

Déclaration de Toronto : Protéger le droit à l'égalité et à la non-discrimination dans les systèmes reposant sur l'apprentissage automatique

Préambule

1. Avec l'essor, en termes de capacités et d'utilisation, des systèmes reposant sur l'apprentissage automatique, il devient nécessaire d'étudier les incidences de cette technologie sur les droits humains. Nous reconnaissons que l'apprentissage automatique et les systèmes connexes peuvent servir à promouvoir les droits humains, mais nous sommes de plus en plus préoccupés par le fait que ces systèmes pourraient potentiellement favoriser la discrimination, intentionnelle ou non, à l'égard de certaines personnes ou certains groupes de personnes. Il est urgent de déterminer les effets qu'auront ces technologies sur les peuples et leurs droits. À l'ère de l'apprentissage automatique, qui sera tenu de rendre des comptes pour les atteintes aux droits humains?
2. Alors que le débat sur l'éthique et l'intelligence artificielle se poursuit, la présente déclaration vise à attirer l'attention sur le cadre juridique et normatif international applicable et reconnu en matière de droits humains. Ces instruments universels, contraignants et susceptibles de donner lieu à des actions en justice, offrent des moyens concrets de protection contre la discrimination, de promotion de l'inclusion, de la diversité et de l'équité, et de sauvegarde de l'égalité. Les droits humains sont « universels, indissociables, interdépendants et intimement liés ».¹
3. La présente déclaration s'appuie sur des débats, des principes et des articles portant sur les préjudices imputables à cette technologie. Le travail considérable

¹ Comité des droits de l'homme des Nations unies, Déclaration et Programme d'action de Vienne, 1993, https://treaties.un.org/doc/source/docs/A_CONF.39_11_Add.2-F.pdf.

accompli dans ce domaine par un grand nombre de spécialistes a contribué à sensibiliser le public et à alimenter la discussion sur les risques de discrimination associés aux systèmes reposant sur l'apprentissage automatique.² Nous souhaitons ajouter à ce travail en réaffirmant le rôle du droit relatif aux droits humains et des normes associées dans la protection des personnes et des groupes contre la discrimination, dans quelque contexte que ce soit. Le droit et les normes relatifs aux droits humains mentionnés dans la présente déclaration constituent une base solide pour l'élaboration de cadres éthiques applicables à l'apprentissage automatique, notamment la mise en place de dispositions régissant l'obligation de rendre des comptes et les moyens d'obtenir réparation.

4. Du maintien de l'ordre aux espaces de discussion en ligne, en passant par les systèmes de protection sociale et la prestation de soins de santé, pour ne donner que quelques exemples, les systèmes ayant recours à l'apprentissage automatique peuvent largement et rapidement renforcer ou modifier les structures de pouvoir dans des proportions inédites. Cela peut avoir des conséquences négatives considérables sur les droits humains et notamment sur le droit à l'égalité. De plus en plus de données démontrent que les systèmes reposant sur l'apprentissage automatique, qui peuvent être opaques et comporter des mécanismes difficiles à expliquer, peuvent contribuer à des pratiques discriminatoires ou répressives s'ils sont adoptés et utilisés sans les garanties nécessaires.
5. Les États et les acteurs du secteur privé doivent encourager le développement et l'utilisation de l'apprentissage automatique et des technologies associées lorsque ceux-ci aident la population à exercer ses droits humains et à en bénéficier. Par exemple, dans le domaine de la santé, les systèmes reposant sur l'apprentissage automatique peuvent être porteurs d'améliorations dans le diagnostic et le traitement de certaines maladies et permettre d'élargir l'accès aux soins. De manière plus générale, en ce qui concerne l'apprentissage automatique et l'intelligence artificielle, les États doivent défendre le droit positif de la population à jouir des progrès

² Voir, par exemple : FAT/ML, *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms* ; IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design* ; Déclaration de Montréal pour un développement responsable de l'intelligence artificielle ; et les principes d'Asilomar, *The Asilomar AI Principles*, élaborés par le Future of Life Institute.

scientifiques et technologiques et démontrer ainsi qu'ils reconnaissent les droits économiques, sociaux et culturels.³

6. La présente déclaration porte principalement sur le droit à l'égalité et à la non-discrimination. De nombreux autres droits humains peuvent être affectés par l'utilisation, parfois abusive, de systèmes reposant sur l'apprentissage automatique. C'est notamment le cas des droits au respect de la vie privée et à la protection des données, à la liberté d'expression et d'association, à l'égalité devant la loi et à l'accès à un recours utile, ainsi que du droit à participer à la vie culturelle. Les systèmes qui prennent des décisions et traitent des données peuvent aussi porter atteinte aux droits économiques, sociaux et culturels. Ils peuvent par exemple affecter l'offre de services fondamentaux comme les soins de santé ou l'éducation et limiter les perspectives telles que l'accès à l'emploi.
7. Bien que la présente déclaration soit axée sur les technologies fonctionnant grâce à l'apprentissage automatique, une grande partie des normes et principes qu'elle expose sont également applicables aux technologies regroupées sous le terme plus large d'intelligence artificielle et aux systèmes de traitement de données associés.

³ Pacte international relatif aux droits économiques, sociaux et culturels (PIDESC), article 15, www.ohchr.org/FR/ProfessionalInterest/Pages/CESCR.aspx.

SOMMAIRE

Préambule.....	1
Faire appel au cadre du droit international relatif aux droits humains.....	5
Le droit à l'égalité et à la non-discrimination.....	6
Prévenir la discrimination.....	6
Protéger les droits de l'ensemble des personnes et des groupes : promouvoir la diversité et l'inclusion.....	7
Obligations des États en matière de droits humains.....	8
Utilisation de systèmes reposant sur l'apprentissage automatique par les pouvoirs publics.....	9
Promouvoir l'égalité.....	13
Obliger les acteurs du secteur privé à rendre des comptes.....	13
Responsabilités des acteurs du secteur privé : diligence requise en matière de droits humains.....	14
Le droit à un recours effectif.....	18
Conclusion.....	19

Faire appel au cadre du droit international relatif aux droits humains

8. **Les États ont pour obligation de promouvoir, protéger et respecter les droits humains. Les acteurs du secteur privé, notamment les entreprises, ont quant à eux pour responsabilité de respecter ces droits en toutes circonstances. Nous publions la présente déclaration afin de rappeler cette obligation et cette responsabilité.**
9. De nombreux débats ont lieu actuellement aux niveaux supranational, national et régional, au sein des entreprises technologiques, des établissements d'enseignement, de la société civile et ailleurs, autour de l'éthique dans le domaine de l'intelligence artificielle et de la manière de placer l'humain au cœur de ces technologies. Ces questions doivent être analysées sous l'angle des droits humains afin d'évaluer les potentielles atteintes aux droits fondamentaux qui sont ou pourraient être causées ou facilitées par ces technologies, et de prendre les mesures concrètes nécessaires pour éviter tout risque de préjudice.
10. Le droit relatif aux droits humains est un système de valeurs universel fondé sur la primauté du droit. Il offre des moyens reconnus pour veiller à ce que les droits fondamentaux soient respectés, y compris les droits à l'égalité et à la non-discrimination. Le caractère universel de cet ensemble de normes contraignantes susceptibles de donner lieu à des actions en justice est particulièrement adapté aux technologies qui ne connaissent pas de frontières. Le droit relatif aux droits humains établit des normes et fournit des mécanismes obligeant les acteurs des secteurs public et privé à rendre des comptes lorsqu'ils manquent à leurs obligations et responsabilités respectives de protéger et respecter les droits fondamentaux. Il impose que toute personne dont les droits ont été niés ou bafoués ait accès à un recours utile et à des réparations.
11. Il est urgent que les pouvoirs publics et les acteurs du secteur privé qui conçoivent, développent et déploient ces systèmes étudient les risques posés par les systèmes reposant sur l'apprentissage automatique et y remédient. Il est indispensable

d'identifier les préjudices potentiels, d'y remédier et de mettre en place des mécanismes visant à obliger les responsables présumés de ces préjudices à rendre des comptes. Les mesures adoptées par les gouvernements doivent être contraignantes et suffisantes pour protéger et promouvoir les droits humains. Des universitaires, des juristes et d'autres spécialistes issus de la société civile doivent pouvoir participer de manière constructive à ces débats et apporter leur avis et leurs conseils concernant l'utilisation de ces technologies.

Le droit à l'égalité et à la non-discrimination

12. **La présente déclaration se concentre sur le droit à l'égalité et à la non-discrimination, principe essentiel qui sous-tend tous les droits humains.**
13. En droit international, on entend par discrimination « toute distinction, exclusion, restriction ou préférence fondée notamment sur la race, la couleur, le sexe, la langue, la religion, les opinions politiques ou autres, l'origine nationale ou sociale, la fortune, la naissance ou toute autre situation, et ayant pour effet ou pour but de compromettre ou de détruire la reconnaissance, la jouissance ou l'exercice par tous, dans des conditions d'égalité, de l'ensemble des droits de l'homme et des libertés fondamentales⁴ ». Cette liste n'est pas exhaustive et le haut-commissaire aux droits de l'homme a reconnu la nécessité de prévenir la discrimination contre d'autres groupes⁵.

Prévenir la discrimination

14. **Pour se conformer au droit et aux normes en vigueur en matière de droits humains, les gouvernements et le secteur privé ont l'obligation et la responsabilité, respectivement, de prévenir activement la discrimination. Lorsque la prévention est insuffisante ou n'est pas satisfaisante et qu'une**

⁴ Comité des droits de l'homme des Nations unies, Observation générale n° 18, doc. ONU HRI/GEN/1/Rev.9, Vol. I, 1989, § 7.

⁵ Haut-Commissariat aux droits de l'homme, *Lutte contre la discrimination à l'égard des lesbiennes, gays, bisexuels, transgenres et intersexués. Normes mondiales de conduite à l'intention des entreprises*, <https://www.unfe.org/fr/standards/>.

discrimination a lieu, le système doit être remis en question et les préjudices réparés immédiatement.

15. S'ils emploient de nouvelles technologies, les pouvoirs publics et le secteur privé devront probablement trouver des moyens novateurs de protéger les droits humains à mesure que surgiront de nouveaux obstacles à l'égalité et à la représentation de certaines personnes et de certains groupes, et que les répercussions de ces technologies sur ces personnes ou ces groupes se feront sentir.
16. Les formes actuelles de discrimination structurelle peuvent être reproduites et aggravées dans certaines situations spécifiques à l'utilisation de ces technologies. Par exemple, les objectifs de systèmes reposant sur l'apprentissage automatique peuvent créer des indicateurs de réussite autoréalisateurs et renforcer les inégalités, tout comme l'utilisation de jeux de données non représentatifs ou entachés de préjugés peut avoir des conséquences néfastes.
17. Tous les acteurs, qu'ils soient publics ou privés, doivent prévenir et atténuer les risques de discrimination dans la conception, le développement et l'emploi des technologies fonctionnant grâce à l'apprentissage automatique. Ils doivent en outre s'assurer que des mécanismes donnant accès à un recours utile soient en place avant le déploiement de ces technologies et maintenus tout au long du cycle de vie du système.

Protéger les droits fondamentaux de l'ensemble des personnes et des groupes : promouvoir la diversité et l'inclusion

18. La présente déclaration réaffirme que l'inclusion, la diversité et l'équité sont des éléments indispensables à la protection et la défense du droit à l'égalité et à la non-discrimination. Lors du développement et du déploiement de systèmes reposant sur l'apprentissage automatique, tout doit être envisagé pour prévenir la discrimination, en particulier contre les groupes marginalisés.
19. Si la collecte de données peut contribuer à limiter la discrimination, rassembler des informations sur les discriminations dont font l'objet certains groupes de personnes

peut s'avérer particulièrement difficile. Ces groupes doivent bénéficier d'une protection supplémentaire, notamment en ce qui concerne les données sensibles.

20. Les préjugés implicites et non intentionnels intégrés dans le système au moment de la conception sont eux aussi source de discrimination lorsque la conception, le développement et l'utilisation finale des systèmes reposant sur l'apprentissage automatique sont en grande partie pilotés par un secteur particulier de la société. À l'heure actuelle, cette technologie est majoritairement développée, employée et analysée par des entreprises basées dans certains pays ou certaines régions. Les personnes qui en sont à l'origine lui apportent leurs propres préjugés et ont peu de chances de bénéficier largement de contributions de la part de groupes diversifiés en matière de couleur de peau, de culture, de genre ou de milieu socio-économique.
21. Pour assurer l'inclusion, la diversité et l'équité, il est essentiel qu'un ensemble diversifié de personnes, y compris d'utilisateurs finaux, soient consultées et participent activement à la conception et à la mise en pratique des systèmes reposant sur l'apprentissage automatique ; cela contribuera à ce que ces systèmes soient créés et utilisés dans le respect des droits humains, et notamment des droits des groupes marginalisés particulièrement exposés à la discrimination.

Obligations des États en matière de droits humains

22. C'est aux États que revient la responsabilité première d'assurer la promotion, la protection, le respect et la mise en œuvre des droits humains. En vertu du droit international, les États ne doivent pas participer à des actions ou pratiques discriminatoires ou portant atteinte aux droits fondamentaux, ni soutenir de telles actions ou pratiques, lorsqu'ils conçoivent ou déploient des systèmes reposant sur l'apprentissage automatique dans un contexte public ou dans le cadre de partenariats public-privé.
23. Les États doivent se conformer aux lois et réglementations nationales et internationales, telles que les textes relatifs à la protection de données et au respect

de la vie privée, qui codifient et permettent de traduire dans les faits les obligations en matière de protection contre la discrimination et contre d'autres atteintes aux droits humains.

24. Les États ont l'obligation positive de protéger la population contre les discriminations pratiquées par les acteurs du secteur privé et de promouvoir l'égalité et les autres droits fondamentaux, notamment en adoptant des lois contraignantes.
25. Les obligations des États réaffirmées dans cette section s'appliquent également à l'utilisation de l'apprentissage automatique dans le cadre de partenariats public-privé.

Utilisation de systèmes reposant sur l'apprentissage automatique par les pouvoirs publics

26. **Les États doivent s'assurer que les mesures existantes visant à prévenir la discrimination et d'autres atteintes aux droits humains soient mises à jour de manière à prendre en compte les risques posés par les technologies reposant sur l'apprentissage automatique et à y remédier.**
27. Des systèmes fonctionnant grâce à l'apprentissage automatique sont de plus en plus souvent déployés ou mis en place par les pouvoirs publics dans des secteurs essentiels à l'exercice et la jouissance des droits humains, à l'état de droit, aux garanties d'une procédure régulière, à la liberté d'expression, à la justice pénale, aux soins de santé, à l'accès à des prestations sociales et au logement. Dans de tels contextes, cette technologie peut avoir des avantages, mais elle risque aussi fortement d'engendrer des discriminations ou d'autres atteintes aux droits fondamentaux. En cas de préjudice, il est indispensable que les États offrent de réelles possibilités d'indemnisation.
28. Comme l'a confirmé le Comité des droits de l'homme, l'article 26 du Pacte international relatif aux droits civils et politiques « interdit toute discrimination en droit ou en fait dans tout domaine réglementé et protégé par les pouvoirs publics⁶ ». C'est

⁶ Comité des droits de l'homme, Observation générale n° 18, 1989, §12

ce qu'énoncent plus en détail les traités relatifs à certaines formes de discrimination, dont les États signataires se sont engagés à s'abstenir de tout acte ou pratique discriminatoire et à faire en sorte que les autorités et institutions publiques « se conforment à cette obligation⁷ ».

29. Les États doivent s'abstenir d'utiliser ou de demander au secteur privé d'utiliser des outils discriminatoires, sources de discrimination ou qui portent atteinte aux droits humains.
30. **Les États doivent prendre les mesures suivantes pour atténuer et réduire les préjudices discriminatoires causés par l'utilisation de l'apprentissage automatique dans les systèmes du secteur public :**

i. Identifier les risques

31. Tout État souhaitant déployer des technologies fonctionnant grâce à l'apprentissage automatique doit étudier avec attention les éventuels risques, discriminatoires ou autres, en matière de droits humains induits par ces systèmes avant de les développer ou de les acquérir, si possible, ou avant de les utiliser. Lorsque ces technologies sont déployées, les États doivent étudier ces risques de façon continue tout au long du cycle de vie des systèmes. Cela peut consister à :
- a) mener des études d'impact régulières avant toute adjudication, pendant la phase de développement, puis à intervalles réguliers tout au long du déploiement et de l'utilisation afin d'identifier les sources potentielles de résultats discriminatoires ou portant atteinte aux droits fondamentaux. Il peut s'agir, par exemple, d'évaluer la conception des algorithmes, les processus de contrôle ou le traitement des données⁸ ;

⁷ Par exemple, la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, article 2(a), et la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes, article 2(d).

⁸ L'institut AI Now a présenté un cadre pratique pour l'évaluation de l'impact des algorithmes par les organismes publics, <https://ainowinstitute.org/aiareport2018.pdf>. L'article 35 du Règlement général sur la protection des données de l'Union européenne prévoit l'obligation de mener une analyse d'impact relative à la protection des données ; l'article 25, quant à lui, impose l'application, à dessein et par défaut, de principes relatifs à la protection des données du stade de la conception du produit, service ou système jusqu'à la fin de son cycle de vie.

- b) prendre les mesures appropriées pour atténuer les risques identifiés lors des évaluations d'impact – par exemple, en réduisant toute discrimination ou sous-représentation involontaires dans les données ou les systèmes ; en effectuant des vérifications dynamiques et des essais préalables à la mise en circulation ; en faisant en sorte d'inclure les groupes potentiellement touchés et des spécialistes du domaine dans la prise de décisions aux étapes de la conception, des essais et de l'évaluation ; et, si nécessaire, en soumettant les systèmes à une expertise indépendante ;
- c) effectuer des tests et des contrôles réguliers de ces systèmes, vérifier les indicateurs de réussite à la recherche d'éventuels biais et boucles de rétroaction autoréalisatrices, et faire en sorte que les systèmes soient évalués de manière indépendante et globale dans un contexte d'atteintes aux droits humains en environnement réel ;
- d) annoncer les limitations connues du système en question – par exemple en indiquant le niveau de confiance, les défaillances prévisibles et les restrictions d'emploi appropriées.

ii. Garantir la transparence et l'obligation de rendre des comptes

32. Les États doivent faire en sorte et exiger que l'obligation de rendre des comptes soit respectée et qu'il existe un maximum de transparence en ce qui concerne l'utilisation de systèmes basés sur l'apprentissage automatique dans le secteur public. Ils doivent notamment veiller à ce que ces technologies soient explicables et intelligibles, de manière à ce que des entités indépendantes puissent véritablement examiner leur impact sur les personnes ou groupes touchés, que les responsabilités soient établies et que les acteurs soient tenus de rendre des comptes. Les États doivent :

- a) faire savoir publiquement dans quels espaces de la sphère publique sont utilisés des systèmes reposant sur l'apprentissage automatique, expliquer en termes clairs et accessibles comment fonctionne la prise de décisions des systèmes automatisés et reposant sur l'apprentissage automatique, et attester des mesures prises pour identifier, enregistrer et atténuer toute conséquence discriminatoire ou portant atteinte aux droits humains ;

- b) rendre possible des analyses et une surveillance indépendantes en utilisant des systèmes pouvant être examinés ;
- c) éviter d'utiliser des systèmes « à boîte noire » ne pouvant pas être soumis à des normes satisfaisantes en matière de reddition de comptes et de transparence, et s'abstenir d'utiliser de tels systèmes dans des contextes où les risques sont élevés⁹.

iii. Imposer des contrôles

33. Les États doivent prendre des mesures pour veiller à ce que leurs agents soient conscients des risques de discrimination et d'autres atteintes aux droits humains induits par les systèmes reposant sur l'apprentissage automatique et qu'ils y soient attentifs. Les États doivent :
- a) adopter préventivement des pratiques d'embauche axées sur la diversité et mener des consultations sollicitant des perspectives variées, de manière à ce que les personnes participant à la conception, à la mise en œuvre et à l'évaluation des systèmes reposant sur l'apprentissage automatique représentent une multiplicité de milieux et d'identités ;
 - b) veiller à ce que les organismes publics offrent des formations aux droits humains et à l'analyse des données aux fonctionnaires participant à l'achat, au développement, à l'utilisation et à l'évaluation d'outils fonctionnant grâce à l'apprentissage automatique ;
 - c) créer les mécanismes nécessaires pour un contrôle indépendant, qui pourrait notamment être mené par les autorités judiciaires lorsque cela s'avère nécessaire ;
 - d) veiller à ce que les décisions basées sur l'apprentissage automatique respectent les normes internationales en matière de régularité des procédures.
34. La recherche et le développement dans le domaine des systèmes utilisant l'apprentissage automatique étant largement menés par le secteur privé, les pouvoirs publics confient généralement la conception et la mise en place de ces technologies dans la sphère publique à des entreprises privées. Cela ne les affranchit pas pour autant de leurs propres obligations : il reste de leur devoir de

⁹ Institut AI Now (université de New York), AI Now 2017 Report, 2017, https://ainowinstitute.org/AI_Now_2017_Report.pdf.

prévenir la discrimination, de veiller à la réparation des atteintes aux droits humains dans les prestations de services et de faire respecter l'obligation de rendre des comptes.

35. Tout État se procurant des technologies reposant sur l'apprentissage automatique auprès d'une entreprise du secteur privé doit surveiller et contrôler de manière continue et appropriée leur utilisation. L'État doit demander à cette entreprise d'identifier, de prévenir et d'atténuer toute discrimination et autre atteinte aux droits humains avec la diligence requise, et de faire connaître publiquement ses efforts en ce sens.

Promouvoir l'égalité

36. **Les États ont le devoir de prendre des mesures préventives pour éliminer les discriminations¹⁰.**

37. Dans le contexte de l'apprentissage automatique et, plus largement, des avancées technologiques, l'une des priorités pour les États est de promouvoir des programmes favorisant la diversité, l'inclusion et l'équité dans les sciences, les technologies, l'ingénierie et les mathématiques. Ces efforts ne sont pas une fin en soi, mais ils peuvent contribuer à atténuer les résultats discriminatoires. Les États doivent en outre investir dans la recherche visant à atténuer les atteintes aux droits humains dues aux systèmes reposant sur l'apprentissage automatique.

Obliger les acteurs du secteur privé à rendre des comptes

38. **Le droit international établit clairement le devoir qu'ont les États de protéger les droits humains. Ce devoir comprend l'obligation de veiller à ce que les acteurs privés respectent le droit à la non-discrimination.**

¹⁰ Selon le Comité des droits économiques, sociaux et culturels des Nations unies, les États parties au PIDESC doivent non seulement s'abstenir de toute mesure discriminatoire, mais aussi « prendre des mesures concrètes, délibérées et ciblées pour mettre fin à la discrimination dans l'exercice des droits consacrés par le Pacte. » - Comité des droits économiques, sociaux et culturels des Nations unies, Observation générale n° 20, doc. ONU E/C.12/GC/20, 2009, § 36.

39. Selon le Comité des droits économiques, sociaux et culturels des Nations unies, « [l]es États parties doivent donc adopter des mesures, y compris législatives, pour empêcher toute discrimination exercée pour des motifs interdits dans la sphère privée par des particuliers ou des personnes morales¹¹. »
40. Les États doivent mettre en place une réglementation conforme au droit relatif aux droits humains pour encadrer l'utilisation de l'apprentissage automatique par le secteur privé dans les contextes présentant un risque de résultats discriminatoires ou préjudiciables aux droits des personnes. Des normes techniques pourraient compléter cette réglementation. En outre, les législations nationales et régionales relatives à la non-discrimination, à la protection des données et au respect de la vie privée, entre autres, peuvent préciser et renforcer les obligations internationales en matière de droits humains applicables à l'apprentissage automatique.
41. Les États doivent garantir l'accès à un recours utile à toutes les personnes dont les droits sont bafoués à cause de l'utilisation de ces technologies.

Responsabilités des acteurs du secteur privé : diligence requise en matière de droits humains

42. Les acteurs du secteur privé ont la responsabilité de respecter les droits humains, et ce, indépendamment des obligations des États¹². Pour exercer cette responsabilité, les acteurs du secteur privé doivent constamment prendre des mesures préventives efficaces pour s'assurer qu'ils n'occasionnent ou ne contribuent à aucune atteinte aux droits humains. C'est ce que l'on appelle la « diligence requise en matière de droits humains¹³ ».

¹¹ Comité des droits économiques, sociaux et culturels des Nations unies, Observation générale n° 20, doc. ONU E/C.12/GC/20, 2009, § 11.

¹² Voir les Principes directeurs relatifs aux entreprises et aux droits de l'homme des Nations unies et les documents de référence.

¹³ Voir la Recommandation CM/Rec(2018)2 du Comité des ministres du Conseil de l'Europe aux États membres sur les rôles et les responsabilités des intermédiaires d'internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680790e37.

43. Les acteurs du secteur privé qui développent et déploient des systèmes reposant sur l'apprentissage automatique doivent se conformer à un cadre de diligence requise en matière de droits humains pour éviter de favoriser ou d'asseoir les discriminations et, plus largement, afin que l'utilisation de leurs systèmes respecte les droits humains.
44. **La diligence requise en matière de droits humains passe par trois étapes fondamentales:**
- i. Identifier les conséquences discriminatoires potentielles
 - ii. Prendre des mesures concrètes pour prévenir et atténuer la discrimination et effectuer un suivi des réponses
 - iii. Faire preuve de transparence quant aux efforts menés pour détecter, prévenir et atténuer la discrimination dans les systèmes reposant sur l'apprentissage automatique
- i. Identifier les conséquences discriminatoires potentielles
45. Au cours du développement et du déploiement de toute nouvelle technologie reposant sur l'apprentissage automatique, les acteurs non étatiques et privés doivent évaluer les risques de discrimination induits par cette technologie. Le risque de discrimination et les préjudices causés ne sont pas équivalents pour toutes les applications, et les mesures nécessaires pour y mettre fin dépendent du contexte. Les acteurs doivent prendre soin d'identifier non seulement la discrimination directe, mais aussi les formes indirectes de traitement différencié pouvant sembler neutre à première vue, mais qui peuvent être source de discrimination.
46. Lorsqu'ils cartographient les risques, les acteurs du secteur privé doivent prendre en compte les risques communément associés aux systèmes reposant sur l'apprentissage automatique – par exemple, l'alimentation des systèmes avec des données incomplètes ou non représentatives, ou les jeux de données reflétant des préjugés historiques ou systémiques. Les acteurs privés doivent consulter les parties intéressées dans un esprit d'ouverture, notamment les groupes touchés, des organisations luttant pour les droits humains, pour l'égalité et contre la

discrimination, ainsi que des spécialistes indépendant-e-s des droits fondamentaux et de l'apprentissage automatique.

ii. Prendre des mesures concrètes pour prévenir et atténuer la discrimination, et effectuer un suivi des réponses

47. Après avoir identifié les risques en matière de droits humains, l'étape suivante consiste à prévenir ces risques. Pour les développeurs de systèmes reposant sur l'apprentissage automatique, cela nécessite de :
- a) rectifier le système pour éviter les discriminations, à la fois au niveau de la conception du modèle et de l'impact du système, et au niveau des décisions concernant les données d'apprentissage à utiliser ;
 - b) rechercher la diversité, l'équité et d'autres modes d'inclusion au sein des équipes de développement dans le domaine de l'apprentissage automatique, afin d'identifier les préjugés intégrés au stade de la conception et de prévenir les discriminations involontaires ;
 - c) soumettre à des contrôles externes indépendants les systèmes risquant particulièrement d'entraîner des atteintes aux droits humains.
48. Dans un contexte où le risque de discrimination ou d'autres violations des droits humains est considéré trop élevé ou impossible à limiter, les acteurs du secteur privé doivent renoncer à déployer un système reposant sur l'apprentissage automatique.
49. Il est en outre essentiel que les acteurs du secteur privé effectuent un suivi des réponses qu'ils apportent aux problèmes qui apparaissent lors de la mise en place du système et de son utilisation, et qu'ils évaluent l'efficacité de ces réponses. Cela implique des inspections régulières et renouvelées de la qualité du système, et des contrôles en temps réel aux stades de la conception, des tests et du déploiement, afin d'évaluer les discriminations potentielles en contexte et en conditions réelles et, le cas échéant, de corriger les erreurs et les préjudices. Ce dernier point est particulièrement important, car les boucles de rétroaction risquent d'exacerber et de pérenniser les discriminations.

iii. Faire preuve de transparence quant aux efforts menés pour détecter, prévenir et atténuer la discrimination dans les systèmes reposant sur l'apprentissage automatique

50. La transparence est un élément clé de la diligence requise en matière de droits humains. Cela implique de « communiquer, en assurant un certain degré de transparence et de responsabilité aux individus ou aux groupes susceptibles d'être touchés et aux autres acteurs pertinents, y compris les investisseurs¹⁴. »
51. Les acteurs du secteur privé qui développent et mettent en place des systèmes reposant sur l'apprentissage automatique doivent communiquer leurs procédures de détection des risques, signaler les risques qu'ils ont identifiés et indiquer les mesures concrètes prises pour prévenir et limiter ceux qui concernent les droits humains.
- Ils peuvent pour cela:
- a) publier des informations concernant les risques et les cas précis de discrimination décelés par l'entreprise, tels que les risques associés à la façon dont a été conçue un système reposant sur l'apprentissage automatique en particulier, ou ceux liés à l'utilisation de ces systèmes dans certains contextes, par exemple ;
 - b) lorsqu'un risque de discrimination existe, publier les spécifications techniques apportant des détails sur l'apprentissage automatique et ses fonctions, accompagnées d'échantillons des données d'apprentissage utilisées et de précisions sur la source des données ;
 - c) établir des mécanismes visant à garantir que lorsqu'un système reposant sur l'apprentissage automatique entraîne une discrimination, les parties concernées, y compris les personnes touchées, soient informées du préjudice et de la manière dont elles peuvent contester une décision ou un résultat.

¹⁴ Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme, principe 21.

Le droit à un recours effectif

52. Le droit à la justice est un élément essentiel du droit international relatif aux droits humains¹⁵. En vertu du droit international, les victimes de violations des droits humains ou d'atteintes à ces droits doivent disposer de voies de recours rapides et efficaces et les personnes présumées responsables de ces violations ou atteintes doivent être amenées à rendre des comptes.
53. Les entreprises et les acteurs du secteur privé qui conçoivent et mettent en place des systèmes reposant sur l'apprentissage automatique doivent prendre les mesures nécessaires pour veiller à ce que les personnes et les groupes aient accès à des voies de recours utiles et efficaces et puissent obtenir réparation. Cela peut impliquer, par exemple, de créer des procédures claires, indépendantes et visibles pour réparer les torts éventuels causés par ces systèmes à des personnes ou à la société, et d'attribuer à certains postes au sein de l'entité la responsabilité d'apporter dans les meilleurs délais une solution aux problèmes pouvant faire l'objet d'un recours accessible et effectif et d'un contrôle judiciaire.
54. L'utilisation de systèmes reposant sur l'apprentissage automatique lorsque les droits de certaines personnes sont menacés peut entraver le droit à un recours. En raison de l'opacité de certains systèmes, les personnes peuvent ignorer comment ont été prises des décisions ayant une incidence sur leurs droits fondamentaux, et ne pas être en mesure de juger si le processus de prise de décision était discriminatoire ou non. Dans certains cas, les acteurs publics ou privés peuvent être eux-mêmes incapables d'expliquer comment les décisions sont prises.
55. Les enjeux sont particulièrement graves lorsque les systèmes reposant sur l'apprentissage automatique qui recommandent, prennent ou appliquent des

¹⁵ Voir, par exemple : Déclaration universelle des droits de l'homme, art. 8 ; Pacte international relatif aux droits civils et politiques, art. 2(3) ; PIDESC, art. 2 ; Comité des droits économiques, sociaux et culturels, Observation générale n° 3 : La Nature des obligations des États parties (art. 2, para 1, du Pacte), doc. ONU E/1991/23, 1990 ; Convention internationale sur l'élimination de toutes les formes de discrimination raciale, art. 6 ; Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes, art. 2 ; Comité des droits économiques, sociaux et culturels, Observation générale n° 9 : Application du Pacte au niveau national, art. 2, E/C.12/1998/24, 1998, <https://undocs.org/fr/E/C.12/1998/24>.

décisions sont employés au sein du système judiciaire, l'institution-même responsable de la protection des droits, notamment du droit à un recours utile.

56. Les mesures citées plus haut, portant sur le fait de détecter les discriminations, de rassembler des informations à leur sujet et d'y répondre, le tout de manière transparente et responsable, aideront les États à veiller à ce que les personnes puissent bénéficier de recours utiles. Les États doivent en outre :
- a) veiller à ce que l'utilisation des systèmes reposant sur l'apprentissage automatique éventuellement déployés dans le secteur public soit conforme aux normes en matière de procédure légale ;
 - b) agir avec prudence lorsqu'il s'agit d'utiliser des systèmes reposant sur l'apprentissage automatique dans le secteur de la justice, étant donné les risques en matière d'équité des procès et de droits des justiciables¹⁶ ;
 - c) définir des principes clairs en ce qui concerne l'obligation de rendre des comptes pour le développement et la mise en place de systèmes reposant sur l'apprentissage automatique et préciser quels sont les organes ou les personnes portant la responsabilité juridique des décisions prises par le biais de ces systèmes ;
 - d) offrir des voies de recours utiles aux victimes de discriminations liées aux systèmes reposant sur l'apprentissage automatique utilisés par les organes publics ou privés, y compris des réparations comprenant, le cas échéant, une indemnisation, l'imposition de sanctions aux responsables et des garanties de non-répétition. Il se peut que cela soit possible grâce aux lois et réglementations déjà en vigueur, ou bien il peut être nécessaire d'en adopter de nouvelles.

Conclusion

57. Les signataires de la présente déclaration demandent aux acteurs des secteurs public et privé de respecter les obligations et responsabilités qui leur incombent en vertu du droit et des normes relatifs aux droits humains afin d'éviter autant que possible que l'utilisation de systèmes reposant sur l'apprentissage automatique ait

¹⁶ Voir, par exemple : Julia Angwin, Jeff Larson, Surya Mattu et Lauren Kirchner, "Machine Bias", *ProPublica*, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

des conséquences discriminatoires. En cas de discrimination, des mesures doivent être en place pour garantir le droit à un recours utile.

58. Nous demandons aux pouvoirs publics et aux acteurs du secteur privé de travailler main dans la main et de participer activement et résolument à la protection des personnes et des groupes contre la discrimination. Lors de la création et du déploiement de systèmes reposant sur l'apprentissage automatique, les pouvoirs publics et les acteurs du secteur privé doivent prendre des mesures concrètes pour promouvoir l'obligation de rendre des comptes et les droits humains, y compris, mais pas uniquement, le droit à l'égalité et à la non-discrimination, conformément aux obligations et responsabilités qui leur incombent en vertu du droit international relatif aux droits humains et des normes associées.
59. Les avancées technologiques ne doivent pas mettre en péril nos droits humains. Nous sommes à la croisée des chemins : les instances de pouvoir doivent agir maintenant pour protéger les droits humains et préserver les droits auxquels nous pouvons prétendre aujourd'hui et dont bénéficieront les générations futures.

Comité de rédaction

Anna Bacciarelli et Joe Westby, Amnesty International

Estelle Massé, Drew Mitnick et Fanny Hidvegi, Access Now

Boye Adegoke, Paradigm Initiative Nigeria

Frederike Kaltheuner, Privacy International

Malavika Jayaram, Digital Asia Hub

Yasodara Córdova, chercheuse

Solon Barocas, Cornell University

William Isaac, Human Rights Data Analysis Group

Cette déclaration a été publiée en anglais le 16 mai 2018 par Amnesty International et Access Now, et présentée lors de la RightsCon 2018 à Toronto, Canada.