

Rapport technique n°1: Conception de projets de données sensibles au genre

Synthèse des principaux cadres et lignes directrices

Rédigé par Alex Berryhill et Lorena Fuentes, Ladysmith.

Ce dossier a été élaboré en partenariat avec [Gender at Work](#) pour le Programme [Global South AI4COVID](#), financé par le Centre de recherches pour le développement international du Canada (CRDI) et l'Agence suédoise de coopération internationale au développement (ASDI).



Historiquement, la collecte de données a exclu et invisibilisé les expériences des femmes et des communautés marginalisées.¹ Par exemple, l'exclusion du travail non rémunéré des femmes des statistiques sur la population active et des comptabilités nationales a rendu invisibles les contributions des femmes et a conduit à la sous-estimation du travail domestique et de soins à autrui dans les politiques publiques.² Le manque de données sur les femmes — en particulier les femmes noires et indigènes — dans les études sur la santé a donné lieu à des modèles de prédiction des risques inadéquats et à un sous-diagnostic fatal des risques de cancer, maladies cardiovasculaires et tuberculose chez celles-ci.³ De même, les données relatives à Zika, Ebola et maintenant la COVID-19, illustrent les impacts différenciés des maladies infectieuses selon l'âge, le sexe, l'origine ethnique et la profession. Pourtant, en septembre 2020, seuls 27 % des cas confirmés de COVID-19 signalés à l'Organisation mondiale de la santé comportaient des données relatives au sexe et à l'âge, tandis que seule une minorité des cas incluait d'autres facteurs de vulnérabilité.⁴ Ces types d'exclusions historiques et continues ont des effets en chaîne, et contribuent à créer et renforcer les inégalités sociales, économiques et politiques.

Les données, dans leurs diverses formes, façonnent les perceptions et les actions. Les données peuvent aider à identifier les domaines dans lesquels une action est nécessaire, éclairer l'élaboration des politiques, services et programmes, et être utilisées pour évaluer l'impact des actions. Des données biaisées et incomplètes se traduisent donc par des politiques, services et programmes qui ne prennent pas en compte — ou ignorent sciemment — les préjudices ou les risques potentiels pour les femmes et les communautés marginalisées.⁵

¹ [Criado-Perez 2019](#); [D'Ignazio & Klein 2020](#); [Fuentes & Cookson 2020](#)

² [Benería, Berik & Floro, 2016](#)

³ [Criado-Perez 2019](#); [Kalinowski et al 2019](#); [Prince et al 2018](#)

⁴ [Heidari et al 2020](#)

⁵ [Criado-Perez 2019](#)

Les projets de données sensibles au genre cherchent à remédier à ces préjugés et lacunes

historiques. Les projets de données sensibles au genre ne se contentent pas de combler les lacunes en la matière (voir l'encadré 1), mais tiennent compte de la dimension du genre dans toutes les étapes du cycle de vie de données de recherche. Les projets de données sexospécifiques sont des projets fondés sur les droits, et accordent donc une priorité à la protection des droits des sujets en matière de données. En s'appuyant sur des méthodes de recherche féministes, les projets de données sexospécifiques mettent en lumière la multiplicité et la superposition des dimensions contextuelles de l'inégalité (notamment le sexe, l'âge, la sexualité, la race, l'origine ethnique, la géographie, les capacités et le revenu) et y répondent. Ce faisant, les projets de données sensibles au genre jettent les bases de politiques, de services et de programmes plus justes, inclusifs et transformateurs, y compris des interventions de santé publique plus adaptées et efficaces.

La pandémie de COVID-19 a mis en évidence l'importance de la collecte de données sensibles au genre pour une réponse et un rétablissement équitables et justes. D'une part, la pandémie a rendu les méthodes traditionnelles de collecte de données plus difficiles, voire impossibles, ce qui a entraîné l'utilisation de méthodes de collecte à distance qui présentent de nouveaux risques pour les femmes et les communautés marginalisées.⁶ D'autre part, la pandémie de COVID-19 — comme les épidémies de maladies infectieuses passées⁷ — a catalysé des initiatives expérimentales de données pour soutenir une approche de la réponse à la pandémie davantage fondée sur des données probantes.

Ce rapport technique⁸ guide les chercheurs et les producteurs de données vers des cadres de bonnes pratiques établis et émergents pour la conception de projets de données sensibles au genre. Plutôt que de se concentrer sur le « quoi », cette note fournit une série de questions d'orientation qui indiquent aux chercheurs et aux producteurs de données **comment rendre opérationnels les principes et les perspectives sensibles au genre, féministes et intersectionnelles**, tout au long du cycle de vie du projet. Cette note cherche également à souligner que ces principes ne sont pas seulement des slogans accrocheurs : ils peuvent être mis en pratique de manière à améliorer les processus et les résultats des projets de recherche, notamment en veillant à ce que les efforts d'exploitation des données et l'IA maximisent le bien commun — et minimisent les dommages.

Encadré 1 : Que sont les lacunes dans les données sur le genre ?

Une théorie dominante du changement dans le domaine du développement international veut que les progrès en matière d'inégalités des genres dépendent de la réduction du « fossé des données sur le genre », qui fait généralement référence au manque de statistiques ventilées par sexe (par

⁶ [Singh et al 2021](#)

⁷ [Kozlakidis et al. 2020](#)

⁸ Alors que le rapport technique n° 1 aborde les principes fondamentaux de la conception de projets de données sensibles au genre, les prochains rapports techniques traiteront en profondeur de la manière de rendre les données **accessibles, responsables et exploitables**.

exemple, sur les taux de mortalité lors de catastrophes climatiques), ou bien aux statistiques reflétant les dynamiques considérées comme spécifiques à la vie des femmes et des filles (par exemple, sur la façon dont une crise sanitaire affecte les taux de violence sexiste ou de mariage des enfants).⁹

Pourtant, l'amalgame entre données sexospécifiques et statistiques réduit le manque de données sexospécifiques à un simple problème de « comptage » qui peut être résolu uniquement par des méthodes de recherche quantitative,¹⁰ négligeant les méthodes qualitatives qui peuvent et doivent être intégrées dans les projets de données – même, et peut-être surtout, lorsque ces projets s'appuient sur des données quantitatives et des données dérivées de technologies nouvelles ou non conventionnelles.

En effet, les projets de données sensibles au genre doivent intégrer des informations provenant de sources de données quantitatives et qualitatives mais également tenir compte des dynamiques de pouvoir sous-jacentes qui façonnent un projet tout au long de son cycle de vie, notamment comment (et par qui) les priorités et les méthodes de collecte de données sont déterminées, et comment (et par qui) les données sont traitées, analysées et utilisées.

Encadré 2: Possibilités pour une IA sensible au genre (ou même, féministe)?

L'analyse prédictive et l'IA ont été présentées comme « l'avenir de la politique publique » et la clé d'une gouvernance plus efficiente, efficace et fondée sur les données.¹¹ Pourtant, dans un monde où les données sont incomplètes et biaisées, et d'un point de vue féministe, l'IA présente un certain nombre de problèmes éthiques. En effet, au cours des dernières années, de nombreuses études ont montré comment des systèmes d'IA biaisés automatisent le racisme et la misogynie, souvent avec une réglementation et une surveillance minimales.¹²

Pourtant, les mouvements féministes ne se contentent pas de critiquer l'utilisation actuelle de ces nouvelles technologies : ils présentent également de nouvelles façons de tirer parti de l'IA pour le changement social. Par exemple, le **Processus de Conception d'un Chatbot Féministe** (*Feminist Chatbot Design Process – FCDP*) de Josie Swords s'appuie sur des méthodes de recherche et des pratiques de réflexion féministes pour développer des chatbots qui contribuent à remettre en question – plutôt qu'à renforcer – les inégalités entre les genres.¹³ La recherche montre que lorsque les chatbots sont représentés par une femme, leur conception est basée sur des stéréotypes de genre, ce qui peut reproduire et renforcer des normes et des perceptions néfastes. Pour contrer ces pratiques, le FCDP de Sword propose une série de questions de réflexion qui encouragent les équipes de conception et de développement à tenir compte du genre, des inégalités et des

⁹ Ce cadrage des lacunes dans les données sur le genre a aussi largement négligé l'importance des données sur les populations de genres divers, ou encore sur les hommes et les garçons.

¹⁰ [Fuentes & Cookson 2020](#).

¹¹ [Thapa, 2019](#)

¹² [D'Ignazio & Klein 2020](#); [Cookson, Zulver, Fuentes & Langworthy 2020](#); [Buolamwini and Gebru, 2018](#)

¹³ [Swords, 2017](#)

préjugés personnels et organisationnels dans toutes les phases du processus de conception. Les questions comprennent :

- La technologie n'est pas isolée des questions politiques, sociales, économiques, culturelles, technologiques, juridiques et environnementales. Avez-vous une bonne compréhension de l'**écosystème** dans lequel votre produit s'inscrira?
- Plutôt que de concevoir un chatbot « utilisable universellement » – c'est-à-dire, un utilisateur unique et universel – pouvez-vous identifier un **utilisateur « marginal »** qui bénéficierait de votre chatbot?
- Quelles sont les différentes **méthodes participatives** dont vous disposez pour que votre utilisateur marginal puisse co-créeer ou avoir un apport direct dans le développement de votre chatbot?
- Dans la conception du chatbot, existe-t-il des **hypothèses** sur la façon dont votre utilisateur s'engagera ou agira envers le chatbot?

Bien que ces questions soient spécifiques à la conception d'un chatbot féministe, elles illustrent comment la réflexion personnelle est une première étape, accessible, pour lutter contre les préjugés néfastes qui sous-tendent souvent les technologies basées sur l'IA.

Conception de projets de données sensibles au genre : Questions à prendre en compte ¹⁴

PLAN

- Quelles communautés et perspectives sont incluses au sein de mon équipe de projet de données? **Lesquelles sont exclues?** Comment la composition de notre équipe pourrait-elle influencer notre collecte de données, notre analyse et notre capacité d'action?
- Comment pouvons-nous établir des partenariats ou soutenir de quelque manière que ce soit et amplifier le travail des **organisations locales de base?**
- Quels risques potentiels la forme de collecte de données que nous proposons pourrait-elle présenter pour les personnes concernées et/ou les **communautés bénéficiaires cibles**¹⁵ et comment pouvons-nous les réduire?
- Avons-nous **consulté** de manière significative les personnes concernées par les données et/ou les communautés bénéficiaires cibles (par exemple, via

¹⁴ Veuillez noter qu'il ne s'agit pas d'une liste exhaustive de questions pour les projets de données. D'autres orientations doivent également être consultées. Par exemple, la plupart des projets de données nécessitent l'autorisation d'un comité d'examen institutionnel (IRB). Nous vous recommandons également de consulter le [Data Responsibility Journey](#) de The GovLab, un outil d'évaluation convivial qui décrit les opportunités et les risques à prendre en compte à chaque étape du cycle de vie des données.

¹⁵ La communauté cible fait référence aux communautés qui seront probablement touchées par votre projet de données, ce qui peut inclure ou non les sujets de données spécifiques qui participent directement au projet.

	<p>des méthodes de recherche-action participatives) concernant la conception du projet, y compris ses impacts prévus et ses potentiels effets involontaires néfastes?</p> <ul style="list-style-type: none"> • Comment le projet de données que nous proposons concilie-t-il le respect des perspectives locales, des lois nationales, ainsi que des normes internationales établies?
COLLECTE	<ul style="list-style-type: none"> • Quelles communautés sont et <i>ne sont pas</i> représentées dans notre processus de collecte de données? • Comment pouvons-nous intégrer d'autres sources de données pour compléter et aider à contextualiser la méthode de collecte de données proposée? • Comment le consentement éclairé des sujets des données est-il assuré? • Quels sont les risques éventuels (par exemple, visibilité, exposition) introduits par la collecte et le partage de données sur les communautés marginalisées? • Comment les risques ou les éventuels effets involontaires de la collecte de données sont-ils surveillés et atténués?
PROCESSUS	<ul style="list-style-type: none"> • Comment avons-nous documenté les hypothèses et les choix qui ont alimenté les processus de nettoyage et de catégorisation des données?
ANALYSE	<ul style="list-style-type: none"> • Avons-nous évalué si notre analyse repose sur des variables de substitution susceptibles d'introduire des biais ou des angles morts? • Avant de partager des données ou des résultats (étape suivante), avons-nous demandé aux principales communautés engagées et aux chercheurs de première ligne de donner leur feedback? Comment notre analyse des données reflète-t-elle leurs préoccupations et leurs expériences vécues?
PARTAGE	<ul style="list-style-type: none"> • Même lorsque les données existent, cela ne garantit pas qu'elles seront largement utilisées. Comment avons-nous géré les possibles écarts ou lacunes en matière de capacités techniques parmi nos potentiels utilisateurs de données? Par exemple, avons-nous envisagé de produire des «fiches de données»¹⁶ et des visualisations conviviales pour nous

¹⁶ Pour répondre aux besoins des utilisateurs de données ayant un faible niveau de connaissances en la matière, les chercheurs et les producteurs de données peuvent joindre aux produits de données (portails, visualisations, articles) des "fiches de données" contenant des informations sur les **origines, limites et utilisations appropriées** des données. Il s'agit également d'une stratégie importante pour combler les difficultés d'accès pour les acteurs de la société civile disposant de moins de ressources. Voir [Cookson, Zulver, Fuentes & Langworthy 2020](#).

assurer que les données soient accessibles à un éventail d'utilisateurs de données?

- Quels plans avons-nous établis pour renforcer la communication et la collaboration entre notre équipe (les **producteurs de données**) et les utilisateurs de données?¹⁷ Existe-t-il des possibilités d'itération? Par exemple, avons-nous envisagé des engagements sous forme de webinaires ou la mise en place d'une enquête auprès des utilisateurs sur notre portail de données afin d'obtenir un retour d'information sur le degré d'exploitabilité des données?

Encadré 3 : Le projet Cosas de Mujeres de Ladysmith

Le projet [Cosas de Mujeres](#)¹⁸ de Ladysmith fournit un autre exemple sur la façon de concevoir concrètement des projets de données qui soient davantage sensibles au genre à travers la co-conception, l'itération réfléchie et les normes féministes en matière de soins.¹⁹ Stimulé par les besoins exprimés par les organisations féminines de base, Cosas de Mujeres tire parti de WhatsApp pour connecter les femmes colombiennes et vénézuéliennes aux services d'autonomisation et de protection dont elles ont besoin, tout en générant des données exploitables sur la violence sexiste. Cosas de Mujeres a d'abord été piloté en janvier 2020 à la frontière entre la Colombie et le Venezuela, dans la ville de Cucuta. Depuis, le projet s'est étendu à deux autres villes colombiennes : Cartagena et Bucaramanga.²⁰

Bien que le projet soit basé sur six principes clés de conception, le présent rapport met en évidence ceux qui peuvent être les plus pertinents pour les nouvelles initiatives de données sensibles au genre :

1. **Les technologies sont des outils, pas des solutions.** Si Cosas de Mujeres a besoin d'une plateforme numérique, ce n'est qu'un des nombreux outils que le projet utilise pour lutter contre la violence sexiste. Par exemple, la grande majorité du projet se concentre sur la socialisation et le renforcement de la confiance dans la plateforme par le biais d'ateliers avec des groupes de femmes, ou sur l'engagement avec les fournisseurs de services afin de relier, de manière plus significative, les résultats des données avec les changements de politique.
2. **Répondre aux besoins immédiats des femmes.** Les femmes qui écrivent à Cosas de Mujeres sont souvent confrontées à des situations d'urgence, potentiellement dangereuses. Pour faire respecter les droits des femmes, les projets sensibles au genre doivent privilégier les besoins et le bien-être des femmes sur la collecte de données et la recherche.

¹⁷ Voir [FSCAP/UN Women 2020, Chapter 3](#).

¹⁸ [Zulver, Cookson & Fuentes, 2021](#)

¹⁹ Pour plus d'informations sur les normes féminines de soins dans les programmes de prévention et de réponse à la VBG, voir [UNFPA 2015](#).

²⁰ Voir [Genderdatakit.org](#) pour plus d'informations sur le projet Cosas de Mujeres.

3. Co-concevoir avec les organisations féminines de base. Dans le monde entier, les organisations féminines de base sont à la tête des mouvements sociaux et politiques responsables de la sensibilisation accrue à la violence sexiste. Reconnaisant, respectant et célébrant cet héritage, Cosas de Mujeres donne la priorité aux investissements et à l'établissement de relations avec les organisations féminines de base. Les organisations féminines de base ont co-conçu la plateforme avec Ladysmith, et continuent d'être activement consultées tout au long de la gestion, de l'itération et de l'adaptation de la plateforme.

Encadré 4 : Ressources clés pour une recherche en santé sensible au genre

Le rapport de l'[Organisation mondiale de la santé](#) intitulé Intégrer l'analyse intersectionnelle des genres dans la recherche sur les maladies infectieuses de la pauvreté : une boîte à outils pour les chercheurs en santé (*Incorporating Intersectional Gender Analysis into Research on Infectious Diseases of Poverty : A toolkit for health researchers*) est une ressource essentielle pour les chercheurs en maladies infectieuses qui s'initient à l'analyse intersectionnelle des genres. En plus des modules éducatifs, des définitions des **indicateurs de programmes sensibles au genre**, et des **outils et méthodes de recherche participative** recommandés, la boîte à outils comprend également des études de cas emblématiques dans diverses régions du monde.

Par exemple, la boîte à outils met en lumière une étude qualitative menée au Malawi qui a examiné comment les normes de genre peuvent influencer la façon dont les hommes vivant dans des communautés urbaines marginalisées cherchent (ou ne cherchent pas) à se faire soigner pour la tuberculose.²¹ Par le biais de discussions de groupe, d'entretiens avec des informateurs clés et d'ateliers participatifs avec des acteurs clés de la santé, l'étude a révélé que «le contrôle était une caractéristique essentielle de la virilité, et que les efforts pour y parvenir conduisent les hommes à négliger leur santé». Une autre étude a cherché à identifier les lacunes dans les connaissances, les attitudes et les pratiques communautaires concernant la trypanosomiase humaine africaine (THA) au Soudan du Sud. Par le biais d'enquêtes, de discussions de groupe et d'entretiens avec des informateurs clés, l'étude a révélé que la participation à des activités agricoles était l'un des obstacles au dépistage actif de la THA dans la communauté, tandis que le sexe et l'éducation étaient les principaux déterminants de la connaissance sur la THA. L'étude a donc recommandé des stratégies de communication plus sensibles au genre et souligné la nécessité de s'appuyer sur l'analyse de l'emploi du temps pour programmer les activités de santé.

Ces deux études – ainsi que la boîte à outils dans son ensemble – illustrent l'importance de la **recherche utilisant des méthodes mixtes** qui intègre une **analyse intersectionnelle de genre** tout au long de la collecte, de l'analyse et du rapport des données, afin d'élaborer des interventions de santé plus sensibles au genre et donc plus efficaces.

Pour plus de directives sur la recherche en santé publique tenant compte du genre, veuillez consulter :

²¹ [Chikovore et al 2014](#)

- [Navigating the Ethics of Big Data in Public Health](#) (The Oxford Handbook of Public Health Ethics 2019)
- [Boîte à outils pour l'analyse de genre dans les systèmes de santé](#) (Jhpiego 2016)
- [Scientific Excellence in Applying Sex- and Gender-Sensitive Methods in Biomedical and Health Research](#) (Nieuwenhoven & Klinge 2010)
- [Violence à l'égard des femmes et des filles : collecte de données pendant la pandémie de COVID-19](#) (ONU Femmes 2020)
- [How to do \(or not to do\)...gender analysis in health systems research](#) (Morgan & SSali 2016)

Conception de projets de données sensibles au genre : Cadres clés

Principes du féminisme des données²²

1. **Examiner le pouvoir** : Analyser comment le pouvoir fonctionne dans le monde.
2. **Défier le pouvoir** : S'engager à contester les structures de pouvoir inégales et à œuvrer pour la justice.
3. **Valoriser l'émotion et l'incarnation** : Valoriser les multiples formes de savoir, y compris le savoir qui provient des personnes en tant que corps vivants et sensibles dans le monde.
4. **Repenser la binarité et les hiérarchies** : Remettre en question la binarité du genre, ainsi que les autres systèmes de comptage et de classification qui perpétuent l'oppression.
5. **Adopter le pluralisme** : La connaissance la plus complète provient de la synthèse de perspectives multiples, en donnant la priorité aux modes de connaissance locaux, indigènes et expérimentaux.
6. **Tenir compte du contexte** : Les données ne sont pas neutres ou objectives. Elles sont le produit de relations sociales inégales, et ce contexte est essentiel pour mener une analyse précise et éthique.
7. **Rendre le travail visible** : Le travail de la science des données, comme tout travail dans le monde, est un travail collaboratif. Le féminisme des données rend ce travail visible afin qu'il puisse être reconnu et valorisé.

Pour plus de directives sur la collecte et l'utilisation des données dans une perspective féministe, veuillez consulter :

- [Feminist Data Manifest-No](#) (Cifor et al 2019)
- [Feminist Internet Ethical Research Practices](#) (Association for Progressive Communication 2019)

²² Ces principes proviennent de [D'Ignazio & Klein 2020](#).

	<ul style="list-style-type: none"> • Feminist Data Visualization (D'Ignazio & Klein 2016) • Design Justice Network Principles (2018) • Data collection and COVID-19: What's gender got to do with it? (Undie et al 2020)
<p>Comment concevoir l'IA pour le bien commun²³</p>	<ol style="list-style-type: none"> 1. Falsifiabilité et déploiement incrémental : Nous ne pouvons pas avoir la certitude qu'une application donnée de l'IA pour le bien commun (AI4SG) est sûre si nous ne pouvons pas la tester dans tous les contextes possibles. Par conséquent, les concepteurs d'AI4SG doivent identifier les exigences de réfutabilité et les tester par étapes progressives, du laboratoire au « monde extérieur ». 2. Mesures de protection contre la manipulation des prédicteurs : Adopter des sauvegardes qui (i) garantissent que les indicateurs non causaux ne biaisent pas de manière inappropriée les interventions, et (ii) limitent, le cas échéant, la connaissance de la manière dont les entrées affectent les sorties des systèmes AI4SG, afin d'éviter toute manipulation. 3. Intervention contextualisée du récepteur : Les concepteurs d'AI4SG doivent construire des systèmes de prise de décision en consultation avec les utilisateurs qui interagissent avec ces systèmes et en subissent l'impact ; en comprenant les caractéristiques des utilisateurs, les méthodes de coordination, les objectifs et les effets d'une intervention ; et en respectant le droit des utilisateurs d'ignorer ou de modifier les interventions. 4. Explication contextualisée pour le récepteur et objectifs transparents : Les concepteurs d'AI4SG devraient choisir un niveau d'abstraction pour l'explication de l'IA qui réponde à l'objectif explicatif souhaité et qui soit approprié au système et aux récepteurs ; puis, déployer des arguments rationnels et suffisamment persuasifs pour les récepteurs afin de fournir l'explication ; et s'assurer que l'objectif (le but du système) pour lequel un système AI4SG est développé et déployé soit connu des récepteurs de ses résultats, par défaut. 5. Protection de la vie privée et consentement des personnes concernées : Les concepteurs AI4SG doivent respecter le seuil de consentement établi pour le traitement des fichiers de données personnelles.

²³ Ces principes proviennent de [Floridi, Cowls, King & Taddeo \(2020\)](#).

- 6. Équité situationnelle :** Les concepteurs AI4SG devraient supprimer des ensembles de données où les variables ou variables de substitution ne sont pas pertinentes pour un résultat, sauf si leur inclusion promeut l'inclusivité, la sécurité ou d'autres impératifs éthiques.
- 7. Sémantisation conviviale :** Les concepteurs d'AI4SG ne doivent pas entraver la capacité des gens à sémantiser (c'est-à-dire, à donner un sens à quelque chose).

Pour plus de directives sur la conception, la gestion et l'application des systèmes d'IA, veuillez consulter :

- [A People's Guide to Artificial Intelligence](#) (Onuoha & Nucera 2018)
- [AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations](#) (Floridi et al 2018)
- [Principes de l'OCDE sur l'IA](#)
- [Les réponses de l'OCDE face au coronavirus \(COVID-19\) : Utiliser l'intelligence artificielle au service de la lutte contre le COVID-19](#) (2020)