

Vers une réelle égalité en intelligence artificielle :

une politique d'IA transformatrice pour
l'égalité des genres et la diversité

Novembre 2024



GPAI

THE GLOBAL PARTNERSHIP
ON ARTIFICIAL INTELLIGENCE

Ce rapport a été élaboré avant l'intégration du Partenariat mondial sur l'intelligence artificielle (PMAI) et de l'Organisation de coopération et de développement économiques (OCDE) à la mi-2024. Par conséquent, il n'a pas été soumis à l'approbation des membres du PMAI et de l'OCDE et ne doit pas être considéré comme reflétant leurs positions.

Remerciements

Ce rapport a été élaboré dans le cadre du projet « *Vers une réelle diversité et une égalité des sexes dans l'IA : pratiques prometteuses et recommandations fondées sur des données probantes* », sous la houlette de la codirection du projet et avec les conseils du groupe consultatif, soutenu par le groupe de travail du PMIA sur l'IA responsable. Le groupe de travail du PMIA sur l'IA responsable a convenu de lever la confidentialité de ce rapport et de le diffuser au public.

Codirection :

Paola Ricaurte Quijano^{*},

Tecnológico de Monterrey et Centre Berkman Klein pour l'internet et la société, Université de Harvard

Benjamin Prud'homme[†],

Mila - Institut québécois d'intelligence artificielle

Le rapport a été préparé par **Isadora Hellegren Létourneau**[‡], Mila, avec la contribution de la codirection du projet **Paola Ricaurte Quijano**^{*}, Tecnológico de Monterrey et Centre Berkman Klein pour l'internet et la société, Université de Harvard, et **Benjamin Prud'homme**[†], Mila, et des consultants du projet **Anita Gurumurthy**[‡], **Nandini Chami**[‡], Malavika Rajkumar et **Merrin Muhammed Ashraf**[‡], IT for Change, et **Wanda Muñoz**[‡], Feminist AI Research Network.

Nous tenons à remercier les membres du groupe consultatif pour leur soutien actif tout au long du projet : **Lucia Velasco**^{*}, École de gouvernance européenne et transnationale, Institut universitaire européen ; **Ricardo Baeza-Yates**^{*}, Institut d'IA expérientielle de l'Université Northeastern ; **Hawa Bi Khan**[†], consultante en soins de santé (santé numérique, IA et durabilité) ; **Karen de Brouwer Vásquez**[†], secrétaire aux relations extérieures du gouvernement mexicain ; **Caroline M. Coward**[†], Jet Propulsion Laboratory/Caltech ; **Kudakwashe Dandajena**^{*}, Institut africain des sciences mathématiques (AIMS) et Université du Cap-Occidental ; **Gillian Dowie**[†], IDRC ; **Laurent Elder**[†], IDRC ; **Golnoosh Farnadi**[†], Université McGill et Mila ; **Dafna Feinzholtz**[†], UNESCO ; **Jessica Fjeld**[†], De|Center, et Centre Berkman Klein pour l'internet et la société, Université de Harvard ; **Karine Gentelet**[†], Université du Québec en Outaouais (UQO) et Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) ; **Alison Gillwald**[‡], Research ICT Africa ; **Gloria Guerrero**[‡], ILDA ; **Toshiya Jitsuzumi**^{*}, Université de Chuo ; **Nicole Kaniki**[†], Senomi Solutions Inc ; **Ching-Yi Liu**^{*}, Université nationale de Taiwan ; **Nicole Osayande**[†], Mila ; **Michael O'Sullivan**^{*}, Université d'Auckland ; **Inese Podgaiska**^{*}, Association of Nordic Engineers ; **Catherine Régis**^{*}, Université de Montréal et Mila ; **Michael Running Wolf**[†], Indigenous AI et Mila ; **Juliana Sakai**^{*}, Transparência Brasil ; **Patricia Shaw**[†], Beyond Reach Consulting Limited ; **Prateek Sibal**[†], UNESCO ; **Andrew Spoprlé**[†], iNZight Analytics Ltd ; **Anupama Srikonda**[†], Indian School of Business ; **Ruhiya Kristine Steward**[†], IDRC ; **Elissa Strome**[†], Institut canadien de recherches avancées (CIFAR) ; **Jaco du Toit**[†], UNESCO ; **Eliane Ubalijoro**[†], CIFOR-ICRAF ; **Jamila Venturi**[†], Derechos Digitales ; et **Jeff Ward**^{*}, Animikii Indigenous Technology.

Il nous tient également à cœur d'exprimer notre reconnaissance la plus sincère envers les membres du groupe d'experts-conseils pour leurs précieux points de vue, ainsi que pour leur expertise et leur expérience lors des étapes préparatoires du rapport. Nos remerciements à **Anaelia Altagracia Ovalle**[‡], Université de Californie à Los Angeles ; **Jake Okechukwu Effoduh**[‡], École de droit de Lincoln Alexander, Université métropolitaine de Toronto ; **Maui Hudson (Whakatōhea)**[‡], Te Kotahi Research Institute, Université de Waikato ; **Petra Molnar**[‡], Centre Berkman Klein pour l'internet et la société, Université de Harvard, laboratoire de droit des réfugiés, Université de York ; **Rosario Patricia Galarza Meza**[‡], Alliance internationale du handicap (IDA) ; et **Umut Pajaro Velasquez**[‡], consultant et chercheur en en IA indépendante et gouvernance de l'internet. Nous remercions également **Carolina Botero Cabrera**[‡], Karisma Foundation, et **Mathieu Marcotte**[‡], CEIMIA, d'avoir partagé leur expertise.

Nous tenons à remercier **Ana Gabriela Ayala Núñez**[‡], Tecnológico de Monterrey ; **Ivanna Martínez Polo**[‡], Tecnológico de Monterrey ; **Leslie Evelin Salgado Arzuaga**[‡], Université de Calgary ; **Shazade Jameson**[‡], Université de Tilburg ; **Shamira Ahmed**[‡], Data Economy Policy Hub, Université de Tilburg ; **Florian Lebre**[‡], Université Laval ; **Gargi Sharma**[‡], CLIMA Fund ; **Selene Yang**[‡], Feminist AI Research Network et Geochicas ; et **Razieh Shirzadkhani**[‡], Mila, pour leur contribution essentielle de conseil en recherche au fil des différentes étapes du projet.

Nous témoignons toute notre reconnaissance à nos organisations partenaires **Derechos Digitales**, **Research ICT Africa** et **Data Pop Alliance** pour les efforts déployés dans l'organisation et la réalisation de consultations auprès d'une multitude de parties prenantes, géographiquement dispersées et culturellement diverses. Nous souhaitons insister sur notre gratitude envers toutes les personnes ayant pris part aux consultations régionales ou de groupes spécifiques, aussi bien à titre anonyme que non-anonyme (voir l'annexe).

Nous remercions de tout cœur les organisations de la société civile ayant contribué à faire entendre les voix des communautés marginalisées. Tous nos remerciements à **Otto Mazariegos**[‡], Association nationale des personnes aveugles du Guatemala et Réseau latino-américain des organisations de personnes handicapées et leurs familles (RIADIS) ; **Abia Akram**[‡], STEPS Pakistan ; **Andrea Gunraj**[‡], Fondation canadienne des femmes ; **Dragana Kaurin**[‡], Centre Berkman Klein pour l'internet et la société, Université de Harvard ; **Pierrine Leukes**[‡], Research ICT Africa ; et **Nancy Gros-Louis McHugh**[‡], Commission de la santé et des services sociaux des Premières Nations du Québec et du Labrador.

Le projet n'aurait pas pu aboutir à de tels résultats sans les efforts des soumissionnaires de la proposition initiale : **Wanda Muñoz**[‡], Feminist AI Research Network ; **Paola Ricaurte Quijano**[‡], Tecnológico de Monterrey et Centre Berkman Klein pour l'internet et la société, Université de Harvard ; **Catherine Régis**[‡], Université de Montréal et Mila ; **Inese Podgaiska**[‡], Association of Nordic Engineers ; **Juliana Sakai**[‡], Transparência Brasil ; **Toshiya Jitsuzumi**[‡], université de Chuo ; **Ricardo Baeza-Yates**[‡], Institut d'IA expérientielle de l'Université Northeastern ; **Celine Caira**[‡] et **Luis Aranda**[‡], OCDE ; **Prateek Sibal**[‡] et **Vanessa Dreier**[‡], UNESCO ; et **Benjamin Prud'homme**[‡], Mila.

Nous tenons à remercier **Ekua Quansah**[‡], Quansah Consulting pour son travail de vérification du langage inclusif, **Zofia Laubitz**[‡] pour son travail de relecture, **Leonardo Studio Design**[‡] pour le graphisme du rapport, et **Julie Descantes**[‡] pour la traduction de ce rapport en français.

Le PMIA désire saluer les efforts considérables déployés par ses collègues de Mila, du Centre d'expertise international de Montréal en Intelligence artificielle (CEIMIA) et du groupe de travail sur l'intelligence artificielle responsable du PMIA, Parties recrutées par le Centre de soutien aux expert·e·s. Nous adressons notre reconnaissance à **Niobe Haitas**, Mila, gestionnaire de projet senior pour son travail lors de la phase initiale du projet. Nous remercions profondément **Laëtitia Vu**, **Camille Seguin** et **Stephanie King** du CEIMIA, et **Anna Jahn**[†] de Mila pour leur soutien inconditionnel tout au long du projet. Enfin, nous exprimons toute notre gratitude à la co-direction du groupe de travail **Amir Banifatemi**^{*}, AI Commons, et **Francesca Rossi**^{*}, IBM Research, ainsi qu'à la codirection précédente **Catherine Régis**^{*}, Université de Montréal et Mila, et **Raja Chatila**^{*}, Université de la Sorbonne.

* Expertise

** Observation

† Spécialiste invité·e

‡ Parties recrutées par le Centre de soutien expert pour contribuer aux projets.

Ce travail est distribué selon les termes de la licence Creative Commons Attribution 4.0, permettant la libre utilisation à condition de mentionner la personne qui l'a rédigé et la source. La licence est disponible sur : <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>.

Comment citer ce document

PMIA 2024. Vers une réelle égalité en intelligence artificielle : une politique d'IA transformatrice pour l'égalité des genres et la diversité, rapport, novembre 2024, Partenariat mondial pour l'intelligence artificielle.

Project Supported by



Mila

ceimia

Table des matières

Avant-propos	7
Synthèse	8
Résumé des recommandations clés	10
Introduction	13
PARTIE I. Consultations, défis majeurs et voies potentielles	18
Consultations régionales et de groupes spécifiques	19
Résultats des consultations régionales et de groupes spécifiques	21
Défis majeurs en lien avec l'égalité des genres et la diversité en IA	21
L'accès n'est pas synonyme d'inclusion	22
Invisibilisation et exclusion du savoir	23
Répartition inégale des ressources	25
Voies pour relever les défis en matière d'égalité des genres et de diversité au sein des écosystèmes de l'IA	26
Développer la capacité du secteur public et du grand public	27
Encourager la conception inclusive et l'innovation démocratique	28
Responsabiliser les équipes de développement, les fournisseurs et les opérateurs ...	29
Garantir la participation effective des groupes marginalisés	30
Un bref résumé des défis à relever et des voies à suivre	31
PARTIE II. Cadre conceptuel	32
Cadre conceptuel	32
Une approche transformatrice pour parvenir à l'égalité réelle	33
Une approche sociotechnique de l'IA	36
Le cycle de vie de l'IA	37
Les étapes du cycle de vie des systèmes d'IA	37
Opportunités d'intervention	38
Le cadre pour la politique d'IA transformatrice	40
Dimensions clés	40

PARTIE III. Pratiques prometteuses et recommandations	43
Pratiques prometteuses pour instaurer une égalité réelle en l'IA	44
Ressources pour le développement des capacités et l'éducation du public	45
Dialogue mondial de l'UNESCO	45
Data Justice Policy Brief	46
AI & Equality Human Rights Toolbox	47
Éclaireurs autochtones en IA	48
Conception technologique inclusive et pratiques d'innovation démocratiques	49
La méthodologie Design From the Margins (DFM)	49
Le réseau de recherche féministe sur l'IA (f&A+i>r)	50
L'initiative Fixing the bAIs	51
Indigenous Jobs Map (Carte de l'emploi autochtone)	52
Mesures de responsabilisation	53
L'indice mondial sur l'IA responsable	53
The Migration and Technology Monitor	55
The Algorithmic Justice League	56
Initiatives d'inclusion et de participation réelles	57
Modèle maori de gouvernance des données	57
L'atelier du RIADIS	58
The Gender and Responsible Artificial Intelligence Network (GRAIN)	59
Recommandations clés sur la politique d'IA transformatrice	60
Conception inclusive et innovation démocratique	61
Participation effective à la gouvernance de l'IA	64
Transparence et responsabilisation dans la prévention des préjudices	66
Accès efficace à la justice	68
Notes de fin	71
Références	73
Annexe : liste des personnes participant·e·s aux consultations régionales	80

Avant-propos

L'IA offre des opportunités exceptionnelles de contribuer au bien commun, mais ce vaste éventail de possibilités peut et doit être exploité de manière équitable. Dans le cadre du présent rapport, nous cherchons à tracer la voie vers une égalité réelle dans les écosystèmes d'intelligence artificielle, en réfléchissant aux mesures à prendre pour assurer la participation de toutes les voix. Nous avons à cœur de transformer les modèles de participation à l'IA (qui peut participer, où et comment), en démocratisant son développement, son déploiement et sa gouvernance.

L'évidence est claire : les systèmes d'IA ne sont pas neutres. Ils reproduisent les modèles du monde, les valeurs culturelles, le savoir et les langages des contextes dans lesquels ils sont conçus, en copiant ou en amplifiant les inégalités systémiques fondées sur le genre, la race, l'origine ethnique, les capacités, la classe sociale et l'éducation, entre autres. Les politiques doivent donc empêcher les systèmes d'IA d'affecter différemment les groupes et de creuser les écarts entre les pays et à l'intérieur même de ces derniers. Néanmoins, il n'existe pas de vision unique qui permettrait de comprendre la reproduction de ces inégalités, ni d'approche unique pour combattre leurs causes et apporter des solutions tout au long du cycle de vie de l'IA. Des efforts doivent être déployés à tous les niveaux et dans tous les domaines, pour articuler des réponses aboutissant à des actions transformatrices.

Dans ce rapport, nous identifions les causes fondamentales des inégalités à travers le monde et nous faisons le point sur les dommages actuellement provoqués par l'IA dans la vie de millions de personnes. Nous adoptons une approche systémique, participative et sociotechnique axée sur les droits de la personne et la justice sociale, visant à instaurer une transformation radicale. Les inégalités au sein des écosystèmes de l'IA se traduisent par des données ou des modèles d'algorithmes biaisés, mais elles ne sont pas limitées à cela. C'est pourquoi nous nuancions et soulignons l'importance de la véritable inclusion dans le développement des systèmes d'IA et les prises de décision la concernant. Nous reconnaissons également que les bénéfices du développement technologique, de la richesse, du pouvoir, du savoir et des infrastructures sont actuellement concentrés entre les mains de quelques personnes, tandis que la vaste majorité de la population se contente de subir les désagréments qui y sont associés : exploitation des ressources sur son territoire, emploi précaire, extraction massive des données.

Et surtout, nous nous efforçons d'identifier des manières concrètes de nous projeter différemment dans l'avenir. Ainsi, en plus de présenter les fructueuses initiatives menées dans différents milieux, ce rapport constitue une feuille de route pour les responsables de la prise de décision, avec des recommandations et des exemples de pratiques prometteuses destinées à servir de référence pour instaurer un changement transformateur. En outre, le rapport propose des orientations pour aligner la conception, le développement et le déploiement de l'IA sur les principes de l'intérêt commun et du bien-être des populations locales et mondiales.

En 2023, le groupe de travail du PMIA sur l'IA responsable nous a fait l'honneur de nous confier la codirection du projet *Vers une réelle égalité en intelligence artificielle : une politique d'IA transformatrice pour l'égalité des genres et la diversité*. À travers le présent rapport et les autres activités du projet, nous espérons contribuer à créer des écosystèmes de l'IA véritablement inclusifs, équitables et justes.

Paola Ricaurte Quijano,

Professeure titulaire, Tecnológico de Monterrey et professeure agrégée au Centre Berkman Klein pour l'internet et la société, Université de Harvard

Benjamin Prud'homme,

Vice-président, Politiques publiques, sécurité et affaires mondiales, équipe de direction, Mila

Synthèse

Les progrès rapides de l'intelligence artificielle (IA) transforment les sociétés et stimulent le développement économique. Cette technologie renferme un potentiel incroyable d'amélioration des vies et des modes de subsistance à travers le monde. Toutefois, elle court le risque d'exacerber les inégalités existantes en reflétant et en amplifiant les préjugés présents dans la société, et tout particulièrement ceux affectant les groupes historiquement marginalisés. Des défis comme la discrimination, l'injustice, les préjugés et les stéréotypes nuisibles se posent tout au long du cycle de vie de l'IA, affectant de nombreux aspects de la vie des gens. Il est urgent de mettre en place des cadres réglementaires solides pour atténuer ces disparités, prévenir les préjudices et œuvrer en faveur d'une réelle égalité et d'une réelle diversité dans les écosystèmes d'IA.

Le rapport *Vers une réelle égalité en intelligence artificielle : une politique d'IA transformatrice pour l'égalité des genres et la diversité* a pour objectif de renforcer la capacité des États et autres parties prenantes à favoriser des écosystèmes d'IA inclusifs, équitables et justes. Il examine les pratiques prometteuses, offre des perspectives stratégiques et des recommandations exploitables pour améliorer l'égalité des genres et la diversité dans l'IA et dans l'élaboration des politiques associées. Le *Guide d'application des recommandations pour une politique d'IA transformatrice* fournit davantage de conseils clés sur la mise en œuvre.

Recommandations clés :

1. **Recommandations clés** : intégrer l'action positive et des mesures d'inclusion institutionnelle, et encourager la conception d'une technologie inclusive.
2. **Participation effective à la gouvernance de l'IA** : favoriser et garantir l'implication active des groupes marginalisés dans la gouvernance de l'IA, afin d'assurer la mise en place pour tout le monde d'une meilleure politique en matière d'IA.
3. **Transparence et responsabilisation dans la prévention des préjudices** : établir au préalable des protections et des mécanismes de responsabilisation auprès de toutes les parties prenantes de l'IA afin d'éviter les préjudices et d'assurer l'équité.
4. **Accès efficace à la justice** : mettre en place des mesures pour garantir l'accès à la justice des groupes marginalisés en cas de discrimination et de préjugés véhiculés par l'IA.

Les perspectives et les recommandations s'appuient sur des consultations régionales et de groupes spécifiques, et sur un cadre conceptuel correspondant à une approche de l'IA fondée sur les droits de la personne. Par le biais de cette consultation, le rapport explore l'état actuel de l'égalité entre les genres et de la diversité dans le secteur de l'IA, identifie les disparités persistantes et ébauche des pistes et des pratiques prometteuses pour parvenir à une égalité réelle dans l'IA. Le cadre part du principe que les inégalités sont structurelles, et qu'il est possible de les combattre et de les corriger en prenant des mesures transformatrices adaptées.

La politique d'IA transformatrice traite à la racine les causes de l'inégalité, en vue de parvenir à une égalité réelle dans l'IA et au-delà. L'instauration d'une égalité réelle grâce à une transformation radicale de l'IA permettra de faire progresser les droits de la personne et de stimuler le développement économique et social. Grâce aux politiques de transformation de l'IA, il est possible d'améliorer la qualité, l'utilisabilité et l'efficacité des systèmes d'IA, dans l'optique d'un avenir plus équitable, durable et prospère dans toutes les communautés.



Femmes africaines discutant de l'une des commandes d'un ordinateur. Photo : Iwaria Inc. sur Unsplash.

Résumé des recommandations clés sur la politique d'IA transformatrice

Ce résumé fournit un bref aperçu des mesures concrètes que les responsables des politiques peuvent prendre pour intégrer efficacement les principes d'égalité entre les genres et de diversité dans les cadres politiques, les lois, les réglementations et les pratiques en matière d'IA.. Les recommandations sur la politique d'IA transformatrice sont regroupées en quatre catégories, tel qu'indiqué ci-dessous :

- Conception inclusive et innovation démocratique
- Participation effective à la gouvernance de l'IA
- Transparence et responsabilisation dans la prévention des dommages
- Accès efficace à la justice

Conception inclusive et innovation démocratique

1. Donner un rôle technique et non technique aux groupes marginalisés au sein de l'écosystème de l'IA

Appliquer l'action positive au sein de l'écosystème de l'IA afin d'impliquer les femmes et d'autres groupes historiquement marginalisés dans des rôles techniques et non techniques, dans le but de favoriser la diversité des points de vue. Allouer des ressources destinées à identifier et à supprimer les obstacles à la représentation de la diversité. Cela implique notamment d'assurer une éducation accessible et inclusive au-delà des écosystèmes de l'IA.

2. Investir dans le renforcement des capacités en vue de l'inclusion institutionnelle

Au sein des institutions et des équipes publiques et privées, investir dans le renforcement des capacités et dans la sensibilisation aux expériences et aux droits des groupes historiquement marginalisés. Maintenir un dialogue régulier avec les regroupements et les personnes qui représentent les groupes marginalisés, en vue de la compréhension et de l'élimination des obstacles spécifiques auxquels ils sont confrontés.

3. Permettre le traitement de catégories spéciales de données

Permettre le traitement de catégories spéciales de données dans certaines circonstances exceptionnelles, en cas de motif d'intérêt public important, afin de parvenir à l'égalité et à la non-discrimination. Pour éviter les résultats discriminatoires, les fournisseurs de systèmes d'IA doivent réaliser des essais de détection des biais systémiques et veiller à la représentation de divers ensembles de données. Cette exigence doit être remplie sans enfreindre le droit à la protection des données à caractère personnel.

4. Financer la recherche dans le domaine de la technologie transformatrice et les approches conceptuelles en innovation dans le domaine de l'IA

Financer la recherche, accorder des subventions et témoigner de la reconnaissance publique pour encourager les approches technologiques et conceptuelles transformatrices dans le domaine de l'IA, par exemple celles reposant sur des principes féministes. Ces approches cherchent à pallier les écarts entre l'équité technique et politique. Le soutien aux innovations conformes à ces principes dans les systèmes d'IA permet de favoriser des applications, pratiques et processus plus équitables et justes.

Participation effective à la gouvernance de l'IA

5. Promouvoir une mobilisation efficace du public et la participation sociale

Utiliser différentes méthodologies de mobilisation du public à l'échelle nationale et internationale. Inclure des voix marginalisées dans les discussions nationales autour de la gouvernance de l'IA et élargir la présence de la majorité mondiale dans les forums internationaux sur la gouvernance de l'IA. Permettre la participation de groupes marginalisés en destinant des budgets à leurs frais de participation, et en veillant à ce que les processus d'information et de consultation soient accessibles, gratuits et compréhensibles.

6. Investir dans le développement des capacités des groupes marginalisés

Financer et soutenir des programmes éducatifs, des structures de réseautage et d'autres ressources visant à développer les compétences et la confiance des groupes marginalisés, pour qu'ils participent réellement aux processus au service de leurs besoins ou qu'ils les dirigent activement. Travailler avec les communautés marginalisées et les organisations de représentation des groupes marginalisés afin de les aider à tenir leurs propres réunions de sensibilisation et de consultation sur les enjeux liés à l'IA.

7. Légiférer sur les droits à la participation publique *ex ante*

Fonder les processus de prise de décision en matière d'IA sur les droits à la participation du public *ex ante*, tels que ceux prévus dans la Convention d'Aarhus de la CEE-ONU. L'application de ces principes aux processus de prise de décision de l'AI permet aux parties concernées, ainsi qu'aux organisations de la société civile et au public, de contester les conséquences de la prise de décision algorithmique, par le biais de la discussion et de la réflexion publique.

8. Protéger les droits collectifs en matière de données et d'IA

Revoir les cadres relatifs aux droits affectés par les systèmes et les processus de l'IA, comme ceux concernant les données ou la propriété intellectuelle afin de 1) protéger la souveraineté des données et du savoir des peuples autochtones et des groupes marginalisés, et notamment des minorités linguistiques, religieuses et ethniques ; et 2) garantir le droit de bénéficier des progrès scientifiques.



Une intervenante portant un foulard et vêtue d'un tailleur s'exprime au pupitre d'une conférence politique

Transparence et responsabilisation dans la prévention des préjudices

9. Établir le droit à l'information dans les systèmes d'IA et améliorer la transparence des algorithmes

Établir le droit à l'information dans l'IA. Les personnes devraient avoir le droit d'être informées clairement des situations faisant appel à l'IA, des algorithmes utilisés, des variables d'entrée et des critères inclus dans les processus de prise de décision. L'amélioration de la transparence des algorithmes permettrait aux gens subissant les conséquences négatives des systèmes d'IA d'en contester les résultats. Cela encouragerait également l'innovation technologique, pour faire face à des limitations telles que l'opacité des comportements, et pour améliorer l'interprétabilité et l'explicabilité.

10. Permettre et mener des études d'impact obligatoires sur les droits de la personne (EIDH)

Permettre et mener des études d'impact en offrant des orientations générales sur la manière de les réaliser. Ces études devraient déterminer si les risques de préjudices sont acceptables au vu des lois garantissant les droits fondamentaux et indiquer clairement les mesures à prendre pour éliminer ou prévenir ces risques. Elles devraient par ailleurs examiner et comparer d'éventuelles approches non technologiques afin d'identifier les mesures les moins intrusives du point de vue des droits de la personne.

11. Établir des mesures de responsabilisation pour les systèmes et processus algorithmiques du secteur public

Établir des lignes directrices spécifiques à l'IA pour la passation de marchés, afin de protéger les droits de la personne et garantir le respect de la légalité, en abordant les complexités et les risques introduits par les systèmes et processus algorithmiques et d'IA. Promouvoir des initiatives de libre accès aux données afin de développer des bibliothèques ouvertes d'algorithmes à utiliser dans les systèmes publics. Veiller à ce que les responsables des politiques aient bénéficié d'un renforcement de leurs capacités, de sorte à pouvoir exercer efficacement leur devoir de vigilance pour la passation de marchés dans le domaine de l'IA.

Accès efficace à la justice

12. Renforcer la responsabilité contextuelle en vue de la non-discrimination dans les systèmes d'IA

Renforcer la responsabilité contextuelle en vue de la non-discrimination dans les systèmes d'IA par rapport à d'autres mesures de responsabilisation comme le niveau de transparence, l'interprétabilité et l'explicabilité. Les réglementations sur les produits et sur la responsabilité pour faute doivent être modifiées de sorte à refléter les complexités des systèmes d'IA et de la prise de décision fondée sur les données. La responsabilisation efficace dans le développement et le déploiement de l'IA passe par la prise en compte de certaines caractéristiques comme l'opacité, l'explicabilité, l'autonomie du comportement, l'adaptation continue et la prévisibilité limitée. Il s'agit de tracer la voie vers la responsabilité dans l'IA, afin d'assurer une responsabilisation appropriée des fournisseurs et des déployeurs, publics et privés.

13. Encourager les organismes de promotion de l'égalité à passer à l'action

Encourager les organismes de promotion de l'égalité, et notamment les institutions nationales de défense des droits de la personne et autres organisations au service de l'intérêt public, à se mobiliser pour le bien commun. Permettre à toutes ces organisations de présenter des plaintes auprès des autorités de contrôle, même en tout anonymat. Alléger la charge de la preuve et doter ces organisations de l'autorité légale et de la formation nécessaires à combattre efficacement les discriminations et les dommages causés par les systèmes d'IA et autres processus associés.

14. Alléger la charge de la preuve pour les organismes et les personnes qui portent plainte

Réexaminer et réviser les règles en matière de preuve afin d'alléger la charge pour les organismes et les personnes (Commission mondiale d'éthique des connaissances scientifiques et des technologies, 2005). À l'heure actuelle, les règles de responsabilité associées aux produits demandent souvent aux parties ayant subi un préjudice de démontrer le lien de cause à effet entre les défauts des produits et les dommages particuliers. Envisager de modifier ces règles pour que les plaignant·e·s puissent plus facilement défendre leur cas et réclamer un dédommagement.

Introduction

Les progrès rapides de l'intelligence artificielle (IA) transforment les secteurs économiques et stimulent la croissance. L'IA renferme un potentiel incroyable d'amélioration des vies et des modes de subsistance à travers le monde. Mais elle comporte aussi un risque d'exacerbation des inégalités existantes. Car les systèmes d'IA peuvent refléter et amplifier les préjugés, et reproduire les inégalités, tout particulièrement celles affectant les groupes historiquement marginalisés. Les systèmes et les processus d'IA sont de plus en plus employés dans des domaines décisifs comme la justice pénale, l'éducation, la santé et l'emploi. Il devient donc impératif d'aborder les inégalités croissantes et le manque de diversité au sein des écosystèmes de l'IA, et d'élaborer des politiques en la matière.

Il est indispensable de reconnaître l'égalité entre les genres et la diversité en tant que priorités pour la prise de décision dans l'IA si l'on veut atteindre les objectifs de développement durable (ODD). Les inégalités croissantes et le manque de diversité dans l'IA mettent sérieusement en péril les ODD, dans la mesure où ils reviennent à ignorer les besoins spécifiques, le savoir-faire local et les contextes culturels des groupes historiquement marginalisés et exclus. En négligeant les points de vue de la majorité mondiale dans les débats d'orientation, ce sont les inégalités existantes qui risquent d'être aggravées à l'échelle mondiale (CNUCED, 2021). Il existe un besoin impérieux d'un cadre réglementaire solide capable d'atténuer ces disparités, afin de prévenir les dommages et d'aboutir à un développement bénéfique des systèmes et des processus d'IA dans toutes les collectivités (Chauhan et Kshetri, 2022 ; Dankwa-Mullan et al., 2021 ; Dignum, 2023 ; Joyce et al., 2021 ; UNESCO, 2022 ; Vinuesa et al., 2020).

Le rapport Vers une réelle égalité en intelligence artificielle : une politique d'IA transformatrice pour l'égalité des genres et la diversité offre aux écosystèmes de l'IA, et tout particulièrement aux États dans leur rôle de porteurs de devoirs, des recommandations et des perspectives tirées de pratiques prometteuses sur la manière d'intégrer efficacement les approches d'égalité entre les genres et de diversité tout au long du cycle de vie de l'IA et dans l'élaboration des politiques en la matière.

Systèmes d'intelligence artificielle (IA)

Un système d'IA est un système fondé sur des machines qui, pour des objectifs explicites ou implicites, réalise des inférences, à partir des données qu'il reçoit, pour générer des sorties tels que des prédictions, du contenu, des recommandations ou des décisions pouvant influencer l'environnement physique ou virtuel. Après leur déploiement, les différents systèmes d'IA varient dans leurs niveaux d'autonomie et d'adaptabilité.

(OCDE, 2024b)

Ce rapport s'inscrit dans le cadre d'une initiative du groupe de travail du PMIA sur l'IA responsable et repose sur l'expérience locale et les connaissances de cinq régions du monde. Les consultations régionales et de groupes spécifiques ont mobilisé environ 200 personnes, de plus de 50 pays et un vaste éventail de communautés et d'identités. Les personnes consultées étaient issues du monde universitaire, de la société civile, de l'industrie et des gouvernements. En mettant l'accent sur les consultations, l'objectif était d'écouter et de comprendre une multitude de voix, de points de vue et d'expériences dans leurs contextes spécifiques. Les perspectives et recommandations de ce rapport s'appuient sur les consultations et sur un examen approfondi des documents et de la bibliographie sur le sujet. Ces résultats mettent en avant les défis majeurs identifiés en matière d'égalité des genres et de diversité dans les écosystèmes et les processus d'IA, qui illustrent les inégalités profondes enracinées au cœur des écosystèmes de l'IA.

« Nous devons utiliser comme point de départ les préjudices qui sont déjà subis et qui risquent d'être subis. Il nous faut pour cela écouter les personnes concernées, ainsi que celles qui ont déjà consacré de nombreuses années à identifier ces préjudices et à y répondre. En particulier, les femmes, les groupes minoritaires et les personnes marginalisées sont affectés de manière disproportionnée par les biais dans l'intelligence artificielle. Nous devons multiplier les efforts pour les inclure dans les discussions sur la gouvernance. »

**Volker Türk,
Haut-Commissaire des Nations Unies aux droits de l'homme, 2023**

Les défis majeurs révèlent comment de nouveaux types de préjudices et de discriminations individuels et collectifs, associés aux systèmes et processus d'IA, affectent de manière disproportionnée les groupes historiquement marginalisés et exclus. Il peut s'agir par exemple des personnes aux prises avec des obstacles fondés sur le genre, l'identité de sexe, l'expression du genre et l'orientation sexuelle, y compris les femmes (Costanza-Chock, 2018 ; Donnelly et Stapleton, 2022 ; Hamidi et al., 2018). Les systèmes et processus d'IA ont facilité encore plus la violence fondée sur le genre favorisée par la technologie (VFGFT), comme la diffusion non consensuelle d'images, le cyberharcèlement et la traque en ligne. L'anonymat et l'ampleur des plateformes fondées sur l'IA aggravent ces abus. L'extraction des caractéristiques de genre dans les tâches biométriques soulève également des préoccupations d'ordre éthique et de confidentialité, tout particulièrement pour les communautés LGBTQ+ (Ovalle et al., 2023a). Les systèmes d'AI comme les technologies de scanner corporel qui évaluent les écarts et les risques sont souvent fondés sur des hypothèses hétéronormatives et cisnormatives, ce qui donne lieu à des discriminations contre les personnes transgenres.

Les groupes historiquement marginalisés

Les groupes historiquement marginalisés et exclus désignent, sans s'y limiter : les femmes, les peuples autochtones, les personnes racisées, les personnes handicapées, les personnes confrontées à des obstacles en raison de leur orientation sexuelle, de leur sexe, de leur identité et/ou de leur expression de genre, et de leurs caractéristiques sexuelles (OSIGEGCS)¹, les personnes en situation de déplacement², et celles à bas revenus ou issues de milieux à bas revenus. Il n'existe pas de consensus autour de la définition ou de l'utilisation des termes « marginalisation » ou « groupes historiquement marginalisés ». La marginalisation fait généralement référence à une forme d'exclusion, impliquant des éléments économiques, politiques, culturels ou sociaux et comme « une situation défavorisée extrême et durable ancrée dans de profondes inégalités sociales » (UNESCO, 2009). Le terme « historiquement marginalisés » souligne le fait que la marginalisation de certains groupes s'est prolongée sur plusieurs décennies ou même plusieurs siècles, et que le système qui perpétue cette marginalisation et ses effets sont intergénérationnels.

(Adapté d'après l'UNESCO, 2009)

Le manque de diversité et d'inclusion dans l'écosystème de l'IA représente un autre défi de taille (Cheong et al., 2021 ; Suresh et al., 2022). L'exclusion des rôles techniques et non techniques limite la diversité des points de vue, aboutissant ainsi à des systèmes d'IA qui perpétuent l'exclusion du savoir et renforcent les stéréotypes dangereux et la discrimination (p.ex. Badaloni et Rodm, 2022, Buolamwini et Gebru, 2018 ; Hull, 2023 ; Ricaurte, 2019 ; UNESCO, 2024a). Les systèmes biaisés de prise de décision en matière des prestations sociales risquent, par exemple, d'accorder moins de ressources aux personnes handicapées et de restreindre les possibilités d'emploi des groupes marginalisés.

L'intersectionnalité des identités des gens appartenant à des groupes historiquement marginalisés peut entraîner une combinaison de désavantages, lorsque les effets néfastes des systèmes d'IA s'entrecroisent avec des inégalités sociales plus générales (Buolamwini et Gebru, 2018 ; Forster, 2022 ; Kong, 2022 ; Ricaurte, 2022 ; Ricaurte et Zasso, 2023 ; CNUCED, 2021 ; CDH, 2020). Par exemple, les systèmes d'IA ont déjà pris par erreur les béquilles pour des armes, exposant ainsi les personnes handicapées à un risque de préjudice encore plus élevé, dans des situations où elles sont déjà en danger (CDH, 2021). De même, le développement des technologies de surveillance et des analyses prédictives dans les contextes de migration ont souvent pour cible les communautés marginalisées, ce qui exacerbe encore plus les tensions existantes et contribue à la violation des droits de la personne (Dumbrava, 2021 ; McGregor et Molnar, 2023). Ces exemples mettent en évidence des inégalités profondément enracinées et des défis de taille à relever en matière d'égalité des genres et de diversité au sein des écosystèmes et des processus d'IA, qui demandent des efforts politiques concertés.

Discrimination algorithmique

La discrimination algorithmique survient lorsque les systèmes d'IA, sans justification aucune, traitent ou affectent les personnes différemment en fonction de leur race, couleur, origine ethnique, sexe, religion, âge, handicap ou toute autre catégorie protégée (Maison Blanche, 2022). Cette problématique complexe se trouve à la croisée du déploiement technologique, des pratiques sociales et des objectifs politiques. Elle va au-delà des biais de données et inclut toutes les situations dans lesquelles l'IA perpétue les inégalités par l'intermédiaire des interactions sociales.

(Adapté de l'indice mondial sur l'IA responsable, 2024)



Femme écoutant la présentation de sa collègue. Photo de Christina @ wocintechchat.com sur Unsplash.

Au vu de ces défis, le présent rapport offre une approche complète pour lutter contre les inégalités entre les genres, et le manque de diversité et d'inclusion dans l'IA, par le biais de recommandations sur la politique d'IA transformatrice. Ces recommandations contribuent à et s'appuient sur les principes de l'OCDE en matière d'IA (OCDE, 2024a), les recommandations de l'UNESCO sur l'éthique de l'intelligence artificielle (UNESCO, 2022), les Objectifs de développement durable (ODD) (Nations Unies, 2024), le Pacte numérique mondial (Nations Unies, 2024b) et d'autres travaux appelant à l'alignement des systèmes d'IA sur le cadre des droits de la personne, pour plus d'égalité entre les genres et de diversité au sein des écosystèmes de l'IA et dans l'élaboration des politiques en la matière (Forum économique mondial, 2024).³ La gouvernance internationale des systèmes et des processus d'IA est un moyen stratégique décisif pour atteindre les ODD. Le cadre pour la politique d'IA transformatrice exposé dans le présent rapport repose sur l'hypothèse que les inégalités sont structurelles, et qu'il est donc possible de les combattre en prenant les mesures de transformation adaptées.

Les recommandations de la politique d'IA transformatrice en vue d'une égalité réelle s'inscrivent dans une approche des systèmes et des processus d'IA fondée sur les droits de la personne. Cette approche place les droits de la personne au cœur des écosystèmes de l'IA en supprimant les contraintes et en améliorant les capacités, afin de garantir la pleine jouissance de ces droits au sein des écosystèmes de l'IA et dans l'élaboration de politiques en la matière. Les recommandations appellent à progresser sur la voie d'une égalité réelle en renversant les barrières structurelles qui empêchent les groupes historiquement marginalisés et exclus de profiter des bénéfices du développement de l'IA. De plus, elles cherchent à donner les moyens aux femmes et aux autres groupes marginalisés d'exercer pleinement leurs droits.

Égalité réelle

L'égalité réelle désigne l'égalité de facto (l'égalité de fait ou vraie égalité) ou l'égalité de résultats (Fredman, 2016 ; ONU-Femmes, 2019). Cela implique, entre autres, de veiller à ce que tout le monde bénéficie des mêmes chances et d'un accès égal aux opportunités, ainsi que d'un environnement permettant de parvenir aux mêmes résultats. Tel qu'exposé dans le présent rapport, l'égalité réelle implique une compréhension de l'équité⁴, de l'égalité entre les genres⁵ et de la diversité défendant l'égalité des chances et des résultats par le biais de mesures qui accueillent la différence, qui favorisent l'expression et la participation et qui éliminent les discriminations, la marginalisation, et les inégalités dans la répartition du pouvoir et des ressources. L'égalité réelle constitue aussi bien un objectif en elle-même qu'un programme de transformation et un moyen d'atteindre d'autres objectifs de développement durable, et notamment en matière de santé, de développement économique et d'éducation.

Buckup, 2009 ; OCDE, 2015 ; Ostry et al., 2018)



Protestations de groupes d'activistes LGBTQ+ contre le projet de loi sur la discrimination religieuse à Sidney, Australie.
Photo : Nikolas Gannon sur Unsplash.

L'instauration d'une égalité réelle grâce à une transformation radicale de l'IA permettra non seulement de faire progresser les droits de la personne mais aussi de stimuler le développement économique et social. Il est essentiel d'intégrer des principes d'égalité entre les genres et de diversité en suivant le cadre des droits de la personne pour créer des écosystèmes de l'IA à la fois innovants, inclusifs et alignés sur les besoins sociaux. L'OCDE et tous les États membres de l'UNESCO se sont engagés formellement en ce sens. Grâce aux politiques inclusives pour l'IA, il est possible d'améliorer la qualité, l'utilisabilité et l'efficacité des systèmes et des processus d'IA, en mettant le cap sur un avenir plus équitable, durable et prospère dans l'ensemble des collectivités.

Présentation du rapport

Partie I. Consultations, défis majeurs et voies potentielles

Le présent rapport commence par décrire les consultations régionales et de groupes spécifiques. Réalisées en collaboration avec des organisations partenaires, elles visent à mieux comprendre les besoins non satisfaits, les barrières systémiques et les solutions stratégiques proposées. Cette section est suivie d'un aperçu des résultats des consultations, identifiant les défis majeurs et les voies potentielles vers l'égalité des genres et la diversité au sein des écosystèmes de l'IA et dans l'élaboration de politiques en la matière. Les consultations servent de fondement au contenu du rapport.

Partie II. Cadre conceptuel

Après l'exposition des défis majeurs et des voies potentielles, nous présentons le cadre conceptuel sur lequel repose l'analyse des pratiques prometteuses et le développement de recommandations clés adressées aux responsables des politiques.

Le cadre conceptuel est composé de trois parties. Tout d'abord, le concept d'égalité réelle et de transformation radicale sont expliqués en tant qu'approche. Ensuite, une approche sociotechnique des systèmes d'IA et du cycle de l'IA est exposée, en présentant les possibilités d'intervention au fil des étapes du pipeline de l'IA. Enfin, le cadre pour la politique d'IA transformatrice est présenté avec ses trois dimensions clés, à savoir 1) Remédier aux désavantages systémiques (droit à l'inclusion) ; 2) Comblent le déficit démocratique (droit de participation) ; et 3) Corriger le déni de reconnaissance (droit à la dignité).



Jeune homme vêtu d'un sweater au milieu de la foule.
Photo de Jake Weirick sur Unsplash.

Partie III. Pratiques prometteuses et recommandations

La section suivante met en avant les pratiques prometteuses qui incluent la promotion des droits et la participation effective ou l'inclusion des groupes marginalisés. Ces informations sont fournies à titre indicatif, dans l'optique d'offrir aux responsables des politiques et autres parties intéressées des exemples concrets de mesures à reproduire ou à adapter à leurs contextes.

Pour finir, le rapport offre une présentation détaillée des recommandations clés en vue d'une politique d'IA transformatrice. Les recommandations sur la politique d'IA transformatrice sont regroupées en quatre catégories qui reflètent les dimensions principales du cadre :

- Conception inclusive et innovation démocratique
- Participation effective à la gouvernance de l'IA
- Transparence et responsabilisation dans la prévention des préjudices
- Accès efficace à la justice

À travers ces catégories, les recommandations clés sur la politique d'IA transformatrice s'inscrivent dans une approche globale visant l'instauration d'une égalité réelle.



PARTIE I.

Consultations, défis majeurs et voies potentielles

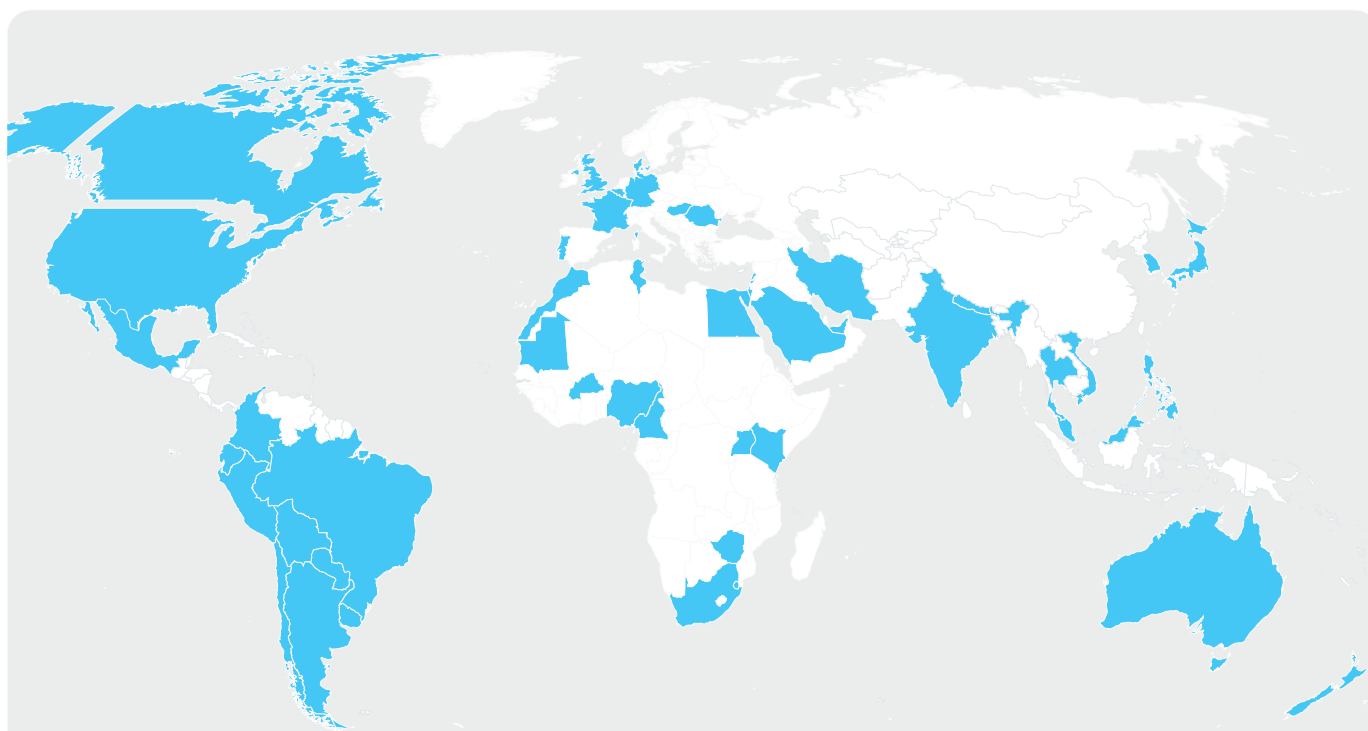


Trois femmes de différentes générations vêtues d'habits traditionnels colorés, assises les unes derrière les autres devant un mur bleu. Photo de Srimathi Jayaprakash sur Unsplash.

Consultations régionales et de groupes spécifiques

Le projet a fait appel à des spécialistes de cinq régions du monde, soit environ 200 personnes représentant plus de 50 pays et un vaste éventail de communautés et d'identités. Les membres du groupe, qui étaient issus du monde universitaire, de la société civile, de l'industrie et du gouvernement, ont partagé leurs points de vue lors de consultations régionales ou de groupes. Ces consultations sont décrites dans la présente section, en les accompagnant de réflexions et en décrivant leurs limites.

L'équipe du projet a collaboré avec des organisations de recherche régionales et de la société civile (organisations partenaires) pour mener à bien les consultations régionales : Derechos Digitales, Research ICT Africa et Data Pop Alliance, ainsi que plusieurs personnes associées de recherche de différentes régions du monde. Les organisations partenaires ont identifié les parties prenantes et les organisations clés dans chaque région, en vue d'une organisation optimale des consultations régionales. L'accent ayant été mis sur une couverture globale, internationale, intersectionnelle et multipartite, les gens ayant collaboré ont tout mis en œuvre pour travailler de manière inclusive avec un vaste éventail de parties prenantes.



Carte du monde représentant les régions dans lesquelles la consultation a eu lieu. Illustration de Leonardo Studio Design.⁶

Les consultations se sont déroulées dans cinq régions du monde :

- Afrique subsaharienne
- Moyen-Orient et Afrique du Nord
- Amérique latine
- Amérique du Nord et Europe
- Asie-Pacifique



50+ PAYS
200+
PARTICIPANT.E.S
(180 RÉGIONAUX.LES + 20+ DE
GROUPES SPÉCIFIQUES)

En plus des consultations régionales, l'équipe du projet avait à cœur de recueillir les points de vue de groupes marginalisés et des personnes s'identifiant comme appartenant à ces groupes, et notamment, sans s'y limiter, les femmes, les peuples autochtones, les personnes racisées, les personnes handicapées, les personnes confrontées à des obstacles en raison de leur orientation sexuelle, de leur sexe, de leur identité et/ou de leur expression de genre, et de leurs caractéristiques sexuelles (OSIGEGCS)¹, les personnes en situation de déplacement, celles à bas revenus ou issues de milieux à bas revenus, et les personnes parlant des langues autochtones et minorisées, ou impliquées dans leur revitalisation, situées dans différentes régions et contextes représentatifs des groupes clés.

La consultation s'est déroulée sous plusieurs formes : entretiens individuels, tables rondes, contributions écrites... Il s'agissait de recueillir un large spectre de voix et de points de vue. La participation se composait d'un vaste éventail de communautés et d'identités, et incluaient des membres de différents groupes clés. De plus, un groupe d'experts-conseils (GEC) et un groupe consultatif du projet (GCP) ont apporté leur contribution. Pendant les étapes préparatoires, le GEC a joué un rôle consultatif auprès de l'équipe du projet. Il était composé de personnes handicapées, de plusieurs personnes confrontées à des barrières en raison de leur OSIGEGCS, de membres des populations autochtones et des groupes racisés, et de personnes représentant des groupes en situation de déplacement, issus de différentes parties du monde. Les membres du GEC ont apporté des points de vue essentiels, variés et intersectionnels. Pour valider les étapes méthodologiques des projets, un GCP destiné à une consultation régulière a été créé. Les membres de ce groupe incluaient des spécialistes du PMIA et des spécialistes externes. L'équipe du projet a analysé les perspectives et les discussions soulevées au cours des entretiens, en tant que données qualitatives, afin d'y dégager des thèmes dominants, des convergences et des divergences d'opinion, des avis unanimes parmi les personnes et les groupes participant aux consultations, ainsi que des pratiques prometteuses, des apprentissages à tirer et un partage des connaissances.

Réflexions et limites

Des sujets sensibles sur le plan culturel et légal

Le nombre de personnes participantes disposées à ou capables de participer aux entretiens a été conditionné par certains sujets pouvant être considérés comme polémiques, tabous ou illégaux dans certaines régions, tels que ceux en lien avec l'orientation sexuelle et l'expression de genre. L'attitude des populations locales, le degré de tolérance et les lois pouvaient en effet compromettre l'acceptation et la sécurité de ces personnes.

La barrière de la langue

Le projet a été mené principalement en anglais, avec la possibilité d'y participer en espagnol et en français. Cela a dans un sens limité l'égalité des chances en matière de participation.

Accès et durée

Les consultations ayant été virtuelles, l'accès à un appareil connecté était nécessaire, ce qui a empêché la participation de certaines personnes, par conséquent considérées comme victimes de marginalisation numérique. Les questions techniques des questionnaires impliquaient une limitation de l'accessibilité en fonction du niveau de connaissances préalables sur l'IA. La brièveté des délais a également restreint la participation et la qualité des résultats.

Des déséquilibres dans la représentation

La participation de membres de la société civile et du monde universitaire a été supérieure à celle des membres des gouvernements ou de l'industrie. La plupart des personnes ayant participé présentaient par ailleurs un haut niveau d'études. La voix des groupes marginalisés ou moins instruits n'a donc pas été représentée de façon adéquate. La représentation des personnes et groupes non binaires a été moins importante que celle d'autres identités.

Limitations de la représentation de certains pays et régions

Certains pays et régions n'ont pas été inclus dans le rapport pour des questions de temps, d'accessibilité et de contraintes de ressources. En particulier, les pays des Caraïbes n'ont pas été inclus dans le processus de consultation.

Résultats des consultations régionales et de groupes spécifiques

Cette section est consacrée aux principales conclusions tirées de la cartographie empirique et détaillée des opinions, points de vue, expériences et bonnes pratiques des différentes parties prenantes en matière d'IA, d'égalité des genres et de diversité, en lien avec leur expérience des barrières systémiques. Des solutions stratégiques pour instaurer l'égalité des genres et la diversité au sein des écosystèmes de l'IA seront également proposées. Dans le cadre de la synthèse de ces conclusions, une attention toute particulière a été portée à la représentation de la diversité des voix et points de vue des circonscriptions et régions. Le résumé suivant ne représente pas les points de vue nuancés de toutes les personnes et groupes participants. Il s'agit d'un aperçu des expériences décrites en lien avec les défis majeurs et les voies potentielles pour aborder l'égalité des genres et la diversité tout au long du cycle de vie de l'IA et dans l'élaboration de politiques en la matière. Les résultats sont classés tels que suit. La première partie de cette section offre une ébauche et une description des trois principaux défis qui se posent pour instaurer l'égalité des genres et la diversité au sein des écosystèmes de l'IA, d'après les déclarations des personnes et groupes participants. Dans la seconde partie de cette section sont exposées quatre voies pour relever les défis en matière d'égalité des genres et de diversité dans l'IA. Cette section se clot sur un bref résumé des difficultés qui se posent et des voies à suivre.

Défis majeurs en lien avec l'égalité des genres et la diversité en IA

Les consultations régionales et de groupes spécifiques ont révélé trois grandes difficultés en vue de l'instauration de l'égalité des genres et de la diversité dans les écosystèmes de l'IA.

■ L'accès n'est pas synonyme d'inclusion

Un thème récurrent lors des consultations régionales, et tout particulièrement en Afrique subsaharienne, a été le besoin de distinguer le simple accès aux produits et services de l'IA d'une véritable inclusion dans les écosystèmes de l'IA. Les personnes interrogées ont été catégoriques sur le sens de la véritable inclusion : une transformation systémique de la conception et du développement de l'IA, qui placent aux commandes les groupes historiquement marginalisés.

■ Invisibilisation et exclusion du savoir

Les modèles d'IA ne sont pas représentatifs de la diversité et perpétuent l'exclusion du savoir, entre autres types d'exclusion, aboutissant ainsi à une injustice épistémique. Les consultations en Amérique latine, en Afrique subsaharienne et en Asie, ainsi qu'auprès des communautés autochtones d'Amérique du Nord, ont permis de décrire l'incompatibilité des modèles dominants dans le développement de l'IA avec l'ontologie du savoir de ces populations.

■ Répartition inégale des ressources

La répartition des ressources au sein des écosystèmes économiques de l'IA à travers le monde est inégale. Toutes les consultations régionales ont identifié en tant que défi majeur l'exclusion systémique des groupes historiquement marginalisés des bénéfices de l'économie de l'IA, et tout particulièrement parmi les populations de la « majorité mondiale ». L'exploitation par le travail et l'extraction massive de données contribuent à perpétuer ces inégalités et l'injustice environnementale est ancrée au cœur de la chaîne de valeur de l'IA.



Groupe de femmes en Sierra Leone. Photo d'Annie Spratt sur Unsplash.



L'accès n'est pas synonyme d'inclusion

Inclusion

L'inclusion consiste à garantir la participation totale et véritable des groupes marginalisés dans tous les secteurs, à tous les niveaux, ce qui produit un impact positif mesurable sur leur qualité de vie et leur accès aux droits de la personne au même titre que le reste de la population. L'inclusion demande a) des mesures d'autonomisation et de soutien des groupes marginalisés, et b) des mesures structurelles, politiques et institutionnelles pour corriger les injustices systémiques à tous les niveaux. Les initiatives inclusives accueillent obligatoirement la diversité et créent un environnement de respect mutuel, d'appartenance, de confiance, de soutien et d'échange.

(Département des affaires économiques et sociales des Nations Unies, 2009)



« L'inclusion dans l'IA ne consiste pas seulement à accueillir ou à donner accès. Cet accès doit permettre l'épanouissement ou apporter des bénéfices. Donc si l'accès n'entraîne pas de bénéfices, c'est un accès excluant. »

*Angella Ndaka,
Centre for African Epistemic Justice, Kenya*

« [L'égalité des genres et la diversité sont] généralement [réduites à] l'augmentation du recrutement de femmes. Mais l'environnement dans lequel elles sont intégrées n'est pas accueillant. »

*Luisa Olaya,
GIZ (FAIR Forward), Allemagne*



« Nous devrions faire preuve de vigilance lors de la création de solutions de genre et de diversité génériques, parce que chaque pays ou région possède ses propres différenciations sociales et ses priorités en matière d'inclusion. »

*Anonymous,
Asie*



Un thème récurrent lors des consultations régionales, et tout particulièrement en Afrique subsaharienne, a été le besoin de distinguer le simple accès aux produits et services de l'IA d'une véritable inclusion dans les écosystèmes de l'IA. Le fait de se concentrer uniquement sur l'accès dessert la véritable inclusion et réduit les programmes de transformation des dynamiques de genre à des initiatives isolées qui ne sont ni complètes, ni durables, ni sources de transformation. Les personnes interrogées lors des consultations régionales ont en revanche été catégoriques sur le sens de la véritable inclusion : une transformation systémique de la conception et du développement de l'IA, qui placent aux commandes les groupes historiquement marginalisés. Les consultations révèlent la perception selon laquelle des mesures localement adaptées visant à améliorer l'autodétermination des groupes marginalisés au sein des écosystèmes de l'IA sont essentielles pour combattre les disparités sociales et les injustices structurelles, de sorte à dépasser les approches décontextualisées en matière d'égalité des genres et de diversité pour la gouvernance de l'IA.



Invisibilisation et exclusion du savoir



« Le manque de représentation de la diversité risque d'aboutir à des équipes de développement de l'IA qui ignorent les besoins et les points de vue des femmes, des personnes de couleur et d'autres groupes marginalisés. »

*Anonymous,
Asie*

« Le problème le plus urgent est la représentation des personnes non-binaires, trans et autochtones à des postes de développeuses et de dirigeantes dans le secteur technologique. »

*Marcos Cornelio Sánchez Ramírez,
PIT Policy Lab, Mexique*



« Il est vraiment important d'aligner nos technologies sur notre histoire et notre culture, et la souveraineté et la réciprocité constituent des valeurs essentielles pour l'IA. »

*Florian Lebret,
communautés autochtones, Canada*

Les modèles d'IA ne sont pas représentatifs de la diversité et perpétuent l'exclusion du savoir, entre autres types d'exclusion, ce qui donne lieu à une injustice épistémique : l'exclusion du savoir, des cultures et des langues. Les consultations régionales en Amérique latine, en Afrique subsaharienne et en Asie, au même titre que les discussions avec les communautés autochtones d'Amérique du Nord, ont révélé l'incompatibilité des modèles dominants du développement de l'IA avec l'ontologie du savoir de ces communautés. Les personnes et les groupes participants ont exprimé une forte inquiétude face au renforcement des hiérarchies intersectionnelles de différences sociales dans les écosystèmes de l'IA. Les systèmes d'IA intègrent les valeurs et les pratiques culturelles des pays et des espaces dans lesquels ils sont conçus et développés, ce qui est une source d'incohérence lorsque ces mêmes technologies sont appliquées dans des sociétés et cultures différentes. Les ensembles de données sur lesquels ces systèmes reposent ne représentent souvent pas la diversité, menant ainsi à l'invisibilisation et au manque de représentation des groupes historiquement marginalisés. Et les préjugés sociaux existants ont toutes les chances d'être renforcés. Par ailleurs, les biais présents chez les équipes de développement de l'IA risquent de provoquer l'effacement culturel ou de violer les droits de la personne. Le dialogue avec les peuples autochtones a révélé leur grande préoccupation face à l'incompatibilité entre les modèles de développement dominants de l'IA et l'ontologie de leur propre savoir. Lors des consultations régionales en Asie, les contraintes de ressources ont été désignées comme le plus grand obstacle à la lutte contre les injustices épistémiques dans les systèmes d'IA.

Injustice épistémique

L'injustice épistémique désigne le traitement injuste accordé à certaines personnes ou groupes, sur la base de la remise en question de Leur capacité à détenir ou à produire le savoir (Fricker, 2007; Hull, 2023). Ce concept met en évidence la manière dont les biais et les préjugés peuvent nier ou discréditer les connaissances, les expériences et les points de vue des personnes ou groupes marginalisés.

- L'injustice herméneutique correspond à une brèche dans la compréhension collective qui empêche les personnes marginalisées de tirer du sens de leurs expériences. Cela est dû au fait que leurs expériences sociales ne sont pas bien comprises ou reconnues par la société dans son ensemble, souvent faute de partager un langage ou des concepts en commun.
- Les injustices testimoniales se produisent lorsque la crédibilité d'une personne qui s'exprime est injustement jugée sous l'influence de préjugés. Par exemple, lorsqu'un membre d'un groupe marginalisé n'est pas cru ou est ignoré en raison de son identité, quelle que soit la validité des informations qu'il transmet.
- L'injustice contributive est une conséquence de l'injustice testimoniale : les personnes marginalisées ne participent pas à la création et à la validation du savoir. Leurs perspectives et contributions ne sont pas valorisées ni reconnues, ce qui limite leur influence dans les processus de prise de décision et empêche une compréhension globale de la société.



Protestation d'un groupe d'activistes avec un mégaphone, concept de grève et de manifestation. Photo de Halfpoint sur iStock.



Répartition inégale des ressources



« Lorsque nous parlons du cycle de vie de l'IA, nous parlons aussi de l'extraction de minéraux dans les mines de cobalt du Congo. Dans quelle mesure réfléchissons-nous aux implications en matière de genre du développement de l'IA dans ce contexte ? »

*Kelly Stone, AI Observatory,
Afrique du Sud*

« Tous ces beaux discours sur les start-ups et la libération économique masquent une autre réalité, à savoir que [...] nous nous trouvons en présence d'une autre forme de colonisation [...]. »

*Maha Jouini,
African Center for Artificial Intelligence
and Digital Technology, Mauritanie*



La répartition inégale des ressources au sein de l'écosystème économique international de l'IA et l'exclusion systématique des bénéfices de l'économie de l'IA, subie par les femmes et les autres groupes historiquement marginalisés, tout particulièrement parmi les populations de la majorité mondiale, ont été identifiés comme un défi majeur lors de toutes les consultations régionales. L'exploitation par le travail et l'extraction massive de données continuent à perpétuer ces inégalités et l'injustice environnementale est ancrée au cœur de la chaîne de valeur de l'IA.

L'exploitation généralisée par le travail des personnes concernées appartenant à la majorité mondiale est apparue comme une sérieuse inquiétude lors de la consultation régionale en Asie. Les tâches pénibles dans les chaînes de valeur de l'IA telles que l'étiquetage de données, l'annotation des données et des images, la transcription, la traduction et la modération de contenu sont accomplies de manière disproportionnée par du personnel contractuel recruté dans la majorité mondiale. Les consultations en Amérique latine ont également révélé que certaines formes de travail du secteur de l'IA sont moins bien perçues que d'autres, et que ces tâches pénibles sont généralement réservées à des personnes marginalisées de la majorité mondiale. Or, ces contributions, tout comme celles des femmes et plus généralement celles des autres groupes marginalisés, ont tendance à être invisibilisées dans l'écosystème de l'IA.

L'extraction massive de données a été aussi identifiée comme une inquiétude : les membres de la majorité mondiale ainsi que les personnes et groupes marginalisés dans le monde minoritaire en sont réduits à fournir des données brutes et à simplement consommer les produits et services finaux de l'IA. La consultation régionale en Amérique latine a mis en lumière que la règle sur le consentement libre et éclairé généralement adoptée dans le cadre législatif de la protection des données à caractère personnel ne serait pas adaptée pour répondre à la problématique de l'extraction massive de données. Ce type d'approche contractualiste ne permet pas de rendre compte des distorsions découlant du déséquilibre de forces entre les fournisseurs de système d'IA et les personnes concernées. Alors que les multinationales du numérique contrôlent les infrastructures numériques décisives sur lesquelles repose une grande partie de la vie économique et sociale, les personnes se voient souvent contraintes à accepter des conditions de service injustes, avec notamment les modèles commerciaux basés sur la surveillance des données.

La consultation régionale au Moyen-Orient et en Afrique du Nord a révélé des inquiétudes autour de la perpétuation de dépendances économiques néocoloniales par le biais de la chaîne de valeur de l'IA. La possibilité pour les pays de la majorité mondiale de développer l'IA selon leurs propres conditions, pour leur propre bénéfice et dans le respect de leurs priorités est sévèrement limitée. Les personnes et les groupes participants ont exprimé l'idée qu'à long terme, la répartition actuelle des ressources de l'IA court le risque d'intensifier les inégalités à l'échelle de la planète. Enfin, tel que mis en exergue dans les consultations régionales en Afrique subsaharienne et auprès des personnes autochtones participantes, des inégalités environnementales sont ancrées au cœur de la chaîne de valeur de l'IA, en lien avec l'extraction des ressources et les autres activités affectant de manière disproportionnée les communautés locales de la majorité mondiale.

Voies pour relever les défis en matière d'égalité de genre et de diversité au sein des écosystèmes de l'IA

Les consultations régionales ont permis d'identifier quatre voies principales pour aborder la problématique de l'égalité réelle.

- **Développer la capacité du secteur public et du grand public**
Développer et renforcer les connaissances et la compréhension des systèmes d'IA par le secteur public et le grand public, tout comme leur capacité à interagir avec lesdits systèmes.
- **Encourager la conception inclusive dans le secteur privé**
Encourager le secteur privé à prendre des mesures pour l'intégration de processus de conception inclusive tout au long du cycle de vie de l'IA.
- **Responsabiliser les équipes de développement, les fournisseurs et les opérateurs**
Instaurer des cadres réglementaires légalement contraignants pour responsabiliser les entités du secteur public et privé qui conçoivent, détiennent, déploient et exploitent les systèmes d'IA face aux préjudices et dommages provoqués et aux droits enfreints.
- **Garantir la participation effective des groupes marginalisés**
Amplifier la voix et élargir les espaces de participation au développement de l'IA et aux processus de prise de décision en la matière.



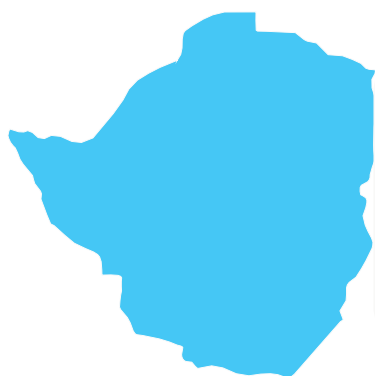
Conversation d'un groupe diversifié de cinq personnes assises à une table. Photo de Christina Morillo sur Unsplash.



Développer la capacité du secteur public et du grand public

« Les systèmes algorithmiques ou d'IA sont perçus comme étant destinés à un groupe réduit spécialistes. La manière dont ils nous affectent et dont ils sont enracinés dans notre quotidien est très peu connue. »

*Dr Preeti Raghunath,
Université de Sheffield, Malaisie*



« La plus grande ressource que nous pouvons apporter aux responsables des politiques est la sensibilisation. Il faut les informer et leur faire comprendre en quoi consiste l'IA, et ce qu'elle a à offrir à la société. »

*Kudakwashe Dangajena,
Union africaine, Zimbabwe*

Les consultations régionales ont bien montré que pour combattre les inégalités existantes, il est décisif de renforcer la capacité des responsables des politiques et du grand public à participer aux processus de prise de décision en matière d'IA. Le renforcement des capacités des fonctionnaires et l'éducation civique contribueront à les doter des connaissances et de la compréhension des écosystèmes de l'IA et de la capacité à interagir ces derniers.

Le développement de la capacité du secteur public a été identifié comme un facteur majeur pour l'intégration efficace de mesures visant à atteindre une égalité réelle dans la gouvernance de l'IA. D'après plusieurs personnes et groupes participants, les programmes d'éducation civique devraient accorder la priorité à la compréhension par le public des enjeux décisifs de l'éthique de l'IA et des principes normatifs pour une IA fondée sur les droits de la personne, tout en luttant activement contre le pessimisme technologique. Ceci implique de créer des mécanismes plus inclusifs et participatifs pour impliquer le public dans l'élaboration des trajectoires que suivra l'IA.



Homme et femme assis devant un ordinateur. Photo de Desola Lanre-Ologun sur Unsplash.



Encourager la conception inclusive et l'innovation démocratique



« L'IA devrait être open source. Les communautés devraient pouvoir ouvrir, voir, vérifier le code et observer ce qu'il fait. »

*Gerardo López Gómez,
Derechos Digitales, Équateur*

« Il n'y a pas de mal à s'inspirer d'un concept, mais encore faut-il l'ancrer, le contextualiser, achever sa conception, le développer, le tester, le déployer, l'enseigner et le peaufiner dans le pays appelé à en bénéficier. Je ne pense pas qu'il existe un raccourci à tout ce processus. »

*Kelly Stone,
AI Observatory, Afrique du Sud*



« L'intégration de la diversité et de l'égalité des genres demande d'adopter une stratégie intersectionnelle, axée sur l'éducation et sur la collaboration et s'appuyant sur des politiques, qui privilégie la création d'une IA éthique, inclusive et équitable. »

*Anonymous,
Asie*

Les consultations ont permis d'identifier le besoin d'encourager les processus de conception inclusifs, tout au long du cycle de vie de l'IA. Les consultations régionales en Asie, en Afrique subsaharienne et en Amérique latine soulignent l'importance d'une conception et d'un développement technologique inclusifs pour parvenir à une égalité des genres réelle et des résultats en matière de diversité dans les écosystèmes de l'IA. Cela demande d'encourager toutes les parties prenantes à miser sur un développement éthique, inclusif et équitable de l'IA. Cela implique également d'inciter les équipes de développement de l'IA à personnaliser leurs modèles en fonction des besoins et des conditions locales, de sorte à prendre en compte le contexte social, culturel et environnemental unique dans lequel ils sont déployés. Un aspect de ce type de contextualisation est apparu comme particulièrement critique lors de la consultation régionale en Asie : la diversité linguistique.

Les pratiques de conception inclusive peuvent être encouragées par le biais de politiques publiques démocratisant l'innovation. Ces politiques devraient fournir l'accès à des ressources, des formations et des infrastructures aux différentes communautés, afin de leur permettre de développer des technologies adaptées à leurs besoins, en imposant leurs propres conditions.



Responsabiliser les équipes de développement, les fournisseurs et les opérateurs

« Je ne crois pas qu'il soit possible pour les entreprises de technologie de s'autoréglementer et de s'autoresponsabiliser. Ce sont nos institutions démocratiques qui doivent leur demander de rendre des comptes. »

*Anonymous,
Thaïlande*



« Les États doivent fixer des règles minimales de transparence, de vérifiabilité et de participation sociale pour la passation de marchés et l'achat de technologie. »

*José Alfredo Hau Caamal,
Mozilla Foundation e Ação Educativa, Brésil*



Tous les participant-e-s considèrent comme une grande priorité de demander des comptes aux équipes de développement, aux fournisseurs et aux opérateurs des systèmes d'IA. Les consultations régionales ont révélé que les participant-e-s sont unanimes quant au besoin de mettre en place des cadres réglementaires contraignants en vue de leur responsabilisation. Les entités du secteur public et privé qui conçoivent, détiennent, déploient et exploitent les systèmes d'IA doivent rendre des comptes face aux préjudices causés et aux droits enfreints.

Les vérifications externes et les études d'impact sur les droits de la personne ont été considérées comme des outils importants d'intégration de la diversité et de l'égalité des genres dans le secteur de l'IA, sans être pour autant la panacée. Les risques de les confier exclusivement au secteur privé ont été mentionnés, et les apprentissages tirés des secteurs environnementaux et financiers en matière de corruption et de captation ont été mis en avant. Les États et la société civile devraient être préparés et capables de mener à bien des vérifications et des évaluations dans l'intérêt public.

Tel que souligné par les personnes interrogées en Amérique latine, la transparence et la vérifiabilité du code source sont essentielles pour pouvoir demander des comptes. Lors des consultations régionales au Moyen-Orient-Afrique du Nord et en Amérique latine, le besoin de répondre au risque croissant d'une alliance entre la surveillance publique et privée avec l'IA est apparu comme une priorité. Les personnes interrogées au Moyen-Orient-Afrique du Nord ont insisté sur le rôle des États dans l'établissement d'exigences de diligence voulue en matière de droits de la personne mais aussi de mécanismes de responsabilisation et de réparation, en lien avec le développement, l'utilisation, la validation et l'achat de technologies numériques.



Garantir la participation effective des groupes marginalisés



« La perception dominante en Amérique latine de qui est autorisé ou non à parler de technologie constitue encore un obstacle de taille. Il est encore majoritairement considéré que seuls les ingénieurs peuvent émettre des critiques sur l'IA. »

*Marcos Cornelio Sánchez Ramírez,
PIT Policy Lab, Mexique*

« Les conversations autour de la gouvernance de l'IA étant menées par les individus au pouvoir, elles restent donc limitées à ces sphères, qui s'alimentent mutuellement et créent des opportunités qui leur sont réservées. Au lieu de nourrir l'intérêt du public, la portée de l'IA s'en trouve restreinte. »

*Dr Preeti Raghunath,
Université de Sheffield, Malaisie*



Les consultations ont mis en évidence le besoin de démocratiser le développement de l'IA et les processus de prise de décision via une participation réelle. Les personnes interrogées défendaient une transition qui permettrait de passer d'une approche étroite axée sur les spécialistes à un processus large, à l'échelle de toute la société et ancré dans la participation massive de toute la communauté. Ce changement est essentiel pour combattre les injustices, les préjudices potentiels et l'opacité des technologies de l'IA.

Les personnes et les groupes participants en Asie, en Afrique subsaharienne et en Amérique latine ont insisté sur les bénéfices de la coconception des systèmes d'IA dans l'utilisation des produits finis et dans la promotion de collaborations interdisciplinaires. Il est essentiel d'impliquer le grand public dans ces débats pour développer des systèmes d'IA sensibles aux contextes du monde réel et à la diversité de points de vue, en tenant tout spécialement compte des droits des personnes les plus exposées à un risque de préjudice.

L'ensemble des personnes et des groupes interrogés ont jugé indispensable de mettre en place des mécanismes pluralistes efficaces pour assurer une participation inclusive et réelle. Toutes les parties, à savoir les gouvernements, le secteur privé, la société civile, le monde technique et universitaire et les personnes utilisatrices doivent s'impliquer dans leurs rôles respectifs conformément aux principes de coopération numérique mondiale inscrits dans l'Agenda de Tunis du SMSI. À l'heure actuelle, la société civile et les groupes marginalisés sont sous-représentés. Des mesures spécifiques (financement, budgets de préparation,ancements de partenariats...) sont nécessaires pour permettre la participation effective des groupes marginalisés à l'élaboration des politiques. En outre, la consultation réalisée en Asie a révélé une inquiétude face à une éventuelle captation par l'élite, c'est-à-dire face au détournement par les groupes dominants de fonds publics destinés à servir les intérêts de l'ensemble de la population. Les Principes directeurs relatifs aux entreprises et aux droits de l'homme des Nations Unies (CDH, 2011) ont été mentionnés comme essentiels pour veiller à l'implantation des processus de diligence requise tout au long du cycle de vie de l'IA et pour garantir une participation effective, notamment celles des personnes handicapées, et la réalisation d'études d'impact.

Un bref résumé des défis à relever et des voies à suivre

Les consultations régionales et de groupes spécifiques ont révélé trois grandes difficultés en vue de l'instauration de l'égalité des genres et de la diversité dans les écosystèmes de l'IA : 1) l'accès n'est pas synonyme d'inclusion ; 2) l'invisibilisation et l'exclusion du savoir ; et 3) les inégalités dans la répartition des ressources.

Les consultations régionales ont permis d'identifier quatre voies principales d'aborder la problématique de l'égalité réelle : 1) développer les capacités du secteur public et du grand public ; 2) encourager la conception inclusive dans le secteur privé ; 3) responsabiliser les équipes de développement, les fournisseurs et les opérateurs ; et 4) garantir la participation effective des groupes marginalisés.



Femme en fauteuil roulant dans le couloir d'un bâtiment en brique. Photo de Marcus Aurelius sur Unsplash.



PARTIE II.

Cadre conceptuel



Gros plan sur une main couverte de peinture de toutes les couleurs sur un fond bleu. Photo d'Alexander Grey sur Unsplash.

Une approche transformatrice pour parvenir à l'égalité réelle

Basé sur des approches de transformation des dynamiques de genre, le présent rapport propose une approche transformatrice des politiques en matière d'IA dans le but d'instaurer une *égalité réelle*.

Égalité réelle

L'égalité réelle désigne l'égalité de facto (l'égalité de fait ou vraie égalité) ou l'égalité de résultats (Fredman, 2016 ; ONU-Femmes, 2019). Cela implique, entre autres, de veiller à ce que tout le monde bénéficie des mêmes chances et d'un accès égal aux opportunités ainsi que d'un environnement permettant de parvenir aux mêmes résultats. Tel qu'exposé dans le présent rapport, l'égalité réelle implique une compréhension de l'équité⁷, de l'égalité entre les genres⁸ et de la diversité défendant l'égalité des chances et des résultats par le biais de mesures qui accueillent la différence, qui favorisent l'expression et la participation et qui éliminent les discriminations, la marginalisation, et les inégalités dans la répartition du pouvoir et des ressources.

L'égalité réelle s'inscrit dans le cadre des droits de la personne et trouve son reflet dans plusieurs droits qui se renforcent mutuellement. L'égalité des genres, la participation et l'inclusion réelles visent à garantir que toutes les personnes, quel que soit leur sexe ou OSIGEGCS, âge, handicap, origine ethnique, religion, statut économique ou autre statut social, puissent jouir pleinement de leurs droits. L'égalité des genres est reconnue comme un principe fondamental des droits de la personne reflété dans les ODD. Elle vise à ce que les personnes, quel que soit leur sexe, puissent jouir de l'égalité des chances et ne subissent aucune discrimination.⁹ La participation et l'inclusion réelles sont des droits de la personne établis dans les conventions internationales, impliquant que les personnes jouent un rôle significatif dans les processus de prise de décision qui affectent leur vie. Afin de parvenir à l'égalité des genres, à la participation réelle et à l'inclusion de tous, il est essentiel d'instaurer l'égalité réelle.



Groupe de sept Indiennes cousant en cercle Photo d'AROYBARMAN sur iStock.



Rivière entourée d'une forêt de mangroves à Krabi, Thaïlande Photo de Jitti Narksompong sur iStock.

L'égalité réelle dans le droit international relatif aux droits de la personne

Il existe neuf principaux instruments internationaux en matière de droits de la personne (Bureau du Haut-Commissaire aux droits de l'homme, 2024b) Certains d'entre eux reflétant le principe d'égalité réelle sont présentés ci-dessous.

La Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (Comité pour l'élimination de la discrimination à l'égard des femmes).

Le Comité pour l'élimination de la discrimination à l'égard des femmes a fait évoluer la référence en matière d'égalité des genres d'une égalité de jure aux yeux de la loi et des règlements à une égalité de facto, en combattant les barrières systémiques à la participation sur un pied d'égalité des femmes et des filles dans toute leur diversité et dans toutes les sphères de la vie (IWRAP, sans date ; Shapiro, sans date). Le Comité reconnaît que la discrimination fondée sur le genre est inextricablement liée à d'autres facteurs tels que la race, l'origine ethnique, la religion ou la conviction, la santé, le statut, l'âge, la classe, la caste, l'orientation sexuelle et l'identité de genre (Campbell, 2015).

Le Pacte international relatif aux droits économiques, sociaux et culturels (PIDESC)

Le PIDESC exige une non-discrimination réelle dans la jouissance des droits sociaux et économiques, sans aucun type de discrimination fondée sur la race, la couleur, le sexe, la langue, la religion, les opinions politiques ou d'un autre type, l'origine nationale ou sociale, la fortune, la naissance ou toute autre situation (Bureau du Haut-Commissaire aux droits de l'homme, 1966).

La Convention sur l'élimination de toutes les formes de discrimination raciale

En vertu de la Convention sur l'élimination de toutes les formes de discrimination raciale, les États parties sont tenus de prendre des mesures pour garantir la pleine jouissance, dans des conditions d'égalité, des droits de la personne et des libertés fondamentales par les groupes marginalisés en raison de leur race, leur couleur, leur ascendance ou leur origine nationale ou ethnique et par les personnes y appartenant (Bureau du Haut-Commissaire aux droits de l'homme, 1965).

La Convention relative aux droits des personnes handicapées

La Convention relative aux droits des personnes handicapées fait évoluer l'approche du handicap d'un modèle médical à un modèle axé sur les droits de la personne, incluant les changements systémiques, juridiques, politiques, sociaux, et tout autre changement requis pour l'avènement de sociétés inclusives. La Convention fournit des mesures spécifiques pour que les États, en tant que porteurs de devoirs, éliminent les barrières et permettent aux personnes handicapées d'avoir accès à leurs droits et d'en jouir pleinement (Goldschmidt, 2017).

L'égalité réelle est fondée sur le principe selon lequel le droit à l'égalité associé aux droits de la personne doit tenir compte des personnes désavantagées, dénigrées, exclues ou ignorées. Il va de la responsabilité de tous les États, en tant que porteurs de devoirs, de protéger, promouvoir et garantir le respect des droits de la personne. Les approches fondées sur les droits de la personne pour l'élaboration de politiques, projets, programmes et systèmes incluent généralement les activités principales suivantes :

- Identification des droits de la personne et des instruments associés que l'initiative affecte ou qui affectent l'initiative.
- Autonomisation et accès à la justice des titulaires de droits afin de leur permettre de les revendiquer individuellement ou collectivement.
- Développement de la capacité des porteurs de devoir à protéger et à promouvoir les droits de la personne.
- Participation effective des groupes marginalisés.

Parmi les obligations spécifiques des États, citons l'élimination systémique des barrières à l'égalité des genres et de toutes les formes de discrimination qui contribuent à la marginalisation persistante. Cela inclut également la promotion active des pratiques inclusives, afin de créer un environnement dans lequel tout le monde peut participer et bénéficier pleinement des ressources et des opportunités sociales.¹⁰

L'égalité réelle peut être instaurée en mettant en œuvre des mesures spécifiques en vue d'un changement transformateur (ONU-Femmes, 2015). Les politiques, programmes et budgets efficacement inclusifs destinés à l'égalité entre les genres et la diversité sont des outils qui entraînent une véritable expansion stratégique de l'égalité réelle et garantissent l'accès des groupes marginalisés aux droits de la personne. Le concept d'« égalité réelle » pose des bases solides pour poursuivre le changement transformateur en réponse aux problématiques de l'inégalité et de l'exclusion tout au long du cycle de vie de l'IA et lors de l'élaboration de politiques en la matière. Dans la prochaine section sont décrites les approches sociotechniques applicables aux systèmes d'IA et au cycle de vie de l'IA en vue du succès de la mise en œuvre d'une politique d'IA transformatrice.



Protestations de différents groupes, principalement des femmes et des Amérindiennes au visage orné de peintures. Photo de Pascal Bernardon sur Unsplash.

Dans l'approche de l'égalité réelle, les politiques doivent traiter à la racine les causes des inégalités entre les genres afin de les corriger, et éliminer les barrières structurelles, comme le creusement géopolitique des inégalités, et les déséquilibres de pouvoir. Cette mission demande d'affronter des conjonctions de facteurs tels que l'âge, le sexe, l'identité, l'orientation sexuelle, le handicap, l'origine ethnique, la religion, la migration, la situation économique ou d'autres situations, afin de garantir que toutes les personnes, quelle que soit leur identité, sont incluses dans tous les aspects de la société et ont le pouvoir d'influencer son évolution. En définitive, ces politiques doivent être transformatrices.

Un changement transformateur

Le fait d'œuvrer pour un changement transformateur ou d'adopter une approche transformatrice des dynamiques de genre implique de traiter les causes d'inégalité à la racine, pour rendre les institutions et les relations sociales plus inclusives et équitables, en éliminant les barrières structurelles ou légales et en autonomisant les groupes marginalisés (UNRISD, 2017). Ces actions incluent la modification de la loi et les règlements, l'adaptation et l'amélioration des systèmes et des services, la redistribution des ressources et la rectification des normes, convictions (y compris les stéréotypes), comportements et pratiques nuisibles.

(UNICEF, 2021)

Une approche sociotechnique de l'IA

Dans le cadre des efforts déployés pour parvenir à l'égalité des genres et à la diversité dans l'IA, il convient de tenir compte de la totalité du cycle de vie de l'IA, d'un point de vue sociotechnique. L'approche sociotechnique des systèmes d'IA et de leur cycle de vie reconnaît que les performances et les impacts des systèmes d'IA découlent de la conception technique, de l'infrastructure et des dynamiques et incitations sociales. Les systèmes d'IA sont généralement décrits comme des systèmes fondés sur des machines ou comme un ensemble de techniques informatiques capable d'avoir un impact sur son environnement. Les définitions classiques des systèmes d'IA les décrivent comme des systèmes fondés sur des machines qui peuvent influencer leur environnement (OCDE, 2024b). Ces descriptions reconnaissent le rôle décisif que les systèmes d'IA ont ou auront sur l'environnement, les pratiques sociales, les normes, les droits et les structures (UNESCO, 2024c). Dans les politiques en matière d'IA, les approches basées sur un modèle adoptent généralement cette vision des systèmes d'IA et de leur cycle de vie.

Systèmes d'intelligence artificielle (IA)

Un système d'IA est un système fondé sur des machines qui, pour des objectifs explicites ou implicites, réalise des inférences, à partir des données qu'il reçoit, pour générer des sorties tels que des prédictions, du contenu, des recommandations ou des décisions pouvant influencer l'environnement physique ou virtuel. Après leur déploiement, les différents systèmes d'IA varient dans leurs niveaux d'autonomie et d'adaptabilité.

(OCDE, 2024b)

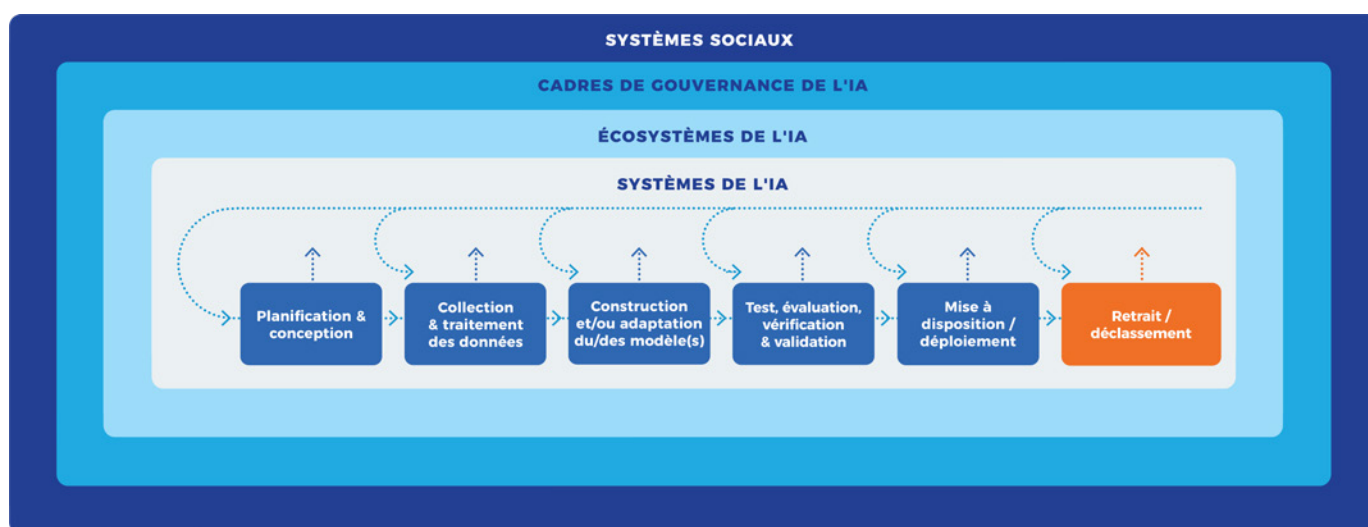
Les politiques en matière d'IA reposent souvent sur des approches centrées sur un modèle, dans un souci de diversité, d'équité et d'inclusion. Ces approches, telle que la suppression technique des biais, visent à corriger les inexactitudes dans les données provoquées par les préjugés sociaux et à améliorer les systèmes d'IA en vue de parvenir à une représentation juste, à un traitement équitable et à une absence de discrimination (Cachat-Rosset et Klarsfeld, 2023). Par exemple, en éliminant les biais dans les données menant à la sous-évaluation des femmes, il est possible de corriger les préjugés dans les notes des évaluations des performances. De même, le choix de meilleurs indicateurs indirects peut réduire la discrimination raciale dans les algorithmes sur le risque de récidive ou dans les modèles destinés à prévoir la récidive (la « caractéristique cible »). Par exemple, l'utilisation de « rapports d'arrestation » comme indicateur indirect pour prédire la récidive, reflète l'existence d'un profilage racial biaisé chez les forces de l'ordre (Benjamin, 2019). L'intégration d'ajustements mathématiques, une conception consciencieuse, la collecte de données variées et représentatives et la réalisation d'essais et d'un suivi collaboratifs permettraient de combattre l'inégalité des genres et la discrimination sociale (Hellman, 2023 ; Jora et al., 2022 ; Solaiman et al., à paraître). Pour autant, l'exactitude des données ne peut pas à elle seule garantir l'équité lorsque des biais systémiques sont reflétés dans la caractéristique cible (Joyce et al., 2021 ; Kong, 2022 ; Ovalle et al., 2023a). Si les approches centrées sur un modèle corrigent les inexactitudes dans les données, l'approche sociotechnique aborde quant à elle les causes fondamentales d'inégalité ancrées dans la dimension sociotechnique du cycle de vie du système d'IA (Iason, 2022). Cette approche met l'accent sur l'impact mutuel des éléments sociaux et techniques.



Personnes à un passage piéton. Photo de bee32 sur iStock.

Le cycle de vie de l'IA

Dans le cadre d'une compréhension sociotechnique de l'IA, la société et la technologie sont considérées comme un seul système cohérent composé d'une multiplicité de systèmes, et comme un processus d'itération. Cela inclut les normes, pratiques, matériels et infrastructures dans lesquels et par le biais desquels les systèmes technologiques fonctionnent. Sont également inclus des discours, des imaginaires, des écosystèmes et des intérêts géopolitiques et économiques, et l'IA en tant qu'objet (Ricaurte et Zasso, 2021). Tous ces éléments sociotechniques sont interconnectés tout au long du cycle de vie du système d'IA (Chen et Metcalf, 2024). Le cycle de vie du système d'IA décrit dans ce rapport fait appel à des conceptualisations existantes, auxquelles sont intégrés des systèmes sociotechniques supplémentaires, essentiels pour l'élaboration de politiques en matière d'égalité des genres et de diversité.¹¹



(Image adaptée, source OCDE, 2024a ; Ricaurte, 2024)

Les étapes du cycle de vie des systèmes d'IA

Planification et conception

Cette étape est consacrée à l'énoncé de la problématique. Les parties impliquées définissent les objectifs et la portée du système d'IA, ainsi que les principaux problèmes et opportunités auxquels ledit système devrait répondre.

Collecte des données et (pré)traitement

À cette étape, les équipes de recherche et de développement identifient les sources de données nécessaires à l'entraînement et à la validation des modèles d'IA. Ces sources peuvent inclure des bases de données, des API, des capteurs, des entrées de données manuelles. Il est important d'identifier les pratiques convenant le mieux à la collecte de données, afin de garantir l'inclusion de tous les éléments et processus nécessaires. Il faut ensuite procéder au prétraitement des données, incluant le nettoyage des données, le traitement des valeurs manquantes et leur mise au format correspondant aux modèles d'IA. D'autres processus incluent la normalisation (mise à l'échelle standard des données), la mise à l'échelle des caractéristiques ou feature scaling (ajustement de l'échelle des données) et l'ingénierie de caractéristiques ou feature engineering (création de nouvelles caractéristiques d'après des données existantes).

Élaboration de modèles

Cette étape consiste à modéliser les données acquises en utilisant les modèles d'IA adaptés, en fonction des caractéristiques de la problématique. Par exemple, il faudra utiliser des modèles basés sur la vue pour les données d'image et des modèles basés sur le langage pour les données textuelles.

Essais, vérification et validation

Les essais, l'évaluation, la vérification et la validation constituent des activités essentielles au cours du cycle de vie du système d'IA. Elles demandent l'utilisation de repères d'évaluation pour identifier les erreurs, biais ou limitations présents dans les modèles d'IA. Sont également incluses d'autres activités comme la validation de modèles à l'aide de techniques telles que les tests A/B (méthode de comparaison), visant à assurer l'exactitude et la fiabilité.

Déploiement, exploitation et suivi

Les responsables du déploiement du système d'IA définissent un calendrier pour la mise en fonctionnement du système. À ce stade, la collaboration avec les équipes informatiques et d'administration du système est essentielle pour intégrer l'IA dans les systèmes, bases de données et interfaces utilisateur existants. La mise en place de mécanismes de suivi et de connexion destinés à surveiller les performances du système d'IA et à identifier les problèmes sera également utile à l'étape suivante. Pendant la phase d'exploitation et de suivi, les responsables du déploiement assurent et contrôlent en permanence les performances du système d'IA.

Retrait et mise hors service

La décision de retirer ou de mettre hors service un système d'IA peut être prise à n'importe quel stade du cycle de vie du système.



Vue aérienne d'un nœud autoroutier au trafic dense composé d'un enchevêtrement de routes et de ponts. Photo de Denys Nevozhai sur Unsplash.

Opportunités d'intervention

Le traitement du cycle de vie des systèmes d'IA sous le prisme sociotechnique permet d'identifier des opportunités d'intervention en vue de la création de systèmes plus inclusifs et équitables. Le cycle de vie des systèmes d'IA peut être semé d'inégalités, d'iniquités et de différents types de conséquences négatives ou de préjudices. Bon nombre de ces désagréments découlent des injustices épistémiques, qui sont des situations dans lesquelles les expériences et les perspectives des personnes ou communautés sont niées ou discréditées à des étapes importantes du cycle de vie du système d'IA, en raison de la faible représentation et reconnaissance des divers types de savoirs et d'expertise (Cheong et al., 2021 ; Nihei, 2022). Pour assurer le succès des interventions, il est donc essentiel de comprendre et de maîtriser les éléments sociotechniques en jeu dans tous les contextes liés aux systèmes d'IA.

Transparence, interprétabilité et explicabilité dans les systèmes et processus d'IA

Dans un souci de transparence, les équipes de développement doivent documenter les processus de sélection des ensembles de données, des variables et des modèles utilisés dans le développement, l'entraînement, la validation et l'essai de l'IA. Elles sont tenues de détailler les mesures qui garantissent la qualité des données et des sorties, l'intervention humaine étant requise pour les sorties peu fiables.

La transparence implique aussi d'utiliser des méthodes de documentation pour détecter d'éventuels biais.

La vérification et la validation indépendante devraient être obligatoires pour les systèmes appelés à produire un impact important sur la vie et le bien-être des populations.

L'interprétabilité désigne le fait de fournir des informations claires sur les procédures suivies par les algorithmes, ce qui permet aux parties prenantes de comprendre le fonctionnement interne des systèmes d'IA.

L'explicabilité consiste à détailler les décisions spécifiques prises par les algorithmes. Cela s'avère essentiel dans le cadre des politiques publiques et dans les contextes dans lesquels les décisions des algorithmes peuvent injustement avantager certains groupes. Le fait de garantir aux groupes marginalisés le droit à des explications permet d'enrayer la perpétuation des inégalités. Il est essentiel de distinguer entre les explications authentiques et la rationalisation a posteriori pour assurer la responsabilisation et la confiance au sein des systèmes d'IA.

(Adapté d'après ACM, 2022)

Les systèmes sociotechniques qui interagissent avec le cycle de vie des systèmes d'IA incluent les écosystèmes de l'IA, les cadres de gouvernance de l'IA et les systèmes sociaux associés au contexte économique et au financement, ainsi que les aspects liés aux relations sociales, aux normes, aux valeurs, aux pratiques, aux institutions, aux infrastructures et à l'environnement. Les écosystèmes de l'IA incluent toutes les parties prenantes, participant à ou affectées par le cycle de vie du système d'IA. En vue d'une participation réelle, il est fondamental d'assurer l'implication des principales parties prenantes en tenant compte de leur position. L'implication multilatérale et interdisciplinaire des parties prenantes tout au long du processus de prise de décision évite par exemple que les modèles n'héritent ou n'amplifient des préjugés dangereux et des pratiques discriminatoires. Les cadres de gouvernance de l'IA incluent des cadres, des politiques, des lois et des réglementations à l'échelle nationale, internationale ou mondiale, qui façonnent l'utilisation, le développement, l'expérience et l'impact des systèmes d'IA: L'identification et le développement de cadres, de stratégies et de mécanismes applicables à tous les niveaux, et la résolution des failles pour assurer l'égalité des genres et la diversité sont des tâches importantes en vue de la transparence, de la responsabilisation et de la conformité avec les systèmes et les processus d'IA. Les systèmes sociaux comme le contexte économique et le financement influencent en grande mesure les possibilités de changement. L'identification et la lutte contre les barrières à l'égalité des genres et à la diversité, en assurant un accès égal aux ressources financières, au savoir et aux réseaux, s'avèrent décisives pour parvenir à une inclusion réelle. De même, il est important de travailler à transformer les normes et pratiques sociales et institutionnelles par le biais de la sensibilisation, de l'éducation, des politiques institutionnelles et d'autres outils. Le cadre de transformation de la politique d'IA en vue d'une égalité réelle sera introduit dans la prochaine section en fournissant une base globale pour les interventions liées aux politiques d'IA.

Le cadre pour la politique d'IA transformatrice

Le cadre pour la politique d'IA transformatrice repose sur une approche transformatrice visant à atteindre une égalité réelle dans la totalité du cycle de vie de l'IA et dans l'élaboration de politiques en la matière.¹² La politique d'IA transformatrice place les droits de la personne au cœur des écosystèmes de l'IA en supprimant les contraintes et en améliorant les capacités, afin de garantir la pleine jouissance des droits de la personne au sein des environnements technologiques.

Ce cadre a pour objet d'autonomiser activement les femmes et autres groupes marginalisés, de sorte à leur permettre de jouir pleinement de leurs droits. Cela passe par l'élimination de la discrimination directe et indirecte, et par le soutien de l'action positive pour promouvoir l'égalité réelle (Gurumurthy et Deepti, 2023). Le cadre propose des mesures pour assurer la promulgation de tout changement nécessaire à faire tomber les barrières institutionnelles perpétuant l'exclusion des femmes et autres groupes marginalisés, exclus du développement, de la prise de décision et de la gouvernance en matière d'IA. En reconnaissant l'exclusion structurelle et les inégalités dans la répartition du pouvoir et des ressources, il devient possible d'intervenir dans les modèles de prise de décision fondés sur l'IA et au niveau d'autres aspects de la conception, du développement et du déploiement de l'IA. De cette façon, les chances d'aboutir à des résultats équitables, et donc à une plus grande justice, augmentent.

La prochaine section expose dans les grandes lignes la manière dont un écosystème de transformation de l'IA peut promouvoir l'égalité des genres et la diversité.

Dimensions clés

Le cadre pour la politique d'IA transformatrice visant l'égalité entre les genres et la diversité dans l'IA s'articule autour de trois dimensions principales d'égalité réelle. Il adopte une approche systémique pour intégrer les droits de la personne (Bartoletti et Xenidis, 2023) et approfondit autour de ces trois volets, afin que les politiques et les lois en matière d'égalité des genres et de diversité dans l'innovation en IA puissent être porteuses de transformation. Cette section décrit les trois dimensions principales suivantes :

1. **Résolution des désavantages systémiques :**
le droit à l'inclusion
2. **Comblement du déficit démocratique :**
le droit de participation
3. **Lutte contre le déni de reconnaissance :**
le droit à la dignité

1. Résolution des désavantages systémiques : le droit à l'inclusion

Un cadre pour une politique transformatrice visant l'égalité des genres et la diversité peut aider à combattre les désavantages systémiques dans les écosystèmes de l'IA en promouvant activement le droit à l'inclusion. Les désavantages systémiques peuvent être considérés comme « une privation d'opportunités réelles d'aller au bout de ses choix » en raison d'une répartition injuste des ressources, biens et opportunités au sein d'une société donnée (Fredman, 2016). Les écosystèmes de l'IA sont susceptibles de perpétuer les désavantages systémiques par le biais de l'injustice de la répartition. Pensons par exemple à une attribution injuste des opportunités, en raison de la discrimination raciale des algorithmes de diffusion des publicités de logement, ou à une exclusion des services publics, causée par la discrimination dans l'automatisation des prestations sociales (Stauffer, 2023). Ces écosystèmes peuvent également perpétuer l'injustice dans la redistribution, avec par exemple la mainmise des grandes entreprises sur des données qui pourraient sans cela constituer un espace commun, dans le but de consolider leur monopole intellectuel (Rikap, 2022).

Une politique de transformation peut faire progresser activement le droit à l'inclusion en agissant sur les fronts suivants :

- Élimination des dommages causés par la discrimination directe et indirecte dans les écosystèmes de l'IA.
- Instauration d'obligations pour les équipes de développement et fournisseurs d'écosystèmes de l'IA, pour favoriser les processus de conception justes et inclusifs, et recours à l'action positive pour les femmes et autres groupes marginalisés (Bartoletti et Xenidis, 2023).

- Innovations publiques et privées destinées à bénéficier aux groupes marginalisés. Cela inclut par exemple des budgets consacrés aux innovations en IA d'assistance pour l'accessibilité et la sécurité sociale des populations racisées.
- Protection des espaces communs de données et de l'innovation ouverte en matière d'IA face à la captation privée, via l'utilisation de cadres adaptés en matière de propriété intellectuelle, basés sur le principe de science ouverte pour démocratiser les bénéfices des progrès de la science (Krishna, 2020).

2. Comblement du déficit démocratique : le droit de participation

Un cadre pour une politique transformatrice visant l'égalité des genres et la diversité se doit d'aborder le déficit démocratique dans les écosystèmes de l'IA par le biais du droit de participation et de créer un espace pour faire entendre la voix des groupes marginalisés. Cela permet aux porteurs de devoirs de faire respecter le principe de base de la démocratie représentative : quiconque est soumis à une politique a son mot à dire dans son élaboration (Pande, 2002).

Les biais, la discrimination et l'injustice des modèles d'IA creusent un déficit démocratique au sein des systèmes d'IA et découlent de l'exclusion intersectionnelle et des inégalités perpétuées dans les espaces dans lesquels l'IA est créée (West et al., 2019). Étant donné que les espaces d'innovation dans le domaine de l'IA sont pour la plupart aux mains des multinationales installées dans les pays de la minorité mondiale, les initiatives en matière de diversité et d'équité ont tendance à être réduites à une solution technique superficielle, comme l'ajout de données supplémentaires dans le réseau neuronal ou l'amélioration de la visibilité de nouvelles identités (Molnar, 2024 ; Niesen, 2016 ; Bureau du Haut-Commissariat des Nations Unies aux droits de l'homme, 2023). De plus, les systèmes d'IA présentant le plus haut degré de risque (identification biométrique, reconnaissance des émotions et surveillance automatisée) ont tendance à être déployés sur des groupes déjà marginalisés par nos systèmes : personnes en situation de déplacement (migrantes, réfugiées et autres personnes déplacées) ou personnes en conflit avec la loi, par exemple. Ces dernières sont bien souvent exclues de la protection de l'État garantie à la communauté politique nationale des populations. (#ProtectNotSurveil, 2024).

Deux types de mesures politiques deviennent primordiales :

- Instituer des mécanismes sociaux de contrôle et de retour d'information du public dans les processus privés de fixation des normes de l'IA
- Garantir le droit d'accès du public aux informations et à la justice et son droit de participation à la prise de décision sur les lois et règlements en matière d'IA (Smuha, 2021).

3. Lutte contre le déni de reconnaissance : le droit à la dignité

Le déni de reconnaissance (dénigrement, humiliation et incapacité à valoriser les personnes) constitue une violation du principe selon lequel tous les hommes et les femmes sont égaux (Fredman, 2016). La stigmatisation, les stéréotypes, l'humiliation et la violence « fondés sur le sexe, la race, le handicap, l'orientation sexuelle ou tout autre facteur social de localisation » (Fredman, 2016) représentent une attaque contre l'essence même de l'individualité, dans le sens d'une déclaration individuelle d'être un « être social, qui possède une rationalité, et [est] capable d'une action autonome » (Martineau et al., 2012).



Protestations d'un groupe citoyen principalement composé de femmes pour la défense de leurs droits à Heroica Puebla de Zaragoza, Mexique. Photo de Domingo Alvarez E sur Unsplash.

Les boucles homme-machine dans le développement et le déploiement de l'IA officialisent certaines formes de discrimination, et agissent de manière à réifier les personnes et à augmenter les formes existantes d'inégalités sociales. Ce phénomène est qualifié d'oppression algorithmique par les auteurs féministes (Benjamin, 2019 ; Eubanks, 2017 ; Noble et Roberts, 2019 ; West, 2020). Cette oppression algorithmique se manifeste non seulement dans le refus injuste de laisser les groupes marginalisés accéder aux opportunités (abordé dans la dimension 1) mais aussi comme une injustice de représentation (par exemple la sexualisation des femmes racisées par les générateurs d'image en IA ou l'amplification du sexisme et de la misogynie par les algorithmes des réseaux sociaux).

Le cadre pour la politique d'IA transformatrice visant l'égalité des genres et la diversité se doit de corriger ce déni de reconnaissance et de saluer la lutte morale des groupes marginalisés pour affirmer leur individualité, en protégeant et en promouvant leur droit à la dignité. La dignité se trouve à la base de toute société réellement ouverte à la diversité. Le droit à la dignité reconnu dans la jurisprudence en matière de droits de la personne exige aux États de reconnaître le besoin d'une action concertée —et d'agir en ce sens— contre tout processus de discrédit stigmatisant une personne ou un groupe, pour quelque motif que ce soit, et contre les mécanismes institutionnels qui le perpétuent.

Face à ce constat, il est nécessaire de mettre en place une initiative autour de deux axes :

- Instauration d'une obligation positive pour les fournisseurs de système d'IA et les personnes qui l'utilisent, visant à empêcher la discrimination algorithmique (directe et indirecte).
- Mise en place d'une transparence radicale et de garanties de responsabilisation dans les écosystèmes de l'IA par le biais d'une obligation de vérifier au préalable les algorithmes, et de surveiller que les secrets d'affaires et le droit des brevets ne fonctionnent pas au détriment des mesures de responsabilisation destinées à éliminer le déni de reconnaissance dans les systèmes d'IA.



Mine creusée dans une montagne. Photo de Vlad Chețan sur Pexels.

PARTIE III.

Pratiques prometteuses et recommandations



Femme dans un champ d'herbes hautes et sèches, les bras levés et croisés au-dessus de la tête, vêtue d'habits aux motifs colorés et aux cheveux tressés, se détachant sur un ciel nuageux. Photo by Ian Kiragu sur Unsplash.

Pratiques prometteuses pour instaurer une égalité réelle en l'IA

Cette section rassemble un ensemble de pratiques prometteuses glanées dans le monde entier, qui démontrent comment progresser dans le sens des dimensions clés de l'égalité réelle ébauchées dans le cadre de la politique d'IA transformatrice (lutte contre les désavantages systémiques, comblement du déficit démocratique, correction du déni de reconnaissance). Les apprentissages tirés de chaque pratique sont résumés dans le paragraphe Perspectives stratégiques.

Les pratiques englobent toute une série de mesures qui répondent efficacement aux objectifs d'égalité des genres et de diversité dans les écosystèmes de l'IA. Elles ont été créées par différentes parties prenantes (organismes multi-latéraux, organisations de la société civile, agences publiques au niveau national et organisations en réseau) et identifiées au cours des consultations régionales ou de groupes spécifiques. Elles incluent ou sont dirigées par différents groupes marginalisés, ou œuvrent à promouvoir leurs droits et leur participation réelle, que ce soit au niveau local, régional ou international. Ces informations sont fournies à titre indicatif et non exhaustif, dans l'optique d'offrir aux responsables des politiques et autres parties prenantes des exemples concrets de mesures à reproduire, à soutenir ou à adapter à leurs contextes.

Les pratiques prometteuses sont décrites comme des exemples (sans toutefois s'y limiter) correspondant aux catégories suivantes :

- **Ressources pour le renforcement des capacités et l'éducation du public**
Conseils de spécialistes et ressources pour les responsables des politiques et le public.
- **Conception technologique inclusive et pratiques d'innovation démocratiques**
Modèles technologiques qui intègrent efficacement les objectifs en matière d'égalité des genres et de diversité à différentes étapes du cycle de vie de l'IA, développés par des organisations à but non lucratif et des agences publiques.
- **Mesures de responsabilisation**
Des indices, des initiatives de suivi et des mesures de sensibilisation destinées à promouvoir la responsabilisation des fournisseurs et des déployeurs d'IA à l'égard du respect, de la protection et de la promotion des droits de la personne.
- **Initiatives d'inclusion et de participation réelles**
Initiatives promouvant une participation active et réelle, et la cogouvernance dans les écosystèmes de données de l'IA.



Photo d'une femme tenant une note sur laquelle figure le message « Nous sommes égaux ». Photo de Jacob Wackerhausen sur iStock.

Ressources pour le renforcement des capacités et l'éducation du public

Dialogue mondial de l'UNESCO

Initiative :

[Intelligence artificielle et égalité des genres : principales conclusions du dialogue mondial de l'UNESCO](#)

Pays/région :

international

Organisation :

[Organisation des Nations Unies pour l'éducation, la science et la culture \(UNESCO\)](#)

Description

Le Dialogue mondial de l'UNESCO a pour objet d'intégrer l'égalité des genres aux politiques et aux pratiques en matière d'IA. Ce rapport examine et identifie les lacunes dans l'intégration de l'égalité des genres aux principes qui régissent l'IA. Il offre des recommandations, ainsi qu'un plan d'action pour orienter chaque groupe de parties prenantes sur leur mise en pratique. Le plan d'action s'articule autour de cinq domaines : 1) sensibilisation ; 2) importance de maintenir l'égalité des genres comme une priorité ; 3) développement d'une coalition ; 4) renforcement des capacités et financement ; et 5) recherche, suivi et financement. Cette ressource a été créée en concertation avec les spécialistes en IA mondiale et égalité des genres issus de différents groupes de parties prenantes et financée via un partenariat avec l'agence allemande de coopération internationale (GIZ).

Comment l'initiative fait progresser l'égalité réelle

Cette initiative promeut les interventions en faveur d'une égalité des genres transformatrices. En défendant l'introduction du prisme de l'égalité des genres dans le développement et la gouvernance de l'IA, elle travaille à éliminer les barrières qui perpétuent les inégalités entre les genres. En recommandant des manières d'autonomiser les femmes et d'encourager leur participation, elle traite à la racine les causes des désavantages systémiques. En permettant aux groupes de défense de l'égalité des genres d'identifier des opportunités et de développer des compétences dans une visée d'émancipation, en se concentrant sur l'augmentation des capacités, de la sensibilisation et de la maîtrise de l'intelligence artificielle, elle contribue au comblement du déficit démocratique. Elle cherche à corriger le déni de reconnaissance en combattant les préjugés et en promouvant un discours plus inclusif sur le développement et le déploiement de l'IA.

Perspectives stratégiques

■ Changer de discours, pour passer des « comportements personnels » à un changement du système

Axé sur les impératifs liés à l'égalité des genres, le rapport recommande au secteur de l'IA et aux responsables des politiques de changer d'approche, en délaissant l'attention particulière aux personnes pour se concentrer sur un changement du système. Cette approche leur permettra d'adhérer avec succès aux principes, aux cadres et aux recommandations en matière de transformation des dynamiques de genre.

■ Orienter sur la manière d'appliquer les principes, cadres et recommandations en matière de transformation des dynamiques de genre

Le rapport souligne l'absence de lignes directrices pour mettre en pratique des principes comme l'équité et la transparence. Les responsables des politiques devraient s'appuyer, entre autres, sur ces ressources, pour refléter les principes en action. Ils et elles devraient également élaborer des conseils et veiller à la création d'écosystèmes de l'IA capables de transformer les dynamiques de genre.

Data Justice Policy Brief

Initiative :

[Data Justice Policy Brief](#)

Pays/région :

international

Organisation :

[Partenariat mondial sur l'intelligence artificielle \(PMIA\)](#)

Description

Le document Data Justice Policy Brief a pour objet d'intégrer des principes de justice des données aux politiques et aux pratiques. Il pose un cadre complet à l'attention des responsables des politiques, afin de garantir un développement et un déploiement justes, transparents et inclusifs pour l'IA. Le rapport insiste sur l'importance de relever à la fois les défis sociaux et les défis économiques associés à la prolifération des systèmes et des processus d'IA. Il offre des recommandations pour élaborer des politiques réduisant les biais, promouvant la responsabilisation et garantissant que les bénéfices de l'IA soient répartis équitablement à travers tous les segments de la société. Avec comme objectif ultime la défense d'un panorama de l'IA plus juste et éthique.

Comment l'initiative fait progresser l'égalité réelle

Ce rapport du PMIA aborde les désavantages systémiques en promouvant des politiques d'IA justes et transparentes, fondées sur des principes d'inclusion, d'équité et de justice économique, et de réparation des préjudices. Il comble le déficit démocratique en défendant des processus d'élaboration de politiques inclusifs, impliquant différentes parties prenantes, pour une meilleure représentation et une plus grande responsabilisation. En outre, en veillant à ce que les communautés marginalisées soient impliquées dès la conception et en promouvant la transparence et la responsabilisation dans les pratiques et systèmes en matière de données et d'IA, elle contribue au respect du droit à la dignité et à la correction du déni de reconnaissance.

Perspectives stratégiques

■ **Participation démocratique des communautés affectées, dès la conception**

La justice des données exige aux responsables des politiques d'identifier l'ensemble complet de parties prenantes susceptibles d'être affectées par la collecte et l'utilisation des données, et par les activités fondées sur les données. Les personnes concernées individuelles et collectives, ainsi que les principaux générateurs de données, constituent des parties prenantes essentielles. Leur participation doit être intégrée démocratiquement dans la conception, le développement et le déploiement de systèmes à usage intensif de données, y compris les systèmes et processus d'IA.

■ **Promouvoir la transparence dans les données, ainsi que dans les pratiques et systèmes d'IA**

Il est recommandé dans le rapport que les personnes intervenant dans la collecte et l'utilisation des données, et dans les innovations fondées sur les données, soient tenues de diffuser des informations au public sur le type de données recueillies, l'utilisation qui en est faite, en informant notamment des entrées de données et des algorithmes. Il leur est également recommandé de fournir ces informations directement aux personnes et aux communautés affectées.

AI & Equality Human Rights Toolbox

Initiative :

[AI & Equality Human Rights Toolbox](#)

Pays/région :

international

Organisation :

[Women at the Table](#) et l'École Polytechnique

Fédérale de Lausanne (EPFL), en consultation avec le [HCDH](#)

Description

AI & Equality Human Rights Toolbox [outils pour les droits de la personne et l'égalité dans l'IA] est une plateforme éducative conçue pour catalyser les débats éclairés et la collaboration, dans le but de créer une approche de l'IA fondée sur les droits de la personne. Elle offre une méthodologie, des ateliers et des ressources, ainsi qu'une sélection de formations en ligne aux technologues, aux professionnel-le-s de l'IA, aux spécialistes en sciences sociales, aux activistes, aux responsables des politiques et au public. Cette initiative est centrée sur le contexte et l'objectif, sur la collaboration multidisciplinaire et sur la mobilisation de la communauté pour garantir l'équité et l'inclusion dans les innovations en matière d'IA. Elle a été lancée par Women at the Table en partenariat avec le Bureau du Haut-Commissariat des Nations Unies aux droits de l'homme (HCDH) et l'EPFL. Ses cours en ligne sont proposés sur le portail de formation du Sorbonne Center for AI (SCAI) en partenariat avec l'Université de La Sorbonne.

Comment l'initiative fait progresser l'égalité réelle

Promouvant une approche de l'IA fondée sur les droits de la personne, l'initiative aborde les désavantages systémiques en veillant à inclure les groupes marginalisés dans le développement de l'IA, pour une réduction des biais et une promotion de l'équité. Elle favorise les débats éclairés et la collaboration entre les technologues, responsables des décisions politiques et le public, de sorte à inclure les communautés marginalisées dans l'élaboration des politiques en matière d'IA. Tout cela contribue au comblement du déficit démocratique. Enfin, par le biais de l'éducation et de la collaboration interdisciplinaire, l'initiative remet en cause le mythe de la neutralité de l'IA, en contribuant à créer un environnement dans lequel les technologues, les responsables des politiques et les personnes affectées par les systèmes d'IA partagent un vocabulaire commun, qui reconnaît et intègre tout un éventail de perspectives sociales.

Perspectives stratégiques

■ **L'adoption d'une approche fondée sur les droits de la personne**

L'adoption d'une approche fondée sur les droits de la personne tout au long du cycle de vie de l'IA garantit que les systèmes d'IA sont développés, déployés et vérifiés dans le respect de ces droits, en favorisant les débats éclairés et un vocabulaire en commun sur l'utilisation éthique des données. En établissant les droits de la personne en tant que base de référence, l'accent est mis sur la responsabilité des équipes de développement de l'IA concernant leurs devoirs de transparence et d'explicabilité, ainsi que sur le « droit de savoir » des personnes affectées par les systèmes d'IA.

■ **Une plateforme pour la collaboration multidisciplinaire et l'implication de la communauté**

L'initiative AI & Equality Human Rights Toolbox a pour objet la création d'une communauté internationale de groupes de recherche en IA, spécialistes en sciences sociales et en mégadonnées et activistes qui collaborent et échangent autour la manière dont les systèmes d'IA peuvent contribuer au bien commun et prévenir les dommages. Cette initiative favorise et promeut l'implication de la communauté et le développement de capacités par l'intermédiaire d'une plateforme offrant des ressources comme des cours en ligne et des opportunités de débat. En soutenant des espaces de collaboration multidisciplinaire et d'implication de la communauté basés sur les droits de la personne, les responsables des politiques peuvent aider à développer la maîtrise, la capacité et la sensibilisation à l'IA.

Éclaireurs autochtones en IA

Initiative :

[Éclaireurs autochtones en IA](#)

Pays/région :

Canada

Organisation :

[Mila - Institut québécois d'IA](#) et [Indspire](#)

Description

Éclaireurs autochtones en IA est un programme professionnel transformateur, conçu pour encourager les talents autochtones à apprendre, à développer et à diriger l'évolution de l'IA. Ancré dans la communauté et la cosmovision autochtone, ce programme tend une passerelle entre la vision de ces peuples et les systèmes et processus d'IA. Sous la houlette du Mila - Institut québécois d'IA en partenariat avec Indspire, le programme introduit les personnes et les groupes participants aux aspects techniques et non techniques, en abordant des sujets comme l'IA responsable et la gouvernance éthique des données. Elles reçoivent une bourse et leurs frais de voyage, d'hébergement et de repas sont couverts, pour ouvrir le programme à tous et à toutes, loin de toute contrainte financière. L'initiative Éclaireurs est une source de nouvelles opportunités pour les personnes autochtones innovantes, invitées à façonner l'avenir de l'IA et à créer des solutions efficaces pour leur communauté.

Comment l'initiative fait progresser l'égalité réelle

Les barrières systémiques, et notamment les difficultés financières, l'accès limité aux opportunités de formation et les infrastructures numériques non adaptées compromettent à la fois le développement des compétences et la sensibilisation aux opportunités de carrière dans l'IA. Les Éclaireurs autochtones en IA font progresser l'égalité réelle par le biais d'une initiative ciblée qui élimine les barrières à l'innovation démocratique et à la participation effective à la gouvernance de l'IA. Ce programme vise à favoriser l'inclusion précoce et efficace dans les écosystèmes de l'IA, en encourageant le talent des membres des communautés autochtones et leur offrant la possibilité de façonner les systèmes et processus d'IA dès l'origine, de sorte à combattre les exclusions épistémiques. En comblant proactivement ces lacunes, le programme œuvre en faveur d'un avenir plus équitable et respectueux de la diversité. Cette approche s'attaque aux désavantages systémiques ancrés dans les inégalités historiques et renforce la capacité de participation effective, comblant ainsi le déficit démocratique.

Perspectives stratégiques

■ Des solutions d'IA axées sur la communauté

En plus de créer des parcours professionnels qui améliorent significativement la maîtrise de l'IA chez les talents autochtones, le programme Éclaireurs développe des solutions d'IA adaptées aux besoins uniques de leurs communautés. Éclaireurs établit un nouveau standard d'intégration des perspectives des populations autochtones tout au long du processus de développement de l'IA, de la conception à la gouvernance éthique des données. Cette initiative introduit une approche transformatrice de la formation à l'IA tout en encourageant les communautés autochtones à façonner l'avenir de l'IA d'une manière fidèle à leurs valeurs et à leurs priorités.

■ Une approche globale qui privilégie le savoir autochtone

L'environnement d'apprentissage du programme Éclaireurs s'inscrit dans les perspectives des populations autochtones. Il repose sur les quatre piliers de la sagesse autochtone : responsabilité, respect, réciprocité et pertinence. L'approche profondément holistique d'Éclaireurs met l'accent sur le bien-être des personnes et des groupes participants, en donnant accès aux Anciens et aux gardiens du savoir, ainsi qu'à des séances consacrées aux enseignements culturels. En encourageant les talents autochtones à relever les défis qu'ils sont les mieux placés pour comprendre, le programme favorise le développement de solutions d'IA qui abordent les problématiques sociales importantes pour leurs communautés et qu'elles définissent elles-mêmes.



Conception technologique inclusive et pratiques d'innovation démocratiques

La méthodologie *Design From the Margins* (DFM)

Initiative:

[La méthodologie Design From the Margins \(DFM\)](#)

Organisation:

[The De|Center](#)

Pays/région :

international

Description

Design From the Margins (DFM) est une méthodologie développée avec en ligne de mire la création d'une technologie équitable, en se concentrant sur les expériences vécues par les groupes marginalisés ou décentrés, tout au long du cycle de vie de l'IA. Cette méthodologie est née au The De|Center, un centre interdisciplinaire consacré à la recherche communautaire, à la responsabilisation des entreprises et à l'intervention dans la conception technologique pour réduire les dommages causés par les technologies existantes et émergentes. Intégrant la recherche axée sur les communautés et la réduction des dommages, la méthodologie est profondément impliquée auprès des groupes marginalisés (utilisateurs décentrés), c'est-à-dire les personnes les plus exposées aux conséquences négatives de la technologie. Parmi les succès de cette approche, citons les améliorations réalisées sur des plateformes comme Grindr, Signal et WhatsApp. Cette initiative est financée par des subventions et des dons.

Comment l'initiative fait progresser l'égalité réelle

Fortement impliquée dans la réduction des dommages, DFM œuvre à une transformation des outils technologiques, en évitant qu'ils soient utilisés pour perpétuer une oppression systémique. À ce titre, cette approche combat les désavantages systémiques. Elle implique une recherche communautaire approfondie, et une collaboration et des interventions intégrant les principes de sécurité, de dignité et de droits des groupes marginalisés. DFM comble le déficit démocratique et permet de garantir que les progrès technologiques profitent à tous et à toutes, et tout particulièrement aux personnes victimes d'injustice. Elle reconnaît que lorsque la technologie est conçue pour les membres de la société les plus affectés par les injustices systémiques, elle devient meilleure pour tout le monde. Cette approche combat également le déni de reconnaissance : les voix marginalisées se font entendre et sont placées au cœur de la méthodologie et du travail.

Perspectives stratégiques

■ Créer ou soutenir un fonds d'entraide

Grâce aux fonds d'entraide, les groupes marginalisés peuvent avoir accès à un soutien financier sans condition sous forme d'aide en espèces ou en nature. Les entreprises qui bénéficient de DFM sont encouragées à apporter des fonds. En créant ou soutenant des initiatives de ce type, les responsables des politiques peuvent permettre aux groupes marginalisés de diriger le développement auquel ils aspirent.

■ Transformer les préjudices et les besoins identifiés en changements applicables

La méthodologie DFM permet d'intervenir dans le domaine technologique en transformant les injustices et dommages constatés en changements à introduire dans la technologie et les pratiques d'IA. Cela concerne aussi d'autres besoins exprimés par les groupes marginalisés. Les responsables des politiques peuvent soutenir des initiatives similaires, qui aboutissent à des changements concrets sur la base des expériences vécues par les personnes les plus affectées.

Le réseau de recherche féministe sur l'IA (f<A+i>r)

Initiative:

[The f<A+i>r Network](#)

Organisation:

[<A+> Alliance for Inclusive Algorithms](#)

Pays/région :

international

Description

Le réseau de recherche féministe sur l'IA (f<A+i>r) a pour objet de créer des technologies d'IA inclusives, en combattant les biais systémiques sexistes, racistes et les préjugés intersectionnels. Son répertoire mondial, exclusivement sur invitation, comprend environ 100 universitaires et activistes féministes du domaine de l'IA qui partagent des recherches multidisciplinaires et des innovations en IA féministes, dans le but d'introduire un changement transformateur. f<A+i>r encourage le partage de connaissances Sud-Sud et Nord-Sud, en diffusant et mettant en avant les innovations féministes à travers le monde. Le réseau f<A+i>r network a été créé en 2020 par l' <A+> Alliance for Inclusive Algorithms, avec le soutien du Centre de recherches pour le développement international du Canada (CRDI), la Fondation Oak et le département fédéral suisse des affaires étrangères (DFAE), section droits des femmes/droits humains.

Comment l'initiative fait progresser l'égalité réelle

L'initiative f<A+i>r poursuit l'objectif d'aborder les désavantages systémiques en luttant activement contre les biais sexistes et racistes, ainsi que les préjugés intersectionnels dans les technologies de l'IA. Elle cherche à prévenir les dommages futurs à l'aide de nouvelles données, algorithmes, modèles politiques et systèmes pouvant faire l'objet de recherches et d'essais, en vue d'introduire un changement transformateur. Elle comble le déficit démocratique en favorisant l'inclusion et en soutenant le développement des capacités des équipes d'innovation féministes issues de la majorité mondiale. Cette initiative lutte contre le déni de reconnaissance en produisant des modèles efficaces, innovants, interdisciplinaires qui exploitent les technologies émergentes, en corrigeant les biais et les obstacles menaçant dans la vie réelle les droits, la représentation et l'égalité des femmes et autres groupes marginalisés.

Perspectives stratégiques

■ La promotion des systèmes de transformation de l'IA

En finançant et en soutenant par d'autres moyens le développement et la mise en place de systèmes et processus d'IA intégrant activement les principes techno-féministes et corrigeant les biais systémiques intersectionnels et fondés sur le genre (via la création de nouveaux ensembles de données, algorithmes et modèles), les responsables des politiques soutiennent l'égalité des genres et la diversité dans l'IA.

■ Collaboration et échange des connaissances

Le réseau de spécialistes de la technologie féministes provenant de la majorité mondiale contribue au développement de solutions d'IA adaptées à chaque contexte, qui répondent aux besoins et aux enjeux régionaux. Avec à l'arrivée, des politiques plus inclusives et équitables pour l'IA à travers le monde.

Initiative Fixing the bAIs

Initiative:

[Fixing the bAIs: Using AI to Correct Gender Bias in AI](#)

Organisation:

MullenLowe MENA for Aurora 50

Pays/région :

Émirats Arabes Unis/Moyen-Orient et Afrique du Nord

Description

La campagne Fixing the bAIs vise à combattre et à réduire les biais sexistes dans les systèmes d'IA en entraînant les ensembles de données via l'IA. Les femmes sont largement sous-représentées dans les ensembles de données utilisés pour entraîner l'IA. En utilisant des outils d'IA de génération d'image comme Midjourney, Dalle et Stable Diffusion et en faisant appel à des invites textuelles, la campagne corrige les biais de représentation dans l'entraînement des pools de données. Pour encourager leur diffusion, ces images sont exemptes de droits et de redevances, de sorte à alimenter les différents ensembles de données utilisés pour entraîner l'IA. Cette initiative encourage également la collaboration et le dialogue entre toutes les parties afin de promouvoir des bonnes pratiques et des solutions innovantes pour la réduction des biais dans l'IA.

Comment l'initiative fait progresser l'égalité réelle

La campagne Fixing the bAIs s'attaque aux désavantages systémiques en cherchant à réduire la discrimination et en promouvant un traitement équitable de toutes les catégories démographiques via la correction des biais dans les systèmes d'IA. Elle contribue au comblement du déficit démocratique en améliorant la transparence et la responsabilisation dans le développement de l'IA, et en garantissant la participation des communautés marginalisées. La campagne lutte également contre le déni de reconnaissance via la conception intentionnelle de systèmes d'IA non sexistes : en fournissant des ensembles de données d'entraînement sous licence ouverte présentant une diversité de genres, en remettant en question les rôles et les attentes de genre et en veillant à une meilleure représentation et à la reconnaissance des contributions des groupes marginalisés.

Perspectives stratégiques

■ L'imposition d'ensembles de données diversifiés et d'un entraînement intentionnel en matière de genre

Les responsables des politiques peuvent réduire les préjugés dans les systèmes d'IA en imposant l'utilisation d'ensembles de données diversifiés et de processus d'entraînement intentionnels en matière de genre. En intégrant des pratiques respectueuses de la diversité et de l'inclusivité, en veillant à ce que les ensembles de données d'entraînement représentent différentes catégories démographiques, les systèmes d'IA reflèteront mieux et répondront mieux aux besoins de tous les groupes sociaux.

■ Exigences sur la diffusion publique

Les responsables des politiques peuvent assurer l'amélioration quotidienne et la confiance du public dans les systèmes d'IA en adoptant des réglementations exigeant aux organisations de diffuser publiquement leurs efforts et leurs progrès en matière d'atténuation des biais de l'IA. Des audits réguliers et des rapports publics sur les biais dans les systèmes d'IA, accompagnés de mesures correctives, favoriseraient la transparence et la responsabilisation dans le développement de l'IA.

Indigenous Jobs Map

Initiative:

[Indigenous Jobs Map \(Carte de l'emploi autochtone\)](#)

Pays/région :

Australie

Organisation:

[Australia National Research Agency
\(Commonwealth Scientific and Industrial
Research Organisation: CSIRO\)](#)

Description

L'Indigenous Jobs Map développée par l'organisation CSIRO offre un aperçu complet des perspectives d'emploi dans les communautés autochtones d'Australie. Il s'agit de soutenir le travail et le développement économique autochtones en cartographiant et visualisant en temps réel les données sur l'emploi. L'Indigenous Jobs Map est un outil d'analyse de données mis au point dans le but d'aider l'agence de la recherche australienne à appliquer efficacement son Plan d'action pour la réconciliation, de sorte à améliorer la situation professionnelle des Aborigènes et Insulaires du détroit de Torrès sur le marché du travail national. Cette initiative a pour objet d'offrir plus de transparence et d'informations sur l'emploi, permettant ainsi aux personnes à la recherche d'un emploi et aux entreprises de prendre de meilleures décisions et de mieux s'organiser.

Comment l'initiative fait progresser l'égalité réelle

L'Indigenous Jobs Map favorise l'accès à l'emploi des communautés autochtones, en cartographiant en temps réel les données sur l'emploi, en promouvant le développement économique et en réduisant les disparités en matière de travail. Tout cela permet de combattre les désavantages systémiques. En fournissant des informations transparentes et accessibles sur l'emploi, cette initiative garantit l'inclusion de la voix et des besoins des autochtones dans les processus de prise de décision, en introduisant plus d'inclusivité dans l'organisation et les politiques. Elle contribue donc au comblement du déficit démocratique. La carte met en avant les contributions et tout le potentiel des communautés autochtones, en combattant les stéréotypes et en veillant à la prise en compte et à la valorisation de leurs compétences et de leurs opportunités. Cela permet de corriger le déni de reconnaissance.

Perspectives stratégiques

■ Action positive ciblée fondée sur les données pour la création de stratégies d'emploi

L'Indigenous Jobs Map utilise des données détaillées pour créer des stratégies d'emploi adaptées aux communautés autochtones, aidant ainsi les responsables des politiques à combattre les disparités dans l'emploi et à promouvoir la croissance économique. Cette initiative souligne l'importance d'une approche axée sur l'égalité réelle dans l'IA, en montrant comment l'analyse de données peut permettre aux responsables de la prise de décision d'orienter le ciblage de l'action positive.

■ Amélioration de la transparence et de l'accès

Cette initiative fournit une plateforme transparente et accessible pour visualiser les opportunités d'emploi. Les responsables des politiques peuvent utiliser des outils de ce type pour faciliter l'accès des communautés autochtones à l'information sur l'emploi, ce qui facilite la prise de décisions éclairées et l'organisation. Cette conception technologique inclusive est conçue de manière à combler les lacunes informatives et à favoriser la présence des Autochtones sur le marché du travail.



Mesures de responsabilisation

L'indice mondial sur l'IA responsable

Initiative:

[The Global Index on Responsible AI](#)

Organisation:

[Global Center on AI Governance \(GCG\)](#)

Pays/région :

international

Description

L'indice mondial sur l'IA responsable est une initiative visant à évaluer et à promouvoir des pratiques responsables en matière d'IA à travers le monde. Il offre une évaluation complète des politiques et pratiques d'IA des différents pays, du point de vue de l'inclusivité, de la responsabilisation et de l'éthique. Servant d'outil multidimensionnel de référence, il fournit des données représentatives aux responsables des politiques, aux équipes de recherche et aux journalistes, pour suivre et mesurer les progrès des initiatives des différents pays allant dans le sens de la jouissance et de la garantie des droits de la personne dans l'IA. Cet indice est élaboré par le GCG avec le soutien du Centre de recherches pour le développement international du Canada (CRDI), le gouvernement canadien et l'Agence américaine pour le développement international (USAID).

Comment l'initiative fait progresser l'égalité réelle

Cet indice encourage les pays à adopter des politiques qui réduisent les biais et la discrimination dans les systèmes d'IA, pour veiller à ce que les groupes marginalisés tirent des bénéfices des technologies de l'IA, en combattant les désavantages systémiques. À l'échelle des pays, il mesure l'engagement et les pratiques qui soutiennent une IA promouvant l'égalité des genres et l'autonomisation des femmes au lieu de les compromettre. Contribuant au comblement du déficit démocratique, cet indice a été développé en consultation avec les parties affectées et est fondé sur les principes d'accessibilité et d'ouverture, d'inclusion et de participation, et sur la volonté de refléter équitablement les contextes et les réalités locales. Corrigant le déni de reconnaissance, il met en avant les bonnes pratiques pour garantir la représentation équitable de tous les groupes sociaux dans le développement et le déploiement de l'IA, en combattant les biais et en promouvant l'inclusion.

L'IA responsable

L'IA responsable est définie par l'indice mondial sur l'IA responsable comme la conception, le développement, le déploiement et la gouvernance de l'IA d'une manière qui respecte et protège les droits de la personne et qui soutient les principes éthiques à chaque étape du cycle de vie de l'IA et de sa chaîne de valeur. Cela exige à tous les membres de l'écosystème de l'IA d'assumer la responsabilité des impacts humains, sociaux et environnementaux de leurs décisions. La conception, le déploiement et la gouvernance responsables de l'IA sont proportionnels à la finalité poursuivie et répondent aux besoins technologiques des individus et des sociétés que l'IA est appelée à servir.

(Adapté de l'indice mondial sur l'IA responsable, 2024)

Perspectives stratégiques

■ Référence pour la réglementation

Cette initiative fournit des points de référence qui tiennent compte des responsabilités des parties prenantes tout au long du cycle de vie et à travers tout l'écosystème de l'IA. Les responsables des politiques peuvent s'y reporter pour promulguer des réglementations imposant la transparence dans les opérations d'IA et des mécanismes de responsabilisation clairs, visant à garantir une utilisation responsable des systèmes d'IA, qui ne perpétue pas les préjugés.

■ Coopération internationale via un suivi concerté

Cette initiative promeut la coopération internationale via un suivi concerté. Les responsables des politiques peuvent renforcer et soutenir le rôle des communautés mondiales dans le suivi concerté et la responsabilisation des membres de l'écosystème de l'IA. La coopération internationale autour de l'IA responsable constitue un engagement partagé entre les différents pays du monde.



Plan serré sur deux mains entrelacées, dont l'une porte une bague. Photo de Nathan Dumlao sur Unsplash.

The Migration and Technology Monitor

Initiative:

[The Migration and Technology Monitor](#)

Organisation:

[Refugee Law Lab](#)

Pays/région :

international

Description

The Migration and Technology Monitor (MTM) est un collectif de journalistes, universités et communautés internationales qui travaillent sur le terrain pour partager la réalité des personnes migrantes affectées par l'évolution des technologies frontalières. Cette communauté est développée par et pour les personnes en situation de déplacement. Elle a pour mission d'observer les technologies de surveillance, l'automatisation et l'utilisation de l'IA dans le contrôle et le suivi des personnes en déplacement, ainsi que dans les décisions prises à leur sujet. The MTM soutient les communautés migrantes en organisant des rencontres propices à développer une narration sur l'impact de la surveillance frontalière. Cette initiative vise à partager les connaissances et à promouvoir le travail participatif, en associant une analyse rigoureuse aux politiques et à la sensibilisation, afin de faire réagir le public, les médias et les responsables des politiques.

Comment l'initiative fait progresser l'égalité réelle

The MTM lutte contre les désavantages systémiques en incluant les personnes en situation de déplacement comme les personnes réfugiées dans la conception, le développement et le déploiement des systèmes d'IA. Elle atténue les préjudices systémiques en mettant en avant les expériences vécues et en dénonçant les pratiques discriminatoires aux frontières et dans les situations de crise humanitaire. En impliquant les personnes migrantes, bien souvent exclues des processus démocratiques, dans les conversations politiques sur la technologie, The MTM lutte contre le déficit démocratique et incarne la devise « rien sur nous sans nous ». ¹³ En outre, The MTM corrige le déni de reconnaissance en mettant en lumière les expériences des personnes migrantes, en assurant leur représentation équitable dans les discussions sur les politiques et en réfutant les discours excluants.

Perspectives stratégiques

■ **Transparence des rapports et de la surveillance**

Rendre compte des pratiques discriminatoires et des conséquences négatives subies par les personnes en situation de déplacement permet de promouvoir la responsabilisation. Les responsables des politiques peuvent promouvoir la responsabilisation en exigeant l'émission de rapports détaillés et une surveillance indépendante des outils d'IA utilisés aux frontières et dans les situations de crise humanitaire, afin de garantir qu'ils respectent les normes relatives aux droits de la personne.

■ **Suivi mené par les communautés**

En soutenant le suivi mené par les communautés des systèmes d'IA qui affectent les personnes migrantes, il est possible d'offrir aux gens un rôle direct dans l'évaluation et la responsabilisation. Les responsables des politiques peuvent améliorer la responsabilisation et la réactivité dans le déploiement de l'IA en établissant des mécanismes d'audits et de boucles de rétroaction auprès des communautés.

The Algorithmic Justice League

Initiative:

[The Algorithmic Justice League](#)

Organisation:

[The Algorithmic Justice League](#)

Pays/région :

États-Unis/Amérique du Nord

Description

L'Algorithmic Justice League (AJL) conjugue l'art, la recherche et le plaidoyer pour alerter sur les biais et les préjugés de l'IA et pour développer la capacité de les réduire. Dans le cadre de campagnes de sensibilisation du public, et de programmes pédagogiques et de développement des capacités, cette organisation soutient les personnes les plus affectées par les systèmes d'IA, ainsi que les responsables des politiques, les groupes de recherche et les personnes professionnelles du secteur afin de veiller à ce que les systèmes d'IA soient développés et déployés en respectant les droits de la personne et en promouvant la justice sociale. L'un de ses domaines d'intervention essentiels est la lutte contre l'injustice intersectionnelle du sexisme et du racisme dans les écosystèmes particuliers des technologies de reconnaissance faciale. L'AJL est financée par la Fondation Ford, la Fondation MacArthur et les contributions individuelles.

Comment l'initiative fait progresser l'égalité réelle

En défendant une IA juste et équitable, l'AJL réduit la discrimination systémique des communautés marginalisées. Elle combat les désavantages systémiques et le déni de reconnaissance en révélant et en atténuant les biais dans les technologies de l'IA, afin que ces systèmes ne perpétuent pas les inégalités existantes. À travers d'audits et de campagnes de plaidoyer autour du racisme, du sexisme, du capacitisme et des autres formes de discrimination, elle met en évidence les enjeux au centre des préoccupations comme le droit à l'autonomie et à la dignité. Elle promeut la transparence et la responsabilisation dans le développement de l'IA, en favorisant le dialogue et la participation du public. Cette approche permet de démocratiser l'élaboration des politiques : toutes les voix sont incluses dans les discussions autour du rôle de la technologie dans la société.

Perspectives stratégiques

■ Sensibilisation du public par le biais de l'art et de l'élaboration de récits

En plus de sa contribution aux discussions et de la production de matériel pédagogique, l'AJL sensibilise le public aux impacts des systèmes et processus d'IA par le biais de l'art et de l'élaboration de récits recueillant les témoignages de personnes dont la vie a été directement affectée par des algorithmes injustes. En soutenant et lançant des initiatives de sensibilisation du public sur les effets néfastes des systèmes et des processus d'IA, les responsables des politiques peuvent rendre les processus de gouvernance de l'IA plus efficaces et inclusifs.

■ Plaidoyer et production collective de documents

L'AJL invite le public à présenter et à partager son vécu et ses expériences des systèmes et processus d'IA, comme les programmes de reconnaissance faciale dans les contrôles de sécurité pour les transports. Ces témoignages permettent une meilleure compréhension de l'expérience des groupes marginalisés face à des problématiques documentées comme la discrimination raciale, les enjeux en matière de confidentialité et les conséquences du choix d'opposition. Les responsables des politiques peuvent lancer ou soutenir des initiatives invitant le public à partager ses expériences, afin de parvenir à une meilleure compréhension des préjugés et d'ouvrir la voie pour que justice soit rendue.

Initiatives d'inclusion et de participation réelles

Modèle maori de gouvernance des données

Initiative:

[Modèle maori de gouvernance des données](#)

Organisation:

[Fiducie caritative Te Kāhui Raraunga](#)

Pays/région :

Nouvelle Zélande

Description

L'initiative du modèle maori de gouvernance des données lancée par Te Kāhui Raraunga a pour objet l'application des principes et des valeurs maoris à ce processus. Les données maories constituent un *taonga* (bien précieux) qui exige des modèles de protection ancrés dans la culture. Ce modèle de gouvernance défend la souveraineté des données, en insistant sur le contrôle, la détention et l'application des données par les communautés maories. Il offre une orientation sur la gouvernance des données maories à l'échelle du système, en accord avec les responsabilités du gouvernement en vertu du *Tiriti o Waitangi* (traité de Waitangi). Cette initiative est née d'un processus de coconception entre les directions maoris, les organisations maories possédant des intérêts dans les données et des spécialistes maoris en données individuelles et des agences de la Couronne.

Comment l'initiative fait progresser l'égalité réelle

Cette initiative promeut la souveraineté des données, pour que les communautés maories contrôlent leurs données et en tirent des bénéfices. En alignant les pratiques en matière de données sur les principes maoris, elle soutient le développement et réduit les désavantages systémiques ancrés dans les inégalités historiques. Des leaders, organisations et spécialistes maoris interviennent dans le processus de coconception, garantissant ainsi l'intégration des voix maories à la prise de décision en matière de données. Cette approche inclusive renforce la participation démocratique et s'inscrit dans le respect des obligations du traité de Waitangi. En canalisant et en faisant entendre les voix maories, l'initiative garantit la reconnaissance et le respect des points de vue et des droits des Maoris en lien avec la gouvernance de données, corrigeant ainsi les représentations trompeuses et l'exclusion historiques, et l'injustice épistémique.

Perspectives stratégiques

■ Processus de coconception fondé sur les traités autour du principe « Rien sur nous sans nous »

Cette initiative met l'accent sur un processus de coconception fondé sur les traités, dont les résultats sont définis par la communauté maorie, détentrice de l'autorité de prise de décision. En reconnaissant son droit à gérer son patrimoine commun de données, en lien avec son droit à l'autodétermination et en tant que principe légitimant au niveau structurel, elle garantit la participation et l'inclusion réelles des communautés maories dans la gouvernance des données.

■ Amplification des voix autochtones

Pour éviter l'injustice liée au manque de représentation et l'appropriation culturelle, il est essentiel de donner aux communautés autochtones leur mot à dire dans la gestion de leurs ressources collectives de données. En corrigeant les failles des études passées et en privilégiant les voix maories dans le processus de coconception, cette initiative veille à ce que la gouvernance des données reflète les points de vue et les droits des communautés maories. Cette pratique améliore la représentation et va dans le sens d'une utilisation exacte et respectueuse des données maories.

L'atelier du RIADIS

Initiative:

[Atelier de RIADIS pour sensibiliser à l'intelligence chrétienne pour artificielle \(IA\) et aux technologies d'assistance pour les personnes handicapées en Amérique latine](#) (en espagnol)

Pays/région :

Amérique latine

Organisation:

[RIADIS](#) financé par la Mission chrétienne pour les personnes aveugles

Description

Le Réseau latino-américain d'organisations non gouvernementales de personnes handicapées et leurs familles (RIADIS) est un réseau régional consacré à la promotion et à la défense des droits de la personne et à l'inclusion des personnes handicapées. Il mise sur le plaidoyer, le renforcement des capacités et la collaboration pour influencer les politiques et sensibiliser aux droits et aux besoins de ce collectif. L'atelier a pour objet de présenter et de partager des bonnes pratiques, des initiatives et les recherches menées dans le domaine de l'IA et des technologies assistées pour les personnes handicapées. Il a permis de faire connaître des outils comme les convertisseurs parole-texte et d'aborder les risques représentés par les systèmes d'IA sur le marché du travail pour les personnes handicapées si elles ne sont pas consultées tout au long du cycle de vie de l'IA.¹⁴

Comment l'initiative fait progresser l'égalité réelle

L'initiative du RIADIS expose des stratégies pratiques pour placer les besoins et les défis qui se posent aux personnes handicapées au cœur de l'innovation en IA. En faisant entendre leur voix à chaque étape du cycle de l'IA, le RIADIS contribue à lutter contre l'exclusion systémique. Les ateliers ciblés offrent aux personnes handicapées les connaissances et les outils nécessaires pour participer aux discussions et à la prise de décision, en les encourageant à défendre une conception inclusive et à mettre les équipes de développement en IA face à leurs responsabilités. En plaçant au centre les besoins et la voix des personnes handicapées, le RIADIS œuvre à corriger le déni de reconnaissance historique les concernant. L'initiative promeut leur visibilité et garantit que leurs contributions sont valorisées et reconnues.

Perspectives stratégiques

■ Autonomisation par l'éducation

Ces ateliers destinés à faire participer les personnes handicapées à la prise de décision en matière d'IA donnent aux personnes participantes les moyens de comprendre les problématiques, de défendre leurs besoins et d'exiger des équipes de développement de l'IA une réelle inclusivité dans la conception. Les responsables des politiques peuvent reproduire cette approche en finançant et en soutenant des programmes éducatifs visant à développer les compétences et la confiance des groupes marginalisés, en vue de leur participation effective aux discussions sur les politiques et aux processus de conception technologique.

■ Adoption de cadres éthiques spécifiques au handicap

L'adoption de cadres éthiques spécifiques au handicap pour la conception et le développement de l'IA permet de veiller à ce que les besoins et les droits particuliers des personnes handicapées soient pris en compte tout au long du cycle de vie de l'IA. Les responsables des politiques devraient ordonner l'intégration de ces cadres aux processus de développement de l'IA, afin d'y introduire des considérations éthiques qui abordent spécifiquement la problématique du handicap.¹⁵

The Gender and Responsible Artificial Intelligence Network (GRAIN)

Initiative:

[The Gender and Responsible Artificial Intelligence Network \(GRAIN\)](#)

Pays/région :

Sénégal, Nigéria, Ouganda

Organisation:

[Initiative Prospective Agricole et Rurale \(IPAR\)](#),
[Centre for the Study of the Economy in Africa \(CSEA\)](#) et [Sunbird AI](#)

Description

Le GRAIN a pour mission d'introduire l'égalité des genres et les pratiques d'IA responsables en Afrique subsaharienne. Le réseau met en contact les organisations, universités et hubs pour collaborer autour de l'IA et de l'inclusion des genres. Il s'agit d'apprendre, de partager, de défendre et de proposer des solutions pour associer le développement de l'IA à l'égalité des genres, en soutenant un développement inclusif dans la région. S'inscrivant dans le cadre du programme Intelligence artificielle pour le développement de l'Afrique, cette initiative est menée par un consortium de groupes de réflexion et d'organisations non gouvernementales (AI4D Africa). Elle bénéficie du soutien financier du Centre de recherches pour le développement international du Canada (CRDI) et de l'Agence suédoise de coopération internationale au développement (Sida).

Comment l'initiative fait progresser l'égalité réelle

Le GRAIN réduit les désavantages systémiques en rassemblant les organisations, universités et hubs et en veillant à ce que le développement de l'IA en Afrique subsaharienne inclut une diversité de points de vue. Cette initiative s'attaque à des défis comme la sous-représentation des femmes, le manque d'attention accordé aux langues africaines et le développement de solutions non discriminantes. Elle promeut la collaboration et la prise de décision inclusive et multidisciplinaire en impliquant plusieurs parties dans les discussions sur l'IA et le genre. Cette approche contribue donc au comblement du déficit démocratique. Elle s'attache à défendre des systèmes d'IA qui reconnaissent et respectent les identités et les contributions de toutes les personnes, en favorisant un environnement technologique plus inclusif et attentif au contexte et en remédiant au déni de reconnaissance historique des communautés marginalisées.

Perspectives stratégiques

■ Développement d'une politique inclusive en matière d'IA

Le GRAIN met en évidence le besoin de politiques en matière d'IA qui luttent contre la discrimination intersectionnelle et le manque de représentation, en garantissant que le développement de l'IA inclut une diversité de points de vue, et notamment en lien avec les femmes d'Afrique subsaharienne et avec les langues africaines. Les responsables des politiques devraient adopter des cadres imposant de prendre en compte l'intersectionnalité et le contexte local dans le développement de l'IA pour réduire les désavantages systémiques.

■ Collaboration multidisciplinaire et participative

Lors des discussions en matière d'IA et de genre, le GRAIN met l'accent sur une collaboration inclusive et multidisciplinaire impliquant différentes parties. Les responsables des décisions politiques devraient encourager les échanges avec les parties affectées dans les politiques relatives à l'IA, afin de créer des systèmes d'IA plus représentatifs et attentifs au contexte.

Recommandations clés sur la politique d'IA transformatrice

Le rapport *Vers une réelle égalité en intelligence artificielle : une politique d'IA transformatrice pour l'égalité des genres et la diversité* offre un cadre complet et des recommandations clés pour parvenir à une égalité réelle dans les écosystèmes de l'IA et dans l'élaboration des politiques en la matière. Les recommandations développées ci-dessous ébauchent des mesures concrètes que les responsables des décisions politiques peuvent prendre pour intégrer efficacement les principes d'égalité des genres et de diversité dans le cadre d'action, la législation, la réglementation et les pratiques en matière d'IA.

Les recommandations de la politique d'IA transformatrice s'appuient sur les résultats des consultations régionales et de groupes spécifiques et sur les perspectives stratégiques tirées des pratiques prometteuses ; elles reflètent les dimensions essentielles du cadre pour une politique d'IA transformatrice. Les problématiques qui ressortent des consultations (la différence entre accès et inclusion, l'exclusion du savoir et la répartition inégale des ressources) sont la preuve des inégalités profondes enracinées au cœur de l'écosystème de l'IA. Tous ces défis à relever demandent des politiques concertées qui 1) renversent les barrières structurelles empêchant les groupes historiquement marginalisés et exclus de profiter des bénéfices du développement de l'IA et 2) cherchent activement à donner les moyens à ces groupes d'exercer pleinement leurs droits. Les recommandations sur la politique d'IA transformatrice sont regroupées en quatre catégories qui reflètent ces défis :

- **Conception inclusive et innovation démocratique**
- **Participation effective à la gouvernance de l'IA**
- **Transparence et responsabilisation dans la prévention des préjudices**
- **Accès efficace à la justice**

À travers ces catégories, les recommandations clés pour une politique d'IA transformatrice adoptent une approche globale visant à l'instauration d'une égalité réelle. En premier lieu, la conception inclusive et l'innovation démocratique sont essentielles pour combattre les désavantages systémiques auxquels font face les femmes et d'autres groupes marginalisés, et pour faire respecter le droit à l'inclusion. Les responsables des politiques peuvent y contribuer activement par le biais de l'action positive : investissements dans le renforcement des capacités pour l'inclusion institutionnelle, autorisation du traitement des catégories spéciales de données et financement de la recherche en technologie transformatrice et des approches conceptuelles. En second lieu, la **participation effective à la gouvernance de l'IA** comble le déficit démocratique et garantit le droit de participation. Elle demande de mobiliser efficacement le public et de favoriser la participation sociale, en investissant dans le développement des capacités des groupes marginalisés,

en légiférant pour les droits à la participation du public et en protégeant les droits collectifs en matière de données et d'IA. En troisième lieu, la **transparence et la responsabilisation sont essentielles en vue de la prévention des dommages**, en comblant le déficit démocratique et en corrigeant le déni de reconnaissance. Elle donne aux personnes qui défendent l'égalité réelle et aux autres parties prenantes les moyens d'examiner les systèmes et les processus d'IA, de détecter les biais, et de demander des comptes aux fournisseurs et déployeurs publics et privés concernant les effets néfastes ou discriminatoires. Ces recommandations incluent la garantie du droit à l'information, l'amélioration de la transparence de l'algorithme et l'instauration d'une responsabilisation claire chez toutes les personnes impliquées dans les écosystèmes de l'IA, par le biais d'évaluations de l'impact sur les droits de la personne et de lignes directrices à respecter lors de la passation de marchés publics. Enfin, la **garantie d'un accès efficace à la justice** permet de lutter contre le déni de reconnaissance et de faire respecter le droit à la dignité. En renforçant la responsabilité contextuelle, en soutenant les organismes de promotion de l'égalité et en allégeant la charge de la preuve, les responsables des politiques peuvent améliorer l'accès à la justice des personnes et des groupes victimes de discrimination et des dommages provoqués par les systèmes et processus de l'IA. En couvrant toutes les dimensions du cadre pour une politique d'IA transformatrice, ces catégories permettent d'atteindre une véritable égalité.

Ces recommandations sont destinées à être examinées et appliquées en contexte. Pour assurer le succès de leur mise en œuvre, les responsables des politiques doivent prendre en compte les éléments de contexte (géographiques, sociaux, culturels, historiques, économiques, juridiques et politiques) qui façonnent les écosystèmes de l'IA dans les différentes régions et adapter les recommandations en conséquence. Cela implique aussi de tenir compte des variations dans le cadre juridique et réglementaire, des capacités techniques, des ressources disponibles et de la coopération avec les parties prenantes, ainsi que d'identifier et d'affronter les obstacles à la mise en œuvre. Vous trouverez des conseils de mise en œuvre dans le Guide d'application des recommandations de la politique d'IA transformatrice. L'engagement de respecter les politiques de transformation de l'IA pour l'égalité des genres et la diversité ouvre la voie à une égalité réelle dans les écosystèmes de l'IA et dans l'élaboration de politiques en la matière.

Conception inclusive et innovation démocratique

La conception inclusive et l'innovation démocratique sont essentielles pour combattre les désavantages systémiques auxquels font face les femmes et d'autres groupes marginalisés, et pour faire respecter le droit à l'inclusion. Le manque de diversité et de représentation, de sensibilisation et de compréhension des différentes expériences et perspectives et des processus permettant une réelle inclusion dans les écosystèmes de l'IA représente un défi majeur en vue de parvenir à une véritable égalité. Les responsables des politiques ont la possibilité de contribuer activement à la conception inclusive et aux processus d'innovation démocratique en misant sur l'action positive, sur les investissements dans le renforcement des capacités en vue de l'inclusion institutionnelle, sur l'autorisation du traitement des catégories spéciales de données et sur le financement de la recherche en technologie transformatrice et en approches pour la conception.

1. Donner un rôle technique et non technique aux groupes marginalisés au sein de l'écosystème de l'IA

Appliquer l'action positive au sein de l'écosystème de l'IA afin d'impliquer les femmes et d'autres groupes historiquement marginalisés dans des rôles techniques et non techniques, dans le but de favoriser la diversité des points de vue. Allouer des ressources destinées à identifier et à supprimer les obstacles à la représentation de la diversité. Cela implique d'assurer une éducation accessible et inclusive au-delà des écosystèmes de l'IA. Les équipes diversifiées et les approches interdisciplinaires permettent d'apporter des points de vue plus larges au développement de l'IA, pour des systèmes plus inclusifs, équitables et transformateurs (Buhl, 2023). Pour un respect optimum des normes en matière d'égalité des genres et de diversité, il est nécessaire que les processus de conception de l'IA obtiennent l'approbation de conseils consultatifs et de programmes de certification.



Femme maorie lors d'une discussion d'affaires avec une autre femme, devant une tablette mobile dans des bureaux en Nouvelle Zélande. Photo de comers74 sur iStock.

2. Investir dans le renforcement des capacités en vue de l'inclusion institutionnelle

Au sein des institutions et des équipes publiques et privées, investir dans le développement des capacités et dans la sensibilisation à l'expérience et aux droits des groupes historiquement marginalisés comme les personnes confrontées à des obstacles en raison de leur sexe, de leur orientation sexuelle, de leur identité et/ou de leur expression de genre, ainsi que les personnes racisées, les personnes handicapées ou les Autochtones (Bartoletti et Xenidis, 2023). Maintenir un dialogue régulier avec les groupes marginalisés, en vue de la compréhension et de l'élimination des obstacles particuliers auxquels ils sont confrontés. Grâce au développement des capacités et à la sensibilisation à l'inclusion institutionnelle, les institutions seront plus à même d'être inclusives face aux groupes marginalisés, aussi bien en tant qu'employeurs qu'en tant que fournisseurs de services. Développer et revoir régulièrement les lignes directrices applicables aux infrastructures de communication et aux communications écrites et orales, afin de vérifier qu'elles ne comportent aucune barrière à l'accessibilité ni de discrimination en raison de l'utilisation d'un certain langage ou de certaines expressions, et qu'elles n'omettent pas de représenter la société dans toute sa diversité. Cela implique par exemple de réaliser des évaluations de l'accessibilité et de tenir des listes de contrôle relatives à l'inclusion pour tous les forums et toutes les activités en ligne. Ces mesures contribuent à une inclusion réelle, en promouvant des pratiques inclusives et une meilleure compréhension de la manière dont les institutions peuvent tenir compte des différents points de vue et expériences. Chaque membre de l'équipe de recherche ou du personnel devrait être capable d'identifier les changements à introduire spécifiquement dans les pratiques, protocoles et lignes directrices pour satisfaire aux objectifs d'inclusion et de diversité de l'institution.

3. Permettre le traitement de catégories spéciales de données

Permettre le traitement de catégories spéciales de données dans certaines circonstances exceptionnelles, en présence d'un motif d'intérêt public important, dans une visée d'égalité et de non-discrimination. Cet objectif devrait être satisfait sans enfreindre aucun droit à la protection des données à caractère personnel, telles que les dispositions de l'alinéa 44c du règlement sur l'intelligence artificielle de l'UE. Pour éviter les résultats discriminatoires, les fournisseurs de systèmes d'IA doivent réaliser des essais de détection des biais systémiques et veiller à la représentation d'ensembles de données variés. Cette obligation peut impliquer de mener un traitement de données à caractère personnel sensibles pour évaluer comment des attributs protégés comme la race et le sexe peuvent correspondre à des variables substitutives dans le modèle et donc perpétuer la discrimination (Deck et al., 2024 ; Dwork et al., 2012). Ce traitement peut être restreint en vertu de lois sur la protection des données. Il est essentiel d'évaluer l'importance de l'intérêt public dans une optique d'égalité et de non-discrimination, tout en respectant les droits à la protection des données à caractère personnel et en promouvant les droits de la personne. Bien qu'inégalement répartis à travers le monde, ces solides cadres de protection des données sont nécessaires pour empêcher la discrimination et l'exploitation dans le contexte de l'IA. Soutenir la recherche pour élargir les points de vue collectifs sur le patrimoine commun de données et les approches d'intérêt public afin de guider la mise en œuvre de cette recommandation (Cofone, 2023).

4. Financer la recherche dans le domaine de la technologie transformatrice et les approches conceptuelles en innovation dans le domaine de l'IA

Financer la recherche, accorder des subventions et témoigner de la reconnaissance publique pour encourager les approches technologiques et conceptuelles transformatrices dans le domaine de l'IA, par exemple celles reposant sur des principes féministes. Ce type d'approches cherche à pallier les écarts entre l'équité technique et politique. En soutenant les innovations respectueuses de ces principes dans les systèmes d'IA, il est possible d'aboutir à des applications, des pratiques et des processus plus équitables et justes.



Femme australienne contemplant un drapeau aborigène en compagnie de deux femmes plus jeunes. Townsville, Australie. pendant les célébrations de NAIDOC week. Photo de Ann Smith sur iStock.

Les principes féministes appliqués à la conception technologique en matière d'innovation dans le domaine de l'IA incluent les mesures suivantes :

Intégration des connaissances inhérentes reflétant une diversité d'identités

Respecter l'autonomie et la dignité de toutes les personnes, et tout particulièrement des membres des communautés marginalisées, en érigeant ce respect en principe fondamental du développement de l'IA.

- Appliquer le principe « Rien sur nous sans nous », en tant que norme de référence pour la pleine participation et l'égalisation des opportunités pour, par et avec les groupes historiquement exclus. Reconnaître cette application comme s'inscrivant dans le cadre du principe « ne pas nuire ».
- Coconcevoir les systèmes d'IA avec les personnes confrontées à des obstacles en raison de leur sexe, de leur orientation sexuelle, de leur identité et/ou de leur expression de genre, afin de garantir des innovations inclusives en matière de genre.

Adopter des indicateurs d'équité intersectionnelle dans les modèles d'IA

Mettre en œuvre des indicateurs d'équité intersectionnelle pour égaliser les performances des systèmes d'IA dans les sous-groupes intersectionnels. Ces indicateurs devraient :

- Tenir compte des différents attributs protégés.
- Protéger toutes les valeurs d'intersection de ces attributs, p. ex. les femmes racisées.
- Poursuivre la protection des valeurs des attributs individuels, p. ex. les femmes.
- Garantir la protection des groupes marginalisés lourdement affectés par les discriminations sociales. Rectifier les différences systématiques provoquées par l'oppression structurelle, plutôt que de les codifier (Buolamwini, 2023 ; Buolamwini et Gebru, 2018 ; Ovalle et al., 2023b).

Développer et promouvoir les langues et cultures autochtones et minoritaires dans le développement de l'IA

- Encourager aussi bien les initiatives à grande échelle que les projets plus modestes axés sur les communautés afin d'encourager la participation active des communautés à la création d'ensembles de données d'IA, de schémas algorithmiques et de formulations de cas d'utilisation. Établir des mesures d'incitation et des infrastructures au niveau national et sous-national dans les langues locales.¹⁶



Files de jetons noirs composant en anglais le mot « féministe » sur une surface en bois. Photo de Jen Theodore sur Unsplash.

Participation effective à la gouvernance de l'IA

La mise en place d'une participation effective à la gouvernance de l'IA permet de remédier au déficit démocratique et de faire respecter le droit de participation, avec à l'arrivée de meilleures politiques pour tout le monde. La confusion entre accès et inclusion est courante. Mais l'accès en tant que tel ne garantit pas la participation ni l'inclusion effective. Lorsque les responsables des politiques sont à l'écoute et dialoguent avec les personnes les plus affectées par les systèmes et processus d'IA et celles habituées à identifier et combattre leurs effets néfastes, les politiques, lois et réglementations sont plus à même de prévenir et d'atténuer efficacement les préjudices. Parmi d'autres recommandations en ce sens, citons la mobilisation efficace du public et la promotion de la participation des communautés, les investissements dans le renforcement des capacités des groupes marginalisés, l'établissement d'une législation défendant le droit de participation du public et la protection des droits collectifs en matière de données et d'IA.

5. Promouvoir une mobilisation efficace du public et la participation sociale

Favoriser l'équité et l'inclusion dans la conception, le développement, le déploiement et la gouvernance des systèmes d'IA, en faisant appel à différentes méthodologies de mobilisation du public, au niveau national et international. Il s'agit d'inclure les voix marginalisées dans les discussions nationales autour de la gouvernance de l'IA et d'élargir la majorité mondiale dans les forums consacrés à ce sujet. Utiliser des scrutins délibératifs, les jurys communautaires, les assemblées citoyennes, les conférences de consensus, les mini-groupes de délibération, la délibération en ligne, la budgétisation participative, les pratiques itératives et participatives de recherche et de conception, les assemblées communautaires et les groupes de référence communautaires (OCDE, 2021).¹⁷ Échanger avec les communautés marginalisées et les organisations les représentant afin de comprendre et d'éliminer les obstacles spécifiques auxquels elles sont confrontées (Nations Unies, 2019, 2021). Allouer des budgets pour couvrir les frais de participation des personnes représentant les groupes marginalisés, pour rémunérer leur implication et leur expertise, et pour prendre en charge les coûts d'aménagement raisonnables au profit de l'accessibilité (p. ex. interprètes en langue des signes ou des langues autochtones). Ces initiatives doivent aussi garantir que les processus d'information et de consultation sont accessibles, gratuits et compréhensibles pour les groupes marginalisés, en utilisant un langage et des canaux et formats de communication appropriés.

6. Investir dans le renforcement des capacités des groupes marginalisés

Financer et soutenir des programmes éducatifs, des structures de réseau et d'autres ressources visant à développer les compétences et la confiance des groupes marginalisés, pour qu'ils participent véritablement aux processus au service de leurs besoins ou qu'ils les dirigent activement. Travailler avec les communautés et groupes marginalisés, et leurs organisations de représentation, afin de les aider à tenir leurs propres réunions de sensibilisation et de consultation sur les enjeux liés à l'IA, qui serviront à améliorer leur compréhension et leur mobilisation. Investir dans des séances de sensibilisation du public sur l'IA, les droits individuels, les institutions pertinentes et les mécanismes pour engager un redressement et demander réparation.

7. Légiférer sur les droits à la participation du public *ex ante*

Fonder les processus de prise de décision en matière d'IA sur les droits à la participation du public *ex ante*, tels que ceux prévus dans la Convention d'Aarhus de la CEE-ONU. La Convention d'Aarhus garantit des droits sociaux d'accès à l'information, de participation du public à la prise de décision et de recours à la justice pour les questions d'environnement (CEE-ONU, 1998). L'application de ces principes aux processus de prise de décision dans l'AI permet aux parties concernées, ainsi qu'aux organisations de la société civile et au public, de s'opposer aux conséquences de la prise de décisions algorithmique par le biais de la discussion et de la réflexion publique (CEE-ONU, sans date). Cette recommandation assoit le rôle décisif du public et ouvre la porte à un examen et une responsabilisation plus conséquents dans les processus de prise de décision en matière d'IA.

8. Protéger les droits collectifs en matière de données et d'IA

Revoir les cadres relatifs aux droits affectés par les systèmes et les processus de l'IA, comme ceux relatifs aux droits des données ou de propriété intellectuelle afin de 1) protéger les données et la souveraineté du savoir des peuples autochtones et des groupes marginalisés, et notamment des minorités linguistiques, religieuses et ethniques ; et 2) garantir le droit de bénéficier des progrès scientifiques. Cette recommandation contribue à une répartition plus équitable des bénéfices. Instaurer de solides garanties institutionnelles pour protéger les ensembles de données du secteur social, tout particulièrement lorsqu'il existe un risque d'accaparement des fonctions de développement principales par le biais des modèles d'IA (p. ex. dans la santé, l'éducation et la protection sociale). Garantir que les cadres de gouvernance en matière de données et d'IA protègent l'innovation publique, avec des réformes du droit de la propriété intellectuelle en faveur d'écosystèmes de l'IA non commerciaux qui contribuent au bien commun et aux droits collectifs. Le droit de toutes les personnes et communautés à bénéficier des progrès scientifiques doit être étendu aux écosystèmes de l'IA. Cette exigence acquiert un sens tout particulièrement dans le contexte de la marginalisation des cultures locales et des systèmes de connaissance dans les modèles dominants de données et d'IA. Les régimes d'accès et d'utilisation applicables aux ensembles de données publics devraient avoir en ligne de mire des objectifs de démocratie, de développement et de droits de la personne, sans menacer aucunement le bien-être social et environnemental (Gurumurthy et Deepti, 2023).

Les paragraphes suivants sont consacrés à la présentation de voies potentielles vers la protection des droits en matière de données collectives et d'IA :

■ Instaurer l'accès conditionnel aux données publiques

L'accès au domaine public et aux données gouvernementales ouvertes devrait être conditionnel, avec une limitation de la finalité et des clauses de restriction dans le temps bien claires. Établir de solides garanties constitutionnelles pour les ensembles de données du secteur social, par exemple dans la santé, l'éducation et la protection sociale, afin de veiller à ce que les modèles d'IA respectent les principes de service public et protègent les communautés marginalisées (Tomasso Fia, 2021).

■ Appliquer des limites d'utilisation raisonnables

Imposer des limites d'utilisation raisonnables sur le mode d'apprentissage des modèles d'IA et leur utilisation des données d'entraînement, afin de les empêcher de les réutiliser et d'en tirer profit.

■ Proposer des licences collectives

Développer des propositions de licences collectives pour parvenir à un équilibre entre les droits moraux des entités créatrices et les systèmes de valeur qui considèrent l'espace intellectuel commun comme un patrimoine public.

■ Garantir la réciprocité dans les pools de données

Établir des garanties de réciprocité dans les pools communs de données, en veillant à ce que les équipes de développement de modèles privés qui utilisent des couches de données publiques soient tenues de rendre en retour et d'enrichir les espaces communs de données.



Membres des communautés amérindiennes lors de l'événement Chumash Day Pow Wow and Intertribal Gathering. Photo de Hanna Tor sur iStock.

Transparence et responsabilisation dans la prévention des dommages

La transparence et la responsabilisation sont essentielles en vue de prévenir les dommages, de combler le déficit démocratique et de corriger le déni de reconnaissance. Elles permettent aux parties prenantes d'examiner les systèmes et les processus d'IA, de détecter les biais, et de demander des comptes aux fournisseurs et déployeurs publics et privés concernant les effets néfastes ou discriminatoires. Elles donnent les moyens aux groupes marginalisés de remettre en question les pratiques discriminatoires et de favoriser la confiance du public en garantissant que les processus de prise de décision sont ouverts et compréhensibles. Les recommandations suivantes permettent de garantir le droit à l'information, d'améliorer la transparence de l'algorithme et d'instaurer une responsabilisation claire chez toutes les personnes impliquées dans les écosystèmes de l'IA par le biais d'évaluations de l'impact sur les droits de la personne et de lignes directrices pour la passation de marchés publics..

9. Établir le droit à l'information dans les systèmes d'IA et améliorer la transparence des algorithmes

Établir le droit à l'information dans l'IA. Les gens devraient avoir le droit d'être informés clairement des situations dans lesquelles l'IA est utilisée, des algorithmes employés, des variables d'entrée et des critères inclus dans les processus de prise de décision. Cette recommandation place plus haut la barre des exigences en matière de transparence algorithmique et permet aux personnes ayant subi l'impact négatif des systèmes d'IA de contester leurs résultats (Grochowski et al., 2021 ; Kossow et al., 2021). Les informations fournies doivent être suffisamment simples pour être comprises par le grand public. Pour empêcher que la propriété intellectuelle ou les secrets commerciaux ne soient utilisés pour contourner la responsabilisation, les obligations de transparence devraient prévaloir sur l'exercice de ces derniers. (Fink et Finck, 2022). L'exigence d'une transparence accrue des algorithmes va également dans le sens d'une plus grande innovation technologique pour faire face à des limitations comme l'opacité des comportements et pour améliorer l'interprétabilité et l'explicabilité (Bommasani et al., 2024 ; Winfield et al., 2021).

Entre autres voies potentielles supplémentaires vers l'amélioration du droit d'information sur l'IA et de la transparence des algorithmes, citons :

- Examiner et envisager la révision de la loi sur le secret commercial, afin d'établir des limites claires et des exceptions dans l'intérêt du public et pour parvenir à un équilibre entre droits des propriétaires et transparence (Kilic, 2024).
- Envisager l'établissement d'un nouveau standard d'adéquation *post facto* pour la transparence algorithmique. Une telle norme demanderait que les inférences des systèmes et processus d'IA dans la fonction publique fournissent des informations claires et accessibles sur les données, le modèle et les critères utilisés dans les processus de prise de décision. Cette norme garantit la traçabilité de la justification des décisions, améliorant ainsi l'explicabilité et permettant aux audits publics d'atténuer les conséquences négatives (Mathews et Sinha, 2020).
- Envisager une combinaison d'explications du modèle (c'est-à-dire allier un système d'IA opaque à un modèle transparent et interprétable qui capte pleinement la logique du système opaque), d'inspection du modèle (c'est-à-dire une représentation rendant possible la compréhension de certaines propriétés spécifiques d'un modèle opaque ou ses prédictions) et d'explication des résultats (c'est-à-dire l'explication des résultats d'une IA opaque dans un cas concret).
- Envisager d'interdire les modèles en boîte noire, et tout particulièrement dans les cas à haut risque et dans les processus publics de prise de décision, et soutenir le recours à des modèles interprétables (Rudin et Radin, 2019).

10. Permettre et mener des études obligatoires d'impact sur les droits de la personne (EIDH)

Permettre et réaliser des études d'impact en offrant des orientations générales sur leur développement. Les fournisseurs et déployeurs du secteur privé et public sont tenus de respecter les droits de la personne et les principes afférents, et notamment en appliquant la diligence voulue et en réalisant des études d'impact tout au long du cycle de l'IA (Pacte numérique mondial, 2024). Ces études devraient déterminer si les risques de préjudices sont acceptables au vu des lois garantissant les droits fondamentaux et indiquer clairement les mesures à prendre pour éliminer ou prévenir ces risques (Jakubowska et al., 2024). Elles devraient par ailleurs examiner et comparer d'éventuelles approches non technologiques, afin d'identifier les mesures les plus respectueuses des droits de la personne. Si les études révèlent des inquiétudes graves, les déployeurs doivent s'abstenir de toute utilisation.

Certaines voies potentielles vers des EIDH obligatoires pour les systèmes d'IA sont exposées ci-dessous :

- Exiger des études d'impact complètes *ex ante*, en mettant explicitement l'accent sur le genre et la diversité.
- Imposer de soumettre les résultats de ces études à un organisme réglementaire ou à une agence de surveillance, qui les examinera et les approuvera avant le déploiement du système d'IA en question. Dans le cadre de l'examen des processus, il convient de respecter les normes de représentation adéquates, afin de tenir dûment compte des aspects liés à l'égalité réelle.
- Exiger l'interdiction de certains systèmes et processus d'IA si leurs effets, potentiels ou avérés, ne sont pas justifiés conformément à la législation internationale en matière de droits de la personne. (McGregor et Molnar, 2023).
- Établir des pénalités ou des sanctions pour les organisations qui ne mènent pas les études d'impact requises.
- Établir une présomption légale en vertu de laquelle le déploiement d'un système d'IA sans étude d'impact complète constitue une preuve de discrimination à première vue.
- Établir des mécanismes d'examen et de réévaluation des risques périodiques.



Jeune Indienne de la génération Y expliquant les détails d'un projet de collaboration à sa collègue. Assises à un bureau, elles partagent des idées, conversations et tâches. Photo de fizkes sur iStock.

11. Établir des mesures de responsabilisation pour les systèmes et processus algorithmiques du secteur public

Établir des lignes directrices spécifiques à l'IA pour la passation de marchés, afin de protéger les droits de la personne et le respect de la légalité dans ces processus, en abordant les complexités et les risques introduits par les systèmes et processus algorithmiques et d'IA (Hickok, 2022). Promouvoir des initiatives de libre accès aux données afin de développer des bibliothèques ouvertes d'algorithmes à utiliser dans les systèmes publics. Veiller à ce que les responsables des politiques aient bénéficié d'un renforcement de leurs capacités, de sorte à pouvoir exercer efficacement leur devoir de vigilance en lien avec l'IA lors de la passation de marchés. Envisager d'établir des mécanismes pour les moratoires ou les interdictions dans les systèmes d'information automatisés (SIA) et les processus susceptibles de menacer les droits de la personne.

Accès efficace à la justice

L'accès efficace à la justice est indissociable de l'égalité réelle. Il implique la responsabilisation, la correction du déni de reconnaissance et le respect du droit à la dignité. En renforçant la responsabilité contextuelle, en soutenant les organismes de promotion de l'égalité et en allégeant la charge de la preuve, les responsables des politiques peuvent améliorer l'accès à la justice des personnes et des groupes victimes de discrimination et de préjudices en lien avec les systèmes et processus de l'IA.

12. Renforcer la responsabilité contextuelle en vue de la non-discrimination dans les systèmes d'IA

Renforcer la responsabilité contextuelle en vue de la non-discrimination dans les systèmes d'IA proportionnellement à d'autres mesures de responsabilisation comme le niveau de transparence, l'interprétabilité et l'explicabilité. Les réglementations sur les produits et sur la responsabilité pour faute doivent être modifiées de sorte à refléter les complexités des systèmes d'IA et de la prise de décision fondée sur les données. Une responsabilisation efficace dans le développement et le déploiement de l'IA passe par la prise en compte de certaines caractéristiques comme l'opacité, l'explicabilité, l'autonomie du comportement, l'adaptation continue et la prévisibilité limitée. Tracer la voie vers la responsabilité dans l'IA, afin de garantir que les fournisseurs et les déployeurs, publics comme privés, rendent des comptes.

Certaines pistes potentielles sont indiquées ci-dessous :

- Introduire des conditions d'utilisation et encourager les bacs à sable réglementaires pour les systèmes d'IA dans les domaines inexplorés. Cela permettrait de fixer des limites claires en faveur de la sécurité de l'utilisation et de réduire les dommages tout en permettant l'innovation et les tests itératifs.
- Envisager un cadre de responsabilité mixte pour les fournisseurs et déployeurs de systèmes d'IA, qui équilibrerait la responsabilité fondée sur la faute et la responsabilité stricte en lien avec d'autres mesures de responsabilisation.
- Évaluer les réglementations de la responsabilité axées sur le logiciel (EPRS, 2024).
- Exiger aux fournisseurs et aux déployeurs publics et privés de réparer tout dommage détecté et toute conséquence discriminatoire découlant de processus associés au développement et au déploiement du système d'IA dans le cadre d'une utilisation donnée.
- Fournir une indemnisation au titre des dommages identifiés (matériels et non matériels), subis à cause de la discrimination algorithmique.

13. Donner les moyens de passer à l'action aux organismes de promotion de l'égalité

Donner les moyens aux organismes de promotion de l'égalité, et notamment aux institutions nationales de défense des droits de la personne et autres organisations au service de l'intérêt public, d'engager des actions pour le bien commun. Permettre à ces organismes de présenter des plaintes auprès des autorités de supervision, même sans être identifiables (Equinet, 2023).¹⁸ Alléger la charge de la preuve et doter ces organismes de l'autorité légale et de la formation nécessaire pour affronter efficacement les discriminations et dommages causés par les systèmes d'IA et les processus associés.

14. Alléger la charge de la preuve pour les organismes et les personnes qui portent plainte

Réexaminer et modifier les règles d'allègement de la charge pour ces organismes et ces personnes (Commission mondiale d'éthique des connaissances scientifiques et des technologies, 2005). À l'heure actuelle, les règles de responsabilité associées aux produits demandent souvent aux parties ayant subi un préjudice de démontrer le lien de cause à effet entre les défauts des produits et les dommages particuliers. Compte tenu de la complexité et de la prévisibilité limitée des systèmes d'IA, cette exigence est difficile à satisfaire. Envisager de modifier ces règles pour qu'on puisse mieux se défendre et réclamer un dédommagement.

Alléger la charge de la preuve en cas de discrimination et de dommages liés à l'IA - Perspectives tirées de la directive proposée par l'UE sur la responsabilité en matière d'IA

■ Introduction d'une « présomption de causalité »

La directive proposée par l'UE sur la responsabilité en matière d'IA introduit une présomption de causalité, qui assume un lien de cause à effet entre le non-respect d'un devoir de vigilance prévu dans la législation de l'UE ou nationale (la faute) et le résultat ou manque de résultat issu du système d'IA ayant causé le dommage.

Ces principes équilibrent les intérêts des organismes et des personnes qui portent plainte et des mis en cause dans les affaires relatives à la responsabilité associée aux produits, sans augmenter les risques de responsabilité susceptibles d'entraver l'innovation et l'adoption de produits et services fondés sur l'IA. La charge de démontrer la validité et l'équité des systèmes pèse sur les fournisseurs, opérateurs et personnes qui utilisent les systèmes d'IA, réduisant ainsi pour les parties ayant subi un préjudice la charge de devoir prouver le lien entre les opérations des systèmes d'IA et les conséquences négatives (Commission européenne, 2022 ; Parlement européen, 2023).

■ Accorder aux tribunaux nationaux l'autorité d'ordonner la divulgation de la preuve (au sujet des systèmes d'IA présentant un haut degré de risque et soupçonnés de provoquer des préjudices)

Étant donné la complexité des systèmes d'IA, il n'est pas aisé pour les personnes ayant subi un préjudice de satisfaire aux conditions permettant de réclamer des responsabilités. Le fait d'accorder aux tribunaux nationaux l'autorité d'ordonner la divulgation de la preuve en lien avec les systèmes d'IA à haut degré de risque soupçonnés de provoquer des préjudices pourrait renforcer la responsabilisation (Commission mondiale d'éthique des connaissances scientifiques et des technologies, 2005). Cette initiative s'inscrit dans un principe de précaution, à condition que les tribunaux restent fidèles aux principes de nécessité et de proportionnalité, et évitent les requêtes génériques.



Silhouette d'une personne le point brandi sur fond de soleil couchant dégradé. Photo de Miguel Bruna sur Unsplash.

Conclusions et voies à explorer

Les recommandations clés sur la politique d'IA transformatrice présentées dans le présent rapport offrent des pistes exhaustives pour atteindre une égalité réelle au sein des écosystèmes de l'IA et pour élaborer des politiques en la matière. Classées de manière à refléter les dimensions clés du cadre pour la politique d'IA transformatrice, ces recommandations s'articulent autour de quatre catégories : la conception inclusive et l'innovation démocratique, la participation effective à la gouvernance de l'IA, la transparence et la responsabilisation en vue de la prévention des préjudices et l'accès efficace à la justice. Ensemble, ces piliers créent une base solide pour les écosystèmes de l'IA et l'élaboration des politiques associées, et ouvrent la voie vers une égalité réelle.

En vue du succès de leur mise en œuvre, ces recommandations devront être accompagnées d'investissements volontaires, d'allocations de budget, de renforcement des capacités, de collaborations avec les parties prenantes et d'innovations réglementaires, de sorte à surmonter les difficultés et les complexités de ce domaine en pleine évolution. Et surtout, leur application requiert une implication permanente, pour affronter les obstacles technologiques, juridiques, financiers, etc., et faire entendre les voix habituellement exclues, dans l'optique d'un changement transformateur positif pour tout le monde.

L'exploration des nouvelles pistes et les investissements constants en recherche interdisciplinaire, intersectorielle et multipartite contribueront à faire apparaître des systèmes et processus d'IA plus inclusifs, équitables et transformateurs. En améliorant notre compréhension de la manière dont les systèmes et les processus d'IA affectent d'autres cadres de droits, en élargissant les points de vue collectifs sur, par exemple, les droits liés à la protection de la vie privée, le droit à la liberté d'expression, le droit à l'éducation et à la santé ou les droits en matière d'environnement et de travail et en poursuivant la recherche autour du patrimoine commun de données et l'intérêt public, nous progresserons vers de futures applications des politiques transformatrices. Les essais itératifs et l'innovation permanente dans des environnements réglementaires sûrs permettront d'augmenter l'apprentissage et de développer des cadres d'action efficaces. En outre, une collaboration nationale et internationale sera nécessaire à travers les secteurs et les régions pour adapter les politiques aux différents contextes locaux, en veillant à ce que les systèmes d'IA et la gouvernance soient à la fois équitables à l'échelle internationale et sensibles aux variations régionales des environnements juridiques, culturels et économiques. Les investissements en recherche pour mieux évaluer l'impact des politiques, lois et réglementations au sein d'un environnement d'IA en mutation constante permettront aux responsables politiques d'améliorer et d'harmoniser les politiques et pratiques transformatrices.



Image détaillée en noir et blanc de mains entrelacées, l'une portant un bracelet et l'autre une bague. Photo Aarón Blanco Tejedor sur Unsplash.

Pour finir, si elles sont appliquées en s'engageant pour une approche sensible au contexte et une innovation permanente, ces recommandations laissent entrevoir un avenir où les bénéfices des systèmes et processus d'IA seront partagés équitablement, et où l'égalité des genres, la diversité et l'inclusion seront une réalité pour tous et pour toutes.

Notes de fin

- ¹ Dans ce rapport, le terme OSIGEGCS est utilisé pour inclure toutes les orientations sexuelles, identités de genre, expressions de genre et caractéristiques sexuelles, y compris les traits intersexués. Les termes OSIG (orientation sexuelle et identité de genre) ou OSIEG (orientation sexuelle, identité de genre et expression de genre) sont parfois également utilisés dans d'autres documents. Ce rapport considère que cet acronyme désigne tous les humains ayant des orientations sexuelles et des identités de genre, y compris les personnes cisgenre et les personnes hétérosexuelles. Quant à lui, le terme LGBTQI+ met l'accent sur des personnes particulières aux identités marginalisées (p. ex. les personnes transgenre, non binaires, lesbiennes, etc.). Pour en savoir plus, vous pouvez consulter Bureau du Haut-Commissariat des Nations Unies aux droits de l'homme (2024a) et Smith (2023).
- ² Les personnes en situation de déplacement incluent les personnes réfugiées, demandeuses d'asile, migrantes et déplacées à l'intérieur de leur propre pays (PDIP).
- ³ Voir, par exemple, la demande d'action des femmes pour une IA éthique appelant tous les pays à rejoindre les activités de l'UNESCO en soutien à la mission de mise en place de mesures systématiques pour une IA inclusive en matière de genre. (UNESCO, 2024b).
- ⁴ L'équité consiste à reconnaître que chaque personne a des circonstances différentes et à attribuer les ressources et les possibilités exactement nécessaires pour atteindre un résultat égal. Voir Journée internationale de la femme (2023).
- ⁵ Le présent rapport adopte une approche non binaire de l'égalité des genres et fait référence aux obstacles rencontrés non seulement par les femmes, mais aussi par les personnes ayant une orientation sexuelle, des identités et expressions de genre et des caractéristiques sexuelles différentes (OSIGEGCS).
- ⁶ Représentation de dix pays en Afrique (Burkina Faso, Cameroun, Kenya, Mauritanie, Nigeria, Afrique du Sud, Swaziland, Tunisie, Ouganda et Zimbabwe) ; dix pays d'Amérique latine (Argentine, Bolivie, Brésil, Chili, Colombie, Équateur, Mexique, Paraguay, Pérou et Uruguay) ; dix pays d'Amérique du Nord et d'Europe (Belgique, Canada, Danemark, France, Allemagne, Hongrie, Portugal, Roumanie, Royaume-Uni et États-Unis) ; neuf pays de la région du Moyen-Orient et de l'Afrique du Nord (Égypte, Iran, Israël, Liban, Mauritanie, Maroc, Arabie saoudite, Tunisie et Émirats arabes unis) ; onze pays de la région Asie-Pacifique (Australie, Inde, Japon, Malaisie, Népal, Nouvelle Zélande, Philippines, Singapour, Corée du Sud, Thaïlande et Vietnam).
- ⁷ L'équité consiste à reconnaître que chaque personne a des circonstances différentes et à attribuer les ressources et les possibilités exactement nécessaires pour atteindre un résultat égal.
- ⁸ Le présent rapport adopte une approche non binaire de l'égalité des genres et fait référence aux obstacles rencontrés non seulement par les femmes, mais aussi par les personnes ayant des OSIGEGCS différentes.
- ⁹ Ce principe est inscrit dans plusieurs instruments internationaux relatifs aux droits de la personne, comme la Déclaration universelle des droits de l'homme et la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (Bureau du Haut-Commissariat des Nations Unies aux droits de l'homme, 1979).
- ¹⁰ Voir la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (ONU Femmes, 2009), art. 2(3) et 7, et Bureau du Haut-Commissariat des Nations Unies aux droits de l'homme (2006) pour consulter des exemples.

- ¹¹ La description de cette section fait appel à la visualisation du cycle de vie du système d'IA par l'OCDE et est guidée par la description du cycle de vie de l'IA et ses différentes étapes par l'UNESCO, de la recherche et l'origine au démontage et à la mise hors service (UNESCO, 2024c, p. 10). Toutefois, elle met également l'accent sur la collecte des données, le prétraitement des données et la modélisation statistique (Ovalle et al., 2023b). La description des systèmes sociaux s'inspire des dimensions clés du Cadre de l'OCDE pour la classification des systèmes d'IA (OCDE, 2022).
- ¹² Le cadre est développé sur les cadres d'IA existants et va un peu plus loin en revendiquant le devoir de l'État concernant l'élimination active de la discrimination et la lutte contre l'injustice structurelle.
- ¹³ La devise « Rien sur nous sans nous » repose sur le principe de participation et a été utilisée par les organisations de personnes handicapées au fil des ans dans le cadre d'un mouvement international pour atteindre la pleine participation et l'égalisation des opportunités pour, par et avec les personnes handicapées. <https://www.un.org/esa/socdev/enable/iddp2004.htm>.
- ¹⁴ Conformément aux recommandations émises par le Rapporteur spécial sur les droits des personnes handicapées
- ¹⁵ Conformément aux recommandations émises par le Rapporteur spécial sur les droits des personnes handicapées
- ¹⁶ Voir, par exemple, l'initiative espagnole de développement d'un modèle de fondation entraîné en espagnol et dans les langues co-officielles, en partenariat avec des institutions comme le Centre de calcul intensif de Barcelone et l'académie espagnole de la langue, qui constitue un exemple de transparence et d'accessibilité (Computerworld Espagne, 2024).
- ¹⁷ Le principe 1.3 de l'OCDE sur la transparence et l'explicabilité souligne le besoin de faire face aux inquiétudes en matière de transparence via la promotion d'un discours public et la création d'organismes dédiés, si nécessaire, pour favoriser la compréhension et la sensibilisation aux systèmes d'IA. L'acceptation et la confiance dans les technologies de l'IA s'en trouvent renforcées.
- ¹⁸ Le pouvoir légal de prendre note d'une problématique sans recevoir de plainte officielle.



Références

ACM (2022), Statement on Principles for Responsible Algorithmic Systems, ACM Technology Policy Office, Washington, DC, <https://www.acm.org/binaries/content/assets/public-policy/final-joint-ai-statement-update.pdf>.

Badaloni, S. and A. Rodà (2022), « Gender knowledge and Artificial Intelligence », document présenté lors du 1st Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming, BEWARE-22, aux côtés d'AlxIA 2022, Université d'Udine, Udine, Italie.

Bartoletti, I. and R. Xenidis (2023), Étude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination, Commission pour l'égalité de genre (GEC) et Comité directeur sur l'anti-discrimination, la diversité et l'inclusion (CDA-DI) Conseil de l'Europe, Strasbourg, <https://edoc.coe.int/fr/intelligence-artificielle/11646-etude-sur-limpact-des-systemes-dintelligence-artificielle-leur-potentiel-de-promotion-de-legalite-y-compris-legalite-de-genre-et-les-risques-quils-peuvent-entraîner-en-matiere-de-non-discrimination.html>

Benjamin, R. (ed.) (2019), *Captivating technology: Race, carceral technoscience, and liberatory imagination in everyday life*, Duke University Press, Durham, NC.

Bommasani, R. et al. (2024), "The Foundation Model Transparency Index v1.1", <https://crfm.stanford.edu/fmti/paper.pdf>.

Buckup, S. (2009), *The price of exclusion: The economic consequences of excluding people with disabilities from the world of work*, Employment Working Paper No. 43, ILO, https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcm_s_119305.pdf.

Buhl, N. (2023), Human-in-the-Loop Machine Learning (HITL) Explained, Encord Blog, <https://encord.com/blog/human-in-the-loop-ai/>.

Buolamwini, J. (2023), *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*, Random House, New York.

Buolamwini, J. and T. Gebru (2018), "Gender shades: Intersectional accuracy disparities in commercial gender classification", in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Association for Computing Machinery, New York, <https://proceedings.mlr.press/v81/buolamwini18a.html>.

Cachat-Rosset, G. and A. Klarsfeld (2023), "Diversity, equity, and inclusion in Artificial Intelligence: An evaluation of guidelines", *Applied Artificial Intelligence*, Vol. 37(1), article 2176618, <https://doi.org/10.1080/08839514.2023.2176618>.

Campbell, M. (2015), "CEDAW and women's intersecting identities: A pioneering new approach to intersectional discrimination", *Revista Direito GV*, Vol. 11(2), pp. 479–504, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2726534.

Chauhan, P. and N. Kshetri (2022), "The role of data and artificial intelligence in driving diversity, equity, and inclusion", *Computer*, Vol. 55, pp. 88–93, <https://ieeexplore.ieee.org/abstract/document/9755238>.

Chen, B.J. and J. Metcalf (2024), *Explainer: A Sociotechnical Approach to AI Policy*, Data & Society Policy Brief, https://datasociety.net/wp-content/uploads/2024/05/DS_Sociotechnical-Approach_to_AI_Policy.pdf.

Cheong, M.C.Y. K. Leins and S. Coghlan (2021), "Computer science communities: Who is speaking, and who is listening to the women? Using an ethics of care to promote diverse voices", in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, <https://api.semanticscholar.org/CorpusID:231639365>.



- Cofone, I. (2023), *The Privacy Fallacy: Harm and Power in the Information Economy*, Cambridge University Press, Cambridge, UK, <https://www.cambridge.org/core/books/privacy-fallacy/547578F2A1AE0C40963105CE066B412E>.
- Computerworld Spain (2024), “Spain will create foundational AI model in local languages”, *Computerworld*, <https://www.computerworld.com/article/1612218/spain-will-create-foundational-ai-model-in-local-languages.html>
- Costanza-Chock, S. (2018), “Design justice, A.I., and escape from the matrix of domination”, *Journal of Design and Science*, <https://doi.org/10.21428/96c8d426>.
- Dankwa-Mullan, I. et al. (2021), “A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle”, *Journal of Health Care for the Poor and Underserved*, Vol. 32, pp. 300–317, <https://muse.jhu.edu/article/789672>.
- Deck, L. et al. (2024), “Implications of the AI Act for Non-Discrimination Law and Algorithmic Fairness”, <https://arxiv.org/html/2403.20089v1>.
- Dignum, V. (2023).” Responsible artificial intelligence: Recommendations and lessons learned”, in D. O. Eke, K. Wakunuma, and S. Akintoye (eds.), *Responsible AI in Africa: Challenges and Opportunities*, Springer International Publishing, Cham, Switzerland, https://doi.org/10.1007/978-3-031-08215-3_9.
- Donnelly, N. and L. Stapleton (2022), “The social impact of data processing: The case of gender mapped to sex”, *IFAC-PapersOnLine*, Vol. 55(39), pp. 117–122, <https://doi.org/10.1016/j.ifacol.2022.12.021>.
- Dumbrava, C. (2021), « L'intelligence artificielle aux frontières de l'Union européenne: Aperçu des applications et questions clés », analyse approfondie, Think Tank, Communauté européenne, [https://www.europarl.europa.eu/think-tank/fr/document/EPRS_IDA\(2021\)690706](https://www.europarl.europa.eu/think-tank/fr/document/EPRS_IDA(2021)690706).
- Dwork, C. et al. (2012), « Fairness through awareness », dans *ITCS '12: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, <https://doi.org/10.1145/2090236.2090255>.
- Equinet (2023), Ensuring European AI that Protects and Promotes Equality for All, <https://equineteurope.org/wp-content/uploads/2023/07/Recommendations-amendments.pdf>.
- Eubanks, V. (2017), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York.
- Commission européenne (2022), *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, Commission européenne, Bruxelles, <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52022PC0496>.
- Parlement européen (2023), Briefing, EU Legislation in Progress, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI\(2023\)739342_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf).
- Fink, M. and M. Finck (2022), “Reasoned A(I)dmistration: explanation requirements in EU law and the automation of public administration”, *European Law Review*, Vol. 47(3), pp. 376–392.
- Forster, M. (2022), *Refugee Protection in the Artificial Intelligence Era. A Test Case for Rights*, Chatham House, London.
- Fredman, S. (2016), “Substantive equality revisited”, *International Journal of Constitutional Law*, Vol. 14(3), pp. 712–738, <https://doi.org/10.1093/icon/mow043>.
- Fricker, M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford Academic, Oxford.



Goldschmidt, J.E. (2017), “New perspectives on equality: Towards transformative justice through the Disability Convention?”, *Nordic Journal of Human Rights*, Vol. 35(1), pp. 1–14, <https://doi.org/10.1080/18918131.2017.1286131>.

Global Center on AI Governance (2024), Methodology: Our Multidimensional Framework, *Global Index on AI*, <https://www.global-index.ai/methodology>

Grobelnik, M. et al. (2024), “What is AI? Can you make a clear distinction between AI and non-AI systems?”, *The AI Wonk*, OECD.AI Policy Observatory, <https://oecd.ai/en/wonk/definition>.

Grochowski, M. et al. (2021), “Algorithmic transparency and explainability for EU consumer protection: Unwrapping the regulatory premises”, *Critical Analysis of Law*, Vol. 8(1), pp. 43–63, <https://doi.org/10.33137/cal.v8i1.36279>.

Gurumurthy, A. and B. Deepti (2023), “Reframing AI governance through a political economy lens”, *IT for Change*, https://itforchange.net/sites/default/files/add/IT%20for%20Change_Reframing%20AI%20Governance%20through%20a%20Political%20Economy%20Lens.pdf.

Hacker, P. (2024), Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence: Complementary Impact Assessment, *European Parliamentary Research Service*, [https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU\(2024\)762861_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf).

Hellman, D. (2023), “Big data and compounding injustice”, *Journal of Moral Philosophy*, Vol. 21(1-2), pp. 62–83, <https://doi.org/10.1163/17455243-20234373>.

Hickok, M. (2024), “Public procurement of artificial intelligence systems: New risks and future proofing”, *AI & Society*, Vol. 39, pp. 1213–1227, <https://doi.org/10.1007/s00146-022-01572-2>.

Hull, G. (2023), “Dirty data labeled dirt cheap: Epistemic injustice in machine learning systems”, *Ethics and Information Technology*, Vol. 25(3), article 38, <https://doi.org/10.1007/s10676-023-09712-y>.

Iason, G. (2022), “Toward a theory of justice for artificial intelligence”, *Daedalus*, Vol. 151(2), pp. 218–231, https://doi.org/10.1162/daed_a_01911

International Women’s Day (2023), “#EmbraceEquity for IWD 2023 and beyond?”, <https://www.internationalwomens-day.com/Missions/18707/Equality-versus-Equity-What-s-the-difference-as-we-EmbraceEquity-for-IWD-2023-and-beyond>.

IWRAW (sans date), Substantive equality, <https://cedaw.iwraw-ap.org/cedaw/cedaw-principles/cedaw-principles-overview/substantive-equality/>.

Jakubowska, E. et al. (2024), “EU’s AI Act fails to set gold standard for human rights”, Joint analysis, <https://www.amnesty.eu/wp-content/uploads/2024/04/EUs-AI-Act-fails-to-set-gold-standard-for-human-rights.pdf>.

Jora, R.B. et al. (2022), « Role of Artificial Intelligence (AI) in meeting Diversity, Equality and Inclusion (DEI) goals », dans *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1.

Joyce, K. et al. (2021), “Toward a sociology of artificial intelligence: A call for research on inequalities and structural change”, *Socius*, Vol. 7, article 2378023121999581, <https://doi.org/10.1177/2378023121999581>.

Kilic, B. (2024), Into Uncharted Waters: Trade Secrets Law in the AI Era, CIGI Papers No. 295, *Centre for International Governance Innovation*, <https://www.cigionline.org/static/documents/no.295.pdf>.

Kong, Y. (2022), « Are ‘intersectionally fair’ AI algorithms really fair to women of color? A philosophical analysis », dans *FACCT ’22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, <https://doi.org/10.1145/3531146.3533114>.



Kossow, N., S. Windwehr and M. Jenkins (2021), “Algorithmic transparency and accountability”, Transparency International Anti-Corruption Helpdesk Answer, Transparency International, https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf.

Krishna, V.K. (2020), “Open science and its enemies: Challenges for a sustainable science–society social contract”, *Journal of Open Innovation: Technology, Market, and Complexity*, Vol. 6(3), article 61, <https://doi.org/10.3390/joit-mc6030061>.

Martineau, W., N. Meer and S. Thompson (2012), “Theory and practice in the politics of recognition and misrecognition”, *Res Publica*, Vol. 18, pp. 1–9, <https://doi.org/10.1007/s11158-012-9181-7>.

Mathews, H.V. and A. Sinha (2020), “Use of Algorithmic Techniques for Law Enforcement”, *Economic and Political Weekly*, Vol. 55(23), <https://www.epw.in/journal/2020/23/special-articles/use-algorithmic-techniques-law-enforcement.html>.

McGregor, M. et P. Molnar (2023), *Digital Border Governance: A Human Rights Based Approach*, Université d’Essex et Haut-Commissaire aux droits de l’homme des Nations Unies, <https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf>.

Molnar, P. (2024), « The deadly digital frontiers at the border », *Time*, 21 mai, <https://time.com/6979557/unregulated-border-technology-migration-essay/>.

Nihei, M. (2022), « Epistemic injustice as a philosophical conception for considering fairness and diversity in human-centered AI principles », *Interdisciplinary Information Sciences*, vol. 28(1), p. 35–43, <https://doi.org/10.4036/iis.2022.A.01>.

Noble, S. et S. Roberts (2019), « Technological elites, the meritocracy, and postracial myths in Silicon Valley », dans *RACISM POSTRACE*, R. Mukherjee, S. Banet-Weiser et H. Gray (eds.), Racism Postrace, Duke University Press, Durham, NC. Rapport #: 6. Extrait de <https://escholarship.org/uc/item/7z3629nh>.

OCDE (2021), Gouvernement ouvert et participation citoyenne, <https://www.oecd.org/fr/themes/gouvernement-ouvert-et-participation-citoyenne.html>.

OCDE (2022), Cadre de l’OCDE pour la classification des systèmes de l’IA, OCDE. Observatoire des politiques de l’IA, https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/02/oecd-framework-for-the-classification-of-ai-systems_336a8b57/cb6d9eca-en.pdf.

OCDE (2024a), Présentation des principes sur l’IA, OCDE.AI Policy Observatory, <https://oecd.ai/fr/ai-principles>.

OCDE (2024b), « Explanatory memorandum on the updated OECD definition of an AI system », *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, <https://doi.org/10.1787/623da898-en>.

Bureau du Haut-Commissaire des Nations Unies aux droits de l’homme (1965), *Convention internationale sur l’élimination de toutes les formes de discrimination raciale*, <https://www.ohchr.org/fr/instruments-mechanisms/instruments/international-convention-elimination-all-forms-racial>

Bureau du Haut-Commissaire des Nations Unies aux droits de l’homme (1966) *Pacte international relatif aux droits économiques, sociaux et culturels*, <https://www.ohchr.org/fr/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights>

Bureau du Haut-Commissaire des Nations Unies aux droits de l’homme (1979), *Convention sur l’élimination de toutes les formes de discrimination à l’égard des femmes* New York, 18 décembre 1979, <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women>.



Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme (2023), *Digital Border Governance: A Human Rights Based Approach*, <https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf>.

Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme, LGBTI, https://www.ohchr.org/fr/topic/lgbti-people?gad_source=1.

Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme (2024b), *Les principaux instruments internationaux relatifs aux droits de l'homme et les organes chargés de suivre leur mise en œuvre* <https://www.ohchr.org/fr/core-international-human-rights-instruments-and-their-monitoring-bodies>.

Ostry, J.D. et al. (2018), « Economic gains from gender inclusion: New mechanisms, new evidence », IMF Staff Discussion Note, IMF, <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2018/10/09/Economic-Gains-From-Gender-Inclusion-New-Mechanisms-New-Evidence-45543>

Ovalle, A. D. Liang and A. Boyd (2023a), « Should they? Mobile biometrics and technopolicy meet queer community considerations », dans *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, et* avec <https://dl.acm.org/doi/abs/10.1145/3617694.3623255>.

Ovalle, A. et al. (2023b), « Factoring the matrix of domination: A critical review and reimagination of intersectionality in AI fairness », dans *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, <https://doi.org/10.1145/3600211.36047>.

Pande, R (2003). "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India.", *American Economic Review*, Vol. 93 (4), <https://www.aeaweb.org/articles?id=10.1257/000282803769206232>.

#ProtectNotSurveil coalition (2024), Joint statement – A dangerous precedent: How the EU AI Act fails migrants and people on the move, <https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/>.

Ricaurte, P. (2024), "A sociotechnical approach to the AI Lifecycle", *Feminist AI Research Network*, Forthcoming.

Ricaurte, P. (2019), "Data epistemologies, the coloniality of power, and resistance", *Television & New Media*, Vol. 20(4), pp. 350–365, <https://doi.org/10.1177/1527476419831640>.

Ricaurte, P. (2022). Ethics for the majority world: AI and the question of violence at scale. *Media, Culture & Society*, 44(4), 726-745. <https://doi.org/10.1177/01634437221099612>.

Ricaurte, P. and M. Zasso (2023), "AI, ethics, and coloniality: A feminist critique", in M. Cebral-Loureda, E.G. Rincón-Flores and G. Sanchez-Ante (eds.), *What AI Can Do*, Chapman and Hall/CRC, New York.

Rikap, C. (2022), "The expansionary strategies of intellectual monopolies: Google and the digitalization of healthcare", *Economy and Society*, Vol. 52(1), pp. 110–136, <https://doi.org/10.1080/03085147.2022.2131271>.

Rudin, C. and J. Radin (2019), "Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition", *Harvard Data Science Review*, Vol. 1, <https://doi.org/10.1162/99608f92.5a8a3a3d>.

Shapiro, J. (n.d.), "CEDAW as a tool for promoting substantive gender equality," <http://www.cedaw.org.tw/en/upload/media/Capacity%20Building/1-1The%20Mechanism%20of%20CEDAW%20Committee.pdf>.

Smith, R.A. (2023), "From LGBTQIA+ to SOGIESC: Reframing sexuality, gender, and human rights", Open Global Rights, New York.



- Smuha, N.A. (2021), "Beyond the individual: governing AI's societal harm", *Internet Policy Review*, Vol. 10(3), <https://doi.org/10.14763/2021.3.1574>.
- Solaiman, I. et al. (forthcoming), "Evaluating the social impact of generative AI systems in systems and society", in P. Hacker, A. Engel, S. Hammer and B. Mittelstadt (eds.), *Oxford Handbook on the Foundations and Regulation of Generative AI*, Oxford University Press, Oxford.
- Stauffer, B. (2023), Automated Neglect: How The World Bank's Push to Allocate Cash Assistance Using Algorithms Threatens Rights, Human Rights Watch, <https://www.hrw.org/report/2023/06/13/automated-neglect/how-world-banks-push-allocate-cash-assistance-using-algorithms>.
- Suresh, H. et al. (2022), "Towards intersectional feminist and participatory ML: A case study in supporting femicide counterdata collection", in *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, <https://api.semanticscholar.org/CorpusID:249872586>.
- Türk, V. (2023), "Artificial intelligence must be grounded in human rights, says High Commissioner", Office of the High Commissioner for Human Rights, Geneva.
- UNCTAD (2021), *Technology and Innovation Report 2021*, <https://unctad.org/page/technology-and-innovation-report-2021>.
- UNECE (sans date), Introduction, <https://unece.org/environment-policy/public-participation/aarhus-convention/introduction>.
- UNECE (1998), *UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters*, <https://unece.org/environment-policy/public-participation/aarhus-convention/text>.
- UNESCO (2009), « Concept paper on marginalization », document présenté lors de la dixième séance du Groupe de travail sur l'Éducation pour tous (EFA), Paris.
- UNESCO (2022), *Recommandation sur l'éthique de l'intelligence artificielle*, UNESCO, Paris, https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre.
- UNESCO (2024a), « IA générative : Une étude de l'UNESCO révèle la présence d'importants stéréotypes de genre » <https://www.unesco.org/fr/articles/ia-generative-une-etude-de-lunesco-revele-la-presence-dimportants-stereotypes-de-genre>.
- UNESCO (2024b), Statement W4EAI, Women for Ethical AI Action Request, SHS/2024/PI/H/1, <https://unesdoc.unesco.org/ark:/48223/pf0000388754/PDF/388754eng.pdf.multi>.
- UNESCO (2024c), Recommandation sur l'éthique de l'intelligence artificielle, https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre.
- UNESCO (2024d), Ethical Impact Assessment, Global AI Ethics and Governance Observatory, <https://www.unesco.org/ethics-ai/en/eia>.
- UNESCO, IDB and OECD (2022), *The Effects of AI on the Working Lives of Women*, UNESCO, Paris.
- UNHRC (2011), *Principes directeurs relatifs aux entreprises et aux droits de l'homme : mise en œuvre du cadre de référence « protéger, respecter et réparer » de Nations Unies* A/HRC/17/31, https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_FR.pdf.
- UNHRC (2020), *Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*, A/HRC/44/57, <http://undocs.org/en/A/HRC/44/57>.
- UNHRC (2021), *Report of the Special Rapporteur on the Rights of Persons with Disabilities*, A/HRC/49/52, <http://docs.org/en/A/HRC/49/52>.



UNICEF (2021), *UNICEF Gender Policy and Action Plan 2022-2025: Gender-Transformative Programming*, <https://www.unicef.org/lac/en/media/43146/file>.

United Nations, (2006), *Convention on the Rights of Persons with Disabilities and Optional Protocol*, <https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>.

Nations Unies (2019), *Stratégie des Nations Unies pour l'inclusion du handicap*, Nations Unies, New York, https://www.un.org/en/content/disabilitystrategy/assets/documentation/UN_Disability_Inclusion_Strategy_french.pdf.

United Nations (2021), *Report of the Secretary-General's Task Force on Addressing Racism and Promoting Dignity for All in the United Nations Secretariat*, United Nations, New York, https://hr.un.org/sites/hr.un.org/files/sap_final_report_0.pdf.

Nations Unies (2024), *Les 17 objectifs*, <https://sdgs.un.org/fr/goals>.

United Nations (2024b), *Global Digital Compact*, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf.

UNRISD (2017), *Transformative Policies for Sustainable Development: What Does It Take?*, <https://www.unrisd.org/en/library/publications/transformative-policies-for-sustainable-development-what-does-it-take>.

UN Women (2009), *Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)*, <https://www.un.org/womenwatch/daw/cedaw/text/econvention.htm#intro>.

UN Women (2015), "Chapter 1: Substantive equality for women: The challenge for public policy", in *Progress of the World's Women 2015–2016*, UN Women, <https://progress.unwomen.org/en/2015/pdf/ch1.pdf>.

UN Women (2019), *CEDAW-based Legal Review: A Brief Guide*, UN Women in Ukraine, Kyiv, https://eca.unwomen.org/sites/default/files/Field%20Office%20ECA/Attachments/Publications/2019/CEDAW%20Legal_ENG_compressed%20%281%29.pdf.

Vinuesa, R. et al. (2020), "The role of artificial intelligence in achieving the Sustainable Development Goals", *Nature Communications*, Vol. 11(1), article 1, <https://doi.org/10.1038/s41467-019-14108-y>.

West, S.M. (2020), "Redistribution and recognition: A feminist critique of algorithmic fairness", *Catalyst: Feminism, Theory, Technoscience*, Vol. 6(2), <https://doi.org/10.28968/cftt.v6i2.33043>.

West, S.M., M. Whitaker and K. Crawford (2019), *Discriminating Systems: Gender, Race, and Power in AI*, AI Now Institute, <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2>.

Maison Blanche (2022), *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, White House Office of Science and Technology Policy, <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>.

Winfield, A.F.T. et al. (2021), "IEEE P7001: A Proposed Standard on Transparency", *Frontiers in Robotics and AI*, Vol. 8, article 665729, <https://doi.org/10.3389/frobt.2021.665729>.

Commission mondiale d'éthique des connaissances scientifiques et des technologies (2005), *The Precautionary Principle*, UNESCO, Paris, <https://unesdoc.unesco.org/ark:/48223/pf0000139578>.

Forum économique mondial (2024), *AI Value Alignment: Guiding Artificial Intelligence Towards Shared Human Goals*, White Paper, World Economic Forum, https://www3.weforum.org/docs/WEF_AI_Value_Alignment_2024.pdf.



Annexe : liste des participant·e·s aux consultations régionales

Voici une liste des participant·e·s aux consultations régionales ayant donné leur consentement pour que leur nom figure dans le rapport.

Les consultations régionales ci-dessous ont été menées par Data Pop Alliance.

Asie

1. **Apekchya Rana Khatri**,
UNFPA Népal
2. **Gaurav Sharma**,
GIZ Inde
3. **Dr. Preeti Raghunath**,
Université de Sheffield
4. **Philippe Doneys**,
Asian Institute of Technology
5. **Hazel T. Biana**,
Chercheuse, Centre de recherche sur le développement social, Université De La Salle
6. **Manila Hoang**,
Ministre des Affaires étrangères du Vietnam
7. **Eleonore Fournier-Tombs**,
Université des Nations Unies
8. **Omran Najjar**,
Humanitarian OSM StreetMap Team

Moyen-Orient et Afrique du Nord

1. **Mohammad Reza Bateni**,
Université technologique d'Amirkabir
2. **Hayfa Khalfaoui**
3. **Maha Jouini**,
African Center for AI and Digital Technology
4. **Sukaina Al-Nasrawi**,
Commission économique et sociale des Nations Unies pour l'Asie occidentale (CESAO)

5. **Fouad Mrad**
6. **Grace S. Thomson**,
Center for AI and Digital Policy (CAIDP)
7. **Ibrahim Helmy Emara**,
Université de Tanta
8. **Dr. Houda Chihi**,
Tunisia Telecom
9. **Limor Ziv**,
Humane AI
10. **Prof. Hazem Abdelazim**,
Université ESLSCA
11. **Arwa ElGhadban**,
Université technique de Munich
12. **Ihab Shouly**

Amérique du Nord et Europe

1. **Mariana Roza Paz**,
Datasphere Initiative
2. **Yavar Baradaran-Khosrghoshahi**
3. **Till Koebe**,
Campus informatique de La Sarre
4. **H. Gijs van den Dool**,
Freelancer/Chercheur indépendant
5. **Luisa Olaya**,
GIZ (Fair Forward)
6. **Ruth Schmidt**,
GIZ (Fair Forward)



Les consultations régionales ci-dessous ont été menées par Derechos Digitales.

Amérique latine et Caraïbes

1. Maria Encalada Córdova
2. Tarcizio Roberto da Silva
3. Luz Elena González
4. Edson Prestes
5. Marianela Milanes
6. Silvana Bahia
7. Gisele da Silva Craveiro
8. Eduardo Carrillo
9. Lutiana Valadares Fernandes Barbosa
10. Zinnya del Villar Islas
11. Cristina Pombo
12. Viviane Ceolin Dallasta Del Grossi
13. Danielle Costa Carrara Couto.
14. Luis Fernando Arias
15. Silvana Fumega
16. Valentina Rozo Ángel
17. Paz Peña
18. Daniela Camila de Araújo
19. Fernanda K. Martins
20. Virginia Pardo
21. Hugo Alvaro Miranda Colque
22. Virginia Brussa
23. Paola Ricaurte Quijano
24. Fernanda Campagnucci
25. Gustavo Macedo
26. Maximiliano Maneiro Vaz
27. Andres Lombana Bermudez
28. Ivana Feldfeber
29. Augusto Luciano Mathurin
30. Iván Terceros

31. Eliana Quiroz
32. María Lorena Flórez Rojas
33. Natali Viezna Figueroa Figueroa
34. Juan de Brigard
35. Juan Pablo Marín Díaz
36. Marcela Mattiuzzo
37. Thiane de Nazaré



Les consultations régionales ci-dessous ont été menées par Paola Ricaurte Quijano.

Amérique latine et Caraïbes

1. Gerardo López Gómez
2. José Alfredo Hau Caamal
3. Verónica Aguilar Martínez
4. Jhonnatan Rangel
5. Marcos Cornelio Sánchez Ramírez
6. Miriam Hernández
7. Kupijy Vargas
8. Manuela Yah Panti
9. Yamili Chan Canul
10. Andrés Ta Chikinib
11. Mariana Díaz

Les consultations régionales ci-dessous ont été menées par ICT Research Africa.

Afrique subsaharienne

1. Sandra Aceng
2. Adedeji Adeniran
3. Adebisi Adewusi
4. Shariffa Amolo
5. Houda ChiChi
6. Kudakwashe Dandajena
7. Laetitia Delaunay-Badolo
8. Khetsiwe Dlamini
9. Sinclair Ebimowei
10. Alison Gillwald
12. Rigobert Kenmogne
13. Jonas Kgomo
14. Slindile Khumalo
15. Cheruiyot Kiprono

16. Anja Kovacs
17. Michael Magoronga
18. Lorraine Matanda
19. Tholang Mathapo
20. Bhekani Mbuli
21. Katleho Mokoena
22. Margaret Muga
23. Judith Murungi
24. Ernest Mwebaze
25. Angella Ndaka
26. Nelson Nkari
27. Adaeze P. Nwoba
28. Dorcas Nyamwaya
29. Durnebi Okwasah
29. Titilola Helen Olojede
30. Eniola Olowu
31. Lavina Ramkisoorn
32. Kelly Stone
33. Oluwatoyin Taiwo
34. Cynthia Tapera
35. Moses Thiga
36. Israel Olatunji Tijani
37. Simone Toussi
38. Josephine Zingani
39. Mpho Moyo
40. Pierrine Leukes