets de recherche et d’innovation axés sur l’IA plus pertinents, plus durables et plus efficaces.

**Annexe A : Ressources recommandées**

* + « Algorithmic Origins of Bias » ([Mandal 2021](https://www.womenatthetable.net/research-advocacy/algorithmic-accountability/algorithmic-origins-of-bias/))
  + Boîte à outils des droits de l’homme - Women at the Table <AI & Equality> ([ici)](https://aiequalitytoolbox.com/)
  + Le projet Gendered Innovations in Science and Engineering de Stanford ([ici)](http://genderedinnovations.stanford.edu/index.html)
  + « We Shape Our Tools, Thereafter Our Tools Shape Us : Artificial Intelligence, Automated Decision-Making & Gender » ([Women at the Table 2019)](https://www.womenatthetable.net/research-advocacy/we-shape-our-tools/).
  + AI4COVID Fiche Technique 1 : Conception de projets de données qui intègre la problématique genre ([Ladysmith 2021)](https://covidsouth.ai/research/download/English-Technical-Brief-1-Designing-gender-responsive-data-projects)
  + AI4COVID Fiche Technique 2 : Un guide pour une recherche en santé qui intègre la problématique genre [(Ladysmith 2021)](https://covidsouth.ai/research/download/English-Technical-Brief-2-A-guide-for-more-gender-responsive-health-research)
  + AI4COVID Fiche Technique 3 : Engagement des parties prenantes pour une recherche en santé qui intègre la problématique genre ([Ladysmith 2021)](https://covidsouth.ai/research/download/English-Technical-Brief-3-Stakeholder-engagement-for-gender-responsive-health-research)
  + AI4COVID Fiche Technique 4 : Relier les données sur le genre à l’action ([Ladysmith 2022)](https://covidsouth.ai/research?lang=en&page=3)
  + « Sept principes féministes intersectionnels pour des données COVID-19 équitables et exploitables » ([D’Ignazio et Klein 2020)](https://covidsouth.ai/research/download/Seven-intersectional-feminist-principles-for-equitable-and-actionable-COVID-19-data)
  + « Relever le défi de l’inégalité entre les sexes par la recherche transformative en matière de genre : leçons tirées de la recherche en Afrique, en Asie et en Amérique latine » ([Njuki et al 2022)](https://doi.org/10.1080/02255189.2022.2099356)
  + Transformer les relations entre les sexes : Insights from IDRC research (CRDI [2019)](https://issuu.com/idrc_crdi/docs/wd_13_000_gender_e-file_en?e=34655515/70235030).

*Pour plus de ressources, nous vous recommandons de visiter* [*ce dossier Google Drive*](https://drive.google.com/drive/folders/1B4O89ZUCTk0k1n4r1giH9WHmuVyVBlUA) *organisé par le CRDI, ainsi que le* [*site web du programme AI4D.*](https://africa.ai4d.ai/research/)  *Le CRDI et l’équipe de soutien au genre d’AI4D continueront à ajouter des ressources accessibles au public sur ces deux pages tout au long du programme AI4D.*

**Annexe B : 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. Pour des ressources et des discussions supplémentaires sur l'IA et les droits de l'homme, consultez la boîte à outils Droits de l'homme <AI & égalité> de Women at the Tabl[e ici.](https://www.womenatthetable.net/project/ai-equality-human-rights-toolbox/)  [↑](#footnote-ref-2)
3. [Zulver, J.M., et al. (2021).](https://doi.org/10.1080/14616742.2021.1894208) "COVID-19 et la violence sexiste : réflexions d'un projet de "données pour le développement" à la frontière entre la Colombie et le Venezuela". *Revue internationale féministe de politique* 23(2) : 341-349. [↑](#footnote-ref-3)
4. Voir Stanford’s Gendered Innovations in Science and Engineering Project’s discussion of intersectional approaches ([here](http://genderedinnovations.stanford.edu/methods/intersect.html)) and their resources to support Intersectional Design methods ([here](https://intersectionaldesign.com/about/)) [↑](#footnote-ref-4)
5. [Fuentes, L. and Cookson, T.P. (2020).](https://doi.org/10.1080/0966369X.2019.1681371) “Counting gender (in)equality? a feminist geographical critique of the ‘gender data revolution’.” *Gender, Place & Culture* 27(6): 881-902 [↑](#footnote-ref-5)
6. Par exemple, voir le cadre analytique de Gender at Work ([ici)](https://genderatwork.org/analytical-framework/). [↑](#footnote-ref-6)
7. Pour en savoir plus sur la façon de relier les données (sur le genre) à l'action, voir l[a fiche technique n°4d](https://covidsouth.ai/research?lang=en&page=3)u programme AI4COVID du CRD[I : Relier les données sur le genre à l'action (](https://covidsouth.ai/research?lang=en&page=3)Ladysmith 2021). [↑](#footnote-ref-7)
8. Voir Cornwall, A., Harrison, E., and Whitehead, A. 2007. *Feminisms in Development: Contradictions, Contestations & Challenges.* London: Zed Books [↑](#footnote-ref-8)
9. [Njuki, J., et al. (2022)](https://doi.org/10.1080/02255189.2022.2099356). "Meeting the challenge of gender inequality through gender transformative research: lessons from research in Africa, Asia, and Latin America." *Canadian Journal of Development Studies* [↑](#footnote-ref-9)
10. [Cookson, T.P & Fuentes, L. (2021).](https://ssir.org/articles/entry/without_actionable_data_gender_equality_will_remain_out_of_reach) “Without Actionable Data, Gender Equality Will Remain Out of Reach.” *Stanford Social Innovation Review.* [↑](#footnote-ref-10)
11. Une fois qu'une technologie basée sur l'IA est ouverte au public, les individus peuvent également trouver des moyens nuisibles de l'utiliser, même si ceux-ci n'étaient pas prévus par ses créateurs originaux. Par exemple, l'IA conversationnelle Tay de Microsoft, lancée en tant que bot sur Twitter en 2016, a rapidement été instrumentalisée par des utilisateurs de l'alt-right pour faire des tweets misogynes et racistes ([Vincent 2016)](https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist). De même, le modèle de langage GPT-3 d'OpenAI a appris à générer des réponses comme s'il était un membre du groupe d'extrême droite QAnon [(McGuffie & Newhouse 2020)](https://arxiv.org/pdf/2009.06807.pdf) [↑](#footnote-ref-11)
12. Voi[r Larrazabal, A. (2020).](https://www.pnas.org/doi/10.1073/pnas.1919012117)  Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis; and [Adewole, S. (2018).](https://jamanetwork.com/journals/jamadermatology/article-abstract/2688587) Machine Learning and Health Care Disparities in Dermatology. [↑](#footnote-ref-12)
13. Comme l'illustre la [Fiche Technique No. 3 d](https://covidsouth.ai/research/download/English-Technical-Brief-3-Stakeholder-engagement-for-gender-responsive-health-research)u programme AI4COVID du CRDI, les intervenants communautaires ont une connaissance unique des lacunes en matière de données à combler en priorité pour maximiser l'impact. En travaillant avec des dirigeants communautaires au Cameroun, par exemple, le Consortium Afrique-Canada pour l'intelligence artificielle et l'innovation des données (ACADIC) a constaté que les données sur les points chauds du COVID-19 n'incluaient pas les communautés ayant un pourcentage plus élevé de personnes déplacées à l'intérieur de leur propre pays, ce qu'ils savaient ne pas être représentatif de la réalité locale. Cela a conduit à des campagnes qui ont augmenté l'accès aux centres de dépistage et ont donc eu un impact sur la politique et la mobilisation des ressources pour ces communautés. [↑](#footnote-ref-13)
14. [Mehrabi, et al. (2019)](https://doi.org/10.48550/arXiv.1908.09635). “A Survey on Bias and Fairness in Machine Learning.” *ACM Computing Surveys* 54(6): 1-35 [↑](#footnote-ref-14)
15. [Dickey, M.R. (2020)](https://techcrunch.com/2020/09/21/twitter-and-zoom-algorithmic-bias-issues/). “Twitter and Zoom’s algorithmic bias issues”. [↑](#footnote-ref-15)
16. [Women at the Table. (2019).](https://www.womenatthetable.net/research-advocacy/we-shape-our-tools/) “We Shape Our Tools, Thereafter Our Tools Shape Us: Artificial Intelligence, Automated Decision-Making & Gender.” [↑](#footnote-ref-16)
17. Comme le robot Twitter Tay de Microsoft (mentionné ci-dessus), qui a fermé 24 heures après son lancement en raison de sa transformation rapide par des tweets discriminatoires ([Vincent 2016)](https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist). [↑](#footnote-ref-17)