



Assemblée générale

Distr. générale
17 juillet 2024
Français
Original : anglais

Assemblée générale
Soixante-dix-neuvième session
Point 71 b) de l'ordre du jour provisoire*
Promotion et protection des droits humains :
Questions relatives aux droits humains, y compris
les divers moyens de mieux assurer l'exercice
effectif des droits humains et des libertés
fondamentales

Droits de l'homme et solidarité internationale

Note du Secrétaire général

Le Secrétaire général a l'honneur de transmettre à l'Assemblée générale le rapport de l'Experte indépendante sur les droits de l'homme et la solidarité internationale, Cecilia M. Bailliet, conformément à la résolution [53/5](#) du Conseil des droits de l'homme.

* [A/79/150](#).



Rapport de l'Experte indépendante sur les droits de l'homme et la solidarité internationale, Cecilia M. Bailliet

Intelligence artificielle et solidarité internationale – vers une conception de l'intelligence artificielle centrée sur l'être humain, au service de la solidarité internationale

Résumé

Le présent rapport est le premier établi à l'intention de l'Assemblée générale par l'Experte indépendante sur les droits de l'homme et la solidarité internationale, Cecilia M. Bailliet. Il est soumis en application de la résolution [53/5](#) du Conseil des droits de l'homme. L'Experte indépendante y examine le problème de l'aggravation des inégalités dans le monde qui se pose à l'heure actuelle, et explique comment la conception d'une stratégie de solidarité internationale en matière d'intelligence artificielle peut clarifier les devoirs des États, des entreprises et de la société civile en vue de renforcer l'égalité d'accès à la technologie et de lutter contre la discrimination par la prise en compte des groupes et des personnes en situation de vulnérabilité. Dans ce rapport, elle soutient que la gouvernance de l'intelligence artificielle devrait appuyer les mécanismes procéduraux de diligence raisonnable mis en place par les États et les entreprises, l'objectif étant d'associer les parties prenantes directes et indirectes au traitement des données et à la prise de décisions tout au long du cycle de vie des systèmes d'intelligence artificielle.

I. Introduction

1. La solidarité internationale est un principe du droit international et une valeur universelle qui suscite l'intérêt des décideurs politiques soucieux de résoudre les problèmes mondiaux, tels que ceux liés aux pandémies, aux changements climatiques, aux inégalités et aux énergies propres¹. Dans le monde entier, on voit émerger une volonté de s'éloigner du néolibéralisme inégalitaire pour tendre vers un modèle reposant sur la gestion responsable, la solidarité et l'action collective, porté par une révolution fondée sur des valeurs². Il est essentiel que l'intelligence artificielle (IA) soit utilisée pour unir l'humanité, et non pour la diviser. Conformément aux droits humains, l'IA devrait contribuer à promouvoir des actions collectives de solidarité internationale par-delà des frontières³. En 2022, l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) a publié une recommandation sur l'éthique de l'intelligence artificielle dans laquelle elle souligne qu'il est nécessaire de renforcer la solidarité pour favoriser un accès équitable aux technologies de l'IA et aborder les défis qu'elles posent dans des cultures et systèmes éthiques divers et interdépendants. En outre, la solidarité peut contribuer à réduire les risques d'utilisation abusive de l'IA, à exploiter pleinement les possibilités offertes par l'IA et à faire en sorte que les stratégies nationales relatives à l'IA soient guidées par des principes éthiques.

2. Comme l'indique une étude, l'IA est de plus en plus intégrée dans des sociétés où des « villes intelligentes » sont créées pour faciliter la possibilité de contrôler et de surveiller tous les aspects de la vie des citoyens, dans des contextes où l'enregistrement biométrique se généralise. Elle est également utilisée comme outil dans le cadre du développement international et des interventions humanitaires menées pour faire face aux déplacements et à d'autres crises⁴. L'utilisation de l'IA par les institutions publiques et leurs sous-traitants (y compris les entreprises et les organisations de la société civile) pour traiter des affaires et des plaintes et prendre des décisions s'y rapportant, ainsi que l'exploitation des données personnelles, devraient augmenter de manière significative partout dans le monde. La surveillance par l'IA porte particulièrement atteinte aux pauvres, car les institutions de l'État utilisent l'IA pour détecter les cas de recours abusif aux prestations sociales et contrôler l'immigration irrégulière, et comme outil de répression policière excessive dans les quartiers marginalisés⁵. On ne saurait trop insister sur l'importance de reconnaître les facteurs de vulnérabilité intersectorielle face à la discrimination causée par l'IA, notamment la race, l'appartenance ethnique, la religion, le genre, le lieu de résidence, la nationalité et le statut socio-économique⁶.

3. En outre, l'*Artificial Intelligence Index Report 2024* a conclu que « les développeurs d'IA manquaient de transparence, notamment en ce qui concerne la divulgation des données et des méthodes d'entraînement, et qu'un tel manque d'ouverture entravait les efforts menés pour mieux comprendre la robustesse et la sécurité des systèmes d'IA⁷ ». Par ailleurs, les entreprises du secteur de

¹ Voir Global Nation, *Global Solidarity Report 2023* (septembre 2023).

² Kurt April, « AI-Induced Solidarity Economy: The Need for Stewardship Orientation », *Effective Executive*, vol. 26, n° 3 (2023).

³ Contribution du Forum sur la gouvernance d'Internet. Voir également <https://intgovforum.org/en/content/pnai-report>.

⁴ Linnet Taylor, « What is data justice? The case for connecting digital rights and freedoms globally », *Big Data and Society*, vol. 4, n°2 (juillet-décembre 2017).

⁵ Ibid.

⁶ Ibid.

⁷ Institute for Human-Centered AI, Université Stanford, *Artificial Intelligence Index Report 2024* (Université Stanford, Californie, 2024).

l'informatique et des communications collaborent davantage avec les organisations de la société civile du monde du Nord, ce qui risque d'affaiblir les liens de solidarité avec les organisations du monde du Sud⁸. L'action menée pour corriger ce déséquilibre est compliquée par la participation limitée des groupes du monde du Sud aux réseaux qui examinent la responsabilité de l'IA dans les violations des droits humains.

4. Aux fins de l'élaboration du présent rapport, l'Experte indépendante a invité les États membres, les organisations de la société civile, les entreprises et les universitaires à lui soumettre des contributions. Une consultation a été organisée auprès des acteurs de la société civile en juillet 2024. Un examen de la littérature universitaire et des rapports des organes conventionnels et des organes créés par la Charte des Nations Unies a été mené, ainsi qu'une analyse des normes juridiques internationales, régionales et nationales.

II. Critères de solidarité pour une intelligence artificielle responsable : transparence, équité, non-discrimination et inclusion

5. Étant donné que le traitement des données et les pouvoirs de décision sont de plus en plus délégués à l'IA, il est fondamental de faire preuve de transparence pour garantir l'équité, la non-discrimination et l'inclusion. L'*Artificial Intelligence Index Report 2024* donne la définition suivante de l'équité : « Le fait de créer des algorithmes équitables, en évitant les biais ou la discrimination, et en tenant compte des divers besoins et situations de toutes les parties prenantes, de manière à s'aligner sur des normes sociétales plus larges en matière d'équité⁹ ». Dans le rapport, il est également indiqué ce qui suit : « Les résultats montrent que la plupart des entreprises ont pleinement appliqué au moins une mesure en faveur de l'équité, mais que les efforts visant à mettre en place une intégration globale sont toujours insuffisants. Au niveau mondial, les entreprises avaient en moyenne adopté 1,97 mesure d'équité sur les cinq proposées », d'après les données recueillies dans le cadre de l'enquête sur l'état mondial de l'IA responsable. La World Benchmarking Alliance a noté qu'en 2023, seulement un quart des 200 entreprises les plus influentes au monde dans le secteur de la technologie respectaient les critères minimaux concernant l'adoption de principes d'IA éthique en matière d'inclusion numérique¹⁰.

6. Dans l'« Appel de Rome pour une éthique de l'IA », il est indiqué que l'inclusion est un principe clé, et lors de l'événement intitulé « L'éthique de l'IA au service de la paix », organisé à Hiroshima en juillet 2024, le cheikh Abdallah Bin Bayyah, Président du Forum d'Abou Dhabi pour la paix, a souligné que « la coopération, la solidarité et la collaboration étaient nécessaires pour faire face à l'évolution de l'intelligence artificielle, qui fait coexister divers intérêts, effets néfastes et avantages, afin de s'assurer que ses systèmes et produits ne sont pas seulement avancés sur le plan technique, mais aussi fondés sur des valeurs morales¹¹ ». Le projet de document du pacte numérique mondial en date du 26 juin 2024 mentionne spécifiquement un objectif d'inclusion : « Notre coopération

⁸ Voir <https://www.business-humanrights.org/en/from-us/briefings/dismantling-the-facade-a-global-south-perspective-on-the-state-of-engagement-with-tech-companies/dismantling-the-facade-a-global-south-perspective-on-the-state-of-engagement-with-tech-companies/>.

⁹ *Artificial Intelligence Index Report 2024*.

¹⁰ Voir <https://www.worldbenchmarkingalliance.org/impact/investor-statement-for-ethical-ai-2024/>.

¹¹ Voir <https://www.romecall.org/> et <https://www.romecall.org/ai-ethics-for-peace-hiroshima-july-9th-2024/>.

contribuera à combler les fossés numériques, à l'intérieur des États et entre eux, et à promouvoir un environnement numérique qui favorise la diversité¹² ».

7. Conformément aux Principes directeurs relatifs aux entreprises et aux droits de l'homme, les entreprises du secteur de la technologie devraient procéder régulièrement à des études d'impact sur les droits humains, recenser les risques dans le cadre de leurs processus de contrôle de la qualité, et associer les parties prenantes externes à la réalisation desdites études d'impact¹³. On pourrait proposer l'adoption d'une stratégie de solidarité internationale en matière d'IA qui mettrait l'accent sur la nécessité d'établir des procédures concrètes visant à lutter contre la discrimination et à renforcer l'inclusion. Il conviendrait de prendre en compte la solidarité intergénérationnelle dans les réglementations en matière d'IA¹⁴.

8. La solidarité sociale peut être invoquée pour justifier le partage des avantages et des coûts liés au déploiement de l'IA, promouvoir la diversité des trajectoires de l'IA et favoriser la transparence et le respect des règles en vue de corriger les asymétries d'information propres à l'IA¹⁵. L'Experte indépendante souligne qu'il faut prendre en compte la solidarité lorsque l'on évalue la conception et la mise en œuvre de l'IA.

A. Des algorithmes inclusifs et non discriminatoires pour une solidarité en matière d'intelligence artificielle

9. Il est essentiel de réglementer l'IA afin de remédier aux biais et à la discrimination et de garantir la sécurité. L'IA peut être un moyen d'émancipation car elle permet d'accéder à l'éducation dans des zones reculées. Elle peut également faciliter l'accessibilité linguistique grâce aux fonctions de traduction et être utilisée pour lutter contre les stéréotypes et les discours de haine, d'où la nécessité de continuer à soutenir la recherche sur l'incidence sociale de l'IA. Les institutions qui conçoivent des systèmes d'IA utilisés dans les processus décisionnels (notamment dans les domaines judiciaire et administratif) devraient solliciter la contribution des groupes vulnérables et des acteurs de la société civile œuvrant pour la démocratie, et ce à toutes les étapes, de la planification à la mise en application, afin de prévenir les violations des droits humains et d'atténuer les effets préjudiciables. En outre, il importe de créer un mécanisme de contrôle indépendant pour la protection des données, qui soit capable de réglementer la collecte et le traitement des données personnelles aux différentes étapes de la planification et du déploiement de l'IA, et d'élaborer des lignes directrices à cet égard. L'Experte indépendante soutient le modèle de disparité démographique conditionnelle, qui sert de norme de référence

¹² Voir le projet de document en date du 26 juin, consultable à l'adresse suivante : <https://www.un.org/techenvoy/fr/global-digital-compact>.

¹³ Kate Jones, *AI Governance and Human Rights: Resetting the relationship*, étude, Programme de droit international (Londres, Institut royal des affaires internationales, 10 janvier 2023).

¹⁴ Sébastien Fassiaux, « Preserving consumer autonomy through European Union regulation of artificial intelligence: a long-term approach », *European Journal of Risk Regulation*, vol. 14, édition spéciale n° 4 (décembre 2023). Voir également Jon Truby et al., « A sandbox approach to regulating high-risk artificial intelligence applications », *European Journal of Risk Regulation*, vol. 13, n°2 (juin 2022), article reconnaissant la solidarité comme principe dans le cadre de la réglementation de l'Union européenne relative à l'IA.

¹⁵ Juan C. Mateos-García, « The Complex Economics of Artificial Intelligence » (2 décembre 2018) (document de travail non publié).

pour l'évaluation des cas de discrimination automatisée fondée sur des données statistiques¹⁶.

10. L'auteur de l'une des contributions énonce l'idée d'une IA inclusive qui respecte le principe de solidarité internationale, en créant des plateformes visant à amplifier la voix des populations traditionnellement sous-représentées dans les dialogues mondiaux (IA appliquée à la traduction ou utilisation des médias sociaux par les activistes pour faciliter la mise en réseau), et en donnant à la société civile les moyens de demander des comptes aux pouvoirs publics, par exemple dans les cas de corruption¹⁷.

B. Normes nationales relatives aux droits de l'homme et à la non-discrimination

11. Au niveau national, il existe de nombreuses normes (dont la majorité ont une visée aspirationnelle, sont non contraignantes ou se trouvent à l'état de projet) qui s'appuient sur les principes d'égalité, d'inclusion et de non-discrimination, mais bien souvent, des compétences institutionnelles distinctes et des mécanismes procéduraux concrets n'ont pas été mis en place pour garantir leur application et offrir des voies de recours en cas de violation.

12. L'Australie a élaboré des principes éthiques volontaires et non contraignants en matière d'IA selon lesquels les systèmes d'IA doivent profiter aux individus, à la société et à l'environnement et respecter les droits de l'homme, la diversité et l'autonomie des personnes¹⁸. De la même manière, la Chine a adopté des mesures provisoires relatives à la gestion des services d'intelligence artificielle générative qui visent à prévenir la discrimination fondée sur l'origine ethnique, la religion, la nationalité, l'origine géographique, le genre, l'âge, la profession et l'état de santé, ainsi que des mesures destinées à améliorer la diversité des données d'entraînement et à favoriser les contenus non discriminatoires¹⁹. L'Inde a mis au point une stratégie nationale pour l'intelligence artificielle qui porte sur l'égalité et la non-discrimination²⁰. Elle prévoit que les personnes concernées par une décision doivent être traitées sur un pied d'égalité par les systèmes d'IA si elles se trouvent dans les mêmes circonstances, que ces systèmes ne doivent pas refuser d'accorder une possibilité à une personne qualifiée sur la base de son identité et qu'ils ne doivent pas aggraver les divisions historiques et sociales néfastes fondées sur la religion, la race, la caste, le genre, l'ascendance, le lieu de naissance ou de résidence dans des domaines tels que l'éducation, l'emploi ou l'accès à l'espace public. En outre, cette stratégie comprend un objectif consistant à empêcher que des personnes soient injustement exclues de services ou d'avantages. Le Brésil a présenté un projet de réglementation de l'IA contenant une norme de non-discrimination qui prévoit le droit de contester les décisions et de demander une intervention humaine et la correction

¹⁶ Sandra Wachter, Brent Mittelstadt et Chris Russell, « Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI », *Computer Law and Security Review*, vol. 41 (2021).

¹⁷ Contribution de Jake Okechukwu Effoduh, professeur adjoint à la Lincoln Alexander School of Law, Toronto Metropolitan University, Canada.

¹⁸ Voir <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>.

¹⁹ Voir <https://www.chinalawtranslate.com/en/generative-ai-interim/>.

²⁰ Voir <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf> ; <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf> et <https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf>.

des biais discriminatoires directs, indirects, illégaux ou abusifs²¹. L'Espagne a mis en place une autorité indépendante pour l'égalité de traitement et la non-discrimination qui exerce des fonctions de contrôle et encourage l'utilisation d'une IA éthique, fiable et respectueuse des droits fondamentaux²².

13. Les États-Unis d'Amérique ont élaboré un document intitulé « Blueprint for an AI bill of rights: making automated systems work for the American people » (plan pour une déclaration des droits sur l'IA : mettre les systèmes automatisés au service du peuple américain), qui soutient l'adoption d'un décret portant création d'une fonction indépendante d'évaluation et de communication de l'information afin de garantir la non-discrimination²³. Le Kenya dispose d'un projet de code de bonnes pratiques en matière d'IA qui prévoit la création d'un mécanisme visant à faire appliquer les normes de non-discrimination, notamment par la réalisation d'évaluations de l'équité et l'établissement de rapports sur les mesures prises pour remédier aux biais et lutter contre les effets discriminatoires²⁴.

C. Risques liés aux applications biaisées de l'intelligence artificielle

14. Il est admis qu'« il est fondamental d'assurer un accès abordable à Internet et de permettre une utilisation éclairée de cet outil pour favoriser le plein exercice de tous les droits humains et de toutes les libertés fondamentales, de la démocratie, du développement et la justice sociale²⁵ ». Ce constat a donné lieu à une série d'initiatives privées de promotion des droits numériques portant sur l'infrastructure, l'application et l'usage d'Internet, le but étant d'améliorer l'accessibilité. Les biais de l'IA soumettent des individus ou des groupes à des stéréotypes ou à des préjugés, des comparaisons entre eux étant établies au moyen d'algorithmes. Ces biais peuvent être présents dans la conception d'un système d'IA, dans la collecte et l'interprétation des données et dans les interactions directes et indirectes avec les parties prenantes. La Rapporteuse spéciale sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée a confirmé que « les problèmes liés aux données, les problèmes liés à la conception des algorithmes, l'utilisation de l'intelligence artificielle à des fins intentionnellement discriminatoires et les questions liées au principe de responsabilité » représentaient un défi permanent (voir [A/HRC/56/68](#)). Dans un rapport sur l'IA datant de 2019, l'UNESCO souligne que la solidarité est un élément indispensable à l'édification d'une « société du savoir » et explique qu'il est urgent de s'attaquer à la discrimination causée par l'IA, qu'elle résulte d'une programmation intentionnelle ou non, des données d'entraînement biaisées présentes dans les algorithmes d'apprentissage automatique, ou d'autres facteurs²⁶ :

²¹ Voir le projet de loi n° 2338 de 2023, disponible à l'adresse suivante :

<https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>.

²² Contribution de l'Espagne.

²³ Voir <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf> ; <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

²⁴ Voir https://www.dataguidance.com/sites/default/files/kebs-tc_094_n66_public_review_kenya_standard_dks_3007_ai_code_of_practice.pdf.

²⁵ George A. Walker, « Technology Law, Rights and Ethics – One Choice, One Future », *The International Lawyer*, vol. 56, n° 1 (2023).

²⁶ Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO), *Steering AI and Advanced ICTs for Knowledge Societies: A Rights, Openness, Access and Multi-stakeholder Perspective* (Paris, 2019), p. 63.

De nombreuses formes de discrimination peuvent être indirectes. Par exemple, un algorithme qui se base sur les habitudes d'utilisation du téléphone portable pour déterminer la solvabilité d'une personne est discriminatoire s'il attribue un risque de crédit élevé aux femmes appartenant à des communautés qui : i) utilisent peu le téléphone portable ; ii) ne possèdent pas de téléphone portable. La condition appliquée peut sembler juste et équitable, mais elle désavantage un groupe particulier. Les algorithmes peuvent causer et exacerber des formes de discrimination multiples. Les préjugés sociaux et politiques existants sont systématisés de nombreuses façons dans les algorithmes d'apprentissage automatique. En outre, il serait utile d'étudier les nouvelles formes de discrimination que l'IA pourrait entraîner, telles que l'exclusion pour des raisons de corrélations statistiques qui ne correspondent pas nécessairement à des caractéristiques sociales notables, mais qui sont néanmoins fortement liées à l'identité personnelle.

15. L'*Intelligence Index Report 2024* décrit la manière dont la segmentation des textes en unités lexicales à des fins d'analyse a une incidence négative sur les langues non occidentales, pour lesquelles l'IA est moins efficace²⁷. Certaines organisations de la société civile s'attaquent directement aux biais algorithmiques, à l'exclusion et à la discrimination contre les minorités raciales, ethniques ou religieuses, ainsi qu'à la discrimination liée au genre ou à d'autres caractéristiques identitaires, et s'emploient à améliorer l'accessibilité²⁸. Amnesty International et la National Association for the Advancement of Colored People ont exprimé leurs préoccupations concernant les outils de police prédictive²⁹, l'emploi de systèmes automatisés pour déterminer l'accès aux soins de santé et aux services sociaux, la surveillance des mouvements de réfugiés et de migrants, et les conséquences pour les groupes racialisés de l'utilisation d'outils d'IA biaisés dans la reconnaissance faciale et la détection des fraudes³⁰. Les personnes exposées à ces processus ne sont parfois pas au courant de l'utilisation de l'AI et n'en ont pas été informées, et elles n'ont souvent aucun recours contre le traitement discriminatoire dont elles sont victimes de la part des institutions de l'État. Les groupes de la société civile soulignent qu'il importe d'associer les parties prenantes à l'élaboration de la réglementation en matière d'IA. Il a été suggéré que l'Union européenne continuait d'utiliser l'IA, malgré les critiques, dans des systèmes de surveillance discriminatoires tels que les dispositifs d'évaluation des risques et d'analyse prédictive pour faciliter les renvois sommaires³¹, tandis que dans le contexte des migrations, l'IA est développée pour différents objectifs : a) la vérification de l'identité à l'aide des technologies de reconnaissance faciale ; b) l'estimation des risques ; c) la prise de décisions relatives à la détention ; d) la surveillance. Le risque que représente le fait de réduire au minimum les contrôles humains a été mis en évidence dans de tels contextes³².

²⁷ *Artificial Intelligence Index Report 2024*.

²⁸ Voir <https://digitalrightsfoundation.pk/wp-content/uploads/2021/03/Policy-brief-2.pdf> ; <https://www.hrw.org/report/2023/12/21/metabas-broken-promises/systemic-censorship-palestine-content-instagram-and> ; <https://www.apc.org/en/member/7amleh-arab-center-social-media-advancement> ; <https://smex.org/> ; <https://www.derechosdigitales.org/> et <https://www.accessnow.org/>.

²⁹ Pour plus d'information sur la manière dont la police prédictive entraîne une surveillance et un ciblage policier disproportionnés des communautés noires, voir : <https://naacp.org/resources/artificial-intelligence-predictive-policing-issue-brief>.

³⁰ Voir <https://www.amnesty.org/en/latest/campaigns/2024/01/the-urgent-but-difficult-task-of-regulating-artificial-intelligence/>.

³¹ Contribution d'Indira Boutier, Glasgow Caledonian University.

³² Ibid. Voir également Petra Molnar et Lex Gill, *Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System*, Programme

16. Étant donné que les institutions publiques et les entreprises qui utilisent l'IA pour le traitement des données et la prise de décision ne sont pas transparentes quant au développement, à l'entraînement, à l'utilisation et à l'application de cette technologie, les personnes ou les groupes qui dénoncent des violations des droits humains sont souvent incapables de produire des preuves établissant un lien avec l'IA et une violation commise ; or, celle-ci peut découler de la mise en œuvre d'une boucle de rétroaction rapide entre les différents systèmes étatiques connectés via des systèmes d'IA, qui donnerait lieu à un traitement ou à une prise de décision préjudiciable de la part de l'institution publique. Les criminels utilisent également l'IA pour s'approprier frauduleusement les transferts de fonds entre les migrants et leur famille, dans le pays d'origine comme dans le pays de résidence.

17. L'Assemblée générale a reconnu qu'il était nécessaire de lutter contre la discrimination algorithmique³³, de détecter les vulnérabilités, d'améliorer l'accessibilité et de prévoir des recours en cas de violation des droits de l'homme³⁴. L'utilisation de l'IA par des institutions publiques chargées des questions de migration, de criminalité, de santé et de prise en charge des personnes âgées a attiré l'attention des organes conventionnels des Nations Unies et des titulaires de mandat au titre d'une procédure spéciale. Le Comité pour l'élimination de la discrimination raciale a publié des observations finales dans lesquelles il s'est inquiété de l'effet discriminatoire de l'utilisation de l'IA dans le cadre des procédures d'asile³⁵. Le Comité pour l'élimination de la discrimination à l'égard des femmes a publié des observations finales dans lesquelles il a recommandé de prendre des mesures pour garantir que les systèmes de surveillance biométrique (algorithmique) et de profilage utilisés par les forces de l'ordre dans le cadre des enquêtes et de la lutte contre la criminalité soient exempts de stéréotypes de genre et pour éliminer les préjugés reproduits par l'intelligence artificielle et les technologies algorithmiques³⁶. L'Experte indépendante chargée de promouvoir l'exercice par les personnes âgées de tous les droits de l'homme a apprécié l'occasion qui lui a été offerte, lors de sa visite, de mieux comprendre l'utilité de l'intelligence artificielle dans les domaines du vieillissement, des soins et des services de santé, mais elle a également recommandé un examen de l'utilisation de la collecte de données, afin de garantir que celle-ci respecte les obligations en matière de non-discrimination à l'égard des personnes âgées d'origines diverses, ce qui suppose l'adoption d'une perspective solidaire en matière d'IA³⁷.

18. Les groupes de la société civile ont souligné que l'IA avait des conséquences négatives sur la capacité des militants LGBTQI de participer à des actions de solidarité.

international sur les droits humains (Faculté de droit, Université de Toronto) et Citizen Lab (Munk School of Global Affairs and Public Policy, Université de Toronto), 2018, p. 31-34.

³³ Résolution [78/265](#) de l'Assemblée générale, par. 6 h).

³⁴ Résolution [78/213](#) de l'Assemblée générale.

³⁵ [CERD/C/DEU/CO/23-26](#) par. 45 et 46.

³⁶ [CEDAW/C/ITA/CO/8](#), par. 26. Voir également les observations finales similaires figurant dans le document [CEDAW/C/DEU/CO/9](#), par. 27 et 28. Le Comité a exprimé ses préoccupations concernant l'IA et la violence de genre dans les observations finales contenues dans le document [CEDAW/C/ESP/CO/9](#), par. 21, 23, 31 et 33.

³⁷ [A/HRC/45/14/Add.1](#), par. 93.

III. Initiatives régionales concernant l'utilisation de l'intelligence artificielle dans les procédures institutionnelles

19. L'Organisation de coopération et de développement économiques (OCDE) a mis à jour ses principes sur l'intelligence artificielle en 2024 et souligné qu'il importait d'améliorer l'inclusion des populations sous-représentées, de réduire les inégalités économiques, sociales, de genre et d'autres inégalités, et de promouvoir la diversité, l'équité, la justice sociale, la transparence, le pouvoir d'action et le contrôle humain³⁸. L'intérêt croissant des institutions publiques pour l'utilisation de l'IA dans le traitement des dossiers et la prise de décisions pose un problème : bon nombre de ces systèmes sont conçus par des entreprises, ce qui peut entraîner un manque de transparence. Certains risques tels que les « hallucinations », contenus trompeurs générés par l'IA, peuvent être méconnus des institutions publiques, qui ne sont pas toujours au fait de la nécessité impérative d'effectuer des contrôles humains pour surveiller l'utilisation de l'information et détecter les biais algorithmiques. L'OCDE propose un catalogue d'outils et d'indicateurs au service de la fiabilité de l'IA, qui fournit des exemples variés d'approches techniques, éducatives et procédurales visant à favoriser la solidarité en matière d'IA³⁹.

20. Dans sa loi sur l'intelligence artificielle, l'Union européenne aborde la solidarité internationale en définissant les pratiques interdites en matière d'IA, aux alinéas b), c) et g) du paragraphe 1) de l'article 5. La Commission européenne a adopté une proposition de directive relative à la responsabilité destinée à être appliquée par les tribunaux nationaux, qui établit des normes en matière de preuve⁴⁰. En 2018, elle a approuvé la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, dans laquelle elle préconise la réalisation d'audits externes sur le traitement des données afin de repérer les discriminations⁴¹. Ayant constaté le risque élevé associé à l'utilisation de l'IA dans la gestion des migrations, des demandes d'asile et des contrôles aux frontières, ainsi que dans l'administration de la justice et les procédures démocratiques, elle a adopté la loi sur l'intelligence artificielle en 2024. Cette dernière prévoit la création d'un Bureau européen de l'intelligence artificielle (Bureau de l'IA) au sein de la Commission afin de s'assurer que les fournisseurs de modèles d'IA à usage général mettent effectivement en œuvre les dispositions de la loi et les respectent. Le Conseil européen de l'intelligence artificielle sera composé de représentants des États membres et d'entreprises. Le Bureau de l'IA peut inviter les fournisseurs de modèles d'IA à usage général et les autorités nationales compétentes à participer à l'élaboration de codes de bonnes pratiques, tandis que la société civile, les entreprises du secteur, les universités, les fournisseurs en aval et les experts indépendants peuvent soutenir le processus⁴². Cette norme manque de rigueur car elle ne donne pas suffisamment de détails concernant la capacité de la société civile à contribuer directement aux études d'impact.

³⁸ Voir <https://oecd.ai/en/ai-principles> ; <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

³⁹ Voir <https://oecd.ai/en/>.

⁴⁰ Voir <https://artificialintelligenceact.eu/the-act/>. Voir également la 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 relative à la responsabilité en matière d'IA). Disponible à l'adresse suivante : <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52022PC0496>.

⁴¹ Voir <https://www.coe.int/fr/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>.

⁴² Voir <https://artificialintelligenceact.eu/fr/high-level-summary/>.

21. Dans un document stratégique, la Cour de justice européenne a déterminé que l'utilisation future de l'IA pourrait entraîner des risques de discrimination : « L'un des principaux risques liés à l'adoption de la technologie de l'IA est l'introduction de préjugés involontaires lors de l'entraînement des modèles d'IA, qui pourrait avoir des effets discriminatoires non intentionnels⁴³ ». Le Sous-Comité aux droits de l'homme du Parlement européen a réalisé une analyse de l'IA dans laquelle il souligne le risque de renforcement de la discrimination structurelle : « Les systèmes d'IA peuvent perpétuer et accentuer les préjugés, et donner lieu à des discriminations dans des domaines divers tels que l'emploi, le maintien de l'ordre et l'évaluation du crédit. De nombreux éléments montrent que l'IA peut aggraver les disparités socioéconomiques en reflétant les préjugés présents dans les données d'entraînement ou la conception⁴⁴ ».

22. Le 17 mai 2024, le Comité des Ministres du Conseil de l'Europe a adopté la Convention-cadre sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'état de droit, qui sera ouverte à la signature en septembre 2024. L'article 10, qui porte sur l'égalité et la non-discrimination, présente l'idée d'adopter une approche fondée sur le cycle de vie afin de détecter et de corriger les préjugés de manière systématique, et indique qu'il faudrait évaluer la discrimination dans les systèmes d'IA tout au long de la période d'utilisation, conformément au droit applicable⁴⁵.

1. Chaque Partie adopte ou maintient des mesures visant à garantir le respect de l'égalité, y compris l'égalité de genre, et l'interdiction de la discrimination dans les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle, comme le prévoit le droit international et interne applicable.

2. Chaque Partie s'engage à adopter ou à maintenir des mesures visant à supprimer les inégalités dans les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle, afin d'obtenir des résultats impartiaux, justes et équitables, dans le cadre des obligations nationales et internationales qui lui incombent en matière de droits de l'homme.

23. D'autres initiatives régionales ont également été lancées pour promouvoir l'inclusion numérique, comme l'adoption de la résolution 580 (LXXVIII) 2024 sur les censures d'Internet et les élections en Afrique par la Commission africaine des droits de l'homme et des peuples⁴⁶. En novembre 2023, la Cour interaméricaine des droits de l'homme a organisé un dialogue avec des juges, au cours duquel ils ont examiné le document de l'UNESCO intitulé « Manuel de formation mondial : l'IA et l'état de droit pour le pouvoir judiciaire ». Le Manuel décrit les problèmes d'exclusion :

Le problème est que [...] les données peuvent être biaisées. Par exemple, les essais cliniques excluent souvent les femmes et les personnes de couleur, ce qui entraîne une représentation inadéquate des données. Cela peut avoir de graves conséquences si des algorithmes formés à l'aide de ces données sont utilisés pour analyser les images de la peau ou donner la priorité aux soins des patients. Par conséquent, il est crucial de veiller à ce que les algorithmes d'IA

⁴³ Voir https://curia.europa.eu/jcms/upload/docs/application/pdf/2023-11/cjeu_ai_strategy.pdf.

⁴⁴ Voir [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf) ; voir également Ahmet Bilal Aytekin, « Algorithmic bias in the context of European Union anti-discrimination directives », article présenté lors de l'atelier européen sur l'équité algorithmique, organisé du 7 au 9 juin 2023 à Winterthur (Suisse), et https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf.

⁴⁵ Voir <https://rm.coe.int/1680afae3d>.

⁴⁶ Voir <https://achpr.au.int/en/adopted-resolutions/580-internet-shutdowns-elections-africa-achpres580-lxxvii>.

soient formés à l'aide de données représentatives, afin d'éviter de tels biais et d'assurer des résultats équitables pour tous⁴⁷.

24. Le Manuel aborde également la question de la représentativité de l'IA :

Les fractures numériques dans de nombreux pays du Sud ont conduit à une « invisibilité des données », ce qui peut avoir un impact sur les groupes historiquement marginalisés comme les femmes, les castes, les communautés tribales, les minorités religieuses et linguistiques et les travailleurs migrants. L'utilité et la validité des algorithmes d'IA développés sur des données facilement disponibles peuvent être limitées par des biais perpétués par l'invisibilité des données. Cela souligne les exigences de transparence et de responsabilité algorithmiques.

Le Manuel met en évidence le problème de la discrimination par procuration, qui se pose par exemple lorsque les banques évaluent les demandes de prêt en se basant sur les codes postaux, le niveau d'éducation ou le revenu, qui peuvent être associés à des minorités ethniques ou raciales, le risque inhérent à cette méthode étant de perpétuer des politiques et des pratiques discriminatoires intersectionnelles. Il suggère que l'IA peut contribuer à repérer les biais qu'entretiennent les acteurs publics et privés, y compris les membres du pouvoir judiciaire, à l'égard des personnes et des groupes marginalisés.

25. Lors du dialogue interaméricain avec les juges, les discussions ont également porté sur l'outil d'évaluation de l'impact éthique de l'IA de l'UNESCO, dans lequel il est demandé que soit prise en compte « la diversité de l'équipe de projet d'IA, notamment – mais pas exclusivement – le genre, l'âge, la race, la couleur, l'ascendance, la langue, la religion, l'origine nationale, l'origine ethnique, l'origine sociale, la situation économique ou sociale, le handicap et l'orientation sexuelle, ainsi que la manière dont cela reflète la complexité et la diversité des utilisateurs potentiels, et les risques éventuels d'introduction de préjugés⁴⁸ ».

A. La solidarité au service de l'habileté numérique et de la lutte contre la désinformation

26. Les appels en faveur de la réglementation des « campagnes de désinformation électorale lancées par des extrémistes » se multiplient⁴⁹. La qualité des médias générés par l'IA s'étant améliorée, il est devenu difficile de détecter la désinformation et les faux contenus, et des informations toujours plus nombreuses font état de coupures d'Internet par les États, ce qui a des effets disproportionnés sur l'accès à l'information. Il est nécessaire de mettre en place une stratégie de solidarité internationale en faisant appel à la coopération sur les plans législatif, technique et éducatif, afin d'améliorer l'habileté numérique⁵⁰. Il a été proposé de faire jouer la solidarité réactive pour faire face aux situations d'urgence liées à l'utilisation malveillante de l'IA, par exemple « lorsque des milliers de vidéos hyper-truquées (« deepfake ») montrant des violences ethniques circulent lors d'une journée électorale dans un pays où un génocide a été perpétré par le passé⁵¹ ».

⁴⁷ Voir <https://nataliazuazo.com/2023/11/20/ai-and-the-rule-of-law-at-the-inter-american-court-of-human-rights/>. Voir également https://unesdoc.unesco.org/ark:/48223/pf0000387331_fre.

⁴⁸ Voir <https://unesdoc.unesco.org/ark:/48223/pf0000386276>.

⁴⁹ Contribution de Jake Okechukwu Effoduh.

⁵⁰ Ibid.

⁵¹ Miguel Luengo-Oroz, « Solidarity should be a core ethical principle of AI », *Nature Machine Intelligence*, vol. 1, (novembre 2019). Voir également Patrik Hummel et Matthias Braun, « Just

27. En 2022, l'Union européenne a adopté un règlement sur les services numériques, entré en vigueur en 2024, dans lequel elle demande aux grandes plateformes en ligne et aux grands moteurs de recherche en ligne d'utiliser des codes de conduite, et notamment de mettre en œuvre des mesures visant à lutter contre la désinformation (y compris la diffusion d'informations trompeuses et de fausses publicités politiques) dont le champ d'application serait ciblé et limité au strict nécessaire⁵². Le règlement soutient le code de bonnes pratiques renforcé contre la désinformation publié en 2022 (applicable aux entités autorégulées), qui comprend des orientations visant à démonétiser la diffusion d'éléments de désinformation, à garantir la transparence des publicités à caractère politique, à renforcer la coopération avec les spécialistes de la vérification des faits et à permettre aux chercheurs d'accéder plus facilement aux données⁵³. Cependant, une critique a été formulée, selon laquelle « bien que ces évolutions puissent avoir un effet unificateur, l'approche continentale en matière de désinformation est actuellement fragmentée⁵⁴ ». En outre, certains s'inquiètent des contradictions possibles entre les codes de conduite et les arrêts de la Cour européenne des droits de l'homme⁵⁵.

28. La loi sur l'intelligence artificielle de l'Union européenne comprend également des dispositions visant à réduire l'influence et l'impact des hypertrucages. Conformément au paragraphe 4 de l'article 50, le déployeur, à savoir la personne qui utilise un système d'IA, doit généralement indiquer qu'un contenu a été généré par une IA. Cependant, il convient d'examiner de près l'efficacité de la loi, dont le texte prévoit des exceptions à cette obligation de transparence.

29. L'Agence de développement de l'Union africaine-Nouveau Partenariat pour le développement de l'Afrique préconise la mise en place de protections juridiques et de cadres réglementaires pour lutter contre la discrimination algorithmique⁵⁶. Le Conseil exécutif de l'Union africaine est favorable à l'élaboration d'une stratégie continentale basée sur le cadre conceptuel qu'il a adopté en matière d'IA⁵⁷. Dans la région, la tendance est à la mise en œuvre de « mesures technologiques » visant à censurer des contenus ou à bloquer l'accès à ces derniers afin d'empêcher la diffusion d'informations fallacieuses ou de discours haineux, et à l'application de « mesures juridiques » visant à examiner, rédiger et adopter des projets de loi afin de réglementer les informations fallacieuses et les discours haineux. Dans les deux cas, il existe un risque que ces mesures soient appliquées de manière trop large ou qu'elles ciblent l'opposition politique, ce qui pourrait porter atteinte à l'indépendance des élections.

data? Solidarity and justice in data-driven medicine », *Life Sciences, Society and Policy*, vol. 16, n° 8 (2020).

⁵² Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques).

⁵³ Voir <https://digital-strategy.ec.europa.eu/fr/library/2022-strengthened-code-practice-disinformation>.

⁵⁴ Contribution de Dorothy Andersen, chercheuse en droits humains au sein du Programme de droit international et de droit comparé de la Faculté de droit de l'Université George Washington (Washington).

⁵⁵ Paolo Cavaliere, « The Truth in Fake News: How Disinformation Laws Are Reframing the Concepts of Truth and Accuracy on Digital Platforms », *European Convention on Human Rights Law Review*, vol. 3, n° 4 (novembre 2022). Voir également Ethan Shattock, « Fake News in Strasbourg: Electoral Disinformation and Freedom of Expression in the European Court of Human Rights », *European Journal of Law and Technology*, vol. 13, n°1 (2022). Voir également <https://edmo.eu/wp-content/uploads/2022/01/Case-law-for-policy-making-Report-2022.pdf>.

⁵⁶ Voir <https://www.nepad.org/publication/ai-and-future-of-work-africa-white-paper>.

⁵⁷ Voir <https://au.int/en/newsevents/20240419/multistakeholder-consultative-sessions-development-continental-strategy>.

30. L'Arabie saoudite a informé l'Experte indépendante de la création de l'Autorité nationale chargée des données et de l'intelligence artificielle, qui représente une avancée importante pour le grand modèle de langage en arabe (connu sous le nom d'ALLaM). Cette initiative vise à exploiter les technologies d'IA et les applications numériques pour favoriser la diversité culturelle au profit de l'ensemble de l'humanité. L'un des projets mis en œuvre, SauTech, est axé sur la localisation des technologies d'IA et la sauvegarde de la culture locale, et permet de reconnaître et de transformer la parole en texte dans un large éventail de dialectes arabes afin de formuler des solutions adaptées au contexte local⁵⁸. L'Autorité saoudienne chargée des données et de l'intelligence artificielle s'emploie à mieux faire connaître l'IA grâce à des programmes de formation. Dans sa contribution, la Colombie a également décrit les formations organisées à différents niveaux dans le domaine de l'IA⁵⁹.

B. Solidarité entre les genres et intelligence artificielle

31. On s'intéresse de plus en plus aux possibilités offertes par une approche de solidarité internationale pour lutter contre la discrimination fondée sur le genre dans le domaine de l'IA⁶⁰. L'*Artificial Intelligence Index Report 2024* a révélé que les pays européens comptaient moins de femmes diplômées que d'hommes dans le domaine de l'informatique, et que les progrès accomplis pour réduire les disparités fondées sur le genre étaient lents⁶¹. Le Comité pour l'élimination de la discrimination à l'égard des femmes a publié des observations finales relatives à plusieurs rapports d'États parties, dans lesquelles il s'est dit préoccupé par la sous-représentation des filles et des femmes dans le domaine de l'intelligence artificielle et par le fait qu'aucune mesure concrète n'avait été prise pour empêcher que des stéréotypes sexistes et des algorithmes reflétant des préjugés discriminatoires et violents à l'égard des femmes soient pris en considération dans la programmation des systèmes d'IA, et a demandé la mise en place de garanties effectives⁶².

32. Dans sa contribution, l'Arabie saoudite a indiqué que les femmes représentaient plus de 50 % du personnel technique de l'Autorité nationale chargée des données et de l'intelligence artificielle⁶³. En collaboration avec Google Cloud, le Centre international de recherche et d'éthique de l'intelligence artificielle du pays a lancé l'initiative Elevate, qui vise à former 25 000 femmes dans le domaine des données et de l'IA. Au Pérou, l'Institut pour la démocratie et les droits de l'homme a indiqué que la Présidente péruvienne avait soutenu un programme de formation en faveur du développement de projets commerciaux grâce à l'IA. Les citoyens inscrits au programme depuis mai 2024 ont eu la possibilité de demander des bourses, dont 50 % ont été attribuées à des femmes en situation de vulnérabilité⁶⁴.

⁵⁸ Contribution de l'Arabie Saoudite.

⁵⁹ Contribution de la Colombie.

⁶⁰ Keratso Georgiadou, « Solidarity, Gender, Dialogue in the AI era », *Kritische Pädagogik und Bildungsforschung: Anschlüsse an Paulo Freire*, Wassilios Baros, Rita BrachesChyrek, Solvejg Jobst et Joachim Schroeder (Eds), (Wiesbaden, Allemagne, Springer Fachmedien, 2024), p. 477-489.

⁶¹ *Artificial Intelligence Index Report 2024*.

⁶² Voir [CEDAW/C/ITA/CO/8](#), [CEDAW/C/DEU/CO/9](#), par. 27 et 28 ; [CEDAW/C/ESP/CO/9](#), par. 21, 23, 31 et 33 ; [CEDAW/C/TJK/CO/7](#), par. 47, 48 et 55. Voir également les observations finales figurant dans le document [CEDAW/C/TUR/CO/8](#), par. 46 ; [CEDAW/C/CRI/CO/8](#), par. 27, 28 et 37.

⁶³ Contribution de l'Arabie Saoudite.

⁶⁴ Contribution de l'Institut pour la démocratie et les droits de l'homme du Pérou.

33. Le Centre de ressources sur les entreprises et les droits de l'homme décrit les problèmes que pose actuellement l'IA, qui fait obstacle à la solidarité à l'égard des femmes⁶⁵ :

Les femmes et les personnes ne correspondant pas à l'image traditionnellement associée à leur genre continuent de subir de manière disproportionnée l'impact des défauts de conception, de développement et de déploiement des technologies, comme en témoignent les failles et les biais présents dans les outils technologiques les plus récents, y compris l'IA générative, qui amplifient le sexisme et les stéréotypes de genre, facilitent le contrôle social des femmes, privent les patientes de moyens d'action, opèrent des discriminations à l'égard des candidates féminines à un emploi, facilitent le ciblage des défenseuses des droits humains et empêchent les femmes d'accéder sur un pied d'égalité aux prestations sociales, entre autres préjudices.

34. Les femmes sont de plus en plus la cible de discours haineux générés par l'IA, de cyberharcèlement, de contenus sexuels non consentis, d'actes de vengeance pornographique et de harcèlement, autant de facteurs qui nuisent à leur capacité de s'épanouir sur les plans personnel et professionnel. Les organisations de la société civile ont présenté une initiative novatrice visant à mettre fin à la violence fondée sur le genre, dans le cadre de laquelle une formation à la sécurité numérique a été organisée à l'Université internationale de Kampala, en 2023⁶⁶. Elles ont également souligné que les médias sociaux étaient essentiels pour favoriser la solidarité internationale et avaient permis d'exiger des poursuites contre les auteurs de féminicides et une réforme législative au Kenya, de déclarer comme urgence nationale les féminicides et la violence à l'égard des femmes, et de créer une commission chargée de lutter contre ces crimes et de briser le cycle de l'impunité⁶⁷. Elles ont demandé que les entreprises du secteur de la technologie, les organisations non gouvernementales et les pouvoirs publics renforcent leur collaboration internationale afin de créer des outils de modération de contenu efficaces et éthiques, qui soient alimentés par l'IA et adaptables à différentes langues et à divers contextes culturels, de manière à recenser et à signaler les actes de violence en ligne contre les femmes.

C. Prise en compte de la solidarité dans la conception des systèmes d'intelligence artificielle

35. Dans la deuxième version du projet de document du pacte numérique mondial, il est indiqué que le pacte sera mis en œuvre dans un esprit de solidarité mondiale et il est expressément fait mention de l'inclusion, l'accent étant mis sur la nécessité pour les entreprises du numérique et les développeurs de collaborer avec des utilisateurs de tous horizons et de toutes aptitudes, afin de prendre en compte leurs perspectives et leurs besoins dans le cycle de vie des technologies numériques, de répartir équitablement les avantages numériques et de favoriser l'accessibilité numérique ainsi que la diversité linguistique et culturelle dans l'espace numérique⁶⁸. En outre, la communauté internationale demande que les besoins des personnes en situation de vulnérabilité et des personnes vivant dans des zones mal desservies, rurales et reculées soient pris en compte dans l'élaboration et la mise en œuvre des stratégies nationales et locales de desserte numérique, et souligne qu'il importe de cibler et d'adapter l'action menée pour renforcer les capacités des femmes et des filles, des

⁶⁵ Voir https://media.business-humanrights.org/media/documents/BHRRC_Submission_Gender_and_Role_of_Business_2023_j2HkLgP.pdf.

⁶⁶ Contribution du consortium « Our Voices, Our Futures ».

⁶⁷ Ibid.

⁶⁸ Disponible à l'adresse suivante : <https://www.un.org/techenvoy/fr/global-digital-compact>.

enfants et des jeunes, ainsi que des personnes âgées, des personnes handicapées et des personnes en situation de vulnérabilité, et de faire en sorte qu'ils participent véritablement à la conception et à l'exécution des programmes. Elle s'emploie également à élaborer et à mener des enquêtes nationales sur l'inclusion numérique en veillant à ce que les données soient systématiquement ventilées par âge, par handicap et par genre, afin de repérer les lacunes en matière d'apprentissage et de déterminer les priorités en fonction du contexte.

36. Le pacte numérique mondial soutient l'adoption de lois nationales sur l'IA et la mise en place de mécanismes de contrôle et de recours efficaces. Il invite également les entreprises du secteur de la technologie et les développeurs d'IA à reconnaître l'obligation de respecter les droits humains, à faire preuve de diligence raisonnable et à réaliser des études d'impact. Il y est demandé aux entreprises technologiques et aux développeurs d'IA de collaborer pour mettre au point des dispositifs d'application du principe de responsabilité propres au secteur (en consultation avec les pouvoirs publics et d'autres parties prenantes) permettant d'améliorer la transparence de leurs systèmes et processus et de définir les responsabilités, et de s'engager à respecter les normes et à établir des rapports publics pouvant faire l'objet d'un audit. Le pacte préconise que les plateformes de médias sociaux établissent des mécanismes sûrs, sécurisés et accessibles (y compris des mécanismes spécialement adaptés aux enfants et aux personnes handicapées), afin que les utilisateurs et les personnes qui les défendent puissent signaler les violations. Il y est indiqué que les entreprises du secteur de l'IA et les plateformes de médias sociaux devraient améliorer la transparence et la responsabilité de leurs systèmes, et notamment faire en sorte que leurs conditions d'utilisation, leurs politiques de modération des contenus et leurs algorithmes, ainsi que les modalités de traitement des données personnelles des utilisateurs soient accessibles dans les langues locales, afin de permettre aux utilisateurs de faire des choix éclairés et de donner ou de retirer leur consentement en toute connaissance de cause.

37. L'UNESCO insiste sur le rôle de l'inclusion dans la création de mécanismes transparents associant les acteurs de l'IA⁶⁹ :

- Le Gouvernement encourage-t-il d'autres acteurs à prendre part à la gouvernance nationale de l'IA ?
- Existe-t-il des associations actives de professionnels, de consommateurs et d'autres groupes de parties prenantes dans le domaine de l'IA ?
- Le Gouvernement invite-t-il activement d'autres groupes de parties prenantes à participer à l'élaboration d'une politique mondiale de gouvernance de l'IA ?

Afin de garantir l'inclusion, l'Organisation recommande l'élaboration d'indicateurs permettant d'évaluer dans quelle mesure les différents groupes de parties prenantes (les femmes et les personnes ne correspondant pas à l'image traditionnellement associée à leur genre, les jeunes et les groupes marginalisés) sont impliqués dans l'élaboration des politiques relatives à l'IA et participent aux forums, aux processus et aux mécanismes internationaux et régionaux sur l'IA.

38. Il a été suggéré que « dans le système sociotechnique, les technologies de l'IA pouvaient contribuer à faciliter ou à façonner les interactions sociales et donc affaiblir ou promouvoir la solidarité⁷⁰ ». Il est essentiel d'impliquer les parties prenantes dans la prise de décisions relatives à la conception de l'IA, sachant que leur rôle pourra évoluer au fil du temps, ce qui risque d'entraîner des problèmes imprévus en matière

⁶⁹ UNESCO, *Steering AI and Advanced ICTs for Knowledge Societies*.

⁷⁰ Catharina Rudschies, « Exploring the Concept of Solidarity in the Context of AI: An Ethics in Design Approach », *Digital Society*, vol. 2, n° 1 (2023).

de solidarité. Les droits de l'homme sont un indicateur de solidarité de l'IA⁷¹. Dans certains cas, on peut considérer que l'utilisation de l'IA entraîne des violations du droit à la vie privée en raison de la reconnaissance faciale, des violations du principe de non-discrimination à l'embauche pour les femmes ou les minorités, ou empêche la réalisation de projets de vie à cause du refus de demandes de prêts au logement ou de prêts éducatifs. Les études d'impact sur les droits de l'homme devraient être considérées comme des éléments centraux de la conception solidaire des systèmes d'IA. En outre, il convient de recenser les intérêts et les besoins des utilisateurs et de les prendre en compte dans le processus de conception. Dans le cadre d'un processus participatif, les utilisateurs seraient invités à examiner les choix de conception et à participer à la prise de décision, qui devrait être collective afin de favoriser la solidarité au sein du groupe. Cependant, si l'on se focalise uniquement sur les utilisateurs, on risque d'oublier les parties prenantes indirectes, par exemple les personnes qui subissent des conséquences involontaires. L'adoption d'une perspective solidaire en matière d'IA nécessiterait de s'intéresser à toutes les parties prenantes concernées au sein du système sociotechnique, et pas seulement aux utilisateurs. Au-delà d'une évaluation globale des avantages et des inconvénients, il conviendrait d'analyser la manière dont ceux-ci se répartissent dans la société et les groupes de parties prenantes. La question fondamentale est de savoir comment faire accepter aux entreprises du secteur de la technologie l'obligation de solidarité et le partage des risques, en plus du partage des bénéfices. Dans le cadre d'une approche solidaire de l'IA, une évaluation des risques liés aux droits humains serait menée afin de mettre en évidence le droit à l'épanouissement et de déterminer le partage des responsabilités en cas de dommages, ainsi que la répartition des avantages découlant du traitement des données et de la prise de décision⁷². Par conséquent, il sera également nécessaire de faire preuve de solidarité politique pour favoriser la prise de décision collective et la mise en place d'un cadre de gouvernance solidaire en matière d'IA, afin de faire face aux implications de cette technologie.

39. Dans l'optique d'une approche de solidarité internationale en matière d'IA, il pourra être nécessaire d'élaborer des règlements dans différents domaines, tels que la santé ou l'énergie. Le dialogue sur la conception et le développement de l'IA devrait mettre l'accent sur la collaboration entre les utilisateurs, les développeurs de technologies d'IA et les autres parties prenantes, afin de trouver des solutions en matière d'IA. Il est essentiel de créer des espaces de collaboration pour faciliter le dialogue avec les développeurs de cas d'utilisation de l'IA, de solutions d'IA ou de toute autre application d'IA ou d'apprentissage automatique, afin d'obtenir régulièrement des informations actualisées permettant de suivre l'évolution des risques et avantages liés à l'utilisation des solutions d'IA. Il convient également d'établir des plans de communication et de coopération adaptés et fondés sur les risques, afin que les clients puissent comprendre facilement les explications sur le développement, les performances et la maintenance des solutions d'IA, et évaluer leur conformité avec les meilleures pratiques et les dernières normes réglementaires.

40. Les notions de reconnaissance et de respect doivent être prises en compte dans le cadre d'une gouvernance solidaire des données qui donne « une place conceptuelle aux expériences concrètes de marginalisation et d'injustice dans des contextes de traitement intensif des données⁷³ ». Dans le domaine de l'IA, le principe de solidarité devrait se traduire par des résultats équitables et durables afin de favoriser un partage de la prospérité et des difficultés et de prévenir les inégalités⁷⁴. Des universitaires ont proposé un cadre global de solidarité en matière d'IA fondé sur un système

⁷¹ Ibid., p. 12.

⁷² Ibid., p. 13.

⁷³ Hummel et Braun, « Just data? Solidarity and justice in data-driven medicine ».

⁷⁴ Luengo-Oroz, « Solidarity should be a core ethical principle of AI ».

d'incitation visant à prévenir les inégalités, qui prévoit des mesures telles que le paiement de redevances chaque fois qu'un système d'IA entraîné avec les données d'une personne est utilisé, l'octroi de récompenses aux médecins qui s'appuient sur un modèle d'IA pour établir un diagnostic, ou la rémunération des personnes qui produisent des textes pour les systèmes de rédaction automatique, chaque fois qu'un article écrit par un robot est publié⁷⁵. D'autres recommandent l'adoption d'une approche solidaire permettant de mesurer les répercussions externes sur la société en favorisant la mise en place de garanties relatives au contrôle individuel des données traitées⁷⁶. Une autre proposition consiste à élaborer un cadre de solidarité en matière de données qui faciliterait la visibilité des processus au service du bien public⁷⁷. L'idée serait d'inciter les entreprises et les acteurs publics du secteur des données à partager les risques et les avantages découlant de l'accès, de la production et du partage des données. La solidarité des données est définie comme un moyen de promouvoir la visibilité des processus de traitement des ensembles de données afin de mettre en évidence la discrimination fondée sur la marginalisation et d'y remédier⁷⁸. Les partisans de cette proposition appellent à une action collective, en s'appuyant sur la solidarité comme principe de gouvernance des données aux fins de la création d'ensembles de données publiques, le but étant de commencer à bâtir la confiance et à appliquer le principe de responsabilité. Certains considèrent que la communauté internationale se trouve à un tournant, où la solidarité des données devrait contribuer à promouvoir la création d'infrastructures et être façonnée selon des valeurs démocratiques, telles que l'inclusion⁷⁹. À cet égard, il pourrait être nécessaire de créer une entité indépendante chargée de la gouvernance mondiale, composée de représentants du secteur, des États, de la société civile, des organisations internationales et du monde universitaire afin de promouvoir des règles sur l'IA fondées sur les droits humains⁸⁰. L'article intitulé « Claude's Constitution », publié sur le site *Anthropic*, met en avant des principes encourageant la prise en compte de perspectives non occidentales, en plus des droits humains⁸¹.

41. Dans leurs contributions, certaines entreprises de technologie ont proposé des mesures à mettre en œuvre pour intégrer la solidarité internationale dans la conception des systèmes d'IA :

- a) Concevoir des séances de formation accessibles, tenant compte des spécificités culturelles et inclusives ;
- b) Entraîner les algorithmes en utilisant des ensembles de données diversifiés afin qu'ils soient représentatifs de la population ;
- c) Procéder régulièrement à des audits, réaliser des tests auprès de différents groupes démographiques et appliquer des méthodes d'apprentissage automatique équitables ;
- d) Élaborer des directives pour une IA conforme à l'éthique visant à favoriser un développement et un déploiement équitables, transparents et responsables des algorithmes ;

⁷⁵ Ibid.

⁷⁶ Hummel et Braun, « Just data? Solidarity and justice in data-driven medicine ».

⁷⁷ Mercedes Bunz et Photini Vrikki, « From Big to Democratic Data: Why the Rise of AI Needs Data Solidarity », *Democratic Frontiers: Algorithms and Society*, Michael Filimowicz (Ed.), (Londres, Taylor & Francis, 2022).

⁷⁸ Ibid.

⁷⁹ Ibid.

⁸⁰ Ana Beduschi, « Human rights and the governance of artificial Intelligence », note de recherche (Geneva Academy of International Humanitarian Law and Human Rights, 2020).

⁸¹ Voir <https://www.anthropic.com/news/claudes-constitution>.

- e) Contrôler et adapter en permanence les systèmes d'IA pour prévenir la discrimination ;
- f) Collaborer avec les groupes de défense des droits, les organisations non gouvernementales et les représentants des populations vulnérables afin de comprendre les préoccupations et de recueillir des avis sur l'impact des technologies ;
- g) Dispenser régulièrement des formations aux équipes sur l'équité algorithmique et les implications éthiques de l'IA ;
- h) Mettre en œuvre des politiques de modération qui interdisent la diffusion d'informations fallacieuses, de discours haineux et de contenus préjudiciables ;
- i) Faire appliquer la loi grâce à des outils automatisés et à des modérateurs humains formés à repérer et à supprimer la désinformation ;
- j) Collaborer avec des organisations de vérification des faits réputées ;
- k) Mettre l'accent sur la transparence des algorithmes afin de limiter l'amplification involontaire de la désinformation ;
- l) Concevoir des algorithmes permettant de mettre en avant les sources crédibles et de réduire la visibilité des contenus trompeurs ;
- m) Former les utilisateurs afin de leur donner les moyens de repérer et de signaler la désinformation par l'intermédiaire de campagnes d'information et de notifications sur les plateformes ;
- n) Collaborer avec des organismes gouvernementaux et non gouvernementaux et des institutions universitaires afin d'échanger des idées et des bonnes pratiques permettant de lutter efficacement contre la désinformation.

IV. Bonnes pratiques mises en œuvre au niveau national

42. L'Espagne a déclaré qu'elle avait adopté une stratégie nationale en matière d'intelligence artificielle et créé un organisme chargé de superviser sa mise en œuvre⁸². Elle a prévu d'élaborer un cadre éthique et réglementaire visant à renforcer la protection des droits individuels et collectifs, afin de garantir l'inclusion et le bien-être social. Elle compte également mettre au point une charte des droits numériques et lancer un modèle de gouvernance nationale pour une IA conforme à l'éthique, en faisant appel à des conseils consultatifs sur l'IA et la transformation numérique. L'article 23 de la loi n° 15/2022 du 12 juillet 2022 indique que des mesures doivent être prises pour réduire les biais et améliorer la transparence et la responsabilité lorsque les organismes publics utilisent l'IA dans la prise de décision. Ces mesures consistent notamment à analyser la conception des systèmes et les données d'entraînement et à évaluer les risques de discrimination⁸³. L'Espagne s'emploie en outre à établir des forums de dialogue, à accroître la sensibilisation et la participation nationale et internationale dans le domaine de l'IA, afin de faciliter la communication entre les pouvoirs publics, la communauté scientifique, le secteur privé et la société civile.

43. L'Arabie saoudite a indiqué que l'Autorité nationale chargée des données et de l'intelligence artificielle disposait d'un système pour l'organisation, le développement et le traitement des données d'IA et la prestation de services publics, et s'employait à appuyer de manière fiable la transformation numérique et la

⁸² Voir <https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/National-Strategy-on-AI.pdf>.

⁸³ Contribution de l'Espagne.

préservation des données en assurant la participation des parties prenantes au développement et au déploiement de l'IA⁸⁴. Elle s'attache à promouvoir les sept principes éthiques suivants en matière d'IA : l'équité ; le respect de la vie privée et la sécurité ; l'humanité ; la promotion des avantages sociaux et environnementaux ; la fiabilité et la sûreté ; la transparence et l'explicabilité ; l'obligation de rendre compte et la responsabilité⁸⁵. L'Autorité a annoncé le lancement d'un programme d'incitation visant à aider les entreprises à se conformer volontairement aux principes d'éthique en matière d'IA. Dans un premier temps, l'objectif est de recenser et d'évaluer tous les risques possibles et leur niveau de gravité potentiel.

44. Dans sa contribution, la Malaisie a fait part de son intention de créer un mécanisme de contrôle chargé de la coordination et de la mise en œuvre des systèmes d'IA, qui assurera la fonction d'organe public responsable de toutes les questions liées à l'IA⁸⁶. Elle prévoit de créer un comité qui sera chargé de réaliser des analyses prospectives et de promouvoir les politiques.

45. L'Allemagne utilise l'IA dans le cadre de son programme Match'In pour améliorer l'intégration des réfugiés dans les municipalités en les consultant pour recenser leurs besoins en ce qui concerne le logement, l'emploi ou l'expérience professionnelle, les loisirs, la santé, la situation familiale, la religion et d'autres questions⁸⁷. Cette approche, qui permet de promouvoir la reconnaissance du droit des réfugiés à l'épanouissement, est conforme au principe de solidarité internationale. Il a été noté que le programme Match'In était le seul dispositif conçu de manière à prendre en compte les perspectives des réfugiés⁸⁸.

46. La République dominicaine a indiqué qu'elle avait adopté une stratégie nationale en matière d'intelligence artificielle afin de favoriser l'utilisation prioritaire de cette technologie dans des secteurs clés de l'administration publique tels que la justice, la santé, l'éducation, l'environnement et la sécurité, et pour l'élaboration de modèles d'analyse prédictive visant à concevoir des services liés à la citoyenneté. Elle a pour objectif d'élaborer un code prévoyant des mesures préventives, des garanties procédurales et des mécanismes d'application du principe de responsabilité afin d'assurer une utilisation responsable de l'IA. En outre, des dispositifs de contrôle seront mis en place pour veiller au respect de l'éthique. Le pôle de ressources humaines et d'innovation « YoSoyFuturoRD » donnera la priorité aux secteurs vulnérables. Il a été proposé de créer un écosystème régional solide et collaboratif en matière d'IA afin de contribuer de manière significative aux progrès technologiques, au développement économique et social et à la coopération régionale. Des mécanismes de supervision, d'indemnisation et de recours seront établis afin de protéger les droits des citoyens contre les dommages causés par les systèmes d'IA.

47. En Argentine, la ville de Rawson utilise l'IA pour promouvoir des programmes d'aide sociale et des mesures d'incitation, comme le « Réseau d'économie sociale et solidaire », l'objectif étant de créer des circuits alimentaires durables dans le cadre de la souveraineté alimentaire. Les autorités de la ville de Rawson ont également mis en place des centres de services de quartier (initiative « Municerca »), chargés du traitement des demandes d'amélioration des espaces publics et des rues, de l'enregistrement et de la consultation des dossiers, et de la réception des plaintes municipales.

⁸⁴ Contribution de l'Arabie Saoudite.

⁸⁵ Dans sa contribution, la Colombie a également noté qu'elle respectait des principes tels que la transparence, la protection de la vie privée, le contrôle humain et la non-discrimination, et veillait à ce que les technologies soient inclusives et équitables.

⁸⁶ Contribution de la Malaisie.

⁸⁷ Voir <https://matchin-projekt.de/en/>.

⁸⁸ Voir https://www.rsc.ox.ac.uk/files/files-1/automating-immigration-and-asylum_afar_9-1-23.pdf.

A. Solidarité en matière de santé et intelligence artificielle

48. La solidarité en matière de santé devrait être un élément fondateur de la réglementation sur l'utilisation de l'IA afin de garantir l'accès et l'équité, mais il est difficile d'atteindre cet objectif en raison de la privatisation des services de santé dans de nombreux pays⁸⁹. Les inégalités de santé préexistantes liées à la race, à la classe sociale, à l'âge et à d'autres facteurs pourraient être perpétuées dans les systèmes médicaux d'IA qui ne tiennent pas compte de l'injustice structurelle. Les systèmes d'IA utilisés pour répondre aux besoins en matière de santé doivent être conçus dans une optique de santé de proximité et, plus spécifiquement, être déployés dans des zones ayant peu de ressources, de manière à faciliter la couverture sanitaire dans les contextes où les moyens sont limités. Certaines voix se sont élevées pour proposer une vision solidaire de l'IA en matière de santé, dans le cadre de laquelle les personnes seraient en mesure de partager ou de supprimer des données en contrôlant les flux de données, l'idée étant de créer une infrastructure de supervision et de mettre l'accent sur une gouvernance axée sur les résultats afin de prévenir et de limiter les injustices⁹⁰. Le Brésil a mis en place le programme « Bolsa Família », dans le cadre duquel des outils numériques sont utilisés pour verser des aides aux familles à faibles revenus. Ce programme a permis de réduire la pauvreté et de promouvoir le développement du capital humain en définissant des critères à remplir pour l'obtention des transferts en espèces, tels que la fréquentation scolaire et la réalisation d'exams médicaux. La Malaisie utilise l'IA pour faciliter les diagnostics médicaux, mettre au point des traitements sur mesure, fournir des informations par l'intermédiaire de boîtes de dialogue et contribuer à l'analyse prédictive afin de définir des mesures proactives⁹¹.

49. L'Arabie saoudite a décrit les investissements qu'elle avait réalisés dans les systèmes d'IA solidaires en matière de santé. Elle a créé un centre d'excellence pour l'intelligence artificielle dans le domaine de la santé. Des experts et des ingénieurs saoudiens ont mis au point l'outil « Eyenai », qui vise à révolutionner la médecine diagnostique au niveau régional. La rétinopathie diabétique est une complication des diabètes de type 1 et 2 et l'une des premières causes de cécité en Arabie saoudite. Il est essentiel d'établir un diagnostic précoce pour éviter des complications ultérieures, et Eyenai contribue à cet objectif en permettant une détection et un diagnostic précis. L'Autorité saoudienne chargée des données et de l'intelligence artificielle a mis au point l'application « Tawakkalna », qui a obtenu le Prix des Nations Unies pour le service public en 2022 pour sa contribution novatrice à la lutte contre la pandémie de maladie à coronavirus 2019 (COVID-19).

B. Solidarité entre les travailleurs et intelligence artificielle

50. La fragmentation du travail causée par l'utilisation de l'IA apparaît comme un obstacle qui empêche les travailleurs de s'organiser de manière solidaire : « La facilité avec laquelle une plateforme de mise en relation peut recruter de nouveaux employés, parfois dans le monde entier, génère de la concurrence et de l'isolement, et entrave considérablement la constitution d'une identité collective. En résumé, le

⁸⁹ Nicolas Terry, « Of Regulating Healthcare AI and Robots », *Yale Journal of Law and Technology*, édition spéciale n° 133 (2019).

⁹⁰ Hummel et Braun, « Just data? Solidarity and justice in data-driven medicine ».

⁹¹ Contribution de la Malaisie.

travail via des plateformes fragmente l'identité collective des travailleurs et sabote l'action collective, en particulier l'action basée sur la solidarité et la confiance⁹² ».

51. Cependant, il a été suggéré que des modèles informatiques de solidarité pouvaient aider les travailleurs en temps de crise⁹³. L'une des questions centrales est de savoir si l'IA peut être repensée comme un outil de promotion de la solidarité entre les travailleurs⁹⁴. La Banque de sécurité sociale de l'Uruguay a créé des procédures centralisées et automatisées pour les employeurs et encouragé l'autogestion en proposant des services en ligne, une application mobile et une assistance via différents canaux. Un robot conversationnel (chatbot) aurait répondu à 97 % des demandes et obtenu un taux de satisfaction de 100 %. En conséquence des mesures prises, une réduction significative de 24,4 % du taux d'évasion des cotisations sociales a été enregistrée, 57 % des employeurs se sont inscrits en ligne et 42 % d'entre eux ont effectué des paiements en ligne⁹⁵. En Argentine, le Bureau du Surintendant des risques au travail a mis en service avec succès un chatbot d'IA appelé Julieta⁹⁶. De nouvelles formes de solidarité horizontale, nettement plus ouvertes et diversifiées, ont fait leur apparition. On peut aussi noter la tendance du « discours partagé », qui se caractérise par une « suppression des barrières bureaucratiques entre les militants et les responsables syndicaux, une plus grande transparence et un pouvoir renforcé des membres ordinaires, grâce auquel ils peuvent contester et remettre en question les structures syndicales oligarchiques⁹⁷ ». On constate également l'émergence d'un « pluralisme accéléré », concept qui désigne « la fragmentation continue des politiques de groupes d'intérêts au profit d'alliances plus fluides et davantage axées sur des questions, ainsi que la création d'espaces sûrs visant à favoriser la participation au débat démocratique des femmes et d'autres travailleurs qui étaient auparavant marginalisés dans le cadre des structures syndicales traditionnelles⁹⁸ ». Parmi les exemples d'utilisation de l'IA au service de la solidarité, on peut citer les mesures prises pour renforcer l'accès des travailleuses agricoles aux systèmes d'IA dotés de capacités prédictives afin d'anticiper les récoltes et les conditions météorologiques⁹⁹. Un autre exemple est le mouvement anti-apartheid mené par des travailleurs du secteur de la technologie¹⁰⁰. D'un autre côté, l'Association des infirmiers et infirmières de Californie s'est engagée à demander des garanties en matière d'IA afin de protéger les patients contre la prescription de traitements inadaptés¹⁰¹. Il a cependant été observé que « les employeurs avaient adopté des

⁹² Tammy Katsabian, « The rule of technology - how technology is used to disturb basic labor law protections », *Lewis and Clark Law Review*, vol. 25, n° 3 (2021).

⁹³ Saiph Savage et Mohammad H. Jarrahi, « Solidarity and AI for transitioning to crowd work during COVID-19 », article publié à l'occasion du colloque en ligne organisé en août 2020 sur le thème « The New Future of Work ». Voir également Kurt April, « AI-Induced Solidarity Economy: The Need for Stewardship Orientation ».

⁹⁴ Frances Flanagan et Michael Walker, « How can unions use artificial intelligence to build power? The use of AI chatbots for labour organising in the US and Australia », *New Technology, Work and Employment*, vol. 36, n° 2 (2021).

⁹⁵ Voir <https://www.issa.int/sites/default/files/documents/2024-06/2-AI%20in%20SecSoc%202024.pdf>.

⁹⁶ Ibid.

⁹⁷ Voir Andy Hodder et David Houghton, « Union use of social media: a study of the University and College Union on Twitter », *New Technology, Work and Employment*, vol. 30, n°3 (novembre 2015).

⁹⁸ Voir Anne-Marie Greene et Gill Kirton, « Possibilities for remote participation in trade unions: mobilising women activists », *Industrial Relations Journal*, vol. 34, n°4 (octobre 2003).

⁹⁹ Voir CEDAW/C/OMN/CO/4. Sur l'IA et les prévisions météorologiques, voir : <https://deepmind.google/discover/blog/graphcast-ai-model-for-faster-and-more-accurate-global-weather-forecasting/>.

¹⁰⁰ Voir <https://www.notechforapartheid.com/>.

¹⁰¹ Voir <https://www.nationalnursesunited.org/press/cna-demand-patient-safeguards-against-artificial-intelligence>.

stratégies de contre-mobilisation dans l'espace numérique et que les incertitudes liées au statut juridique du discours sur les médias sociaux avait eu un effet dissuasif supplémentaire sur la prise de parole des employés en ligne¹⁰² ».

52. Par conséquent, le type d'activité syndicale ainsi que les différences politiques et organisationnelles entre les mouvements syndicaux influenceront l'usage qui est fait de l'IA en matière de solidarité internationale en ce qui concerne les questions de contestation du pouvoir, d'identité collective, de cohésion interne, de culture, de stratégie, de gouvernance et de communauté¹⁰³.

V. Mise en place de procédures de dépôt de plainte conformément au devoir de diligence

53. En raison des inquiétudes croissantes en matière de transparence, des voix s'élèvent pour demander que soient examinés les abus des entreprises en ce qui concerne l'utilisation des données personnelles et l'accès à celles-ci. Par exemple, l'entreprise Meta a annoncé qu'elle prévoyait d'établir un réglage par défaut permettant d'utiliser le contenu des utilisateurs pour entraîner des modèles d'intelligence artificielle. Estimant que la procédure de retrait était complexe, le Conseil norvégien des consommateurs a déposé une plainte contre l'entreprise pour violation du règlement général sur la protection des données de l'Union européenne¹⁰⁴. Les Principes directeurs relatifs aux entreprises et aux droits de l'homme établissent une obligation de diligence raisonnable pour les entreprises, qui doivent, conformément au principe 15, adopter des stratégies de prévention et des procédures de recours. Le principe 25 définit un cadre relatif à l'accès à la justice et le principe 29 invite les entreprises à établir des mécanismes de réclamation.

54. Les Principes mondiaux des Nations Unies pour l'intégrité de l'information prévoient la mise en place de mécanismes de contrôle et l'organisation régulière d'audits externes indépendants sur les droits de l'homme, qui porteraient notamment sur les domaines suivants : conditions de service, normes communautaires, politiques de publicité, modération de contenu, procédures de plainte, accès aux données pour les chercheurs, études d'impact sur la vulnérabilité et la marginalisation, l'égalité des genres et les droits de l'enfant. Les résultats de ces audits devront être rendus publics, accessibles et compréhensibles pour tous les utilisateurs¹⁰⁵.

55. Dans sa recommandation sur l'éthique de l'intelligence artificielle, l'UNESCO a indiqué que les États devraient mettre en place des mécanismes de supervision associant toutes les parties prenantes, et encourager celles-ci à promouvoir les droits de l'homme. L'article 14 de la directive (UE) 2024/1760 du Parlement européen et du Conseil en date du 13 juin 2024 sur le devoir de vigilance des entreprises en matière de durabilité établit l'obligation pour les États de veiller à ce que les entreprises mettent en place une procédure de dépôt de plainte équitable, mise à la disposition du public, accessible, prévisible et transparente pour les personnes, les organisations de la société civile, les défenseurs des droits de l'homme, les syndicats et d'autres organismes, en vue de traiter les incidences des activités des entreprises, de leurs

¹⁰² Flanagan et Walker, « How can unions use artificial intelligence to build power? The use of AI chatbots for labour organising in the US and Australia ». Voir également Louise Thornthwaite, « Chilling times: Social media policies, labour law and employment relations », *Asia Pacific Journal of Human Resources*, vol. 54, n° 3 (août 2015).

¹⁰³ Ibid.

¹⁰⁴ Voir <https://www.forbrukerradet.no/side/legal-complaint-against-metas-use-of-personal-content-for-ai-training/>.

¹⁰⁵ Voir <https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf>.

filiales et d'autres entités sur les droits de l'homme et l'environnement, dans le cadre de la chaîne d'activités¹⁰⁶. Les mesures prises doivent respecter les normes de confidentialité et empêcher que des représailles soient exercées contre la personne ou l'entité ayant soumis la plainte. La directive établit un cadre qui complète le mécanisme de communication prévu par le projet révisé de déclaration sur le droit à la solidarité internationale (par. 3 de l'article 6). En outre, le projet révisé de déclaration (par. 3 de l'article 8) propose un cadre flexible qui pourrait être utilisé pour la conception et le partage de politiques et de pratiques de solidarité internationale en matière d'IA au titre de l'Examen périodique universel. Conformément au paragraphe 3 de l'article 8 du projet révisé de déclaration, les États ont le devoir de prendre des mesures, dans la limite de leurs capacités respectives, pour faciliter la protection des espaces de communication réels et virtuels, notamment l'accès à Internet et aux infrastructures, afin de permettre aux individus et aux peuples d'échanger des idées sur la solidarité. Ces normes constituent une base sur laquelle on pourrait s'appuyer pour établir des mécanismes novateurs visant à contrôler la prise en compte de la solidarité internationale dans les systèmes d'IA et à favoriser une plus grande participation de la société civile.

VI. Conclusion

56. L'Experte indépendante partage l'avis selon lequel il est nécessaire de créer un modèle mondial de gouvernance multipartite en vue de prévenir la discrimination et d'autres violations des droits de l'homme liées aux systèmes d'IA et d'y remédier¹⁰⁷. En matière de gouvernance, l'Organe consultatif de haut niveau sur l'intelligence artificielle recommande d'adopter une approche associant les acteurs du secteur public et du secteur privé et de faciliter l'interopérabilité entre les juridictions, aux niveaux international, régional et national, en s'inspirant de loi sur l'intelligence artificielle de l'Union européenne¹⁰⁸. Cependant, l'un des principaux marqueurs de l'évolution du système est l'utilisation de codes et d'accords volontaires, tels que l'accord anti-deepfake visant à lutter contre l'ingérence dans les élections, signé par des entreprises du secteur de la technologie, des développeurs d'IA et des sociétés de sécurité¹⁰⁹. À l'heure actuelle, la communauté internationale ne parvient pas à partager équitablement les avantages et les risques liés à la solidarité en matière d'IA et un profond manque de confiance semble régner entre les organisations de la société civile du monde du Sud et les entreprises technologiques et les développeurs d'IA du monde du Nord. En outre, la majorité des principes éthiques de l'IA sont rédigés par les pays du Nord et ne tiennent pas forcément compte des préoccupations particulières des pays du Sud, alors que ces derniers fournissent les minéraux et l'énergie essentiels au développement des infrastructures de données d'IA¹¹⁰. La concentration du pouvoir entre les entreprises technologiques et les développeurs d'IA accroît le risque que les mesures liées à l'IA portant atteinte à la solidarité soient renforcées, ce qui aggraverait la fracture numérique entre les pays et à l'intérieur de ceux-ci, ainsi qu'entre les différents secteurs de la société. Le Forum sur la gouvernance d'Internet souligne le problème de l'exclusion des représentants du monde du Sud, qui ne participent pas consultations mondiales en raison du manque de financement, de leur

¹⁰⁶ Voir https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=OJ:L_202401760.

¹⁰⁷ Voir https://www.unwomen.org/en/news-stories/explainer/2024/05/artificial-intelligence-and-gender-equality?gad_source=1&gclid=CjwKCAjw4f6zBhBVEiwATEHFVvzWNyAYvJV56epBISUMbQTVZ4hm_tCsn_VJGGvIzoRMMXfutYebdhoCzqEQAvD_BwE.

¹⁰⁸ Voir https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf. Voir également <https://oecd.ai/en/accountability>.

¹⁰⁹ Voir <https://www.techradar.com/pro/top-tech-companies-ai-developers-and-security-firms-sign-anti-deepfake-agreement-to-combat-election-interference>.

¹¹⁰ Contribution du Forum sur la gouvernance d'Internet.

situation précaire en matière de connectabilité et de la priorité accordée à l'anglais dans les débats, entre autres facteurs¹¹¹. La conception de systèmes d'IA favorisant la solidarité internationale peut jouer un rôle déterminant dans l'évolution du développement de l'IA et la mise en place de mécanismes de contrôle visant à faire appliquer la justice procédurale et à promouvoir la participation et l'inclusion de tous les acteurs concernés. Afin de soutenir un changement politique favorisant la mise en œuvre d'une approche de solidarité internationale durable en matière d'IA, l'Experte indépendante a formulé des recommandations qu'il convient d'adopter d'urgence au cours des cinq prochaines années.

VII. Recommandations

57. L'Experte indépendante recommande aux États, aux entreprises et aux acteurs de la société civile :

a) De veiller à ce que les institutions publiques, les entreprises du secteur de la technologie et les groupes de la société civile qui utilisent l'IA pour le traitement des données et la prise de décision favorisent la participation active de tous les individus et groupes, indépendamment de leur race, de leur couleur, de leur ascendance, de leur genre, de leur âge, de leur langue, de leur religion, de leurs opinions politiques, de leur origine nationale, ethnique ou sociale, de leur condition économique ou sociale de naissance, de leur handicap ou de tout autre motif, conformément à la Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle ;

b) D'adopter des réglementations nationales visant à promouvoir la conception de systèmes d'IA qui soient au service de la solidarité internationale, en appliquant une approche non discriminatoire et en veillant à ce que la société civile et les autres acteurs concernés participent à la planification, à la sélection, à la conception et au déploiement des systèmes d'IA utilisés par les institutions publiques et les entreprises technologiques pour le traitement des données et la prise de décision ;

c) De faire en sorte que les pouvoirs publics et les entreprises réalisent régulièrement des études d'impact sur les droits de l'homme, avec la participation directe de la société civile, avant de procéder à la collecte, au traitement ou au déploiement des données, et fassent preuve d'une transparence totale, notamment en ce qui concerne l'entraînement de l'IA ;

d) De veiller à ce que les institutions publiques et les entreprises de technologie qui utilisent l'IA pour le traitement des données et la prise de décision effectuent des audits indépendants, externes et systématiques sur la transparence et des études d'impact sur les droits de l'homme (afin de repérer les biais associés à l'entraînement, aux algorithmes et à la prise de décisions), en reconnaissant la solidarité internationale comme une valeur et un objectif à prendre en compte tout au long du cycle de vie de l'IA¹¹² ;

e) De veiller à ce que la législation nationale et régionale prévoie des mécanismes de plainte et de recours indépendants, transparents, accessibles et efficaces, afin d'assurer le respect du principe de responsabilité en cas de discrimination ou d'exclusion résultant de l'utilisation de l'IA par les institutions

¹¹¹ Ibid.

¹¹² Voir <https://digitalrightscheck.bmz-digital.global/> ; voir également <https://digital-strategy.ec.europa.eu/en/library/ethicsguidelines-trustworthy-ai> et <https://ieeexplore.ieee.org/document/8058187>.

publiques, les entreprises de technologie ou les groupes de la société civile intervenant dans le traitement des données ou la prise de décision ;

f) De reconnaître qu'il devrait incomber aux institutions publiques ou aux entreprises de technologie concernées de démontrer la manière dont l'IA a été utilisée en cas de discrimination ou d'autres violations des droits de l'homme résultant du traitement des données ou de la prise de décision basés sur l'IA ;

g) D'adopter des règlements explicites ayant trait au traitement des données institutionnelles basé sur l'IA, à l'évaluation de l'impact sur la prise de décision et aux mécanismes de protection des droits pour les cas où le développement des systèmes implique le traitement de données sensibles (données sur la santé, par exemple) ou de quantités massives de données (bases de données nationales, par exemple) ;

h) De veiller à ce que toutes les personnes et tous les groupes en situation de vulnérabilité soient informés par les institutions publiques, les entreprises de technologie ou les groupes de la société civile, d'une manière facile à comprendre, que l'IA sera utilisée dans le traitement des données ou la prise de décision, et que le consentement sera demandé au préalable sans coercition directe ou indirecte ;

i) De garantir la mise en place de mécanismes indépendants d'accès à la justice afin de lutter contre la discrimination ou d'autres violations des droits de l'homme liées à l'IA ;

j) De faire en sorte que les pouvoirs publics et les entreprises de technologie fournissent de façon explicite des informations juridiques aux groupes en situation de vulnérabilité sur leurs droits et les mécanismes de recours auxquels ils ont accès en cas de violations liées à l'utilisation de l'IA ;

k) D'investir dans la promotion d'une culture des données auprès du grand public afin de lutter contre les hypertrucages, la désinformation et les discours de haine liés à l'IA, de manière à créer des sociétés plus résilientes ;

l) De reconnaître le droit de toutes les personnes à l'épanouissement comme un principe clé dans la conception de systèmes d'IA au service de la solidarité internationale, grâce à la prise en compte de critères tels que l'accès à l'éducation, au logement, à l'emploi et aux soins de santé ;

m) De respecter l'autonomie des personnes concernées par la collecte et le traitement des données ;

n) De faire en sorte que les pouvoirs publics et les entreprises de technologie soient tenus d'informer les personnes concernées de leurs droits en matière de protection des données et de mettre tout en œuvre pour veiller à ce que chacun puisse exercer ces droits, y compris le droit de suppression, le droit d'accès et le droit de retrait ;

o) De veiller à ce que les pouvoirs publics et les entreprises de technologie qui exploitent des données à des fins autres que celles définies initialement (utilisation secondaire des données) demandent le consentement préalable, libre et éclairé des individus, quelle que soit la manière dont les données ont été obtenues ; les autorités chargées de la sécurité ne devraient pas utiliser des outils d'IA destinés à la lutte contre l'extrémisme ou le terrorisme pour cibler les personnes militant en faveur de la solidarité internationale, conformément aux normes juridiques de nécessité, de proportionnalité, de légalité et de non-discrimination ;

p) D'établir un cadre juridique et des procédures systématiques afin d'évaluer l'intérêt pour le public ainsi que le degré d'urgence et de nécessité du traitement des données ; les personnes devraient conserver un droit de retrait afin de protéger leur vie privée ;

q) De faire en sorte que les États et les entreprises interdisent la création de deepfakes intimes synthétiques ou manipulés, non consensuels, conformément au respect de la solidarité internationale entre les genres dans le domaine de l'IA.

58. L'Experte indépendante recommande à la société civile :

a) De continuer à demander que l'inclusion soit prise en compte dans la création d'outils éthiques de modération de contenus alimentés par l'IA afin de prévenir, de repérer et de supprimer les contenus numériques violents et discriminatoires à l'égard des femmes et des personnes victimes de discrimination intersectorielle ;

b) De continuer à fournir des exemples de discrimination algorithmique à l'Organisation des Nations Unies, aux gouvernements nationaux, aux entreprises de technologie ou aux développeurs d'IA, afin de formuler des recommandations visant à améliorer les pratiques ;

c) De veiller à ce que les syndicats fassent en sorte que les travailleurs participent à la conception et à l'entraînement des systèmes d'IA, ainsi qu'à l'évaluation des risques, soient consultés à ce sujet et aient accès aux outils de formation au numérique.

59. L'Experte indépendante adresse les recommandations suivantes aux entreprises :

a) Les entreprises devraient établir des mécanismes concrets pour faire entendre la voix des groupes traditionnellement marginalisés afin de promouvoir un environnement en ligne sûr et inclusif reconnaissant le principe du respect mutuel de la diversité, et notamment organiser des consultations accessibles dans différentes langues et mettre en place des subventions pour faciliter la participation des parties prenantes du monde du Sud ;

b) Les entreprises de technologie et les développeurs d'IA devraient prendre les mesures suivantes : favoriser la conception de systèmes d'IA qui soient au service de la solidarité internationale et respectent les limites planétaires ; réaliser des évaluations indépendantes, transparentes et continues des risques pour l'environnement ; promouvoir des algorithmes économes en énergie ; favoriser un développement numérique durable en utilisant des énergies renouvelables ; financer le développement d'une IA au service de la solidarité afin de créer des bases de données permettant au public d'accéder à des connaissances sur l'environnement et les changements climatiques (par exemple dans des domaines tels que les fonds marins et les océans) ; gérer les déchets d'équipements électriques et électroniques et promouvoir les principes de l'économie circulaire associant les communautés locales ;

c) Les bailleurs de fonds et les développeurs des systèmes d'IA devraient s'engager à prendre en compte la solidarité internationale en menant des évaluations des droits de l'homme de manière préventive et systématique, afin de recenser les risques liés à l'utilisation de biais algorithmiques en violation des principes de non-discrimination et d'égalité, tout au long du cycle de vie des données, et demander que les parties prenantes directes et indirectes de la société civile (y compris les communautés traditionnellement réduites au silence) participent à un mécanisme de contrôle indépendant, l'objectif étant de repérer les actions néfastes à l'égalité des genres et à la démocratie et de les combattre ;

d) Les entreprises de technologie devraient mettre en place des mécanismes de solidarité préventive et réactive afin de lutter contre les violations des droits de l'homme liées à l'utilisation de l'IA, telles que les campagnes de désinformation et de mésinformation qui entraînent des violences sociétales ou un harcèlement, une surveillance, une discrimination ou une censure visant de manière disproportionnée les communautés traditionnellement réduites au silence¹¹³ ;

e) Les entreprises de technologie devraient mettre en place des équipes d'intervention rapide dotées de ressources suffisantes et du pouvoir de décision nécessaire pour traiter rapidement les plaintes et proposer des solutions.

60. L'Experte indépendante recommande à l'Organisation des Nations Unies :

a) De mettre en place une plateforme de solidarité numérique permettant aux organisations de la société civile (du monde du Nord et du monde du Sud) d'échanger des idées sur la solidarité internationale et d'examiner les politiques et les pratiques, les problèmes et les innovations liés à l'inclusion et à la non-discrimination en matière d'IA avec des représentants d'entreprises de technologie, des développeurs d'IA et des responsables de l'État, conformément aux recommandations formulées dans son rapport sur la société civile et la solidarité internationale¹¹⁴ ;

b) De créer un fonds visant à soutenir la mise en place de grands modèles de langage pour que les données soient accessibles dans toutes les langues, de manière à protéger la diversité culturelle.

¹¹³ Cette recommandation correspond aux Principes mondiaux des Nations Unies pour l'intégrité de l'information : « Améliorer la réponse aux crises. Collaborer avec les parties prenantes opérant dans des zones à haut risque, établir des processus d'alerte précoce et d'escalade avec des taux de réponse accélérés et opportuns dans des contextes de crise et de conflit. Mettre en place des mécanismes permettant un accès important et rapide à des informations fiables et précises qui servent l'intérêt public ».

¹¹⁴ [A/HRC/56/57](#).