

**CONFÉRENCE  
MONDIALE  
SUR LA**

**LIBERTÉ**  
**DE LA PRESSE**

**LE 16 NOVEMBRE, 2020**

**Liberté de la presse et  
intelligence artificielle**

**Julia Haas, Bureau du Représentant pour la  
liberté des médias de l'OSCE**



# LIBERTÉ DE LA PRESSE ET INTELLIGENCE ARTIFICIELLE

*Avis : Les vues et les positions exprimées dans le présent rapport n'engagent que leur auteur et ne représentent pas nécessairement celles du ministère des Affaires étrangères, du Commerce et du Développement ou du gouvernement du Canada. Ce document est la traduction de l'original anglais.*



## Résumé

Le présent article traite des répercussions de l'utilisation de l'intelligence artificielle (IA) sur la liberté d'expression et la liberté de la presse. Bien que l'IA puisse améliorer à bien des égards la communication et l'accès à l'information, y compris dans les médias traditionnels, il est surtout question dans cet article des principales préoccupations que soulève le déploiement de l'IA lorsqu'il ne se fait pas dans le respect des droits de la personne.

L'IA peut servir d'outil pour censurer les médias et surveiller illicitement des citoyens et des journalistes indépendants. Par ailleurs, dans l'environnement virtuel d'aujourd'hui, quelques intermédiaires dominants de l'Internet jouent un rôle de contrôleurs en ce qui concerne la conservation, la diffusion et la monétisation de l'information, y compris des nouvelles. Ces intermédiaires déploient de plus en plus l'IA pour régir les discours public et privé.

Les outils d'IA, qui soutiennent une bonne partie de la diffusion des contenus à l'heure actuelle, font souvent partie intégrante d'un modèle d'entreprise fondé sur la publicité ciblée. L'utilisation de l'IA pour diffuser du contenu en fonction des préférences présumées des utilisateurs repose sur l'élaboration de profils établis à partir d'une grande quantité de données. Pour maximiser leurs revenus, les intermédiaires peuvent prioriser les contenus qui augmentent la participation des utilisateurs au détriment des contenus qui donnent accès à une information diversifiée d'intérêt public ou à un journalisme indépendant de qualité. Cette tendance peut réduire la capacité des utilisateurs d'avoir accès à une information diversifiée et peut influencer leurs réflexions et leurs croyances.

En matière de surveillance du discours, les États et les intermédiaires ont souvent recours à l'IA pour déceler et supprimer les contenus considérés comme illégaux ou indésirables. L'immense quantité de contenu qui circule dépasse toute capacité humaine de surveillance. Bien que le filtrage par l'IA des contenus générés par les utilisateurs puisse sembler intéressant à cet égard, les outils de l'intelligence artificielle sont sujets à erreurs. En plus de déployer eux-mêmes l'IA, les États demandent à des acteurs du secteur privé de surveiller et de supprimer des contenus, en fonction de définitions vagues et dans des délais contraignants. Une telle sous-traitance de la protection des droits de la personne à des acteurs

du secteur privé qui sont motivés par le profit peut inciter ceux-ci à bloquer de façon exagérée des discours légitimes, ce qui soulève des préoccupations supplémentaires quant à la primauté du droit et à la discrimination.

La possibilité d'utiliser l'IA pour faciliter la surveillance et la censure, que ce soit à des fins économiques ou politiques, constitue une menace au droit de chercher et d'obtenir de l'information et met en péril la diversité des médias. Ce danger est amplifié par le pouvoir et l'influence qu'exerce une poignée d'intermédiaires, et par le fait que la plupart des outils d'IA fonctionnent de façon opaque et sans trop de réglementation ou de surveillance.

Cet article montre également de quelle manière les préjugés présents dans les jeux de données et chez les développeurs de ces systèmes risquent de perpétuer les inégalités actuelles. Il décrit la façon dont l'IA touche les médias traditionnels et dont la pandémie de COVID-19 exacerbe les inquiétudes décrites ci-dessus. En guise de conclusion, l'article recommande des orientations que devraient prendre les États et le secteur privé pour garantir le respect des droits de la personne dans la conception et le déploiement de l'IA, et pour que la transparence et la responsabilisation soient assurées à toutes les étapes.

## Liberté de la presse et intelligence artificielle

Les nouvelles technologies offrent des possibilités inédites pour l'exercice de la liberté d'expression et de la liberté de la presse<sup>1</sup>. L'intelligence artificielle (IA) joue un rôle important dans la transformation de notre façon de communiquer ainsi que de consommer du contenu médiatique et d'interagir avec ce contenu. L'IA propose des solutions attrayantes pour filtrer et classer le contenu généré par les utilisateurs et l'information en ligne, lesquels sont apparemment infinis<sup>2</sup>. Comme de nombreux progrès technologiques, l'IA est susceptible de déboucher sur le meilleur, mais elle peut aussi constituer une véritable menace pour les droits de la personne – en particulier la liberté d'expression et la liberté de la presse.

En l'absence d'une définition universellement reconnue, le terme générique « IA » est couramment utilisé pour désigner les processus automatisés et fondés sur des données<sup>3</sup>. Certains outils d'IA peuvent être de simples instructions programmées par des humains, alors que d'autres sont plus avancées et comportent de l'apprentissage automatique. Étant donné que l'IA repose sur des architectures et des données fournies par des humains, les résultats qu'elle produit sont forcément façonnés par des valeurs culturelles ainsi que des expériences et des croyances subjectives, notamment des préjugés fondamentaux<sup>4</sup>.

Certains États déploient l'IA pour surveiller illicitement leur population et pour contrôler les communications publiques d'une manière qui va à l'encontre du droit international en matière de droits de la personne. Par ses possibilités

inédites de surveillance, l'IA peut faciliter la censure et réprimer le journalisme indépendant ainsi que la dissidence, aussi bien en ligne que hors ligne. Par conséquent, certains États se servent de l'IA pour exercer une coercition contre la presse et, en fin de compte, resserrer leur autoritarisme numérique<sup>5</sup>.

Par ailleurs, des acteurs du secteur privé, en particulier les fournisseurs de moteurs de recherche et de plateformes de médias sociaux, utilisent l'IA afin de filtrer les contenus pour déceler ceux qui sont jugés indésirables et les supprimer ou en réduire la visibilité (« modération de contenu »), ainsi que pour classer et diffuser une information sur mesure (« édition de contenu »)<sup>6</sup>. Dans les deux cas, le discours est régulé dans le but de faciliter la communication en ligne, de fournir des services conviviaux et, essentiellement, d'accroître les profits.

Le filtrage et le classement du contenu au moyen de l'IA sont rendus possibles par la surveillance à grande échelle du comportement des utilisateurs. Pour évaluer et prédire la pertinence du contenu, l'IA a besoin d'une grande quantité de données fines, lesquelles facilitent également la publicité, fondement du modèle d'entreprise de nombreux intermédiaires de l'Internet. La marchandisation des données personnelles aux fins du ciblage publicitaire, qui se traduit par des profits, vient encourager des activités exhaustives de collecte et de traitement des données, un phénomène désigné sous le nom de « capitalisme de surveillance »<sup>7</sup>. Les intermédiaires, qui proposent des services « gratuits », tirent leurs profits de l'élaboration de profils et de la commercialisation de l'espace public. Ces pratiques intrinsèquement invasives ouvrent la porte à des risques d'abus de pouvoir

et à un contrôle étatique omniprésent<sup>8</sup>. Même si toutes les formes de surveillance ont un effet paralysant sur la liberté d'expression et sur les médias<sup>9</sup>, l'IA peut imposer des contraintes qui portent atteinte aux journalistes d'enquête et à la protection des sources<sup>10</sup>.

Fréquemment comparée à une boîte noire<sup>11</sup>, l'IA est souvent opaque<sup>12</sup>, et son utilisation reste invisible, ce qui peut donner l'impression trompeuse qu'elle produit des résultats neutres et une représentation objective de la réalité. L'utilisateur peut très bien ignorer qu'un système d'IA est utilisé et la façon dont ce système obtient les résultats d'une recherche ou encore favorise ou supprime du contenu. De la même façon, il n'est pas toujours évident de savoir à quel moment et de quelle manière l'IA est déployée pour entraver le travail des médias, que ce soit par la surveillance ou par d'autres formes d'ingérence<sup>13</sup>. L'opacité et la méconnaissance entourant l'IA constituent les principaux défauts de toute utilisation de cette technologie<sup>14</sup>.

L'IA opaque qui régit la diffusion de l'information au profit d'intérêts commerciaux peut avoir de lourdes conséquences pour le discours public, d'autant plus que le marché est dominé par une petite poignée d'intermédiaires. Ces oligopoles sont devenus les arbitres privés de la parole, qui dictent les modalités d'utilisation de la communication en ligne et de l'accès à l'information à l'échelle mondiale. Ceux qui souhaitent faire partie du cyberspace n'ont guère d'autres choix que d'accepter les règles et la surveillance imposées par ces intermédiaires dominants. En outre, de tels systèmes privés d'IA et l'étendue des empreintes numériques peuvent aussi faciliter la surveillance étatique et la censure politique de la presse<sup>15</sup>.

Les modèles d'entreprise basés sur la publicité, qui sont au cœur de la structure actuelle de l'Internet, ont profondément ébranlé la viabilité des médias traditionnels en changeant les structures mêmes des rapports de pouvoir au détriment d'un journalisme de qualité<sup>16</sup>. L'utilisation des technologies d'IA accentue ce déséquilibre, en particulier dans les pays dont les populations ont un accès limité à Internet ou dont les médias de service public ne sont pas solidement établis<sup>17</sup>.

Que ce soit par la censure ciblée, la surveillance généralisée des journalistes d'enquête ou l'usage de robots pilotés par l'IA pour attaquer et museler certains journalistes, toute utilisation de

l'IA avec l'intention d'entraver le journalisme indépendant constitue une grave menace à la liberté de la presse<sup>18</sup>. Même en l'absence de mauvaise foi, le recours général à l'IA pour surveiller et restreindre certains contenus ou pour diffuser de l'information comporte des risques majeurs. Bien que de nombreuses questions fondamentales quant à la suppression ou à l'organisation des contenus ne mettent pas en cause exclusivement l'IA, l'utilisation de celle-ci pour façonner et modérer l'information à grande échelle amplifie bien des défis actuels et en soulève de nouveaux. Dans les sections qui suivent, il sera question du déploiement de l'IA aux fins de modération et d'organisation du contenu, et de ses répercussions possibles sur la liberté d'expression et la liberté de la presse.

## Modération du contenu

La prédominance de certains contenus, comme les messages trompeurs, haineux ou qui prônent l'extrémisme violent, détériore la qualité du discours public<sup>19</sup>. L'IA sert à évaluer les contenus afin de signaler, de démonétiser ou de supprimer certains d'entre eux, ou d'en réduire la visibilité, ou encore afin d'exclure des comptes en particulier<sup>20</sup>. L'IA est couramment déployée comme outil préalable de modération, sous la forme de filtres de téléversement, et pour analyser du contenu une fois que celui-ci est en ligne ou qu'il a été signalé par des utilisateurs<sup>21</sup>. À ce stade, soit l'IA intervient de façon autonome, soit le contenu est ultimement soumis à une vérification humaine<sup>22</sup>.

La capacité de l'IA à analyser du contenu demeure limitée. L'évaluation du discours dépend grandement du contexte et exige une compréhension des nuances culturelles, linguistiques et politiques<sup>23</sup>. Par conséquent, l'IA manque souvent d'exactitude<sup>24</sup>. Les faux positifs restreignent le discours d'une manière injustifiée alors que les faux négatifs peuvent causer un effet paralysant qui mène à l'autocensure et réduit au silence des personnes marginalisées<sup>25</sup>.

L'utilisation de l'IA par des intermédiaires pour modérer les contenus de façon proactive est une forme d'autorégulation dans laquelle les décisions pilotées par l'IA reposent généralement sur les conditions d'utilisation ou les règles de conduite propres à une communauté d'utilisateurs<sup>26</sup>. Les États demandent de plus en plus souvent aux intermédiaires d'éliminer des messages précis et



leur imposent de supprimer certaines catégories de contenu, souvent en fonction de définitions vagues, ce qui peut mener au blocage de nouvelles d'intérêt public<sup>27</sup>. Cette sous-traitance de l'application de la loi et des responsabilités juridiques force les acteurs du secteur privé à déployer l'IA, surtout lorsque des délais stricts leur sont imposés<sup>28</sup>. Non seulement ce phénomène suscite-t-il d'importantes inquiétudes au sujet de la primauté du droit et de l'application régulière de la loi, mais il entraîne également une dépendance à l'égard de quelques entreprises déjà puissantes<sup>29</sup>. Somme toute, l'IA semble accélérer la tendance vers une surveillance généralisée des communications, ce qui nuit profondément à la liberté de la presse<sup>30</sup>.

Pendant la pandémie de COVID-19, le confinement des modérateurs humains et le besoin grandissant de contrer la désinformation ont amené les États et les intermédiaires à se tourner de plus en plus vers l'IA. La pandémie a illustré l'importance de pouvoir compter sur une information fiable et diversifiée, mais elle a aussi fait ressortir la nécessité de réduire les effets secondaires de l'IA elle-même, car les erreurs qui s'accumulaient tardaient à être corrigées<sup>31</sup>.

## Organisation du contenu

L'abondance du contenu en ligne disperse de plus en plus l'attention des utilisateurs. Les intermédiaires d'Internet se servent de l'IA pour diffuser de l'information en fonction des préférences personnelles présumées de chaque utilisateur<sup>32</sup>. Toutefois, l'intention des intermédiaires lorsqu'ils réalisent leurs prédictions est de monnayer les données aux fins du ciblage publicitaire<sup>33</sup>. Par conséquent, l'organisation des fils d'actualité et des résultats de recherches, lorsqu'elle est alimentée par l'IA, vise à inciter les utilisateurs à s'investir davantage dans ces services respectifs et à leur consacrer plus de temps<sup>34</sup>. Les contenus controversés ou sensationnalistes peuvent attirer davantage l'attention, comme c'est le cas pour les propos misogynes, racistes, haineux ou qui attisent la peur<sup>35</sup>. Ainsi, les systèmes de classement par IA qui accordent plus de poids à la capacité de générer des clics qu'à l'intérêt réel pour de l'information peuvent entraîner une polarisation, une radicalisation et la diffusion de messages trompeurs ou haineux<sup>36</sup>.

Par ailleurs, les médias traditionnels, étant donné qu'ils dépendent de plus en plus de l'accès à leurs contenus et du partage de ceux-ci en ligne, doivent aussi se disputer l'attention des utilisateurs. Ils peuvent ainsi être obligés de mettre l'accent davantage sur l'« infodivertissement » que sur l'intérêt public, ce qui accroît d'autant la pression que subit le journalisme de qualité<sup>37</sup>.

En personnalisant le contenu, y compris les nouvelles, l'IA peut conforter les utilisateurs dans leurs opinions, créant ainsi des chambres d'écho et des bulles de filtrage<sup>38</sup>, et diminuant du même coup la probabilité qu'ils soient exposés à un contenu médiatique diversifié<sup>39</sup>. Il en résulte une perception déformée de la réalité, laquelle peut renforcer les déséquilibres de pouvoir et amplifier le sentiment d'altérité, tout en menaçant gravement la diversité des médias<sup>40</sup>.

La structure actuelle de l'Internet n'offre aux intermédiaires que peu d'incitatifs économiques pour les amener à proposer un contenu diversifié ou, à plus forte raison, factuel. L'IA conçue pour servir des intérêts commerciaux ou politiques favorise inmanquablement certains types de contenus susceptibles d'influencer et de réorienter les comportements à des fins d'optimisation des profits ou de persuasion, ou dans le but de réprimer intentionnellement le journalisme indépendant<sup>41</sup>. Les systèmes d'IA des intermédiaires peuvent être utilisés à mauvais escient par des régimes autoritaires ou autres, par exemple en recourant à des robots qui propagent des messages précis ou qui masquent la visibilité de contenus journalistiques<sup>42</sup>. Les outils d'IA peuvent servir à attaquer les journalistes dans le but de les faire taire, par exemple au moyen de campagnes de harcèlement orchestrées pour simuler un mouvement populaire. Cette méthode est particulièrement répandue pour cibler les femmes journalistes, et les systèmes de diffusion pilotés par l'IA peuvent même donner une impulsion à de telles attaques en les rendant virales<sup>43</sup>.

Bien que l'ampleur des répercussions de l'IA sur le discours public soit encore incertaine<sup>44</sup>, il ne fait aucun doute que l'IA est souvent déployée dans le but d'influencer les perceptions ou le comportement des gens. Les intermédiaires de l'Internet sont devenus des gardiens de l'information qui utilisent l'IA pour gérer le contenu médiatique et les flux d'information, ce qui façonne inévitablement les opinions et le

comportement des utilisateurs<sup>45</sup>. Les structures de l'IA peuvent être utilisées pour exercer un contrôle sur la presse, que ce soit en remplissant une fonction de blocage de l'information sous forme de censure ou, au contraire, de vecteur de propagande ou d'attaques<sup>46</sup>. En l'absence de garanties démocratiques, la surveillance du discours et la diffusion du contenu par IA mettent en péril la liberté de la presse, l'accès à l'information et la liberté d'expression, tout en soulevant des inquiétudes au sujet de la primauté du droit et de la discrimination systémique.

## Recommandations

De tout temps, l'humanité a fait appel aux technologies pour résoudre des problèmes de société. Pourtant, des questions qui soulèvent des controverses depuis longtemps ne peuvent pas être résolues simplement en sous-traitant les processus décisionnels à l'IA<sup>47</sup>. Au-delà de cet aspect, les technologies peuvent servir d'outils de surveillance, de censure et de répression des médias d'une ampleur sans précédent. Même si bon nombre de ces sujets d'inquiétude ne sont pas spécifiques à l'IA, l'utilisation de celle-ci aggrave les menaces qui pèsent déjà sur la liberté d'expression et la liberté de la presse. Pour contrer efficacement ces menaces, il est crucial de tenir compte du contexte sociotechnique dans lequel se déploie l'IA, de ses utilisateurs et des objectifs qu'ils poursuivent. Bien qu'il n'existe aucune solution universelle, on ne saurait évaluer ni traiter l'incidence de l'IA de manière judicieuse sans faire preuve de transparence et de sens des responsabilités<sup>48</sup>.

Puisqu'ils ont une obligation formelle de protéger la liberté d'expression et la liberté de la presse, les États doivent promouvoir un environnement propice à la diversité<sup>49</sup>. Lorsque les autorités publiques déploient elles-mêmes l'IA, elles doivent se conformer aux normes internationales en matière de droits de la personne, en s'assurant que toute restriction de la liberté d'expression ou de la liberté de la presse est nécessaire et proportionnée<sup>50</sup>. Le fait de collecter ou de fusionner à outrance des données dans le cadre de partenariats public-privé va à l'encontre de ces critères. Souvent, cela facilite au contraire un autoritarisme numérique qui pratique la surveillance de masse, cible autant des journalistes que des particuliers et impose une censure sans

précédent<sup>51</sup>. Les États ne doivent pas tirer parti de l'IA pour manipuler l'opinion publique, harceler les journalistes ou exercer d'autres mesures répressives, mais plutôt pour déterminer les limites acceptables de l'utilisation d'une telle technologie.

Les mesures réglementaires et les politiques en matière d'IA doivent reposer sur des faits probants et ne doivent pas entraîner de répercussions négatives sur la liberté de la presse. Les États doivent s'abstenir de déléguer aveuglément la protection des droits de la personne à l'IA<sup>52</sup>. En outre, tous les efforts déployés à cet égard doivent être encadrés par des règles strictes de protection des données<sup>53</sup>. Il ne devrait pas être nécessaire de consentir à des pratiques intrusives de surveillance pour pouvoir participer au discours public en ligne.

Il incombe aussi aux entreprises de respecter les droits de la personne<sup>54</sup>. Elles doivent empêcher que leurs systèmes d'IA soient utilisés à mauvais escient pour réduire au silence les dissidents et la presse. Même si de nombreuses entreprises s'engagent à faire preuve d'éthique, elles ne le font pas forcément en respectant les droits de la personne<sup>55</sup>. Néanmoins, les initiatives privées en matière d'éthique sont importantes, et les codes de déontologie jouent un rôle crucial dans la responsabilité sociale des entreprises. Or, ces codes et principes ne pouvant généralement pas s'appuyer sur des garanties démocratiquement légitimées ni sur des règles d'application de telles garanties, ils ne peuvent à eux seuls assurer une protection efficace<sup>56</sup>.

La transparence est une condition essentielle de toute surveillance publique<sup>57</sup>. Chacun doit être en mesure de savoir de quelle manière ont été prises les décisions qui influent sur sa vie, quelles données ont été traitées et dans quel but<sup>58</sup>. Les autorités de réglementation et la société en général doivent être au courant de l'incidence de l'IA sur le discours public et médiatique. Toutefois, en raison de l'asymétrie fondamentale qui caractérise l'information, ce domaine demeure très peu étudié<sup>59</sup>. Il faut donc encourager la recherche indépendante sur les conséquences sociétales de l'IA. Pour qu'un tel examen soit possible, l'IA doit être explicable et interprétable<sup>60</sup>. Ainsi, les États devraient envisager de rendre obligatoire la divulgation de l'usage de l'IA et de ses fonctions sous-jacentes, tout en étant transparents quant à leurs propres déploiements d'IA. De telles exigences

pourraient se décliner par niveaux, selon l'objectif poursuivi, le rôle de l'intervenant et le stade de développement ou d'utilisation de l'IA, ainsi que le risque de violation des droits de la personne<sup>61</sup>. En outre, il convient d'établir des règles claires pour garantir que l'information sur le déploiement de l'IA peut faire l'objet de comparaison<sup>62</sup>, et que les renseignements personnels sont protégés à toutes les étapes<sup>63</sup>.

La transparence doit aller de pair avec un renforcement de la capacité d'action des utilisateurs. Ces derniers doivent pouvoir exercer un choix et un contrôle sur la collecte, la surveillance et l'analyse des données utilisées pour personnaliser le contenu qui leur est destiné, ainsi que sur la conception des interfaces utilisées par les intermédiaires<sup>64</sup>. Pour augmenter le pouvoir des utilisateurs et accroître la résilience des citoyens, il faut renforcer la littératie numérique<sup>65</sup>.

La transparence est indispensable pour que l'on sache quels outils d'IA sont déployés et comment sont prises les décisions automatisées. Elle est aussi nécessaire pour qu'il soit possible de remettre en question les processus problématiques. Ceux qui tirent profit de l'IA doivent être tenus responsables de toutes les conséquences négatives de son utilisation. Pour assurer cette responsabilisation, des normes de gouvernance strictes sont cruciales. Des règles doivent garantir que les profits que réalise une entreprise lui imposent des responsabilités et que les décideurs peuvent être tenus responsables de leurs gestes<sup>66</sup>. Les États doivent envisager de mettre en place une structure de surveillance de l'IA par niveaux<sup>67</sup>, et étudier des modèles d'autoréglementation et de coréglementation, assortis de mécanismes de règlement des différends, d'organismes consultatifs sur les médias sociaux ou de tribunaux électroniques appelés à trancher rapidement les cas d'infraction<sup>68</sup>.

Pour garantir l'indépendance du contrôle, il faudrait accorder aux institutions nationales de défense des droits de la personne le pouvoir d'exercer elles aussi une supervision de l'IA. Un moyen important d'y parvenir consiste à évaluer périodiquement et avec rigueur l'incidence de l'IA sur les droits de la personne tout au long de son cycle de vie, et de rendre ces évaluations accessibles au grand public<sup>69</sup>. Par ailleurs, les outils d'IA devraient être soumis à des audits périodiques indépendants<sup>70</sup> assortis d'une analyse minutieuse visant à déterminer si l'IA est

utilisée abusivement pour entraver la liberté de la presse.

L'accès à des mesures correctives et à une réparation doit être garanti tant aux journalistes qu'aux utilisateurs dont le contenu a été restreint par l'IA, de même qu'à ceux qui relaient du contenu ou, encore, à ceux qui ont été lésés par des interfaces pilotées par l'IA<sup>71</sup>. Il faut également que ces mesures soient prises en temps opportun et qu'on leur alloue des ressources suffisantes<sup>72</sup>. En ce qui concerne certaines décisions automatisées, il faut prévoir une intervention humaine pour examiner le processus et le renverser au besoin<sup>73</sup>.

Il est possible aussi de s'inspirer de pratiques exemplaires qui ont cours dans d'autres domaines, dont les médias traditionnels, en vue d'assurer la transparence et la responsabilisation<sup>74</sup>.

De plus, il faut contrer le danger persistant de laisser des éléments discriminatoires entacher la conception et le déploiement de l'IA<sup>75</sup>. Seule une démarche globale et interdisciplinaire peut répondre efficacement à cette exigence. Les discussions à cet égard doivent être ouvertes à tous les interlocuteurs et à toutes les sphères de la société, notamment les utilisateurs finaux touchés<sup>76</sup>, la société civile, le milieu universitaire et les médias<sup>77</sup>.

La plupart des problèmes sont étroitement liés au fait qu'une poignée d'entreprises dominantes exercent un pouvoir et un contrôle considérables sur l'écosystème de l'information en ligne. La concentration du pouvoir, que ce soit entre les mains de l'État ou d'une entreprise, comporte toujours un risque considérable d'entrave aux libertés<sup>78</sup>. Les États doivent veiller à ce que le marché de l'IA soit concurrentiel pour inciter les intermédiaires à se doter de modèles d'entreprise différents pour offrir leurs services<sup>79</sup>. Soutenir le développement d'outils d'IA qui ne sont pas basés sur un système d'exploitation des données et de publicité ciblée permettrait de promouvoir la diversité du marché, de démocratiser l'IA et de favoriser des espaces en ligne publics axés sur des valeurs<sup>80</sup>.

Enfin, vu le caractère transnational et hautement interrelié des défis que pose l'IA, il est primordial de faire converger les efforts et de viser des solutions mondiales. Il existe différentes initiatives importantes à cet égard, par exemple celles de l'Organisation pour la sécurité et la coopération en Europe (OSCE), de l'UNESCO,

du Conseil de l'Europe ou de l'Union européenne<sup>81</sup>.

Si l'IA n'est pas une panacée aux problèmes de la société, elle ne devrait pas non plus être le bouc émissaire de toutes les difficultés auxquelles se heurtent la liberté d'expression ou la liberté de la presse. L'IA ne doit faciliter ni l'autoritarisme numérique ni la répression des médias au moyen de la haute technologie. Pour que l'IA permette l'accès à une information diversifiée et favorise la liberté d'expression et la liberté de la presse plutôt que de les

<sup>1</sup> Le droit relatif à « [...] la liberté de rechercher, de recevoir et de répandre des informations et des idées [...] par tout autre moyen de communication » est inscrit à l'article 19 du Pacte international relatif aux droits civils et politiques, ainsi que dans la Déclaration universelle des droits de l'homme de l'ONU et dans divers cadres régionaux sur les droits de la personne. Le Conseil des droits de l'homme des Nations Unies a affirmé que les droits de la personne hors ligne doivent aussi être protégés en ligne; voir à ce sujet la résolution sur « La promotion, la protection et l'exercice des droits de l'homme sur l'Internet », A/HRC/RES/20/8, 2012, <https://undocs.org/fr/A/HRC/RES/20/8>.

<sup>2</sup> On estime qu'en 2020, pour chaque minute qui passe, près de 150 000 messages sont échangés et 147 000 photos sont téléversées sur Facebook, 500 heures de vidéos sont versées sur YouTube et 347 222 histoires sont publiées sur Instagram. Voir Domo, « Data Never Sleeps 8.0 », <https://www.domo.com/learn/data-never-sleeps-8>, (en anglais seulement). Voir aussi E. Llansó, J. van Hoboken et collab., « Artificial Intelligence, Content Moderation, and Freedom of Expression », Transatlantic Working Group (TWG) on Content Moderation Online and Freedom of Expression, 2020, <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf> (en anglais seulement).

<sup>3</sup> Voir p. ex. OCDE, « Recommendations of the Council on Artificial Intelligence », 2020, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> [Recommandation du Conseil sur l'intelligence artificielle]; et Commission européenne, Groupe d'experts de haut niveau sur l'intelligence artificielle, « A Definition of AI: Main Capabilities and Disciplines », 2019, <https://ec.europa.eu/futurium/en/ai-alliance-consultation> (en anglais seulement).

<sup>4</sup> Voir p. ex., J. Bryson, A. Theodorou, « How Society Can Maintain Human-Centric Artificial Intelligence », 2019, <http://www.cs.bath.ac.uk/~jib/ftp/BrysonTheodorou19.pdf> (en anglais seulement).

<sup>5</sup> Voir p. ex., Haut-Commissariat des Nations Unies aux droits de l'homme, A/HRC/39/29, 2018, <https://undocs.org/fr/A/HRC/39/29>; S. Feldstein, « The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression », 2019, <https://carnegieendowment.org/files/201901-Feldstein-JournalOfDemocracy.pdf> (en anglais seulement); H. A. Ünver, « Artificial Intelligence, Authoritarianism and the Future of Political Systems », 2018, [https://edam.org.tr/wp-content/uploads/2018/07/AKIN-Artificial-Intelligence\\_Bosch-3.pdf](https://edam.org.tr/wp-content/uploads/2018/07/AKIN-Artificial-Intelligence_Bosch-3.pdf) (en anglais seulement).

<sup>6</sup> Voir p. ex., B. Bukovska, « Spotlight on Artificial Intelligence and Freedom of Expression #SAIFE », Bureau du

compromettre, il est impératif que tous les interlocuteurs veillent à la mise en place d'un cadre fondé sur les droits de la personne pour assurer la transparence et la responsabilité à l'égard de l'IA. Puisque l'IA touche de plus en plus tous les aspects de nos communications et de notre consommation médiatique, il est plus que temps d'intégrer des mesures de protection quant à son développement et à son utilisation, afin que puisse s'imposer la liberté de la presse.

Représentant pour la liberté des médias de l'OSCE, 2020, [https://www.osce.org/files/f/documents/9/f/456319\\_0.pdf](https://www.osce.org/files/f/documents/9/f/456319_0.pdf) (en anglais seulement).

<sup>7</sup> S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: Public Affairs, 2019 (Traduction française à paraître en novembre 2020 sous le titre *L'âge du capitalisme de surveillance*).

<sup>8</sup> Voir p. ex., N. Maréchal et collab., « It's the Business Model: How Big Tech's Profit Machine is Distorting the Public Sphere and Threatening Democracy », Ranking Digital Rights, 2020, <https://rankingdigitalrights.org/its-the-business-model> (en anglais seulement); J. Gary, A. Soltani, « First Things First: Online Advertising Practices and Their Effects on Platform Speech », Knight First Amendment Institute, 2019, <https://academiccommons.columbia.edu/doi/10.7916/d8-13tk-k917> (en anglais seulement); D. Strumpf, W. Fan, « Who Wants to Supply China's Surveillance State? The West », *The Wall Street Journal*, 2017, <https://www.wsj.com/articles/who-wants-to-supply-chinas-surveillance-state-the-west-1509540111> (en anglais seulement).

<sup>9</sup> Voir p. ex., Comité des droits de l'homme des Nations Unies, Observation générale n° 37, 2020, [https://www.ohchr.org/Documents/HRBodies/CCPR/GC37/FRENCH\\_GC37.doc](https://www.ohchr.org/Documents/HRBodies/CCPR/GC37/FRENCH_GC37.doc); Privacy International, « Two sides of the same coin », 2018, <https://privacyinternational.org/blog/1111/two-sides-same-coin-right-privacy-and-freedom-expression> (en anglais seulement); E. Stoycheff, « Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of the NSA Internet Monitoring », *Journal of Mass Communication Quarterly*, 2016 (en anglais seulement).

<sup>10</sup> J. Posetti, « Protecting Journalism Sources in the Digital Age », UNESCO, 2017, <https://unesdoc.unesco.org/ark:/48223/pf0000248054> (en anglais seulement); B. Heller, « Combating Terrorist-Related Content through AI and Information Sharing », TWG on Content Moderation Online, 2019, [https://www.ivir.nl/publicaties/download/Hash\\_sharing\\_Heller\\_April\\_2019.pdf](https://www.ivir.nl/publicaties/download/Hash_sharing_Heller_April_2019.pdf).

<sup>11</sup> F. Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge, London: Harvard University Press, 2015.

<sup>12</sup> Voir p. ex., Privacy International, ARTICLE 19, « Privacy and Freedom of Expression in the Age of Artificial Intelligence », 2018, <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf> (en



anglais seulement); J. Burrell, « How the Machine 'thinks' Understanding Opacity in Machine Learning Algorithms », *Big Data and Society*, 2016.

<sup>13</sup> Voir p. ex., E. Thorsen, « Surveillance of Journalists/Encryption Issues », *The International Encyclopaedia of Journalism Studies*, 2019.

<sup>14</sup> Voir p. ex., K. Yeung, « Responsabilité et IA : Étude sur les incidences des technologies numériques avancées (dont l'intelligence artificielle) sur la notion de responsabilité, sous l'angle des droits humains », Conseil de l'Europe, MSI-AUT, DGI(2019)05, 2019, <https://rm.coe.int/responsability-and-ai-fr/168097d9c6>.

<sup>15</sup> Amnistie Internationale, « Les géants de la surveillance : le modèle économique de Facebook et Google menace les droits humains », <https://www.amnesty.org/download/Documents/POL3014042019FRENCH.pdf> (Extraits de « Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights », 2019, <https://www.amnesty.org/en/documents/pol30/1404/2019/en>).

<sup>16</sup> Titulaires de mandat de l'ONU, de l'OSCE, de l'Organisation des États américains (OEA) et de la Commission africaine des droits de l'homme et des peuples (CADHP) sur le thème de la liberté d'expression et des médias, « Joint Declaration on Media Independence and Diversity in the Digital Age », 2018, <https://www.osce.org/files/f/documents/1/e/379351.pdf> (en anglais seulement); UNESCO, « Journalisme, liberté de la presse et COVID-19 », *Tendances mondiales en matière de liberté d'expression et de développement des médias*, 2020, [https://unesdoc.unesco.org/ark:/48223/pf0000373573\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000373573_fre).

<sup>17</sup> Voir p. ex., D. Brown, « Big Tech's Heavy Hand Around the Globe: Facebook and Google's dominance of developing-world markets has had catastrophic effects », Human Rights Watch, 2020, <https://www.hrw.org/news/2020/09/08/big-techs-heavy-hand-around-globe> (en anglais seulement); A. Hofseth, « E. Bell thinks public service media today has its most important role to play since World War II », NiemanLab, 2018.

<sup>18</sup> Voir p. ex., « Internet Censorship in China explained », Daxue Consulting, 2020, <https://daxueconsulting.com/internet-censorship-in-china> (en anglais seulement).

<sup>19</sup> Voir p. ex., le projet de l'OSCE sur la sécurité des femmes journalistes en ligne, <https://www.osce.org/fom/sofjo>; le rapport du Rapporteur spécial de l'ONU sur la promotion et la protection du droit à la liberté d'opinion et d'expression [Rapporteur spécial de l'ONU], A/HRC/38/35, 2018, <https://undocs.org/fr/A/HRC/38/35>; J. W. Penney, « Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study », *Internet Policy Review*, 2017; Rapporteur spécial de l'ONU, A/74/486, 2019, [https://www.ohchr.org/Documents/Issues/Opinion/A\\_74\\_486.pdf](https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf) (en anglais seulement).

<sup>20</sup> Selon Facebook, au cours du deuxième trimestre de 2020, 99 % du contenu terroriste et 94,5 % du discours haineux a été détecté grâce à la technologie proactive de détection; voir <https://transparency.facebook.com/community-standards-enforcement#dangerous-organizations> (en anglais seulement). Voir aussi C. Marsden, T. Meyer, « Regulating Disinformation with Artificial Intelligence », Service de recherche du Parlement européen, 2019; S. Spandana, « Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content », New America, 2019.

<sup>21</sup> Voir p. ex., D. Krivokapić, « OSCE RFoM Non-paper on the Impact of Artificial Intelligence on Freedom of Expression », Bureau du Représentant pour la liberté des médias de l'OSCE, 2020,

<https://www.osce.org/files/f/documents/b/a/447829.pdf> (en anglais seulement); E. Llansó, « No amount of « AI » in content moderation will solve filtering's prior-restraint problem », *Big Data & Society*, 2020.

<sup>22</sup> La vérification par des humains a des répercussions psychologiques sur les modérateurs; voir p. ex., S. Roberts, « Meet the people who scar themselves to clean up our social media networks », *Maclean's*, 2018, <https://www.macleans.ca/opinion/meet-the-people-who-scar-themselves-to-clean-up-our-social-media-networks> (en anglais seulement), et le documentaire « The Cleaners » d'H. Block et M. Riesewieck, 2018.

<sup>23</sup> Conseil de l'Europe, « Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications », MSI-NET, DGI(2017)12, 2018, <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> (en anglais seulement). Voir aussi, p. ex., « Caught in the Net: The Impact of Extremist Speech Regulations on Human Rights Content », EFF, *Syrian Archive*, Witness, 2019, <https://syrianarchive.org/en/lost-found/impact-extremist-human-rights#youtube-censorship-of-conflict-documentation-in-syria-yemen-and-ukraine> (en anglais seulement).

<sup>24</sup> Par exemple, Perspective, l'interface de programmation (API) de Jigsaw (filiale de Google) offre une trousse à outils en accès libre pour évaluer la « toxicité » des contenus. Pourtant, l'équipe de développement déconseille de l'utiliser pour la modération automatique en raison de trop nombreuses erreurs; voir <https://conversationalai.github.io>. Voir aussi, p. ex., Cambridge Consultants, « Use of AI in Online Content Moderation », Ofcom, 2019; N. Duarte, E. Llansó, « Mixed Messages? The Limits of Automated Social Media Content Analysis », Center for Democracy & Technology (CDT), 2017, <https://cdt.org/wp-content/uploads/2017/11/Mixed-Messages-Paper.pdf> (en anglais seulement).

<sup>25</sup> Voir p. ex., Algorithm Watch, « Automated moderation tool from Google rates People of Color and gays as 'toxic' », 2020, <https://algorithmwatch.org/en/story/automated-moderation-perspective-bias> (en anglais seulement); Fairspeech Project, « The Race Gap in Speech Recognition Technology », 2020, <https://fairspeech.stanford.edu> (en anglais seulement).

<sup>26</sup> Elles ne sont pas nécessairement conformes aux normes en matière de droits de la personne et peuvent être modifiées de façon opaque. Voir p. ex., Rapporteur spécial de l'ONU, A/HRC/38/35, 2018.

<sup>27</sup> Voir p. ex., titulaires de mandat de l'ONU, de l'OSCE, de l'OEA et de la CADHP sur le thème de la liberté d'expression et des médias, « 20<sup>th</sup> Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade », 2019, <https://www.osce.org/files/f/documents/9/c/425282.pdf> (en anglais seulement).

<sup>28</sup> Différents textes de loi prévoient des laps de temps très courts, p. ex. la directive de l'UE sur le droit d'auteur et les droits connexes dans le marché unique numérique (UE 2019/790), ou la loi de l'Allemagne sur l'application du droit sur les réseaux, 2017.

<sup>29</sup> Voir p. ex., Rapporteur spécial de l'ONU, rapport sur les technologies d'intelligence artificielle et leur incidence sur la liberté d'expression et l'écosystème de l'information,

A/73/348, 2018, <https://undocs.org/A/73/348> (en anglais seulement); H. Bloch-Wehba, « Automation in Moderation », *Cornell International Law Journal*, 2020.

<sup>30</sup> Les États ne doivent pas imposer aux intermédiaires une obligation générale de surveiller l'information qu'ils transmettent, voir p. ex., Rapporteur spécial de l'ONU, A/74/486, 2019, <https://undocs.org/fr/a/74/486>; Conseil de l'Europe, Recommandation CM/Rec(2018)2 du Comité des Ministres aux États membres sur les rôles et les responsabilités des intermédiaires d'internet, 2018, [https://search.coe.int/cm/Pages/result\\_details.aspx?Objectid=0900001680790e37](https://search.coe.int/cm/Pages/result_details.aspx?Objectid=0900001680790e37).

<sup>31</sup> Voir p. ex., UNESCO, « Journalism, press freedom and COVID-19 », 2020; J. York, C. McSherry, « Automated Moderation Must be Temporary, Transparent and Easily Appealable », EFF, 2020, <https://www.eff.org/deeplinks/2020/04/automated-moderation-must-be-temporary-transparent-and-easily-appealable> (en anglais seulement); E. Llansó, « COVID-19 Content Moderation Research Letter, CDT, 2020, <https://cdt.org/insights/covid-19-content-moderation-research-letter> (en anglais, en espagnol et en arabe). On peut aussi trouver de l'information pertinente dans les rapports sur la COVID-19 que produisent des intermédiaires d'Internet.

<sup>32</sup> Voir p. ex., M. DeVito, « From Editors to Algorithms », *Digital Journalism*, 2016; B. Stark, D. Stegmann et collab., « Are Algorithms a Threat to Democracy? The rise of Intermediaries: A Challenge for Public Discourse », Algorithm Watch, 2020, <https://algorithmwatch.org/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf> (en anglais seulement); F. Ricci et collab., *Recommender Systems Handbook*, Berlin: Springer, 2010.

<sup>33</sup> S. Zuboff parle du « commerce de l'avenir des humains », voir *The Age of Surveillance Capitalism*, 2019. (Traduction française à paraître en novembre 2020 sous le titre *L'âge du capitalisme de surveillance*).

<sup>34</sup> Voir p. ex., C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown Publishers, 2016.

<sup>35</sup> Voir, p. ex., la déclaration de l'ancien directeur principal de la technologie de Google, N. Wong, 2018, <https://www.vox.com/2018/9/12/17848384/nicole-wong-cto-lawyer-google-twitter-kara-swisher-decode-podcast-full-transcript> (en anglais seulement), et M. Zuckerberg, « A Blueprint for Content Governance and Enforcement », 2018, <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634> (en anglais seulement).

<sup>36</sup> N. Statt, « Facebook reportedly ignored its own research showing algorithms divided users », *The Verge*, 2020, <https://www.theverge.com/2020/5/26/21270659/facebook-division-news-feed-algorithms> (en anglais seulement); J. Cobbe, J. Singh, « Regulating Recommending: Motivations, Considerations, and Principles », *European Journal of Law and Technology*, 2019; M. Ribeiro et collab., « Auditing Radicalization Pathways on YouTube », 2019; S. Vosoughi et collab., « The spread of true and false news online », *Massachusetts Institute of Technology*, 2018, <https://science.sciencemag.org/content/359/6380/1146> (en anglais seulement).

<sup>37</sup> N. Helberger et collab., « Artificial Intelligence – Intelligent Politics: Challenges and opportunities for media and democracy », Conseil de l'Europe, 2020, [https://rm.coe.int/cyprus-2020-ai-and-freedom-of-](https://rm.coe.int/cyprus-2020-ai-and-freedom-of)

[expression/168097fa82](https://rm.coe.int/cyprus-2020-ai-and-freedom-of-expression/168097fa82) (en anglais seulement); N. Tashidian et collab., « Platforms and Publishers: The End of an Era », *Tow Center for Digital Journalism*, 2020; E. Bell, T. Owen, « The Platform Press: How Silicon Valley reengineered journalism », *Columbia Journalism Review*, 2017; UNESCO, « Étude préliminaire sur l'éthique de l'intelligence artificielle », 2019, [https://unesdoc.unesco.org/ark:/48223/pf0000367823\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000367823_fre); Reporters sans frontières, « Online Harassment of Journalists: Attack of the trolls », 2018, [https://rsf.org/sites/default/files/rsf\\_report\\_on\\_online\\_harassment.pdf](https://rsf.org/sites/default/files/rsf_report_on_online_harassment.pdf) (en anglais seulement).

<sup>38</sup> E. Pariser, « The Filter Bubble », *The Atlantic*, 2010, <https://www.theatlantic.com/daily-dish/archive/2010/10/the-filter-bubble/181427> (en anglais seulement); C. R. Sunstein, *Echo Chambers: Bush v. Gore, Impeachment, and Beyond*, Princeton University Press, 2001; B. Bodó et collab., « Interested in Diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization », *Digital Journalism*, 2018; S. Ignatidou, « AI-driven Personalization in Digital Media », Chatham House, 2019; M. McCarthy, « The Ethical Character of Algorithms – and What It Means for Fairness, the Character of Decision-Making, and the Future of News », Shorenstein Center on Media, Politics and Public Policy, 2019; K. Yeung, « Five Fears About Mass Predictive Personalisation in an Age of Surveillance Capitalism », *International Data Privacy Law*, 2018.

<sup>39</sup> Voir p. ex., N. Helberger, « Challenging Diversity – Social Media Platforms and a New Conception of Media Diversity », in: *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, New York: Oxford University Press, 2018. Voir aussi J. Möller et collab., « Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity », *Information, Communication & Society*, 2018; UNESCO, « Tendances mondiales en matière de liberté d'expression et de développement des médias: rapport mondial 2017/2018 », 2018, <https://unesdoc.unesco.org/ark:/48223/pf0000261965>.

<sup>40</sup> Voir p. ex., S. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York: NYU Press, 2018; UNESCO, *Global Report 2017/2018*; N. Usher et collab., « Twitter Makes It Worse: Political Journalists, Gendered Echo Chambers, and the Amplification of Gender Bias », *The International Journal of Press/Politics*, 2018.

<sup>41</sup> Z. Tufekci parle d'« architectures de persuasion » dans « We're building a dystopia just to make people click on ads », TEDGlobal, 2017. Voir aussi H. Brignull, <https://darkpatterns.org> (en anglais seulement); S. Spandana, « Rising Through the Ranks: How Algorithms Rank and Curate Content in Search Results and on News Feeds », *New America*, 2019, <https://www.newamerica.org/oti/reports/rising-throughranks> (en anglais seulement); Conseil de l'Europe, DGI(2017)12, 2018, « Algorithmes et droits humains », <https://rm.coe.int/algorithms-and-human-rights-fr/1680795681>; Rapporteur spécial de l'ONU, rapport sur les technologies d'intelligence artificielle et leurs implications pour la liberté d'expression et l'écosystème de l'information, A/73/348, 2018, <https://undocs.org/fr/A/73/348>.

<sup>42</sup> S. Feldstein, « We Need to Get Smart About How Governments Use AI », *Carnegie Endowment for International Peace*, 2019, <https://carnegieendowment.org/2019/01/22/we-need-to-get-smart-about-how-governments-use-ai-pub-78179> (en anglais seulement); C. Cadwalladr, E. Graham-Harrison, « The Cambridge Analytica Files », *The Guardian*, 2018, <https://www.theguardian.com/news/series/cambridge-analytica-files> (en anglais seulement); M. Brundage et

---

collab., « The Malicious Use of Artificial intelligence: Forecasting, Prevention, and Mitigation », Future of Humanity Institute, OpenAI, Electronic Frontier Foundation et collab., 2018.

<sup>43</sup> M. Ressa, « Propaganda war: Weaponizing the internet », Rappler, 2016, <https://www.rappler.com/nation/propaganda-war-weaponizing-internet> (en anglais seulement); M. Ferrier, « Attacks and Harassment: The Impact on Female Journalists and Their Reporting », International Women's Media Foundation and TrollBusters, 2018, <https://www.iwmf.org/wp-content/uploads/2018/09/Attacks-and-Harassment.pdf> (en anglais seulement); Reporters sans frontières, « Online Harassment of Journalists: Attack of the trolls », 2018, [https://rsf.org/sites/default/files/rsf\\_report\\_on\\_online\\_harassment.pdf](https://rsf.org/sites/default/files/rsf_report_on_online_harassment.pdf) (en anglais seulement).

<sup>44</sup> Les recherches ne permettent pas de tirer de conclusions à cet égard. Voir p. ex., E. Dubois, G. Blank, « The echo chamber is overstated: the moderating effect of political interest and diverse media », 2018; Stark, Stegmann, « Are Algorithms a Threat to Democracy? », 2020; UNESCO, Rapport annuel 2017/2018.

<sup>45</sup> Voir p. ex., Rapporteur spécial de l'ONU, A/73/348, 2018.

<sup>46</sup> N. Wright, « Artificial Intelligence, China, Russia, and the Global Order », Air University Press, 2019, [https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B\\_0161\\_WRIGHT\\_ARTIFICIAL\\_INTELLIGENCE\\_CHINA\\_RUSSIA\\_AND\\_THE\\_GLOBAL\\_ORDER.PDF](https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B_0161_WRIGHT_ARTIFICIAL_INTELLIGENCE_CHINA_RUSSIA_AND_THE_GLOBAL_ORDER.PDF) (en anglais seulement); S. Feldstein, « The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression », 2019.

<sup>47</sup> Notamment la question de savoir comment déterminer les limites du discours acceptable. Voir D. Kaye, *Speech Police: The Global Struggle to Govern the Internet*, New York: Columbia Global Reports, 2019. Voir aussi p. ex., J. York, C. McSherry, « Content Moderation is Broken, Let Us Count the Ways », EFF, 2019, <https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways> (en anglais seulement).

<sup>48</sup> Pour des initiatives de la société civile, voir p.ex., les principes de Santa Clara sur la transparence et la reddition de compte en matière de modération des contenus, <https://santaclaraprinciples.org> (en anglais seulement), et les principes de Manille sur la responsabilité des intermédiaires, <https://www.manilaprinciples.org> (en anglais seulement).

<sup>49</sup> ONU, OSCE, OEA, CADHP, « 20<sup>th</sup> Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade », 2019.

<sup>50</sup> Voir l'article 19(3) du Pacte international relatif aux droits civils et politiques, qui prévoit l'exercice de restrictions pour des motifs alliant la légalité, la légitimité et la nécessité. Voir aussi, p. ex., Rapporteur spécial de l'ONU, A/73/348, 2018; Conseil de l'Europe, recommandation CM/Rec(2020)1 du Comité des ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, 2020, [https://search.coe.int/cm/pages/result\\_details.aspx?Objectid=09000016809e1124](https://search.coe.int/cm/pages/result_details.aspx?Objectid=09000016809e1124).

<sup>51</sup> Voir p. ex., Haut-Commissariat des Nations Unies aux droits de l'homme, A/HRC/39/29, 2018; N. Wright, « How Artificial Intelligence Will Reshape the Global Order », *Foreign Affairs*, 2018, <https://alainstitute.org/images/Library/ArtificialIntelligenceReshapeGlobalOrder.pdf> (en anglais seulement); Comité des droits de l'homme des Nations Unies, Observation générale n° 37, 2020.

<sup>52</sup> Voir p. ex., M. Cornils, « Designing platform governance: A normative perspective on needs, strategies, and tools to regulate intermediaries », Algorithm Watch, 2020, <https://algorithmwatch.org/wp-content/uploads/2020/05/Governing-Platforms-legal-study-Cornils-May-2020-AlgorithmWatch.pdf> (en anglais seulement).

<sup>53</sup> Voir p. ex., Conseil de l'Europe, « Lignes directrices sur l'intelligence artificielle et la protection des données », Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, 2019, <https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40>; « Big data, artificial intelligence, machine learning and data protection », UK Information Commissioner's Office, 2020.

<sup>54</sup> Haut-Commissariat des Nations Unies aux droits de l'homme, « Principes directeurs relatifs aux entreprises et aux droits de l'homme », 2011, [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_fr.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_fr.pdf); Conseil de l'Europe, commissaire aux droits de l'homme, « Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme », 2019; Rapporteur spécial de l'ONU, A/74/486, 2019.

<sup>55</sup> J. Field et collab., « Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI », Berkman Klein Center for Internet and Society, 2020, <http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420> (en anglais seulement); R. Ochigame, « The Invention of 'Ethical AI': How Big Tech Manipulates Academia to Avoid Regulation », *The Intercept*, 2019, <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence> (en anglais seulement).

<sup>56</sup> Voir p. ex., T. Hagendorff, « The Ethics of AI Ethics: An Evaluation of Guidelines, Minds and Machines », 2020; M. Latonero, « Governing Artificial Intelligence: Upholding Human Rights & Dignity », *Data & Society*, 2018, [https://datasociety.net/wp-content/uploads/2018/10/DataSociety\\_Governing\\_Artificial\\_Intelligence\\_Upholding\\_Human\\_Rights.pdf](https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf) (en anglais seulement).

<sup>57</sup> Voir p. ex., Rapporteur spécial de l'ONU, A/73/348, 2018; « Spotlight on Artificial Intelligence and Freedom of Expression #SAIFE », Bureau du Représentant pour la liberté des médias de l'OSCE, 2020; H. Tworek, « How Transparency Reporting Could Incentivize Irresponsible Content Moderation », TWG on Content Moderation Online, 2019, <https://www.cigionline.org/articles/how-transparency-reporting-could-incentivize-irresponsible-content-moderation> (en anglais seulement).

<sup>58</sup> Par exemple, le programme Ranking Digital Rights publie un rapport annuel sur les renseignements divulgués par les sociétés Internet; voir « 2019 RDR Corporate Accountability Index », 2019, <https://rankingdigitalrights.org/index2019/assets/static/download/RDRindex2019report.pdf> (en anglais seulement).

<sup>59</sup> B. Bodó et collab., « Tackling the algorithmic control crisis – the technical, legal, and ethical challenges of research into algorithmic agents », *Yale Journal of Law and Technology*, 2017; S. Cope, J. York, « Industry Efforts to Censor Pro-Terrorism Online Content Pose Risks to Free Speech », EFF, 2017, <https://www.eff.org/deeplinks/2017/07/industry-efforts-censor-pro-terrorism-online-content-pose-risks-free-speech> (en anglais seulement); C. Colett, S. Dillon, « AI and Gender: Four Proposals for Future Research », University of Cambridge, 2019, <http://lcfi.ac.uk/resources/ai-and-gender-four-proposals-future-research> (en anglais seulement).



<sup>60</sup> C. Rudin, « Stop explaining black box machine learning models for high stake decisions and use interpretable models instead », *Nature Machine Intelligence*, 2019, <https://www.nature.com/articles/s42256-019-0048-x.pdf> (en anglais seulement). Le Règlement général sur la protection des données de l'UE (Règlement (UE) 2016/679) consacre le droit à la transparence en matière d'information; voir p. ex., A. Burt, « Is there a 'right to explanation' for machine learning in the GDPR? », *Privacy Tech*, 2017, <https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr> (en anglais seulement).

<sup>61</sup> Une approche par niveaux pourrait s'attaquer à l'opacité due aux secrets d'État ou aux secrets industriels et réduire le risque que la manipulation mette en péril la transparence de l'IA. Voir p. ex., M. MacCarthy, « Transparency Requirements for Digital Social Media Platforms », 2020, [https://www.ivir.nl/publicaties/download/Transparency\\_MacCarthy\\_Feb\\_2020.pdf](https://www.ivir.nl/publicaties/download/Transparency_MacCarthy_Feb_2020.pdf) (en anglais seulement); TWG on Content Moderation Online, « Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online », 2020, <https://www.annenbergpublicpolicycenter.org/feature/transatlantic-working-group-freedom-and-accountability> (en anglais seulement); titulaires de mandat de l'ONU, de l'OSCE, de l'OEA et de la CADHP sur le thème de la liberté d'expression et des médias, « Joint Declaration on Freedom of Expression and Elections in the Digital Age », 2020, [https://www.osce.org/files/f/documents/9/8/451150\\_0.pdf](https://www.osce.org/files/f/documents/9/8/451150_0.pdf) (en anglais seulement).

<sup>62</sup> Les rapports actuels sur la transparence ne sont pas suffisamment étoffés, et on ne peut pas les comparer entre eux. Voir p. ex., D. Keller, P. Leerssen, « Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation », dans N. Persily, J. Tucker, *Social Media and Democracy: The State of the Field and Prospects for Reform*, Cambridge: University Press, 2019; P. Leerssen, « The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems », *Institute for Information Law, University of Amsterdam*, 2020.

<sup>63</sup> Conseil des droits de l'homme des Nations Unies, résolution sur « Le droit à la vie privée à l'ère du numérique », A/HRC/RES/34/7, 2017, <https://undocs.org/fr/A/HRC/RES/34/7>; Haut-Commissariat des Nations Unies aux droits de l'homme, A/HRC/39/29, 2018; Rapporteur spécial de l'ONU, A/HRC/29/32, 2015, <https://undocs.org/fr/A/HRC/29/32>.

<sup>64</sup> J. Harambam et collab., « Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems », 2019, [https://www.ivir.nl/publicaties/download/paper\\_recsys\\_19.pdf](https://www.ivir.nl/publicaties/download/paper_recsys_19.pdf) (en anglais seulement); Rapporteur spécial de l'ONU, A/73/348, 2018.

<sup>65</sup> Voir p. ex., Conseil de l'Europe, recommandation CM/Rec(2019)10 du Comité des ministres aux États membres visant à développer et à promouvoir l'éducation à la citoyenneté numérique, [https://search.coe.int/cm/Pages/result\\_details.aspx?Objectid=090000168098ddfe](https://search.coe.int/cm/Pages/result_details.aspx?Objectid=090000168098ddfe), 2019.

<sup>66</sup> Voir p. ex., « Getting to the Source of Infodemics: It's the Business Model », *Ranking Digital Rights*, 2020; Access Now, « Protecting Free Expression in the Era of Online Content Moderation », 2019, <https://www.accessnow.org/cms/assets/uploads/2019/05/AccessNow-Preliminary-Recommendations-On-Content-Moderation-and-Facebooks-Planned-Oversight-Board.pdf> (en anglais seulement).

<sup>67</sup> Voir p. ex., Marsden, Meyer, « Regulating Disinformation with Artificial Intelligence », 2019; C. Wendehorst, C. Woopen, « Opinion of the German Data Ethics Commission », 2019, [https://www.bmiv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission\\_EN\\_node.html](https://www.bmiv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html); (en anglais seulement); M. Wieringa, « What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability », *ACM Conference on Fairness, Accountability, and Transparency*, 2020.

<sup>68</sup> TWG on Content Moderation Online, « Freedom and Accountability », 2020; H. Tworek et collab., « Dispute Resolution and Content Moderation: Fair, Accountable, Independent, Transparent, and Effective », 2020, [https://www.ivir.nl/publicaties/download/Dispute\\_Resolution\\_Content\\_Moderation\\_Final.pdf](https://www.ivir.nl/publicaties/download/Dispute_Resolution_Content_Moderation_Final.pdf) (en anglais seulement); ARTICLE 19, « Social Media Councils », <https://www.article19.org/social-media-councils> (en anglais seulement). Pour des initiatives d'autoréglementation, voir p. ex., Global Network Initiative, <https://globalnetworkinitiative.org> (en anglais seulement); IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, <https://ethicsinaction.ieee.org> (en anglais seulement); Partnership for AI, <https://www.partnershipnai.org> (en anglais seulement).

<sup>69</sup> Les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme exigent qu'une diligence raisonnable soit exercée en matière de droits de la personne, et que les incidences réelles ou possibles à cet égard soient cernées et gérées. Voir aussi, D. Reisman et collab., « Algorithmic impact assessments: a practical framework for public agency accountability », *AI Now*, 2018, <https://ainowinstitute.org/aiareport2018.pdf> (en anglais seulement); A. Mantelero, « AI and Big Data: A blueprint for a human rights, social and ethical impact assessment », *Computer Law & Security Review*, 2018.

<sup>70</sup> Le réseau Women Leading in AI, par exemple, propose de mettre en place des certificats pour les systèmes d'IA audités; voir « Women Leading in AI : 10 Principles of Responsible AI », 2019, <https://womenleadinginai.org/wp-content/uploads/2019/02/WLIAI-Report-2019.pdf> (en anglais seulement). Voir aussi J. Kroll et collab., « Accountable Algorithms », *Pennsylvania Law Review*, 2017, [https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn\\_law\\_review](https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn_law_review) (en anglais seulement); C. Sandvig et collab., « Auditing Algorithms », 2014, <http://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf> (en anglais seulement); H. Ziady, « Facebook and YouTube accept hate speech audits to keep advertisers happy », *CNN Business*, 2020, [https://edition.cnn.com/2020/09/23/tech/facebook-youtube-advertisers/index.html?mc\\_cid=eb7d5959b9&mc\\_eid=8db8161786](https://edition.cnn.com/2020/09/23/tech/facebook-youtube-advertisers/index.html?mc_cid=eb7d5959b9&mc_eid=8db8161786) (en anglais seulement).

<sup>71</sup> Voir p. ex., Rapporteur spécial de l'ONU, A/HRC/32/38, 2016; titulaires de mandat de l'ONU, de l'OSCE, de l'OEA et de la CADHP sur le thème de la liberté d'expression et des médias, « Joint Declaration on Freedom of Expression and « Fake News », Disinformation and Propaganda », 2017, <https://www.osce.org/files/f/documents/6/8/302796.pdf> (en anglais seulement); York, McSherry, « Content Moderation is Broken. Let Us Count the Ways, » 2019.

<sup>72</sup> Pour que les mesures correctives et la réparation soient efficaces, l'engagement de la direction et l'affectation de ressources à cette fin sont cruciaux; voir p. ex., C. Silverman et collab., « 'I Have Blood on My Hands': A Whistleblower Says Facebook Ignored Global Political Manipulation », *BuzzFeed News*, 2020, <https://www.buzzfeednews.com/article/craigsilverman/facebook>

---

[ok-ignore-political-manipulation-whistleblower-memo](#) (en anglais seulement).

<sup>73</sup> Voir p. ex., K. Bontcheva et collab., « Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression », International Telecommunications Union (ITU) et UNESCO, 2020,

[https://www.broadbandcommission.org/Documents/working-groups/FoE\\_Disinfo\\_Report.pdf](https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf) (en anglais seulement); R. Gorwa et collab., « Algorithmic content moderation: Technical and political challenges in the automation of platform governance », *Big Data & Society*, 2020; G. Wang, « Humans in the Loop », 2019,

<https://hai.stanford.edu/blog/humans-loop-design-interactive-ai-systems> (en anglais seulement). Le Rapporteur spécial de l'ONU a lancé un appel pour que l'on veille à ce que des humains fassent partie de tout processus lié aux outils d'IA (A/74/486).

<sup>74</sup> En ce qui concerne les médias traditionnels, la prise de décision éditoriale doit habituellement être indépendante des activités commerciales. Parmi les autres domaines d'intérêt, citons la responsabilité associée aux produits; voir Commission européenne, « Livre blanc sur l'intelligence artificielle », ou la réglementation aérospatiale; voir I. Raji et collab., « Closing the AI accountability gap », 2020. Des leçons peuvent aussi être tirées du domaine du droit de l'environnement (études d'impact), de la réglementation automobile (sécurité et limites de vitesse), de la médecine (certifications), du domaine militaire (« véritable contrôle humain »), de la réglementation des télécommunications, etc.; J. Ausloos et collab., « Operationalizing Research Access in Platform Governance: What to learn from other industries? », Algorithm Watch, 2020, [https://algorithmwatch.org/wp-content/uploads/2020/06/GoverningPlatforms\\_IViR\\_study\\_June2020-AlgorithmWatch-2020-06-24.pdf](https://algorithmwatch.org/wp-content/uploads/2020/06/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf) (en anglais seulement).

<sup>75</sup> Voir p. ex., Amnistie Internationale et collab., « Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems », 2018, [https://www.accessnow.org/cms/assets/uploads/2018/08/Th\\_e-Toronto-Declaration\\_ENG\\_08-2018.pdf](https://www.accessnow.org/cms/assets/uploads/2018/08/Th_e-Toronto-Declaration_ENG_08-2018.pdf) (en anglais seulement); Lighthouse3 initiative, liste de femmes actives dans le domaine de l'éthique de l'IA, <https://lighthouse3.com/womeninaethics> (en anglais seulement); AI Now, « 2019 Report », 2019, [https://ainowinstitute.org/AI\\_Now\\_2019\\_Report.pdf](https://ainowinstitute.org/AI_Now_2019_Report.pdf) (en anglais seulement); UNESCO, « Artificial Intelligence and Gender Equality », 2020, <https://en.unesco.org/AI-and-GE-2020> (en anglais et en espagnol); Conseil de l'Europe, Recommandation du Conseil des ministres aux États membres sur le pluralisme des médias et la transparence de leur propriété, CM/Rec(2018)1, 2018, [https://search.coe.int/cm/Pages/result\\_details.aspx?Objectid=0900001680790e36](https://search.coe.int/cm/Pages/result_details.aspx?Objectid=0900001680790e36); R. Avila et collab., « Artificial Intelligence: open questions about gender inclusion », WWW Foundation, <http://webfoundation.org/docs/2018/06/AI-Gender.pdf> (en anglais seulement).

<sup>76</sup> Selon une étude récente du BEUC, le Bureau européen des unions de consommateurs, les utilisateurs ne font pas confiance aux applications d'IA quant à la protection de leurs renseignements personnels et craignent que l'IA soit utilisée pour manipuler leurs décisions. Seuls 20 % d'entre eux pensent que les lois actuelles sont adéquates. Voir BEUC, « Artificial Intelligence: what consumers say », 2020, <https://www.beuc.eu/publications/beuc-x-2020->

[078\\_artificial\\_intelligence\\_what\\_consumers\\_say\\_report.pdf](#) (en anglais seulement).

<sup>77</sup> X. Hu et collab., « Steering AI and advanced ICTs for knowledge societies: A Rights, Openness, Access, and Multi-stakeholder Perspective »; UNESCO, 2019, [https://unesdoc.unesco.org/ark:/48223/pf0000372132\\_locale=en](https://unesdoc.unesco.org/ark:/48223/pf0000372132_locale=en) (en anglais seulement); N. Helberger et collab., « Governing online platforms: From contested to cooperative responsibility », *The Information Society*, 2017; Conseil de l'Europe, CM/Rec(2018)1, 2018; F. Carugati, « A Council of Citizen Should Regulate Algorithms », WIRED, 2020, <https://www.wired.com/story/opinion-a-council-of-citizens-should-regulate-algorithms> (en anglais seulement).

<sup>78</sup> En ce qui concerne les médias traditionnels, voir p. ex., Comité des droits de l'homme des Nations Unies, Observation générale n° 34 pour « empêcher une domination ou concentration induite des organes d'information par des groupes de médias contrôlés par des intérêts privés dans des situations de monopole qui peuvent être préjudiciables à la diversité des sources et des opinions ». ONU, OSCE, OEA, CADHP, « 20<sup>th</sup> Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade », 2019; Z. Teachout, *Break 'Em Up: Recovering Our Freedom from Big Ag, Big Tech, and Big Money*, New York: St. Martin's Press, 2020.

<sup>79</sup> Parmi les idées qui circulent pour renforcer la compétitivité figurent les normes sur l'interopérabilité, la portabilité des données, la décentralisation, les exigences relatives aux sources ouvertes, etc. Voir p. ex., J. Furman et collab., « Unlocking digital competition: Report of the Digital Competition Expert Panel », 2019, <https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel> (en anglais seulement); U. Dolata, *Internet Companies: Market Concentration, Competition and Power*, Berlin: Springer, 2018.

<sup>80</sup> Les espaces en ligne publics axés sur des valeurs pourraient, p. ex., fournir un service d'intérêt public pour les communications indispensables. Voir p. ex., E. Zuckermann, « The Case for Digital Public Infrastructure », Knight First Amendment Institute, 2020, <https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure> (en anglais seulement); I. Moura, « AI is transforming society: Here's what we can do to make sure it prioritizes human needs », Medium, 2020, <https://medium.com/human-machine-collaboration/ai-is-transforming-society-heres-what-we-can-do-to-make-sure-it-prioritizes-human-needs-34794565aad0> (en anglais seulement).

<sup>81</sup> Voir en particulier, OSCE, « Impact of Artificial Intelligence on Freedom of Expression », <https://www.osce.org/fom/ai-free-speech> (en anglais seulement); UNESCO, « Élaboration d'une recommandation sur l'éthique de l'intelligence artificielle », <https://fr.unesco.org/artificial-intelligence/ethics>; Conseil de l'Europe, « Conseil de l'Europe et intelligence artificielle », <https://www.coe.int/fr/web/artificial-intelligence>; Union européenne, « Artificial Intelligence », <https://ec.europa.eu/digital-single-market/en/artificial-intelligence> (en anglais seulement); ITU, « Artificial Intelligence », <https://www.itu.int/en/ITU-T/AI/Pages/default.aspx> (en anglais seulement); et OCDE, « OECD Principles on AI », (<https://www.oecd.org/going-digital/ai/principles>) (en anglais seulement).