Nations Unies A/HRC/56/68

Assemblée générale

Distr. générale
3 juin 2024
Français
Original : anglais

Conseil des droits de l'homme
Cinquante-sixième session
18 juin-14 juillet 2024
Point 9 de l'ordre du jour
Racisme, discrimination raciale, xénophobie et intolérance qui y est associée : suivi et application de la Déclaration et du Programme d'action de Durban

Formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée

Rapport de la Rapporteuse spéciale sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée, Ashwini K. P.*

Résumé

Dans le présent rapport, la Rapporteuse spéciale sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée, Ashwini K. P., rend compte des activités qu'elle a menées au cours de l'année écoulée et examine comment l'idée répandue selon laquelle la technologie est neutre et objective permet à l'intelligence artificielle de perpétuer la discrimination raciale. Elle passe en revue quatre différentes manières dont l'intelligence artificielle peut contribuer à la discrimination raciale, à savoir à travers 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. Elle donne ensuite des exemples illustrant l'application de l'intelligence artificielle dans divers domaines et ses effets discriminatoires sur le plan racial. Elle analyse les initiatives visant à réglementer et à gérer l'intelligence artificielle, puis donne un aperçu des normes du droit international des droits de l'homme applicables. En conclusion, elle formule des recommandations à l'attention des États sur l'approche à adopter pour gérer et réglementer les technologies de l'intelligence artificielle de manière à prévenir et à combattre la discrimination raciale.

* Il a été convenu que le présent document serait publié après la date normale de publication en raison de circonstances indépendantes de la volonté du soumetteur.

I. Introduction

1. Le présent rapport est soumis en application de la résolution 52/36, dans laquelle le Conseil des droits de l'homme a prié la Rapporteuse spéciale sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée de lui soumettre un rapport annuel. La Rapporteuse spéciale y décrit les activités qu'elle a menées au titre de son mandat et y traite du thème de l'intelligence artificielle et de la discrimination raciale.

2. Pour étayer son rapport, la Rapporteuse spéciale a lancé un appel à contributions à l'attention des États Membres de l'Organisation des Nations Unies et d'autres parties prenantes, notamment des organisations de la société civile, des organisations internationales et des institutions nationales des droits de l'homme. Elle remercie sincèrement tous les États Membres et les autres parties prenantes qui lui ont communiqué des informations dont elle a tiré parti pour élaborer le présent rapport. Elle est disposée à maintenir un dialogue permanent avec toutes les parties prenantes sur cet important sujet[1].

II. Résumé des activités

3. En octobre 2023, la Rapporteuse spéciale a présenté ses rapports sur la lutte contre la glorification du nazisme, du néonazisme et d'autres pratiques qui contribuent à alimenter les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée et sur les discours de haine raciale en ligne à l'Assemblée générale à sa soixante-dix-huitième session[2]. Elle s'est en outre rendue aux États-Unis d'Amérique du 31 octobre au 14 novembre 2023[3].

4. En août 2023, la Rapporteuse spéciale a participé à la neuvième session du Groupe d'éminents experts indépendants sur la mise en œuvre de la Déclaration et du Programme d'action de Durban. En janvier 2024, elle a participé à la réunion régionale pour l'Asie et le Pacifique relative à la Décennie internationale des personnes d'ascendance africaine. En février 2024, elle a assisté à la Conférence internationale sur la justice alimentaire du point de vue des droits de l'homme organisée au Qatar et qui portait sur les défis actuels et les enjeux futurs. En avril 2024, elle a pris part à la troisième session de l'Instance permanente pour les personnes d'ascendance africaine et a fait un exposé sur les solutions pour combattre le racisme systémique et les préjudices historiques dans le domaine de l'éducation.

III. Intelligence artificielle et discrimination raciale

5. La Rapporteuse spéciale a choisi de traiter ici la question de l'intelligence artificielle et de la discrimination raciale, qui s'inscrit dans le volet stratégique de son mandat portant sur les liens entre les technologies numériques et la discrimination raciale, tel qu'exposé dans le rapport qu'elle a remis au Conseil des droits de l'homme à sa cinquante-troisième session et qui présente sa vision stratégique et ses priorités initiales[4]. La Rapporteuse spéciale s'appuie sur les travaux réalisés par la précédente titulaire du mandat sur les nouvelles technologies numériques et la discrimination raciale[5] et répond à l'intérêt que le Conseil des

[1] Dans ses travaux de recherche et d'analyse, la Rapporteuse spéciale a bénéficié de l'appui de la Clinique juridique des droits de l'homme de la faculté de droit de Harvard et de la Clinique des droits de l'homme et de règlement des conflits et du Centre de Stanford pour la justice raciale de la faculté de droit de Stanford. Elle remercie sincèrement toutes les personnes ayant pris part à ces travaux pour le soutien inestimable qu'elles lui ont apporté dans le cadre de l'élaboration du présent rapport.

[2] A/78/302 et A/78/538.

[3] Voir A/HRC/56/68/Add.1.

[4] A/HRC/53/60, par. 50 à 53.

[5] Voir A/75/590, A/HRC/44/57 et A/HRC/48/76.

droits de l'homme et l'ensemble des organismes des Nations Unies ont manifesté concernant la gouvernance de l'intelligence artificielle[6].

6. Les développements récents de l'intelligence artificielle générative et l'essor que connaît l'application de l'intelligence artificielle continuent de soulever de graves questions en matière de droits de l'homme, dont des préoccupations concernant la discrimination raciale. L'intelligence artificielle générative est en train de changer le monde et pourrait à l'avenir entraîner des évolutions majeures dans la société. La Rapporteuse spéciale est vivement préoccupée par la vitesse à laquelle l'application de l'intelligence artificielle se propage dans différents domaines. Son inquiétude n'est pas due à un manque de potentiel de l'intelligence artificielle ; de fait, cette dernière présente des possibilités d'innovation et d'inclusion. Cela étant, les technologies dans ce domaine se développent et évoluent à un rythme sensiblement effréné. La Rapporteuse spéciale est préoccupée par le décalage qui existe entre les mesures générales et les mesures juridiques que prennent les États en vue de gérer et de réglementer l'intelligence artificielle et la croissance rapide que connaît cette technologie. Elle est également préoccupée par le fait que les initiatives prises en vue de gouverner et de réglementer l'intelligence artificielle tiennent peu compte de la forte capacité actuelle et éventuellement future qu'a cette dernière de perpétuer et d'aggraver la discrimination raciale systémique, ainsi que de creuser les inégalités dans et entre les régions, les pays et les populations.

7. Comme l'avait souligné la précédente titulaire du mandat, beaucoup pensent à tort que la technologie est neutre et objective :

> Le grand public voit généralement la technologie comme intrinsèquement neutre et objective, et certains observateurs ont souligné que cette présomption d'objectivité et de neutralité restait très présente, même parmi les producteurs de technologie. Pourtant, la technologie n'est jamais neutre : elle traduit les valeurs et les intérêts de ceux qui interviennent dans sa conception et son utilisation et, fondamentalement, elle est pétrie par les structures d'inégalité qui se retrouvent dans la société[7].

8. Dans le présent rapport, la Rapporteuse spéciale examine comment l'idée répandue selon laquelle la technologie est neutre et objective permet à l'intelligence artificielle de perpétuer la discrimination raciale.

A. Comment l'intelligence artificielle peut contribuer à la discrimination raciale

9. L'intelligence artificielle n'est pas une technologie monolithique. Il en existe différents types. L'intelligence artificielle prédictive est considérée comme une forme traditionnelle d'intelligence artificielle dans laquelle les modèles utilisent des données historiques, des constantes et des tendances pour faire des prédictions, en toute connaissance de cause, sur des événements ou des résultats futurs.

10. L'intelligence artificielle utilisée pour déchiffrer des caractères imprimés, identifier des visages, détecter des objets et d'autres informations est une autre forme d'intelligence artificielle traditionnelle ; elle englobe diverses technologies qui permettent de reconnaître et de distinguer des objets, des individus et des tendances parmi les données qui les alimentent.

11. Les systèmes d'intelligence artificielle générative sont des formes plus récentes d'intelligence artificielle. Polyvalents, ils peuvent être utilisés à des fins diverses. Ils englobent une catégorie de systèmes d'intelligence artificielle conçus pour produire divers résultats en s'appuyant sur de vastes jeux de données d'apprentissage, des réseaux neuronaux, une architecture d'apprentissage profond et des prompts saisis par les utilisateurs. Les modèles d'intelligence artificielle générative peuvent produire un large éventail de résultats

[6] Voir, par exemple, Organe consultatif de haut niveau sur l'intelligence artificielle, « Gouverner l'intelligence artificielle au bénéfice de l'humanité » (décembre 2023) ; résolution 53/29 du Conseil des droits de l'homme ; résolutions 78/213 et 78/265 de l'Assemblée générale.

[7] A/HRC/44/57, par. 12.

comme des images, du texte, du son, des vidéos et des données synthétiques. Contrairement aux modèles d'intelligence artificielle qui cherchent à repérer des tendances dans des données existantes, l'intelligence artificielle générative est entraînée à créer de nouveaux points de données qui imitent les tendances observées dans les données utilisées pour entraîner les modèles d'apprentissage automatique, ainsi que les caractéristiques de ces données. L'avènement de l'intelligence artificielle générative va entraîner une multitude de nouvelles applications et de nouvelles questions liées aux droits de l'homme[8].

12. Ces différents types d'intelligence artificielle ont des applications multiples. La Rapporteuse spéciale donne ci-après des exemples détaillés des utilisations qui en sont faites et des conséquences qui en découlent en matière de discrimination raciale. Elle tient à souligner à quel point il est capital d'examiner les facteurs communs qui font que l'intelligence artificielle peut perpétuer la discrimination raciale, en particulier dans le cadre des débats sur les lois et les politiques liées à la gestion et à la réglementation de cette dernière. Dans ces débats, les effets de l'intelligence artificielle doivent impérativement être envisagés sous l'angle du racisme systémique, défini comme « un système complexe et interdépendant de lois, de politiques, de pratiques et d'attitudes, dans les institutions de l'État, le secteur privé et les structures sociétales qui, ensemble, produisent des formes, directes ou indirectes, intentionnelles ou non, en droit ou dans les faits, de discrimination, de différenciation, d'exclusion, de restriction ou de préférence ayant pour fondement la race, la couleur, l'ascendance ou l'origine nationale ou ethnique »[9]. Comme le montre cette définition, le racisme systémique est un phénomène complexe, souvent insidieux et qui touche l'ensemble de la société. Les manifestations de ce racisme dans un domaine sont intimement liées à celles qui se produisent dans d'autres domaines et elles se renforcent mutuellement. En examinant les différentes manières dont l'intelligence artificielle contribue à la discrimination raciale, on peut repérer l'influence qu'elle exerce sur les manifestations du racisme systémique et la façon dont elle les aggrave ainsi que la façon dont elle renforce de manière globale l'oppression systémique dans la société sur la base de critères raciaux et ethniques[10].

1. Problèmes liés aux données

13. L'essor des systèmes d'intelligence artificielle et des algorithmes d'apprentissage automatique a conduit à la numérisation à très grande échelle des données. Les algorithmes utilisent ces données pour prendre des décisions et des mesures dans plusieurs secteurs. Cela étant, les jeux de données utilisés pour entraîner ces algorithmes sont souvent incomplets ou certains groupes de personnes y sont sous-représentés. La surreprésentation ou la sous-représentation de groupes de population particuliers dans les jeux de données d'apprentissage, notamment lorsqu'elle est fondée sur des critères raciaux et ethniques, peut générer un biais algorithmique. De même, si ces jeux de données comprennent des données déjà biaisées, les algorithmes peuvent produire des résultats biaisés.

14. S'ils reposent sur des données d'apprentissage insuffisantes, les algorithmes peuvent faire des prédictions qui sont systématiquement discriminatoires envers les groupes non représentés ou sous-représentés dans les données. Non seulement des biais algorithmiques peuvent se produire si les données sont insuffisantes, mais les algorithmes qui reposent sur des données non représentatives peuvent également produire des résultats faussés. Par exemple, une étude portant sur les bases de données d'images utilisées par les forces de l'ordre aux États-Unis a montré que les personnes d'ascendance africaine risquaient davantage d'être identifiées à tort par les systèmes de reconnaissance faciale utilisés par ces forces. Ce biais était imputable à des erreurs d'identification faciale concernant ce groupe et à la surreprésentation des personnes d'ascendance africaine dans les bases de données photographiques de la police, ce qui traduit des schémas historiques de racisme systémique[11].

[8] Contribution de la Commission australienne des droits de l'homme. Toutes les contributions seront publiées sur le site Web du Haut-Commissariat des Nations Unies aux droits de l'homme.

[9] A/HRC/47/53, par. 9.

[10] A/HRC/44/57, par. 43.

[11] Nicol Turner Lee, Paul Resnick et Genie Barton, « Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms », Brookings Institution, 22 mai 2019.

15. Des biais historiques peuvent influer sur les données. Un élément central de l'apprentissage automatique consiste à faire des prédictions sur l'avenir en s'appuyant sur des données du passé. Cela étant, si les données antérieures manquent d'objectivité à l'égard de certains groupes, notamment parce qu'elles sont fondées sur des critères raciaux et ethniques, les modèles informatiques peuvent reproduire et amplifier ces biais. Si l'on utilise des données biaisées ou erronées pour prendre des décisions dans la vie réelle, on risque de cibler davantage des groupes raciaux et ethniques marginalisés et de leur porter préjudice dans la mesure où les données utilisées dans le contexte de l'intelligence artificielle servent à créer davantage de données sur lesquelles s'appuieront de futures décisions. Ces systèmes qui s'auto-alimentent peuvent reproduire et aggraver les disparités existantes.

16. En dernier lieu se pose la question du respect de la vie privée. Les données utilisées dans les systèmes d'intelligence artificielle comprennent souvent des informations personnelles sur leurs détenteurs. La collecte et le traitement de données sans consentement portent atteinte au droit à la protection de la vie privée. Il arrive en outre que des données collectées dans un cadre particulier, par exemple dans le contexte de la santé (notamment lorsqu'on utilise des applications dans ce domaine), soient partagées sans consentement et utilisées dans d'autres cadres, par exemple à des fins d'application de la loi. Les fuites de données et l'accès non autorisé à des informations personnelles au moyen du piratage informatique soulèvent de nouvelles préoccupations en matière de protection de la vie privée. Pour les personnes appartenant à des groupes marginalisés sur le plan racial, les préoccupations liées au droit à la protection de la vie privée peuvent être amplifiées, les atteintes à ce droit de l'homme pouvant les exposer à un risque d'ostracisme, de discrimination ou de danger physique[12].

2. Problèmes liés à la conception des algorithmes

17. Une deuxième forme courante de biais constatée dans les outils d'intelligence artificielle découle de la conception des algorithmes. Si les choix en la matière véhiculent des biais, l'algorithme peut donner des résultats faussés, et ce même lorsque les données qui l'alimentent sont parfaitement représentatives. Les décisions relatives aux paramètres et au fonctionnement d'un algorithme peuvent introduire des biais. Les concepteurs définissent les variables qui seront utilisées par un algorithme et paramètrent les catégories ou les seuils qui serviront à trier les informations ainsi que les données qui permettront de construire l'algorithme. Les choix qu'ils effectuent portent notamment sur la manière de mesurer des éléments particuliers et de définir ce qui constitue un succès au niveau de l'algorithme. Parfois, le profil ou les perspectives des concepteurs peuvent les amener à intégrer inconsciemment des biais, y compris raciaux, dans la conception de leurs algorithmes[13]. Le manque de diversité dans les secteurs des technologies numériques serait aggravé par le fait qu'aucun processus de consultation inclusif n'est prévu dans le cadre du développement des systèmes d'intelligence artificielle, ce qui alimente les problèmes de conception des algorithmes[14].

18. Les choix de conception des algorithmes peuvent avoir d'importants effets discriminatoires dans la vie réelle. Par exemple, lors de l'élaboration d'un algorithme d'évaluation du risque de crédit, la manière dont le risque est défini et mesuré peut conduire à des résultats discriminatoires. Ainsi, la décision d'un concepteur d'utiliser les cotes de crédit comme indicateur de risque pourrait entraîner des résultats discriminatoires pour les groupes de personnes qui ont tendance à avoir des cotes de crédit plus faibles. Des études ont montré qu'il pouvait y avoir une forte corrélation entre la cote de crédit, la race et d'autres indicateurs démographiques et que l'utilisation des cotes de crédit désavantageait certains groupes[15]. Dans nombre de cas, cette corrélation peut être considérée comme un produit dérivé du racisme et de l'exclusion systémiques existants. Des personnes peuvent se retrouver

[12] Samantha Lai et Brooke Tanner, « Examining the intersection of data privacy and civil rights », Brookings Institution, 18 juillet 2022. Voir aussi la contribution de Privacy International.

[13] Ninareh Mehrabi et al., « A survey on bias and fairness in machine learning », *ACM Computing Surveys*, vol. 54, nº 6 (2022) ; Contribution de The London Story ; A/HRC/44/57, par. 17.

[14] Contribution de NetMission.Asia.

[15] A. R. Lange et Natasha Duarte, « Understanding bias in algorithmic design », Medium, 6 septembre 2017.

désavantagées à la suite du choix qu'a fait un concepteur d'algorithmes d'utiliser les cotes de crédit pour évaluer le risque de crédit, bien qu'en apparence il ne s'agisse pas d'un critère discriminatoire.

3. Utilisation à des fins discriminatoires

19. Dans certains cas, l'intelligence artificielle peut être utilisée à des fins ouvertement racistes lorsqu'elle est déployée de manière sélective contre des groupes ciblés, ce qui entraîne des effets discriminatoires. Selon certaines informations, des services de police utilisent intentionnellement l'intelligence artificielle pour enquêter sur des populations particulières et les soumettre à des contrôles excessifs sur la base de critères discriminatoires sur le plan racial[16]. Une discrimination intentionnelle peut en outre se produire lorsque des gouvernements et d'autres entités exploitent les capacités de la technologie pour surveiller et cibler des groupes ou des individus particuliers et se livrer à un profilage sur la base de leur identité raciale ou ethnique[17].

20. La désinformation est un autre moyen d'utiliser l'intelligence artificielle à des fins ouvertement racistes. Des acteurs politiques peuvent faire appel à l'intelligence artificielle pour générer des textes, des images et des vidéos visant à manipuler l'opinion publique et les processus politiques en leur faveur et à saper la confiance dans les institutions, notamment sur la base de critères raciaux. Des gouvernements auraient en outre utilisé l'intelligence artificielle pour semer la discorde et faciliter la censure en ligne[18].

4. Problèmes liés au principe de responsabilité

21. Le fait que certains outils d'intelligence artificielle prennent des décisions sans aucune intervention humaine entraîne une prise de décisions qui est dissimulée, comme si elle se déroulait dans une « boîte noire » opaque. De plus, un algorithme d'intelligence artificielle peut prendre des décisions de manière indépendante puisqu'une fois exposé à des données, il se met constamment à jour. Au fil du temps, un tel outil peut faire reposer ses décisions sur des facteurs autres que ceux paramétrés au départ. Ces facteurs proviennent de tendances que l'outil a lui-même repérées dans les données. Au fur et à mesure que l'algorithme intègre ces nouvelles tendances dans son code et dans sa prise de décisions, les personnes qui se fient à l'algorithme risquent de ne plus pouvoir « regarder sous le capot » et déterminer les critères que l'algorithme a utilisés pour produire certains résultats. Ainsi, le problème de la boîte noire rend le processus de raisonnement de l'intelligence artificielle insidieux et opaque[19]. En outre, de nombreux algorithmes développés par des entreprises échappent à tout contrôle en raison des lois relatives aux contrats et à la propriété intellectuelle, ce qui aggrave les problèmes liés au principe de responsabilité[20].

22. Ce problème de la boîte noire a des incidences particulièrement préoccupantes dans le contexte du racisme systémique. Comme décrit ci-dessus, le racisme systémique est un fléau insidieux et profondément destructeur qui touche l'ensemble de la société. Les forces qui l'alimentent sont difficiles à repérer et les lacunes qui continuent d'être observées concernant la collecte de données ventilées selon la race et l'origine ethnique aggravent encore la situation[21]. Si aucun mécanisme chargé de veiller à l'application effective du principe de responsabilité n'est créé dans le domaine de l'intelligence artificielle, il est fort probable que cette dernière alimentera le phénomène déjà insidieux et destructeur qu'est le racisme systémique.

---

[16] Voir Amnesty International, *Decode Surveillance NYC: Methodology* (Londres, 2022) ; contribution de NetMission.Asia.

[17] Contribution de NetMission.Asia.

[18] Tate Ryan-Mosley, « How generative AI is boosting the spread of disinformation and propaganda », *MIT Technology Review*, 4 octobre 2023.

[19] Yavar Bathaee, « The artificial intelligence black box and the failure of intent and causation », *Harvard Journal of Law and Technology*, vol. 31, nº 2 (2018) ; A/HRC/44/57, par. 34 ; Renata M. O'Donnell, « Challenging racist predictive policing algorithms under the Equal Protection Clause », *New York University Law Review*, vol. 94, nº 3 (juin 2019).

[20] A/HRC/44/57, par. 44.

[21] A/HRC/47/53, par. 16.

23. Les questions relatives au principe de responsabilité en matière d'intelligence artificielle ont d'importantes répercussions sur la capacité des victimes d'actes de discrimination raciale d'accéder à des recours effectifs. Aujourd'hui, lorsque des personnes appartenant à des groupes raciaux et ethniques marginalisés font l'objet d'un traitement différent en raison de décisions humaines, les tribunaux et d'autres mécanismes veillant à l'application du principe de responsabilité peuvent examiner les mesures prises à leur égard afin de déterminer si elles étaient intentionnelles et justifiables[22]. Lorsque les décisions sont prises par des personnes, il existe souvent des éléments probants qui peuvent être utilisés pour trancher ces questions. Cela n'est souvent pas le cas lorsque les décisions sont le résultat de processus autonomes[23]. Les problèmes liés au phénomène de la boîte noire aggraveront encore les obstacles importants auxquels font déjà face les victimes de discrimination raciale pour accéder à la justice.

B. Utilisation de l'intelligence artificielle et effets discriminatoires

24. Dans la présente partie, la Rapporteuse spéciale donne des exemples de l'utilisation de l'intelligence artificielle dans divers domaines et des effets discriminatoires de cette technologie sur le plan racial. Ces exemples, qui n'ont rien d'exhaustif, visent à montrer clairement que l'intelligence artificielle contribue déjà à la discrimination raciale. Selon la Rapporteuse spéciale, il s'agit de manifestations de la discrimination raciale qui sont étroitement liées entre elles, s'aggravent mutuellement et renforcent de manière globale l'oppression systémique qui touche l'ensemble de la société sur la base de critères raciaux et ethniques.

25. La Rapporteuse spéciale a choisi trois domaines pour illustrer les effets discriminatoires de l'intelligence artificielle : le maintien de l'ordre, la sécurité et le système de justice pénale ; l'éducation ; la santé. En ce qui concerne l'utilisation de l'intelligence artificielle dans d'autres contextes, elle recommande de consulter les rapports établis par la précédente titulaire du mandat sur la montée des frontières numériques et l'état des lieux de la discrimination raciale et xénophobe dans la gestion numérique des frontières et de l'immigration, ainsi que sur l'utilisation des technologies numériques dans le contrôle des frontières et de l'immigration[24]. La Rapporteuse spéciale invite en outre les lecteurs à consulter le rapport sur les discours de haine raciale en ligne qu'elle a soumis à l'Assemblée générale à sa soixante-dix-huitième session et qui porte sur l'utilisation de l'intelligence artificielle dans la modération de contenu publié sur les médias sociaux[25], ainsi que le rapport que le Rapporteur spécial sur les droits de l'homme et l'extrême pauvreté a soumis à l'Assemblée générale à sa soixante-quatorzième session et qui comprend une analyse de l'utilisation de l'intelligence artificielle dans les systèmes de protection sociale[26].

1. Maintien de l'ordre, sécurité et système de justice pénale

a) Identification automatisée

26. Les services de police utilisent des outils d'identification automatisée pour établir des correspondances entre ce qu'ils observent dans un environnement particulier et une concordance potentielle dans une base de données. La reconnaissance faciale est l'un des types d'outils d'identification automatisée les plus courants. Les outils de reconnaissance faciale utilisent des enregistrements vidéo ou des photographies d'une personne et les intègrent à des algorithmes. Ces algorithmes comparent les images à une base de données de photographies prises par la police, de photos de permis de conduire ou d'autres images dans le but d'identifier la personne[27]. Les concepteurs de ces outils entraînent les modèles sur lesquels reposent ces derniers en leur montrant des images de visages à l'aide d'un processus

[22] Bathaee, « The artificial intelligence black box ».
[23] Ibid.
[24] A/75/590 et A/HRC/48/76.
[25] A/78/538.
[26] A/74/493.
[27] Marissa Gerchick et Matt Cagle, « When it comes to facial recognition, there is no such thing as a magic number », American Civil Liberties Union, 7 février 2024.

d'apprentissage automatique, ceci afin que ces outils deviennent capables d'identifier les caractéristiques distinctives des visages[28]. Les jeux de données d'images qui servent à entraîner ces modèles ne sont cependant pas toujours représentatifs sur le plan démographique[29]. Dans une étude portant sur une base de données d'images très populaire, les chercheurs ont constaté que les hommes âgés de 18 à 40 ans y étaient surreprésentés et que les personnes à la peau foncée y étaient sous-représentées[30]. Selon une autre étude portant sur des systèmes de reconnaissance faciale disponibles dans le commerce, les algorithmes utilisés pour analyser les visages en fonction du sexe reposent à l'origine sur des jeux de données comportant une écrasante majorité de visages d'hommes blancs[31]. Le manque de diversité raciale, sexuelle et culturelle dans les jeux de données employés pour entraîner les outils d'intelligence artificielle conduit à l'un des problèmes classiques liés aux données qui est décrit ci-dessus. Les groupes qui sont sous-représentés dans les données d'apprentissage, notamment ceux qui subissent des formes de discrimination intersectionnelle, risquent davantage de ne pas être correctement identifiés par l'algorithme.

27. Des erreurs d'identification liées à la reconnaissance faciale auraient entraîné une hausse du nombre d'arrestations de personnes d'ascendance africaine[32]. Le Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression et la Haute-Commissaire des Nations Unies aux droits de l'homme ont fait observer que les outils de reconnaissance faciale contribuaient souvent à la discrimination illégale et au profilage racial[33]. Malgré ces préoccupations relatives aux droits de l'homme, les services de police de plusieurs pays ont déployé des systèmes de reconnaissance faciale. Le Gouvernement indien aurait par exemple beaucoup investi dans de tels systèmes. Celui qu'utilise la police de Delhi ne serait fiable que dans 2 % des cas et exposerait les minorités à un risque disproportionné d'erreur d'identification et d'arrestation illégale[34]. Les forces de l'ordre brésiliennes auraient accusé et arrêté à tort des personnes sur la base d'outils de reconnaissance faciale défectueux. Selon une étude menée en 2019, 90 % des personnes arrêtées dans des villes brésiliennes après avoir été identifiées à l'aide de la reconnaissance faciale étaient d'ascendance africaine[35].

28. Les systèmes de détection des coups de feu sont un autre type d'outil d'identification automatisée couramment utilisé par les forces de l'ordre dans certains pays. L'un de ces systèmes, baptisé ShotSpotter, consiste à placer dans les quartiers des capteurs contenant un microphone, un système GPS, une mémoire, un processeur et une connexion au réseau mobile[36]. Lorsque les capteurs détectent un bruit susceptible d'être un coup de feu, un algorithme triangule l'emplacement de la source du bruit puis passe en revue les différents bruits avant d'envoyer les fichiers audios pertinents à une personne pour vérification[37]. À en juger par les informations disponibles, les systèmes de détection des coups de feu sont déployés de manière disproportionnée dans les quartiers habités par des groupes marginalisés

[28] Julia Dressel et Andrew Warren, « Breaking down data analytics and AI in criminal justice », Recidiviz, 8 mars 2022.

[29] Contribution de AI for the People.

[30] Khari Johnson, « ImageNet creators find blurring faces for privacy has a 'minimal impact on accuracy' », VentureBeat, 16 mars 2021.

[31] Joy Buolamwini et Timnit Gebru, « Gender shades: intersectional accuracy disparities in commercial gender classification », *Proceedings of Machine Learning Research*, vol. 81 (2018). Voir aussi Gerchick et Cagle, « When it comes to facial recognition, there is no such thing as a magic number » ; contribution de AI for the People ; contribution d'Internet Lab.

[32] Gerchick et Cagle, « When it comes to facial recognition, there is no such thing as a magic number ».

[33] Voir A/HRC/41/35 et A/HRC/48/31.

[34] Amnesty International, « Ban the scan: Hyderabad », disponible à l'adresse suivante : https://banthescan.amnesty.org/hyderabad/.

[35] Contribution d'un groupe d'experts basés au Brésil.

[36] Alisha Ebrahimji, « Critics of ShotSpotter gunfire detection system say it's ineffective, biased and costly », CNN, 24 février 2024.

[37] Jay Stanley, « Four problems with the ShotSpotter gunshot detection system », American Civil Liberties Union, 24 août 2021.

sur le plan racial[38]. De plus, ils présentent un taux d'erreur qui peut être très élevé. L'installation de tels systèmes dans des quartiers où vivent des groupes raciaux et ethniques marginalisés et leur manque de fiabilité aggravent encore les préjugés systémiques qui existent au sein des forces de l'ordre.

29. De nombreux exemples montrent que l'utilisation des technologies d'identification automatisée a bouleversé la vie d'individus. En 2019, aux États-Unis, un homme noir du New Jersey aurait été arrêté illégalement et maintenu en détention pendant dix jours en raison d'une erreur de reconnaissance faciale. Alors qu'elles disposaient d'éléments de preuve à décharge, les autorités ont mis près d'un an à classer l'affaire. L'intéressé risquait jusqu'à vingt-cinq années d'emprisonnement pour les chefs d'accusation retenus contre lui. Cette erreur a eu de lourdes répercussions sur la vie de cet homme[39]. En février 2024, à la suite d'une fausse alerte de ShotSpotter, les forces de l'ordre de Chicago auraient ouvert le feu sur un enfant qui allumait des feux d'artifice[40]. L'adoption par les Forces de défense israéliennes du système Wolf Pack est un autre exemple de l'utilisation de ce type de technologie de l'intelligence artificielle. Il s'agit d'une vaste base de données contenant des images et toutes les informations disponibles sur les Palestiniens de Cisjordanie et qui intègre divers programmes de surveillance tels que Blue Wolf et Red Wolf[41]. Dans la vieille ville d'Hébron, les Forces de défense israéliennes auraient installé des caméras s'appuyant sur l'intelligence artificielle pour identifier des visages. Ces caméras sont reliées au programme Blue Wolf, application mobile qui permet aux soldats de repérer les Palestiniens de Cisjordanie et de les classer au moyen d'une vaste base de données biométriques dans laquelle la plupart des personnes ont été enregistrées sans leur consentement, programme qui entraîne une surveillance permanente des Palestiniens. L'application rigoureuse du système Wolf Pack par les Forces de défense israéliennes aggrave encore l'apartheid dont font l'objet les Palestiniens[42]. Ces exemples montrent les graves incidences que l'utilisation de systèmes d'intelligence artificielle pour prendre des décisions importantes dans des environnements à haut risque a sur les droits de l'homme.

b) Algorithmes de police prédictive

30. La police prédictive est une autre forme de technologie de l'intelligence artificielle couramment utilisée par les services de police. Les outils de police prédictive cherchent à évaluer qui commettra des infractions à l'avenir et où celles-ci pourraient se produire en se fondant sur l'étude de données de localisation et de données personnelles.

31. La police prédictive peut aggraver encore les contrôles excessifs dont ont toujours fait l'objet certaines communautés sur la base de critères raciaux et ethniques[43]. Étant donné que les forces de l'ordre concentrent depuis longtemps leur action sur ces quartiers, les membres des communautés qui y habitent sont surreprésentés dans les registres de la police. Cette surreprésentation a une incidence sur les endroits où les algorithmes prédisent que des infractions seront commises, ce qui fait que les services de police sont davantage déployés dans les zones en question[44]. On retrouve dans la police prédictive certains aspects du

[38] Ibid. ; MacArthur Justice Center, « ShotSpotter is deployed overwhelmingly in Black and Latinx neighborhoods in Chicago », disponible à l'adresse suivante : https://endpolicesurveillance.com/burden-on-communities-of-color/.

[39] Gerchick et Cagle, « When it comes to facial recognition, there is no such thing as a magic number » ; Khari Johnson, « How wrongful arrests based on AI derailed 3 men's lives », *Wired*, 7 mars 2022.

[40] Adam Schwartz, « Responding to ShotSpotter, police shoot at child lighting fireworks », Electronic Frontier Foundation, 22 mars 2024.

[41] Amnesty International, *Apartheid automatisé : Comment la reconnaissance faciale fragmente, ségrègue et contrôle les Palestiniens et les Palestiniennes dans les TPO* (2023), p. 46-51.

[42] Sophia Goodfriend, « Algorithmic State violence: automated surveillance and Palestinian dispossession in Hebron's Old City », *International Journal of Middle East Studies*, vol. 55, nº 3 (2023).

[43] Tim Lau, « Predictive policing explained », Brennan Center for Justice, 1er avril 2020 ; Jon Fasman, « The black box of justice: how secret algorithms have changed policing », *Fast Company*, 9 février 2021.

[44] Kristian Lum et William Isaac, « To predict and serve? », *Significance*, vol. 13, nº 5 (2016) ; contribution de la Commission australienne des droits de l'homme.

problème de la boîte noire car les algorithmes manquent de transparence, notamment en ce qui concerne les données analysées et la manière dont les prédictions sont utilisées[45].

32. Les algorithmes prédictifs centrés sur la localisation s'appuient sur les liens entre les lieux, les événements et les données historiques en matière de criminalité pour prédire quand et où de futures infractions sont susceptibles d'être commises[46]. Les forces de police planifient leurs patrouilles en conséquence. Lorsque les agents qui interviennent dans les quartiers faisant l'objet de contrôles excessifs enregistrent de nouvelles infractions, cela crée un engrenage dans lequel l'algorithme génère des prédictions de plus en plus biaisées concernant ces quartiers. En résumé, les biais du passé entraînent des biais pour l'avenir. Au Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, une étude commandée par le Gouvernement sur les biais algorithmiques dans les activités de police a montré que la désignation de lieux géographiques comme zones sensibles en matière de criminalité pouvait inciter les agents à s'attendre à une hausse de la criminalité dans ces zones. En conséquence, ces derniers étaient plus susceptibles d'interpeller ou d'arrêter des personnes dans ces zones sensibles sur la base de préjugés plutôt que sur la base de véritables impératifs de sécurité publique[47]. En Uruguay, des chercheurs ont découvert que les données utilisées dans les algorithmes prédictifs centrés sur la localisation pouvaient être biaisées, la variable de localisation pouvant indirectement indiquer le milieu socioéconomique ou l'origine ethnique et entraîner une discrimination[48].

33. Les outils de police prédictive centrés sur les personnes cherchent, en s'appuyant sur des données de base concernant les personnes, à prédire qui pourrait à l'avenir commettre une infraction. Les données de base peuvent comprendre l'âge, le sexe, la situation matrimoniale, les antécédents en matière de consommation de substances psychoactives et le casier judiciaire d'une personne. De même que pour les outils centrés sur la localisation, les données relatives aux arrestations passées, souvent entachées de racisme systémique dans le système de justice pénale, peuvent fausser les prédictions qu'établiront les algorithmes. L'utilisation de variables telles que le milieu socioéconomique, le niveau d'instruction et le lieu de résidence peut indirectement indiquer la race et perpétuer des biais historiques[49]. En Australie, la police de Nouvelle-Galles du Sud a utilisé un plan de gestion des cibles suspectes basé sur des algorithmes (*Suspect Target Management Plan*) pour identifier les individus susceptibles de commettre des infractions pénales. Ce plan aurait conduit à un nombre disproportionné de contacts entre la police et les membres des communautés aborigènes et insulaires du détroit de Torres avant d'être mis hors service[50].

c) Algorithmes servant à évaluer les risques de récidive

34. Des outils servant à évaluer les risques de récidive sont utilisés pour orienter les décisions prises à différents stades d'une procédure pénale, par exemple concernant la libération sous caution, la caution, la détermination de la peine et la libération conditionnelle[51]. Ces outils utilisent des données historiques pour évaluer la probabilité que des accusés se comportent d'une certaine manière et se rendent coupables de récidive. Ils produisent des probabilités concernant les risques en utilisant des informations provenant de sources telles que les casiers judiciaires des accusés et les enquêtes menées sur ces derniers[52].

[45] Lau, « Predictive policing explained ».

[46] Will Douglas Heaven, « Predictive policing algorithms are racist. They need to be dismantled », *MIT Technology Review*, 17 juillet 2020.

[47] Ibid. Voir aussi Gouvernement du Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, « Report commissioned by CDEI calls for measures to address bias in police use of data analytics », 16 septembre 2019.

[48] Juan Ortiz Freuler et Carlos Iglesias, « Algorithms and artificial intelligence in Latin America: a study of implementation by governments in Argentina and Uruguay », World Wide Web Foundation, septembre 2018 ; Eticas Foundation, « Uruguay's Ministry of the Interior invests in predictive policing », 13 septembre 2021.

[49] Heaven, « Predictive policing algorithms are racist ».

[50] Contribution de la Commission australienne des droits de l'homme.

[51] Julia Angwin et al., « Machine bias », ProPublica, 23 mai 2016.

[52] Ibid.

35. Les outils de prédiction du risque de récidive posent de nombreux problèmes liés à l'intelligence artificielle qui contribuent à la discrimination raciale. Premièrement, les données sur lesquelles ils reposent sont problématiques. En effet, dans les données issues du système de justice pénale qui sont utilisées pour entraîner leurs algorithmes, on retrouve des inégalités systémiques qui découlent de pratiques policières historiquement racistes[53]. En outre, les choix de conception, tels que la manière dont les variables sont mesurées ou évaluées, peuvent contribuer à la discrimination algorithmique[54]. La manière dont le concepteur d'un algorithme définit ce qui constitue un succès peut avoir des incidences sur ce à quoi l'algorithme accorde la priorité. Si un algorithme est configuré de façon à chercher à réduire au minimum le nombre de nouvelles infractions, il peut établir une corrélation entre des peines plus longues et un taux de récidive plus faible, les personnes ne pouvant pas récidiver pendant leur incarcération. Sur la base de cette corrélation, il peut recommander que des peines plus longues soient prononcées.

36. Des chercheurs ont émis l'opinion selon laquelle les indicateurs de récidive n'étaient pas fiables et les erreurs dans ce domaine touchaient de manière disproportionnée les groupes marginalisés sur le plan racial. Une étude menée aux États-Unis a par exemple montré que les probabilités concernant les risques ne présentaient qu'une fiabilité toute relative pour ce qui était de prévenir les crimes violents. La proportion de personnes d'ascendance africaine identifiées à tort comme de futurs auteurs d'infraction serait presque deux fois supérieure à celle des personnes blanches.

d) Systèmes d'armes autonomes

37. Les systèmes d'armes autonomes comprennent tous les systèmes d'armes dont les fonctions essentielles sont autonomes, notamment les armes létales autonomes et les armes à létalité réduite. Utilisés dans le domaine du maintien de l'ordre et dans le domaine militaire, ils restent largement incontrôlés. Ces systèmes peuvent sélectionner, détecter, repérer et attaquer des cibles sans intervention humaine. Une arme autonome est déclenchée par des capteurs et un logiciel qui associent une personne à un profil de cible déterminé par l'algorithme du système. Ces systèmes ont des effets très graves sur les droits de l'homme, notamment en ce qui concerne le droit à la vie, l'interdiction de la torture et d'autres mauvais traitements et le droit à la sécurité de la personne[55].

38. Selon des déclarations effectuées devant la Première Commission de l'Assemblée générale, les possibilités de mettre en place des garde-fous contre les dangers que représentent les armes autonomes et les applications militaires de l'intelligence artificielle s'amenuisent rapidement alors que le monde se prépare à vivre une « rupture technologique »[56]. Le Rapporteur spécial sur les exécutions extrajudiciaires, sommaires ou arbitraires a déjà recommandé au Conseil des droits de l'homme de demander à tous les États de décréter et d'appliquer des moratoires sur, au moins, l'essai, la production, l'assemblage, le transfert, l'acquisition, le déploiement et l'emploi des robots létaux autonomes[57].

39. L'utilisation de systèmes d'armes autonomes entraîne un risque sérieux de discrimination raciale grave, voire parfois mortelle. Il est probable que les critères utilisés pour sélectionner les cibles comprennent le sexe, l'âge et la race[58]. Les profils de cible comprennent aussi des critères apparemment neutres, tels que le poids ou la signature thermique, mais on retrouve souvent dans les outils les préjugés de leurs programmeurs et de la société. Ces programmeurs peuvent les paramétrer en utilisant des profils de cible

[53] Voir Heaven, « Predictive policing algorithms are racist » ; Michael Mayowa Farayola et al., « Fairness of AI in predicting the risk of recidivism: review and phase mapping of AI fairness techniques », in *Proceedings of the 18th International Conference on Availability, Reliability and Security* (Association for Computing Machinery, 2023).

[54] Mehrabi et al., « A survey on bias and fairness in machine learning ».

[55] Amnesty International, « Autonomous weapons systems: five key human rights issues for consideration » (avril 2015), p. 5.

[56] ONU, « Without adequate guardrails, artificial intelligence threatens global security in evolution from algorithms to armaments, speaker tells First Committee », 24 octobre 2023.

[57] A/HRC/23/47, par. 113.

[58] Ray Acheson, « Gender and bias », disponible à l'adresse suivante : https://www.stopkillerrobots.org/wp-content/uploads/2021/09/Gender-and-Bias.pdf.

intentionnellement discriminatoires[59]. Ainsi, selon des informations, Israël a recours à des systèmes d'armes létales autonomes et semi-autonomes. Il utiliserait notamment des quadrirotors téléguidés pour cibler les Palestiniens et des systèmes qui génèrent automatiquement des cibles, à une vitesse et dans des volumes inédits, pour établir des listes de personnes à abattre[60]. The Gospel et Lavender, deux systèmes faisant appel à l'intelligence artificielle qu'utilisent les Forces de défense israéliennes, auraient intensifié les niveaux de destruction à Gaza, entraînant un nombre important de victimes, en particulier parmi les femmes et les enfants palestiniens[61].

2. Soins de santé

a) Évaluation du risque pour la santé

40. L'intelligence artificielle peut être utilisée pour évaluer, sous la forme d'un score, le risque pour la santé à diverses fins, notamment le diagnostic médical et la planification des soins. Lorsque des algorithmes servent à l'allocation des ressources de soins, ils peuvent avoir, du fait de leur conception et des données utilisées pour entraîner les systèmes d'intelligence artificielle, des effets discriminatoires sur le plan racial. Dans certains cas, des algorithmes sont conçus pour déterminer les personnes qui devraient bénéficier de soins supplémentaires, dont les besoins médicaux sont évalués sur la base des dépenses de santé précédemment engagées. Les données sur lesquelles sont fondées les décisions peuvent être influencées par le fait que des personnes appartenant à des groupes raciaux et ethniques marginalisés n'ont pas suffisamment eu accès, compte tenu du contexte de racisme systémique, aux soins de santé, et que des pathologies différentes ont fait leur apparition, sous l'effet des disparités observées entre les différents facteurs socioéconomiques déterminants de la santé.

41. Aux États-Unis, un calculateur a été mis au point afin d'aider les prestataires de soins de santé à évaluer la probabilité de succès d'un accouchement par voie basse après un accouchement par césarienne. Une étude réalisée en 2019 a révélé des biais dans l'algorithme de base utilisé par le calculateur. En effet, celui-ci appliquait deux facteurs de correction fondés sur la race, ce qui donnait lieu à des taux de succès d'accouchement par voie basse inférieurs pour les femmes d'ascendance africaine et les femmes hispaniques, par rapport aux femmes blanches présentant des caractéristiques similaires. Le calculateur a donc potentiellement aggravé les disparités raciales relevées en ce qui concerne les effets sur le plan de la santé maternelle, en dissuadant les cliniciens de proposer un accouchement par voie basse aux femmes d'ascendance africaine et aux femmes hispaniques, ce qui a entraîné une hausse du taux de césariennes[62].

b) Détection des maladies

42. La détection des maladies, dont le cancer, est une des autres applications de l'intelligence artificielle[63]. Les systèmes en la matière, entraînés au moyen de vastes ensembles de données contenant des milliers, voire des millions d'images, notamment des radiographies, des images et photographies de pathologies, peuvent apprendre à distinguer les lésions normales des lésions cancéreuses[64]. Ce déploiement de l'intelligence artificielle peut contribuer, dans une large mesure, à la détection précoce du cancer, et donc potentiellement sauver des vies tout en améliorant l'efficacité du système de soins de santé.

[59] Bonnie Docherty, « Expert Panel on the Social and Humanitarian Impact of Autonomous Weapons at the Latin American and Caribbean Conference on Autonomous Weapons », Human Rights Watch, 8 mars 2023.

[60] Marwa Fatafta et Daniel Leufer, « Artificial genocidal intelligence: how Israel is automating human rights abuses and war crimes », Access Now, 9 mai 2024.

[61] Yuval Abraham, « 'Lavender': the AI machine directing Israel's bombing spree in Gaza », *+972 Magazine*, 3 avril 2024.

[62] Darshali A. Vyas et al., « Challenging the use of race in the Vaginal Birth after Cesarean Section Calculator », *Women's Health Issues*, vol. 29, n° 3 (2019).

[63] Contribution de Privacy International.

[64] Likhitha Kolla et Ravi B. Parikh, « Uses and limitations of artificial intelligence for oncology », *Cancer*, 30 mars 2024.

Cependant, les personnes appartenant à des groupes raciaux et ethniques marginalisés ne peuvent bénéficier, dans des conditions d'égalité avec les autres, de ces avancées, puisque les algorithmes ne peuvent pas être généralisés aux groupes de patients qui ne sont pas correctement représentés dans les données d'entraînement. Selon des chercheurs, l'utilisation d'algorithmes d'intelligence artificielle pour détecter le cancer de la peau donne de moins bons résultats pour les personnes ayant une peau plus foncée parce que de nombreux ensembles de données d'images publiques utilisés pour les entraîner comportent des biais, la diversité, en termes de couleur de peau et d'origine ethnique, n'étant pas suffisante[65]. Ainsi, une étude sur les lésions cutanées, menée sur 21 ensembles de données en libre accès contenant plus de 100 000 images, a montré que les images de personnes ayant une peau plus foncée étaient largement sous-représentées[66].

c) Dispositifs médicaux faisant appel à l'intelligence artificielle

43. Le développement et l'utilisation de l'intelligence artificielle sont parallèles à ceux que connaissent d'autres technologies dans le secteur des soins de santé, notamment les dispositifs médicaux. Nombre de ces dispositifs faisant appel à l'intelligence artificielle, des préjugés raciaux peuvent en altérer le fonctionnement. Par exemple, au Royaume-Uni, un rapport a montré que des biais étaient ajoutés à chaque stade du développement des dispositifs médicaux, y compris aux stades du développement des algorithmes et de l'apprentissage automatique. Pendant la pandémie de maladie à coronavirus (COVID-19), l'utilisation d'oxymètres de pouls pour mesurer les faibles concentrations d'oxygène dans le sang a conduit à des surestimations du taux d'oxygène chez les personnes ayant une peau plus foncée[67].

3. Éducation

a) Algorithmes mesurant la réussite scolaire et professionnelle

44. Dans des pays comme la Finlande et les États-Unis, des outils d'analyse prédictive sont utilisés dans le secteur de l'éducation pour déterminer la probabilité de réussite des élèves, sur la base de données, d'algorithmes statistiques et de l'apprentissage automatique[68]. Ces algorithmes s'appuient notamment sur les notes obtenues par les élèves et sur des données liées à leur assiduité, leur comportement et leur activité en ligne. Ils sont conçus pour aider les enseignants à guider les élèves dans les décisions qu'ils prennent concernant leurs études et leur orientation. S'ils ont pour mission d'aider les enseignants à améliorer les résultats des élèves, les outils d'analyse prédictive sous-estiment souvent les chances de réussite, sur le plan scolaire et professionnel, des personnes appartenant à des minorités raciales, compte tenu de la manière dont les algorithmes sont conçus et dont les données sont sélectionnées. Les enseignants risquent, sur la foi de ces estimations, de détourner les élèves appartenant à des groupes raciaux et ethniques marginalisés de choix d'études ou d'orientation qui leur permettraient de réaliser pleinement leur potentiel et leur donneraient l'occasion unique de rompre le cycle de l'exclusion, ou de consacrer moins de ressources à ces élèves.

b) Algorithmes de notation

45. En règle générale, les algorithmes de notation s'appuient sur des données rétrospectives pour évaluer les résultats scolaires. Or ces données peuvent être faussées par le racisme systémique qui existe depuis longtemps dans les établissements d'enseignement. Ce biais sera alors reproduit par les algorithmes de notation prédictive, en particulier en

[65] David Wen et al., « Characteristics of publicly available skin cancer image datasets : a systematic review », *The Lancet Digital Health*, vol. 4, nº 1 (2022).

[66] Ibid. Voir aussi la contribution de Privacy International.

[67] Contribution de Privacy International.

[68] Stina Westman et al., « Artificial intelligence for career guidance − current requirements and prospects for the future », *International Academic Forum Journal of Education*, vol. 9, nº 4 (2021) ; Kelli A. Bird, Benjamin L. Castleman et Yifeng Song, « Are algorithms biased in education ? Exploring racial bias in predicting community college student success », *Journal of Policy Analysis and Management*, 31 janvier 2024.

l'absence de toute intervention des enseignants[69]. Les algorithmes de notation peuvent jouer un rôle déterminant dans les perspectives offertes aux élèves, notamment en ce qui concerne l'accès à l'enseignement universitaire ou les possibilités d'emploi après les études. Les décisions automatisées fondées sur des préjugés raciaux risquent donc de limiter les possibilités offertes aux élèves appartenant à de groupes raciaux et ethniques marginalisés, et d'empêcher l'éducation de jouer pleinement son rôle dans la lutte contre le racisme systémique.

46. L'exemple du déploiement au Royaume-Uni d'un algorithme de notation devrait sonner comme une mise en garde. En 2020, les examens de fins d'études secondaires (A-level) ont été annulés en raison de la pandémie de COVID-19. En lieu et place de l'évaluation finale, il a été demandé aux enseignants de prédire les résultats des élèves. L'organisme national de réglementation des notes a ensuite appliqué un algorithme afin de normaliser ces prédictions, sur la base des données rétrospectives de chaque établissement. Quarante pour cent des élèves, dont beaucoup scolarisés dans des quartiers à revenu faible, ont alors vu leurs notes baisser. À l'inverse, après l'application de l'algorithme, un nombre disproportionné d'élèves d'écoles indépendantes et payantes ont vu leurs notes augmenter. Le Gouvernement a mis fin à la polémique en annulant la normalisation par l'algorithme, mais cette affaire a considérablement perturbé les procédures d'admission à l'université[70].

c) Grands modèles de langage utilisés dans le système éducatif

47. Les outils d'intelligence artificielle générative s'appuient sur de grands modèles de langage pour créer des contenus inédits, notamment des textes, de la musique, des images et des vidéos. Utilisés en milieu scolaire, ces modèles peuvent contribuer à améliorer les résultats des élèves de tous âges. D'après certaines études, les modèles de langage favorisent l'anglais, qui est la langue la plus utilisée sur Internet et celle dans laquelle travaillent la plupart des chercheurs et experts en intelligence artificielle. En outre, les ressources de données de haute qualité capables d'entraîner les modèles d'intelligence artificielle ne sont disponibles que dans une poignée des quelque 6 000 langues parlées dans le monde. Pour combler cette lacune, les entreprises ont commencé à développer des modèles de langage multilingues, qui ne sont toutefois pas aussi performants que les modèles en langue anglaise. L'utilisation de grands modèles de langage en milieu scolaire risque de désavantager les élèves dont les origines linguistiques ne sont pas représentées dans les ressources de données sous-jacentes, ce qui pourrait avoir des effets disproportionnés sur le plan racial[71].

48. La question de savoir si les outils d'intelligence artificielle générative basés sur de grands modèles de langage devraient être interdits plutôt qu'être incorporés aux programmes scolaires fait débat. Dans certains établissements d'enseignement, des mesures ont également été prises pour tenter de restreindre l'utilisation par les élèves d'outils d'intelligence artificielle générative qui s'appuient sur de grands modèles de langage. Certains établissements font appel à ces outils pour détecter l'utilisation de l'intelligence artificielle par les élèves. L'emploi de tels outils, qui peuvent comporter des biais algorithmiques, pour lutter contre la tricherie, peut créer d'autres biais préjudiciables aux élèves appartenant à des

[69] Benjamin Herold, « Why schools need to talk about racial bias in AI-powered technologies », *Education Week*, 12 avril 2022.

[70] Bryan Walsh, « How an AI grading system ignited a national controversy in U.K. », Axios, 19 août 2020 ; Daan Kolkman, « 'F**k the algorithm' ? What the world can learn from the UK's A-level grading fiasco », London School of Economics Impact Blog, 26 août 2020.

[71] Felix Richter, « The most spoken languages: on the Internet and in real life », Statista, 21 février 2024 ; Emily M. Bender, « The #BenderRule : on naming the languages we study and why it matters », The Gradient, 14 septembre 2019 ; Gabriel Nicholas et Aliya Bhatia, « Lost in translation : large language models in non-English content analysis », Center for Democracy and Technology, 23 mai 2023 ; A. Bergman et Mona Diab, « Towards responsible natural language annotation for the varieties of Arabic », in *The 60th Annual Meeting of the Association for Computational Linguistics: Findings of ACL 2022* (Association for Computational Linguistics, 2022) ; BigScience Workshop, « A 176B-parameter open-access multilingual language model » (ArXiv, 2022).

groupes raciaux et ethniques marginalisés. Lorsque les établissements n'ont pas mis en place des procédures d'appel équitables, le préjudice causé peut être plus grave encore[72].

d) Reconnaissance faciale en milieu scolaire

49. De nombreux établissements d'enseignement à travers le monde ont adopté les technologies de reconnaissance faciale, même s'il est établi que leur fonctionnement est faussé par des préjugés raciaux, comme cela a été décrit précédemment. Ces établissements font appel à des systèmes de reconnaissance faciale pour automatiser le contrôle de l'assiduité des élèves, renforcer la sécurité des locaux, surveiller les épreuves d'examen et même enregistrer les émotions des élèves afin d'assurer le suivi des apprentissages, bien souvent sans qu'une diligence raisonnable en matière de droits de l'homme ou un contrôle réglementaire soit exercé. Ainsi, au Brésil, de plus en plus d'écoles adoptent des outils de reconnaissance faciale pour rationaliser leurs activités, surveiller l'assiduité des élèves et renforcer la sécurité[73]. Or, d'après les informations disponibles, ni les autorités municipales ni les États n'ont réalisé d'études d'impact sur les droits de l'homme ou analysé les risques de discrimination associés aux logiciels de reconnaissance faciale avant de lancer de tels projets[74].

50. L'utilisation de logiciels de reconnaissance faciale en milieu éducatif a des effets discriminatoires sur le plan racial. Dans certains cas, notamment au Royaume des Pays-Bas, des élèves d'ascendance africaine ont dû s'éclairer le visage afin d'être reconnus par les systèmes d'intelligence artificielle mis en place pour faciliter l'accès à des examens importants. Outre qu'elles portent atteinte au droit des élèves de bénéficier d'une éducation dans des conditions d'égalité, de telles situations créent des tensions et des formes d'exclusion lorsque les élèves appartenant à de groupes raciaux et ethniques marginalisés ont le sentiment que le système n'a pas été conçu pour eux. L'enregistrement et la surveillance des émotions des élèves dans leur établissement ont d'importantes répercussions sur le droit à la vie privée de tous les élèves et peuvent perpétuer les préjugés raciaux. Il a été constaté que ces systèmes interprétaient différemment les expressions faciales des personnes d'ascendance africaine et celles des personnes blanches, attribuant plus fréquemment aux premières des sentiments négatifs, tels que le mépris et la colère[75].

C. Initiatives visant à réglementer et à gérer l'intelligence artificielle

51. Des États ont commencé à prendre des mesures encourageantes pour réglementer et gérer l'intelligence artificielle. La Rapporteuse spéciale souhaite, dans cette partie, appeler l'attention sur certaines de ces initiatives. Dans son analyse, qui ne se veut pas exhaustive, elle s'appuie sur les contributions d'États et de groupes de la société civile, ainsi que sur les travaux qu'elle a effectués concernant les pays et sur les études menées en vue d'établir le présent rapport.

1. Initiatives nationales

52. Les États ont pris diverses mesures pour réglementer et gérer l'intelligence artificielle au niveau national, adoptant des dispositions juridiques contraignantes et des normes générales non contraignantes, voire, dans de nombreux cas, un mélange des deux. Ainsi, au Brésil, des dispositions législatives sur la réglementation de l'espace technologique[76], y compris l'intelligence artificielle, sont en attente d'examen. Le Gouvernement brésilien a également adopté un certain nombre de notes d'orientation, telles que le document intitulé « Racism on the Internet : evidence for the formulation of digital policies », qui aurait été établi par le Ministère de l'égalité raciale et dans lequel sont énoncées des mesures de lutte

[72] Voir Regina Ta et Darrell M. West, « Should schools ban or integrate generative AI in the classroom? », Brookings Institution, 7 août 2023 ; Robert Topinka, « The software says my student cheated using AI. They say they're innocent. Who do I believe? », *The Guardian*, 13 février 2024.

[73] Contribution d'Internet Lab.

[74] Ibid.

[75] Ibid.

[76] Contribution du Brésil.

contre les biais algorithmiques, y compris raciaux[77]. La Rapporteuse spéciale salue les mesures prises pour élaborer des dispositions réglementaires particulières et contraignantes, complétées par les normes générales correspondantes. Toutefois, elle a reçu des informations préoccupantes concernant l'absence de consultation et de participation véritables des personnes d'ascendance africaine dans le cadre de l'élaboration des dispositions législatives sur la réglementation de l'intelligence artificielle, et le manque de cohérence globale entre les différentes normes et les pratiques actuelles de l'État[78].

53. Les États-Unis, qui, semble-t-il, ont pris des mesures pour établir un ensemble de normes contraignantes et non contraignantes sur l'utilisation de l'intelligence artificielle, sont un autre bon exemple. À la suite de sa visite récente dans ce pays, la Rapporteuse spéciale s'est félicitée de la signature du décret n° 14110 relatif au développement et à l'utilisation de l'intelligence artificielle dans un cadre sûr, sécurisé et fiable, ainsi que de la mention qui y est faite des risques d'utilisation biaisée et discriminatoire de l'intelligence artificielle. Lors de l'établissement du rapport, la Rapporteuse spéciale a reçu d'autres informations sur la réglementation de l'intelligence artificielle aux États-Unis, notamment sur les travaux menés aux fins de l'adoption d'une charte des droits sur l'intelligence artificielle, dans le cadre d'une démarche non contraignante d'États, tels que la Virginie, la Californie et New York, qui souhaitent réglementer ce domaine d'activité, et sur les initiatives tendant à ce que les entreprises s'engagent, sur une base volontaire, en faveur du développement de l'intelligence artificielle dans un cadre sûr, sécurisé et transparent[79]. La Rapporteuse spéciale salue ces initiatives, même si elle déplore que malgré les vastes recherches sur les biais algorithmiques importants qui sont associés, sur le plan racial, aux produits numériques commercialisés aux États-Unis, le décret susmentionné ne fasse pas expressément référence à la discrimination et aux préjugés raciaux[80].

54. Selon les informations disponibles, le Canada élabore actuellement un ensemble de normes contraignantes et non contraignantes. La loi sur l'intelligence et les données artificielles, qui est encore à l'état de projet, devrait prévoir des mesures contraignantes de contrôle des systèmes d'intelligence artificielle à haut risque. L'État canadien a en outre mis en place des normes non contraignantes, notamment le Code de conduite volontaire visant un développement et une gestion responsables des systèmes d'IA générative avancés. Il a également mis au point l'Outil d'évaluation de l'incidence algorithmique, qui a pour but d'aider les ministères et les organismes publics à évaluer et à atténuer les risques liés à l'intelligence artificielle, notamment en ce qui concerne la discrimination et les préjugés[81].

55. La Rapporteuse spéciale a reçu des informations concernant d'autres États, tels que l'Australie, la Chine, l'Inde et le Japon, qui auraient pris des dispositions pour gérer et réglementer l'intelligence artificielle, notamment au moyen de mesures de politique générale et, dans certains cas, de textes législatifs contraignants[82].

2. Initiatives régionales

56. Pour ce qui est des initiatives régionales, la Rapporteuse spéciale se félicite des informations reçues de l'Union européenne et de ses États membres concernant l'adoption de la législation sur l'intelligence artificielle[83], qui, selon elle, fixe une norme réglementaire

[77] Contribution d'un groupe d'experts basés au Brésil.

[78] Ibid.

[79] Contribution de NetMission.Asia. Voir aussi Kay Firth-Butterfield, Karen Silverman et Benjamin Larsen, « Understanding the US 'AI Bill of Rights' − and how it can help keep AI Accountable », Forum économique mondial, 14 octobre 2022 ; États-Unis, Bureau des politiques scientifiques et technologiques de la Maison Blanche, « Blueprint for an AI bill of rights : making automated systems work for the American people », livre blanc, octobre 2022 ; États-Unis, Maison Blanche, « Fact sheet : Biden-Harris Administration secures voluntary commitments from eight additional artificial intelligence companies to manage the risks posed by AI », 12 septembre 2023.

[80] A/HRC/56/68/Add.1, par. 54.

[81] Canada, Innovation, sciences et développement économique Canada, « Loi sur l'intelligence artificielle et les données (LIAD) − document complémentaire », 13 mars 2023 ; contribution de NetMission.Asia.

[82] Contribution de NetMission.Asia.

[83] Contributions de l'Union européenne et de l'Espagne.

contraignante qui aura d'importantes retombées dans l'Union européenne, grâce à l'alignement des normes juridiques nationales sur les dispositions communautaires. Elle constate avec satisfaction que le texte de la législation sur l'intelligence artificielle prend en compte le facteur racial, offre des garanties sur le plan des droits de l'homme en ce qui concerne les applications à haut risque de cette technologie, interdit certains emplois de l'intelligence artificielle et prévoit des mécanismes de recours pour les personnes lésées par l'utilisation de systèmes d'intelligence artificielle à haut risque. Elle note également avec satisfaction que le plan d'action de l'Union européenne contre le racisme 2020-2025 semble s'attaquer à la discrimination raciale découlant de l'utilisation des nouvelles technologies, telles que l'intelligence artificielle, ce qui est le signe d'une certaine cohérence stratégique entre les différentes normes de l'Union européenne[84]. En revanche, elle a reçu des informations très préoccupantes selon lesquelles des exceptions aux protections énoncées dans la législation sont prévues dans le contexte de la gestion de l'immigration et des frontières, ainsi que du maintien de l'ordre[85]. Ces exceptions seraient maintenues malgré la discrimination raciale profonde et tenace observée dans ces deux domaines et les écueils inhérents au fait d'autoriser le développement de deux cadres juridiques parallèles[86]. Une telle approche risque d'enraciner les hiérarchies raciales existantes et de perpétuer les violations des droits de l'homme commises dans le contexte de la gestion de l'immigration et des frontières, ainsi que du maintien de l'ordre, dans l'ensemble des États membres de l'Union européenne.

3. Initiatives internationales

57. La Rapporteuse spéciale a connaissance des mesures que l'ONU a prises pour faciliter la gestion de l'intelligence artificielle. Elle se félicite de la création par le Secrétaire général de l'Organe consultatif de haut niveau sur l'intelligence artificielle et de la publication récente du rapport d'activité de cet organe. Toutefois, elle déplore que ce rapport ne mentionne pas expressément le risque de discrimination et de préjugés raciaux. Elle salue le travail que le Haut-Commissariat des Nations Unies aux droits de l'homme (HCDH) a accompli pour intégrer la question des droits de l'homme dans les débats des instances internationales sur les technologies émergentes telles que l'intelligence artificielle, y compris dans le cadre du projet B-Tech. Outre l'action de l'ONU, la Rapporteuse spéciale a connaissance d'autres initiatives internationales, menées notamment par l'Organisation de coopération et de développement économiques et le Groupe des Sept pour promouvoir le dialogue et la gestion[87].

58. Les organisations internationales sont bien placées pour faciliter la coopération internationale, l'assistance technique et la recherche, et faire en sorte que la réglementation dans le domaine de l'intelligence artificielle ne creuse pas davantage les inégalités qui sont déjà flagrantes entre les pays et, dans de nombreux cas, héritées du colonialisme et de l'esclavage. Leurs infrastructures technologiques présentant de grandes disparités, les pays peuvent rencontrer des problèmes différents lorsqu'ils déploient des outils d'intelligence artificielle. L'attention accordée à la technologie de l'intelligence artificielle et les problèmes de discrimination posés par cette technologie se concentrent surtout dans les pays du Nord, ce qui pourrait entraîner une méconnaissance des effets futurs de l'intelligence artificielle sur les minorités culturelles, religieuses et autres dans les pays du Sud[88]. Le risque existe que les pays les plus développés du Nord puissent influencer le débat et le dialogue sur l'intelligence artificielle d'une manière qui perpétue les déséquilibres de pouvoir à l'échelle mondiale et limite la capacité des pays du Sud de tirer parti des bienfaits de cette technologie.

[84] Contribution de l'Union européenne. Voir également https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_FR.html.

[85] Contribution de Privacy International ; Access Now, « The EU AI Act : a failure for human rights, a victory for industry and law enforcement », 13 mars 2024.

[86] Voir A/HRC/48/76.

[87] Contribution de NetMission.Asia.

[88] Danni Yu, Hannah Rosenfeld et Abhishek Gupta, « The 'AI divide' between the Global North and the Global South », Forum économique mondial, 16 janvier 2023.

D. Cadre juridique international des droits de l'homme

59. Les technologies d'intelligence artificielle devraient être fondées sur les normes du droit international des droits de l'homme. C'est dans la Convention internationale sur l'élimination de toutes les formes de discrimination raciale que l'on trouve l'interdiction la plus complète de la discrimination raciale. Comme l'indique son article premier (par. 1), les États ont établi la Convention de sorte à y intégrer une définition large de la discrimination raciale, qui vise toute distinction, exclusion, restriction ou préférence fondée sur la race, la couleur, l'ascendance ou l'origine nationale ou ethnique, qui a pour but ou pour effet de détruire ou de compromettre la reconnaissance, la jouissance ou l'exercice, dans des conditions d'égalité, des droits de l'homme et des libertés fondamentales dans les domaines politique, économique, social et culturel ou dans tout autre domaine de la vie publique.

60. Les États parties à la Convention internationale sur l'élimination de toutes les formes de discrimination raciale se sont engagés à œuvrer en faveur de l'édification d'une communauté nationale et internationale affranchie de toute forme de racisme. Pour faciliter la réalisation concrète de l'égalité raciale, les États parties doivent, conformément à l'article 2 de la Convention, veiller à ne jamais prendre part à un quelconque acte de discrimination raciale ni à promouvoir des programmes conduisant à l'inégalité raciale. En outre, face à des situations de racisme, d'inégalité raciale ou de discrimination raciale, ils ont l'obligation de prendre des mesures efficaces et immédiates. Cette obligation d'agir est absolue. L'obligation des États parties de prévenir les inégalités et la discrimination raciales passe par l'adoption de mesures de prévention et de réparation. Si c'est dans la Convention qu'est énoncée l'interdiction la plus complète de la discrimination raciale, d'autres traités prévoient également une protection contre ces formes de discrimination.

61. Les obligations de réaliser l'égalité raciale et de garantir la non-discrimination s'étendent à tous les domaines d'action et d'influence des États, y compris la conception et l'application des technologies d'intelligence artificielle. Le caractère intentionnel ou non de la discrimination raciale découlant de l'intelligence artificielle ne change rien à l'obligation d'agir qui incombe aux États parties, compte tenu de la portée de l'interdiction de la discrimination raciale énoncée dans la Convention internationale sur l'élimination de toutes les formes de discrimination raciale et d'autres traités relatifs aux droits de l'homme. L'obligation faite aux États parties d'œuvrer en faveur de l'édification d'une communauté nationale et internationale affranchie de toute forme de discrimination raciale influence également la manière dont ils préviennent et combattent les inégalités sur leur territoire national et entre eux, pour ce qui est de la répartition des bienfaits des technologies d'intelligence artificielle.

62. Les États doivent également veiller à ce que tous les groupes raciaux et ethniques jouissent pleinement de leurs droits humains, conformément à l'article 5 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, qui dispose que l'État garantit le droit de chacun à l'égalité devant la loi, y compris le droit à un traitement égal devant les tribunaux et tout autre organe administrant la justice ; le droit à la sûreté de la personne et à la protection par l'État contre les voies de fait ou les sévices de la part, soit de fonctionnaires du gouvernement, soit de toute personne, groupe ou institution ; le droit à la liberté de réunion et d'association pacifiques ; le droit à la santé, aux soins médicaux, à la sécurité sociale et aux services sociaux ; le droit à l'éducation et à la formation professionnelle. Ces droits et les dispositions qui garantissent leur application sans discrimination sont également énoncés dans d'autres traités internationaux relatifs aux droits de l'homme, notamment le Pacte international relatif aux droits civils et politiques et le Pacte international relatif aux droits économiques, sociaux et culturels.

63. D'autres dispositions du droit international des droits de l'homme confèrent aux États la responsabilité de lutter contre les effets discriminatoires des technologies d'intelligence artificielle, comme cela a été décrit précédemment. La collecte et l'utilisation de données, si elles ne sont assorties d'aucune garantie relative aux droits de l'homme, soulèvent, s'agissant du respect de la vie privée, d'importantes préoccupations qui peuvent être plus vives encore dans le cas des personnes appartenant à des groupes raciaux et ethniques marginalisés. La Rapporteuse spéciale souhaite donc rappeler aux États les dispositions de l'article 17 du Pacte international relatif aux droits civils et politiques qui prévoient que nul ne peut faire l'objet

d'immixtions arbitraires ou illégales dans sa vie privée et exigent des États qu'ils garantissent les protections juridiques nécessaires. D'autres dispositions du Pacte s'appliquent également aux manifestations de discrimination raciale liées aux technologies d'intelligence artificielle. L'utilisation de l'intelligence artificielle, y compris dans des contextes tels que l'application de la loi, peut porter atteinte à la liberté et à la sécurité de la personne, et mettre en péril la vie des personnes appartenant à des groupes raciaux et ethniques marginalisés. L'article 6 du Pacte souligne le droit inhérent à la vie de toute personne et fait obligation aux États de garantir des protections juridiques à cet égard. L'article 7 dispose que nul ne peut être soumis à la torture, ni à des peines ou traitements cruels, inhumains ou dégradants. L'article 9 prévoit que tout individu a le droit à la liberté et à la sécurité de sa personne, et que nul ne peut faire l'objet d'une arrestation ou d'une détention arbitraire. L'article 14 indique clairement que tous sont égaux devant les tribunaux et les cours de justice. L'article 26 garantit aux groupes minoritaires une protection contre la discrimination. L'article 2 (par. 1) énonce l'obligation faite aux États parties de garantir l'application sans discrimination de toutes les dispositions du Pacte. Le cadre juridique international des droits de l'homme contient également des dispositions concernant l'utilisation de l'intelligence artificielle dans le cadre du contrôle de l'immigration et du contrôle aux frontières, ainsi que dans le contexte des médias sociaux. Ces questions sont examinées dans de précédents rapports établis au titre du mandat[89].

64. En droit international des droits de l'homme, toutes les personnes qui peuvent faire l'objet de discrimination raciale ont le droit d'accéder à des recours, y compris lorsque la discrimination résulte de l'intelligence artificielle. L'article 6 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale dispose que toute personne a le droit d'accéder à une protection et à des voies de recours effectives, devant les tribunaux nationaux et autres organismes d'État compétents. En outre, selon l'Assemblée générale, le droit à un recours et à la réparation des victimes de violations flagrantes des droits de l'homme comporte cinq éléments principaux : la restitution, l'indemnisation, la réadaptation, la satisfaction et les garanties de non-répétition[90].

65. Les entreprises jouent un rôle important dans la conception et l'application de l'intelligence artificielle. Elles sont les principales actrices de son développement et sont souvent chargées par les pouvoirs publics de déployer cette technologie dans le secteur public. Les Principes directeurs relatifs aux entreprises et aux droits de l'homme précisent les obligations des gouvernements et les responsabilités correspondantes qui incombent à ceux-ci et aux entreprises dans le domaine des droits de l'homme. Les États doivent, en application de ces Principes, assurer une protection contre les atteintes aux droits de l'homme commises par des tiers, y compris des entreprises, se trouvant sur leur territoire et/ou relevant de leur juridiction. Pour ce faire, ils doivent appliquer des politiques, des législations, des réglementations et des décisions efficaces, entre autres moyens d'action. D'après les Principes directeurs, les entreprises sont tenues de prévenir, d'atténuer et de réparer les violations des droits de l'homme qu'elles peuvent causer ou auxquelles elles peuvent contribuer, et d'exercer une diligence raisonnable en matière de droits de l'homme s'agissant de leurs activités commerciales[91]. Selon les mêmes Principes, les gouvernements et les entreprises ont l'obligation de garantir l'accès à des recours en cas de violations des droits de l'homme causées par les activités des entreprises, complétant ainsi le droit de recours prévu par d'autres instruments, comme indiqué précédemment. Le projet B-Tech du HCDH propose des orientations et des ressources pour appliquer les Principes directeurs dans l'espace technologique, et met en particulier l'accent sur l'intelligence artificielle[92].

[89] A/75/590 et A/78/538.

[90] Principes fondamentaux et directives concernant le droit à un recours et à réparation des victimes de violations flagrantes du droit international des droits de l'homme et de violations graves du droit international humanitaire, par. 15 à 23.

[91] Voir également Bureau de la prévention du génocide et de la responsabilité de protéger de l'ONU et Economic and Social Research Council Human Rights, Big Data and Technology Project, Université d'Essex, « Countering and addressing online hate speech : a guide for policy makers and practitioners », document d'orientation, juillet 2023 ; A/74/486, par. 44 et 45.

[92] Voir HCDH, « Projet B-Tech : Le HCDH et la question des entreprises et des droits de l'homme », disponible à l'adresse suivante : https://www.ohchr.org/fr/business/b-tech-project.

IV. Conclusions et recommandations

66. **La titulaire précédente du mandat avait expressément demandé aux États et aux autres parties prenantes, y compris les entreprises, de rejeter une approche « aveugle à la couleur de la peau » de la gouvernance et de la réglementation des technologies émergentes, notamment l'intelligence artificielle. Elle avait exhorté les États à réglementer ces technologies dans un cadre tenant compte du racisme structurel et fondé sur les principales normes relatives aux droits de l'homme. Toutefois, la gestion et la réglementation de l'intelligence artificielle restent très insuffisantes, ne prennent pas suffisamment en compte les préjugés raciaux et ne sont pas conformes aux normes du droit international des droits de l'homme. Cette situation perdure, malgré les appels clairs et opportuns en faveur d'une approche « aveugle à la couleur de la peau » et la prise de conscience plus large, au cours des quatre dernières années, de l'existence d'un racisme systémique. L'idée très répandue selon laquelle la technologie est neutre et objective conduit à accélérer l'intégration de l'intelligence artificielle dans la société malgré ses effets discriminatoires sur le plan racial et sans que l'on ait vraiment déterminé si cela est nécessaire. Si l'intelligence artificielle peut avoir des effets positifs, notamment en matière d'égalité et d'inclusion, elle n'est pas, s'agissant de l'ensemble des questions sociétales, la panacée et doit être bien encadrée si l'on veut parvenir à un équilibre entre ses bienfaits et ses risques.**

67. **Pour parvenir à cet équilibre délicat, une réglementation efficace et complète de l'intelligence artificielle est essentielle. Si une telle réglementation est indispensable, les États et d'autres acteurs peuvent également prendre d'autres mesures pour lutter efficacement contre les effets discriminatoires de ces technologies sur le plan racial. Il est également très important de développer, dans le domaine des technologies émergentes, des programmes éducatifs publics fondés sur les droits de l'homme et de renforcer l'éducation à l'intelligence artificielle. Lorsque les personnes et les groupes comprennent ce qu'est l'intelligence artificielle et connaissent leurs droits dans l'espace numérique, ils peuvent utiliser ces connaissances de manière responsable et agir alors en pleine connaissance de cause pour que le principe de responsabilité soit mieux appliqué aux systèmes d'intelligence artificielle.**

68. **Les États devraient :**

a) S'attaquer de toute urgence à la question de la réglementation de l'intelligence artificielle, en gardant à l'esprit la rapidité avec laquelle ces technologies sont mises au point et les multiples façons dont elles perpétuent déjà la discrimination raciale dans tous les domaines ;

b) Élaborer, s'agissant de l'intelligence artificielle, des cadres réglementaires qui reposent sur une compréhension globale du racisme systémique et sont ancrés dans le droit international des droits de l'homme, y compris l'interdiction de la discrimination raciale. Ces cadres ne devraient pas s'appuyer sur une approche cloisonnée, mais prendre en compte différents instruments juridiques, notamment une législation visant expressément l'intelligence artificielle, des lois sur la protection de la vie privée, des dispositions concernant la liberté d'information, une législation et des réglementations sectorielles de lutte contre la discrimination, le but étant de mettre en place une réglementation complète et efficace qui prévienne et élimine les effets discriminatoires de l'intelligence artificielle sur le plan racial ;

c) Examiner le rôle que les normes non contraignantes peuvent jouer dans les cadres réglementaires relatifs à l'intelligence artificielle. Ces normes peuvent guider de manière concrète l'application des normes juridiques dans la pratique. Toutefois, la réglementation sur l'intelligence artificielle ne devrait pas s'appuyer seulement sur des normes non contraignantes, compte tenu des effets considérables que ces technologies ont sur les droits de l'homme, notamment en ce qui concerne la discrimination raciale ;

d) Inscrire dans les cadres réglementaires l'obligation juridiquement contraignante de procéder à des évaluations complètes de la diligence raisonnable en matière de droits de l'homme, y compris des critères précis d'évaluation des préjugés raciaux et ethniques, lors du développement et du déploiement de toutes les technologies

d'intelligence artificielle. L'évaluation de cette diligence raisonnable doit passer par l'application de protocoles d'essai des données et la définition de seuils de protection contre les biais algorithmiques, notamment les préjugés raciaux et ethniques. Elle doit être réalisée avant le déploiement des nouvelles technologies, en particulier auprès d'organismes publics tels que les établissements scolaires, les services de police et les établissements de santé ;

e) Envisager d'interdire l'utilisation des systèmes d'intelligence artificielle dont il a été démontré qu'ils faisaient peser des risques inacceptables sur les droits de l'homme, en particulier les systèmes qui violent l'interdiction de la discrimination raciale ;

f) Veiller à ce que des dispositions des cadres réglementaires garantissent une transparence totale des processus de prise de décision automatisés et soient assorties du droit d'accès à l'information, lorsque rien ne s'oppose, après une évaluation complète de la diligence raisonnable en matière de droits de l'homme, à l'utilisation de l'intelligence artificielle ;

g) Mettre en place des procédures de recours claires et accessibles visant à évaluer et à éliminer les effets discriminatoires de l'intelligence artificielle sur le plan racial et nécessitant une intervention humaine. Il convient de garantir un accès équitable à ces procédures de recours ;

h) Faire en sorte que les personnes et les groupes concernés aient accès à des recours qui comprennent des mesures de restitution, d'indemnisation, de réadaptation et de satisfaction ainsi que des garanties de non-répétition dans les cas où les technologies d'intelligence artificielle ont entraîné des violations des droits de l'homme, y compris la discrimination raciale ;

i) Éviter toute exception aux normes réglementaires qui pourrait entraîner des violations de l'interdiction de la discrimination raciale prévue par le droit international des droits de l'homme ;

j) Veiller à ce que les parties prenantes de tous les groupes raciaux et ethniques marginalisés et les professionnels des secteurs concernés soient effectivement et véritablement consultés lors de l'élaboration et de la mise en application des réglementations relatives à l'intelligence artificielle, mais également lors du développement et de l'utilisation des technologies d'intelligence artificielle ;

k) Investir dans la collecte de données ventilées dans tous les secteurs concernés afin d'obtenir les informations nécessaires pour fonder la réglementation de l'intelligence artificielle sur une compréhension du racisme systémique, résoudre les problèmes de données rencontrés par les systèmes d'intelligence artificielle et mieux contrôler et évaluer les effets de cette technologie sur les personnes appartenant à des groupes raciaux et ethniques marginalisés ;

l) Adopter une approche fondée sur les normes des droits de l'homme dans la collecte et le stockage de données, en garantissant la ventilation des données, l'auto-identification, la transparence, la confidentialité, la participation et la responsabilisation[93] **;**

m) Mettre en place des mécanismes solides de surveillance et de contrôle continu des outils d'intelligence artificielle, y compris des audits réguliers de leurs effets, afin de garantir le respect des réglementations, de répondre à toute préoccupation soulevée par les personnes ou les communautés concernées et de lutter contre les biais qui pourraient être créés au fil du temps par les modèles d'intelligence artificielle ;

[93] Voir A/HRC/42/59 et A/HRC/44/57.

n) Mener des activités de coopération internationale, de renforcement des capacités et de recherche afin de répartir les bienfaits de l'intelligence artificielle plus équitablement entre les pays et d'éviter que les inégalités héritées du colonialisme et de l'esclavage ne se creusent davantage ;

o) Renforcer les programmes éducatifs publics axés sur les droits de l'homme qui promeuvent une utilisation acceptable et responsable de la technologie de l'intelligence artificielle, afin d'accroître les connaissances en la matière, notamment les éléments qui permettent en particulier de mieux prendre conscience des effets discriminatoires de l'intelligence artificielle sur le plan racial.

69. **Les entreprises devraient :**

a) Procéder à des évaluations de diligence raisonnable en matière de droits de l'homme à tous les stades de la conception, du développement et du déploiement des produits d'intelligence artificielle ;

b) Faire en sorte que les personnes appartenant à des groupes raciaux et ethniques marginalisés, les professionnels des secteurs concernés de la société et les personnes ayant des compétences spécialisées dans le domaine du racisme systémique soient effectivement et véritablement consultés lors de la conception, du développement et du déploiement des produits d'intelligence artificielle ;

c) Concevoir des protocoles visant à garantir une transparence totale et le partage d'informations sur la prise de décision relative aux algorithmes de produits qui ont des effets sur les droits de l'homme ;

d) Veiller à ce que les produits d'intelligence artificielle fassent l'objet d'une surveillance continue qui permette de repérer les préjugés raciaux ;

e) Renforcer la formation de toutes les personnes qui participent à la conception, au développement et au déploiement de l'intelligence artificielle sur le thème de la discrimination raciale, y compris les préjugés implicites et le racisme systémique. Pour élaborer des programmes de formation, il conviendra de s'inspirer à la fois des normes du droit international des droits de l'homme et des travaux de recherche menés sur les effets discriminatoires des technologies d'intelligence artificielle sur le plan racial ;

f) Contribuer aux efforts visant à concevoir des programmes éducatifs publics axés sur les droits de l'homme, afin d'améliorer l'accès aux connaissances dans le domaine de l'intelligence artificielle.

70. **L'ONU et ses mécanismes indépendants relatifs aux droits de l'homme devraient :**

a) Promouvoir un dialogue et un débat fructueux entre les parties prenantes sur les technologies d'intelligence artificielle et leur réglementation ;

b) Axer clairement les travaux de l'Organe consultatif de haut niveau sur l'intelligence artificielle sur les effets discriminatoires que les technologies d'intelligence artificielle ont sur le plan racial ;

c) Veiller à ce que les publications et les orientations relatives aux technologies d'intelligence artificielle soient fondées sur le droit international des droits de l'homme, y compris l'interdiction de la discrimination raciale, et considèrent expressément les préjugés raciaux repérés lors de la conception et du déploiement de ces technologies comme un problème mondial grave ;

d) Jouer un rôle dans le suivi des effets sur les droits de l'homme des technologies d'intelligence artificielle, notamment pour ce qui est de la discrimination raciale ;

e) Soutenir la coopération internationale, le renforcement des capacités et la recherche, afin de tenter de répartir plus équitablement les bienfaits de l'intelligence artificielle entre les différents pays.