



Recommandation sur

l'éthique de l'intelligence artificielle

Adoptée le 23 novembre 2021

Publié en 2022 par l'Organisation des Nations Unies pour l'éducation, la science et la culture
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2022

SHS/BIO/PI/2021/1

Tous droits réservés.



Œuvre publiée en libre accès sous la licence Attribution-NonCommercial-ShareAlike 3.0 IGO (CC-BY-NC-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-nc-sa/3.0/igo/>). Les utilisateurs du contenu de la présente publication acceptent les termes d'utilisation de l'archive ouverte de libre accès UNESCO (www.unesco.org/open-access/terms-use-ccbyncsa-fr).

Photo de couverture :

Irina Bg/Shutterstock.com; Artistdesign29/Shutterstock.com; kynny/gettyimages; gorodenkoff/gettyimages; Gorodenkoff/Shutterstock.com; dieddin/Shutterstock.com; Gorodenkoff/Shutterstock.com; PopTika/Shutterstock.com; Horth Rasur/Shutterstock.com.

Photos intérieures :

p.8 : ESB Professional/Shutterstock.com
p.11 : Ruslana Iurchenko/Shutterstock.com
p.12 : metamorworks/Shutterstock.com
p.14/15 : Alexander Supertramp/Shutterstock.com
p.16 : Wazzkii/Shutterstock.com
p.19 : Mukesh Kumar Jwala/Shutterstock.com
p.21 : supparsorn/Shutterstock.com
p.24 : everything possible/Shutterstock.com
p.29 : Gorodenkoff/Shutterstock.com
p.32 : only_kim/Shutterstock.com
p.34 : SeventyFour/Shutterstock.com
p.35 : ESB Professional/Shutterstock.com
p.37 : KlingSup/Shutterstock.com
p.39 : Miriam Doerr Martin Frommherz/Shutterstock.com

Création graphique : Aurélia Mazoyer

Imprimé par l'UNESCO

Imprimé en France

CLD 1517_21

Recommandation sur

l'éthique de l'intelligence artificielle

Adoptée le 23 novembre 2021

Table des matières

Préambule	5
I. Champ d'application	9
II. Buts et objectifs	13
III. Valeurs et principes	17
III.1 VALEURS	18
III.2 PRINCIPES	20
IV. Domaines d'action stratégique	25
DOMAINE STRATÉGIQUE 1 : ÉVALUATIONS DE L'IMPACT ÉTHIQUE	26
DOMAINE STRATÉGIQUE 2 : GOUVERNANCE ET GESTION ÉTHIQUES	27
DOMAINE STRATÉGIQUE 3 : POLITIQUES EN MATIÈRE DE DONNÉES	29
DOMAINE STRATÉGIQUE 4 : DÉVELOPPEMENT ET COOPÉRATION INTERNATIONALE	30
DOMAINE STRATÉGIQUE 5 : ENVIRONNEMENT ET ÉCOSYSTÈMES	32
DOMAINE STRATÉGIQUE 6 : ÉGALITÉ DES GENRES	32
DOMAINE STRATÉGIQUE 7 : CULTURE	34
DOMAINE STRATÉGIQUE 8 : ÉDUCATION ET RECHERCHE	35
DOMAINE STRATÉGIQUE 9 : COMMUNICATION ET INFORMATION	37
DOMAINE STRATÉGIQUE 10 : ÉCONOMIE ET TRAVAIL	37
DOMAINE STRATÉGIQUE 11 : SANTÉ ET BIEN-ÊTRE SOCIAL	38
V. Suivi et évaluation	40
VI. Utilisation de la présente Recommandation	42
VII. Promotion de la présente Recommandation	43
VIII. Dispositions finales	44

Préambule

La Conférence générale de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO), réunie à Paris du 9 au 24 novembre 2021, à l'occasion de sa 41e session,

Consciente des répercussions positives et négatives profondes et dynamiques de l'intelligence artificielle (IA) sur les sociétés, l'environnement, les écosystèmes et la vie humaine, y compris l'esprit humain, en raison notamment des nouvelles façons dont son utilisation agit sur la pensée, les interactions et la prise de décisions des êtres humains et dont elle retentit sur l'éducation, les sciences sociales et humaines et les sciences exactes et naturelles, la culture et la communication et l'information,

Rappelant que l'UNESCO se propose, aux termes de son Acte constitutif, de contribuer à la paix et à la sécurité et de resserrer, par l'éducation, les sciences, la culture et la communication et l'information, la collaboration entre nations, afin d'assurer le respect universel de la justice, de la loi, des droits de l'homme et des libertés fondamentales reconnus à tous les peuples,

Convaincue que la Recommandation présentée ici, en tant qu'instrument normatif, développée selon une approche globale, fondée sur le droit international et centrée sur la dignité humaine et les droits de l'homme, ainsi que sur l'égalité des genres, la justice sociale et économique et le développement, le bien-être physique et mental, la diversité, l'interdépendance, l'inclusion et la protection de l'environnement et des écosystèmes, peut donner une orientation responsable aux technologies de l'IA,

Guidée par les objectifs et les principes de la Charte des Nations Unies,

Considérant que les technologies de l'IA peuvent rendre de grands services à l'humanité et que tous les pays peuvent en bénéficier, mais qu'elles soulèvent également des préoccupations éthiques de fond, à l'égard, par exemple, des biais qu'elles sont susceptibles de comporter et d'accentuer, lesquels pourraient entraîner discrimination, inégalité, fractures numériques et exclusion, menacer la diversité culturelle, sociale et biologique et entraîner des clivages sociaux ou économiques ; la nécessité d'assurer la transparence et l'intelligibilité du fonctionnement des algorithmes et des données à partir desquelles ils ont été entraînés ; et leurs éventuelles conséquences sur, entre

autres, la dignité humaine, les droits de l'homme et les libertés fondamentales, l'égalité des genres, la démocratie, les processus sociaux, économiques, politiques et culturels, les pratiques scientifiques et d'ingénierie, le bien-être animal, et l'environnement et les écosystèmes,

Considérant également que les technologies de l'IA peuvent creuser les écarts et les inégalités qui existent dans le monde, au sein des pays et entre eux, et qu'il faut maintenir la justice, la confiance et l'équité pour qu'aucun pays ni aucune personne ne soient laissés de côté, qu'il s'agisse de jouir d'un accès équitable aux technologies de l'IA et de profiter de leurs avantages ou de se prémunir contre leurs conséquences négatives, tout en reconnaissant les différences de situation qui prévalent entre les pays et en respectant le souhait de certaines personnes de ne pas prendre part à toutes les innovations technologiques,

Consciente également que tous les pays connaissent une accélération dans l'utilisation des technologies de l'information et de la communication et des technologies de l'IA, ainsi qu'un besoin croissant en matière d'éducation aux médias et à l'information, et que l'économie numérique représente des défis et des possibilités immenses aux niveaux sociétal, économique et environnemental en termes de partage des bienfaits, en particulier pour les pays à revenu faible et intermédiaire, y compris, sans s'y limiter, les pays les moins avancés (PMA), les pays en développement sans littoral (PDSL) et les petits États insulaires en développement (PEID), ce qui exige de prendre en compte, de protéger et de promouvoir les cultures, valeurs et connaissances endogènes pour développer des économies numériques durables,

Reconnaissant que les technologies de l'IA peuvent avoir des effets bénéfiques sur l'environnement et les écosystèmes, et que la matérialisation de ces avantages exige de ne pas ignorer mais au contraire de tenir compte des dommages et des impacts négatifs qu'elles peuvent avoir sur l'environnement et les écosystèmes,

Notant que la prise en compte des risques et des préoccupations éthiques ne devrait pas freiner l'innovation et le développement, mais plutôt offrir de nouvelles possibilités et encourager une recherche et une innovation menées de manière éthique plaçant les droits de l'homme et les libertés

fondamentales, les valeurs, les principes et la réflexion morale et éthique au fondement des technologies de l'IA,

Rappelant également que la Conférence générale de l'UNESCO, à sa 40e session en novembre 2019, a adopté la résolution 40 C/37, par laquelle elle a chargé la Directrice générale « d'élaborer un instrument normatif international sur l'éthique de l'intelligence artificielle, sous la forme d'une recommandation », qui doit lui être présenté à sa 41e session en 2021,

Reconnaissant également que l'essor des technologies de l'IA nécessite de développer l'éducation à l'utilisation des données, aux médias et à l'information en conséquence, de même que l'accès à des sources d'information indépendantes, pluralistes et fiables, y compris dans le cadre des efforts visant à atténuer le risque de mésinformation, de désinformation et de diffusion de discours de haine, ainsi que les préjudices causés par une utilisation abusive des données personnelles,

Constatant que le cadre normatif applicable aux technologies de l'IA et à ses implications sociales trouve son fondement dans les cadres juridiques internationaux et nationaux, les droits de l'homme et les libertés fondamentales, l'éthique, la nécessité d'accéder aux données, aux informations et aux connaissances, la liberté de recherche et d'innovation, le bien-être de l'humanité, de l'environnement et des écosystèmes, et établit un lien entre les valeurs et principes éthiques et les défis et possibilités liés aux technologies de l'IA, sur la base d'une compréhension commune et d'objectifs partagés,

Reconnaissant en outre que les valeurs et principes éthiques peuvent aider à élaborer et à appliquer des mesures politiques et des normes juridiques fondées sur les droits, en fournissant des orientations en vue du développement rapide des technologies,

Convaincue également que des normes éthiques mondialement reconnues pour les technologies de l'IA, qui respectent pleinement le droit international, en particulier le droit relatif aux droits de l'homme, peuvent jouer un rôle essentiel dans l'élaboration de normes relatives à l'IA dans le monde,

Ayant à l'esprit la Déclaration universelle des droits de l'homme (1948); les instruments internationaux relatifs aux droits de l'homme, dont la Convention relative au statut des réfugiés (1951), la Convention concernant la discrimination (emploi et profession) (1958), la Convention internationale sur l'élimination de toutes les formes de discrimination raciale (1965), le Pacte international relatif aux droits civils et politiques (1966), le Pacte international relatif aux droits économiques, sociaux et culturels (1966), la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (1979), la Convention relative aux droits de l'enfant (1989), la Convention relative aux droits des personnes handicapées (2006), la Convention concernant la lutte contre la discrimination dans le domaine de l'enseignement (1960), et la Convention sur la protection et la promotion de la diversité des expressions culturelles (2005), ainsi que tous autres instruments, recommandations et déclarations internationaux pertinents,

Prenant acte de la Déclaration des Nations Unies sur le droit au développement (1986) ; de la Déclaration sur les responsabilités des générations présentes envers les générations futures (1997) ; de la Déclaration universelle sur la bioéthique et les droits de l'homme (2005) ; de la Déclaration des Nations Unies sur les droits des peuples autochtones (2007) ; de la résolution de l'Assemblée générale des Nations Unies sur l'examen des textes issus du Sommet mondial sur la société de l'information (A/RES/70/125) (2015) ; de la résolution de l'Assemblée générale des Nations Unies « Transformer notre monde : le Programme de développement durable à l'horizon 2030 » (A/RES/70/1) (2015) ; de la Recommandation concernant la préservation et l'accessibilité du patrimoine documentaire, y compris le patrimoine numérique (2015) ; de la Déclaration de principes éthiques en rapport avec le changement climatique (2017) ; de la Recommandation concernant la science et les chercheurs scientifiques (2017) ; des indicateurs sur l'universalité de l'Internet (approuvés par le Programme international de l'UNESCO pour le développement de la communication en 2018), incluant les principes ROAM (approuvés par la Conférence générale de l'UNESCO en 2015) ; de la résolution du Conseil des droits de l'homme sur « Le droit à la vie privée à l'ère du numérique » (A/HRC/RES/42/15) (2019), et de la résolution du Conseil des droits de l'homme

« Nouvelles technologies numériques et droits de l'homme » (A/HRC/RES/41/11) (2019),

Soulignant qu'il est nécessaire de prêter une attention particulière aux pays à revenu faible et intermédiaire, y compris, sans s'y limiter, aux pays les moins avancés (PMA), aux pays en développement sans littoral (PDSL) et aux petits États insulaires en développement (PEID), car ils disposent de capacités mais sont sous-représentés dans le débat sur l'éthique de l'IA, ce qui soulève des préoccupations quant à une prise en compte insuffisante des savoirs locaux, du pluralisme culturel, des systèmes de valeurs et des exigences d'équité mondiale dans la gestion des incidences positives et négatives des technologies de l'IA,

Consciente en outre qu'il existe de nombreuses politiques à l'échelle nationale ainsi que d'autres cadres et initiatives élaborés par les entités des Nations Unies concernées, par des organisations intergouvernementales, y compris des organisations régionales, ainsi que celles du secteur privé, des organisations professionnelles, des organisations non gouvernementales et de la communauté scientifique, relatifs à l'éthique et à la réglementation des technologies de l'IA,

Convaincue en outre que les technologies de l'IA peuvent être porteuses d'importants avantages, mais que leur matérialisation pourrait amplifier les tensions en lien avec l'innovation, l'accès asymétrique aux connaissances et aux technologies, notamment l'insuffisance de l'éducation civique et de la culture numérique qui limite la capacité du public à participer aux thèmes relatifs à l'IA, mais aussi les limitations à l'accès à l'information et les écarts en matière de capacités humaines et institutionnelles, les restrictions de l'accès à l'innovation technologique, et le manque d'infrastructures physiques et numériques et de cadres réglementaires adéquats, notamment ceux relatifs aux données, autant de problèmes auxquels il convient de remédier,

Soulignant également que le renforcement de la coopération et de la solidarité mondiales, y compris par le biais du multilatéralisme, est nécessaire pour favoriser un accès équitable aux technologies de l'IA et aborder les défis qu'elles posent dans des cultures et systèmes éthiques divers et interdépendants, pour réduire les risques d'utilisation abusive, pour exploiter pleinement les possibilités offertes

par l'IA, en particulier dans le domaine du développement, et pour garantir que les stratégies nationales relatives à l'IA sont guidées par des principes éthiques,

Tenant pleinement compte du fait que le développement rapide des technologies de l'IA rend difficiles une mise en œuvre et une gouvernance éthiques, ainsi que le respect et la protection de la diversité culturelle, et peut perturber les normes et valeurs éthiques locales et régionales,

1. **Adopte**, en ce vingt-troisième jour de novembre 2021, la présente Recommandation sur l'éthique de l'intelligence artificielle ;
2. **Recommande** aux États membres d'appliquer, à titre volontaire, les dispositions de la présente Recommandation en prenant des mesures appropriées, notamment législatives, conformes aux pratiques constitutionnelles et aux structures de gouvernance de chaque État, en vue de donner effet, dans leurs juridictions, aux principes et normes énoncés dans la Recommandation conformément au droit international, y compris le droit international des droits de l'homme ;
3. **Recommande également** aux États membres d'impliquer toutes les parties prenantes, y compris les entreprises, afin de veiller à ce qu'elles jouent leurs rôles respectifs dans la mise en œuvre de la Recommandation, et de porter la présente Recommandation à la connaissance des autorités, organismes, instituts de recherche et organismes universitaires, et institutions et organisations des secteurs public, privé et de la société civile engagés dans les technologies de l'IA, afin que le développement et l'utilisation des technologies de l'IA soient guidés par des recherches scientifiques de qualité ainsi que par une analyse et une évaluation éthiques.



I.

Champ d'application

1. La présente Recommandation traite des questions éthiques soulevées dans le domaine de l'intelligence artificielle, dans la mesure où celles-ci relèvent du mandat de l'UNESCO. Elle aborde l'éthique de l'IA en tant que réflexion normative systématique, basée sur un cadre global, détaillé, multiculturel et évolutif de valeurs, de principes et d'actions interdépendants de nature à orienter les sociétés pour qu'elles apportent des réponses responsables aux conséquences connues et inconnues des impacts des technologies de l'IA sur les êtres humains, les sociétés, l'environnement et les écosystèmes, et leur offre une base pour accepter ou rejeter les technologies de l'IA. Elle considère l'éthique comme une base dynamique pour l'évaluation et l'orientation normative des technologies de l'IA, en faisant de la dignité humaine, du bien-être et de la prévention des dommages des repères et en s'appuyant sur l'éthique des sciences et des technologies.

2. La présente Recommandation ne cherche pas à donner de définition unique de l'IA, celle-ci étant appelée à évoluer en fonction des progrès technologiques. Son objectif est plutôt de traiter des caractéristiques des systèmes d'IA qui revêtent une importance majeure sur le plan éthique. En conséquence, la présente Recommandation envisage les systèmes d'IA comme des systèmes capables de traiter les données et l'information par un processus s'apparentant à un comportement intelligent, et comportant généralement des fonctions de raisonnement, d'apprentissage, de perception, d'anticipation, de planification ou de contrôle. Trois éléments occupent une place centrale dans cette approche :

- (a) les systèmes d'IA sont des technologies de traitement des informations qui intègrent des modèles et des algorithmes, lesquels génèrent une capacité d'apprentissage et d'exécution de tâches cognitives conduisant à des résultats tels que l'anticipation et la prise de décisions dans des environnements matériels et virtuels. Les systèmes d'IA sont conçus pour fonctionner avec différents degrés d'autonomie, au moyen de la modélisation et la représentation des connaissances, de l'exploitation des données et du calcul de corrélations. Ils peuvent intégrer plusieurs méthodes, telles que, sans s'y limiter :
 - (i) l'apprentissage automatique, y compris l'apprentissage profond et l'apprentissage par renforcement ;
 - (ii) le raisonnement automatique, y compris la planification, la programmation, la représentation des connaissances et le raisonnement, et la recherche et l'optimisation.

Les systèmes d'IA peuvent être utilisés dans des systèmes cyberphysiques, y compris l'Internet des objets, les systèmes robotiques, la robotique sociale et les interfaces homme-machine qui impliquent

des fonctions de contrôle, de perception et de traitement de données recueillies par des senseurs, et le fonctionnement d'actionneurs dans l'environnement du système d'IA ;

- (b) les questions éthiques concernant les systèmes d'IA se rapportent à toutes les étapes du cycle de vie de ces systèmes, compris ici comme allant de la recherche, la conception et le développement au déploiement et à l'utilisation, et incluant la maintenance, l'exploitation, la commercialisation, le financement, le suivi et l'évaluation, la validation, la fin de l'utilisation, le démontage et la mise hors service. En outre, les acteurs de l'IA peuvent désigner tout acteur impliqué dans au moins une étape du cycle de vie des systèmes d'IA, et peuvent renvoyer à des personnes physiques comme à des personnes morales, telles que les chercheurs, les programmeurs, les ingénieurs, les spécialistes des données, les utilisateurs finaux, les entreprises, les universités et les entités publiques et privées, entre autres ;
- (c) les systèmes d'IA soulèvent de nouveaux types de questions éthiques qui comprennent, sans s'y limiter, leur impact sur la prise de décisions, l'emploi et le travail, les interactions sociales, les soins de santé, l'éducation, les médias, l'accès à l'information, la fracture numérique, la protection des consommateurs et des données personnelles, l'environnement, la démocratie, l'état de droit, la sécurité et la police, le double usage et les droits de l'homme et les libertés fondamentales, y compris la liberté d'expression, la protection de la vie privée et la non-discrimination. En outre, de nouveaux défis éthiques sont générés par la reproduction et le renforcement potentiels des biais existants par les algorithmes d'IA, ce qui exacerbe des formes déjà existantes de discriminations, de préjugés et de stéréotypes. Certaines de ces questions ont trait à la capacité des systèmes d'IA d'effectuer des tâches qu'auparavant seuls des êtres vivants pouvaient réaliser, parfois même uniquement des êtres humains. Ces caractéristiques confèrent aux systèmes d'IA un rôle déterminant et inédit dans les pratiques et les sociétés humaines, ainsi que dans leur relation avec l'environnement et les écosystèmes, créant un nouveau contexte dans lequel les enfants et les jeunes vont grandir, développer une compréhension du monde et d'eux-mêmes, comprendre de manière critique les médias et l'information, et apprendre à prendre des décisions. Les systèmes d'IA pourraient, à long terme, disputer aux êtres humains le sentiment d'expérience et la capacité d'agir qui leur sont propres, ce qui susciterait de nouvelles inquiétudes quant à la compréhension qu'ont les êtres humains d'eux-mêmes, leurs interactions sociales, culturelles et environnementales, leur autonomie, leur capacité d'agir, leur valeur et leur dignité, entre autres.



3. La présente Recommandation prête une attention particulière aux implications éthiques plus larges des systèmes d'IA dans les domaines centraux de l'UNESCO : éducation, science, culture et communication et information, lesquelles font l'objet de l'étude préliminaire sur l'éthique de l'intelligence artificielle réalisée par la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST) de l'UNESCO en 2019 :

- (a) l'éducation, parce que vivre dans des sociétés en cours de numérisation exige de nouvelles pratiques éducatives, une réflexion éthique, une pensée critique, des pratiques de conception responsables et de nouvelles compétences, au vu des conséquences sur le marché de l'emploi, l'employabilité et la participation citoyenne ;
- (b) la science, au sens le plus large et incluant tous les domaines d'étude, des sciences exactes et naturelles et des sciences médicales aux sciences sociales et humaines, car les technologies de l'IA apportent de nouvelles capacités et approches de recherche, influencent nos conceptions de la compréhension et de l'explication scientifiques, et créent une nouvelle base pour la prise de décisions ;
- (c) l'identité et la diversité culturelles, car les technologies de l'IA peuvent enrichir les industries culturelles et créatives, mais aussi aboutir à une concentration accrue de l'offre, des données, des marchés et des revenus de la culture entre les mains

d'un petit nombre d'acteurs, avec des répercussions potentiellement négatives sur la diversité et le pluralisme des langues, des médias, des expressions culturelles, la participation et l'égalité ;

- (d) la communication et l'information, car les technologies de l'IA jouent un rôle croissant dans le traitement, la structuration et la transmission de l'information ; les questions du journalisme automatisé et de l'utilisation d'algorithmes pour diffuser des actualités et pour modérer et organiser des contenus sur les réseaux sociaux et les moteurs de recherche ne sont que quelques exemples de problématiques liées à l'accès à l'information, à la désinformation, à la mésinformation, au discours de haine, à l'émergence de nouvelles formes de récits sociétaux, à la discrimination, à la liberté d'expression, à la protection de la vie privée et à l'éducation aux médias et à l'information, entre autres.

4. La présente Recommandation s'adresse aux États membres, à la fois en tant qu'acteurs de l'IA et en tant qu'autorités responsables de l'élaboration de cadres juridiques et réglementaires tout au long du cycle de vie des systèmes d'IA, et de la promotion de la responsabilité des entreprises. Elle fournit également des orientations éthiques à tous les acteurs de l'IA, y compris des secteurs public et privé, en constituant une base pour une évaluation de l'impact éthique des systèmes d'IA tout au long de leur cycle de vie.



II.

Buts et objectifs



5. La présente Recommandation a pour objet de servir de base afin de mettre les systèmes d'IA au service de l'humanité, des individus, des sociétés, de l'environnement et des écosystèmes, ainsi que de prévenir les préjudices. Elle a également pour vocation de favoriser l'utilisation pacifique des systèmes d'IA.
6. En plus des cadres éthiques concernant l'IA déjà établis à travers le monde, la présente Recommandation vise à fournir un instrument normatif accepté au niveau mondial qui se concentre non seulement sur la définition des valeurs et des principes, mais aussi sur leur application pratique, par le biais de recommandations stratégiques concrètes, mettant fortement l'accent sur l'inclusion, sur les questions d'égalité des genres et sur la protection de l'environnement et des écosystèmes.
7. Parce que la complexité des questions éthiques qui entourent l'IA exige la coopération des multiples parties prenantes aux différents niveaux et dans les différents secteurs des communautés internationales, régionales et nationales, la présente Recommandation vise à permettre aux parties prenantes d'assumer leur part de responsabilité par le biais d'un dialogue mondial et interculturel.
8. Les objectifs de la présente Recommandation sont les suivants :
 - (a) offrir un cadre universel de valeurs, de principes et d'actions pour guider les États dans la formulation de leur législation, de leurs politiques ou d'autres instruments concernant l'IA, conformément au droit international ;



- (b) guider les actions des individus, des groupes, des communautés, des institutions et des entreprises du secteur privé afin de garantir la prise en compte de l'éthique à tous les stades du cycle de vie des systèmes d'IA ;
- (c) protéger, promouvoir et respecter les droits de l'homme et les libertés fondamentales, la dignité humaine et l'équité, y compris l'égalité des genres ; protéger les intérêts des générations présentes et futures ; préserver l'environnement, la biodiversité et les écosystèmes ; et respecter la diversité culturelle à tous les stades du cycle de vie des systèmes d'IA ;
- (d) favoriser un dialogue multipartite, pluridisciplinaire et pluraliste ainsi que la recherche du consensus au sujet des questions éthiques en lien avec les systèmes d'IA ;
- (e) promouvoir un accès équitable aux progrès et aux connaissances dans le domaine de l'IA, ainsi que le partage des bienfaits qui en découlent, en accordant une attention particulière aux besoins et contributions des pays à revenu intermédiaire inférieur (PRITI), notamment aux PMA, aux PDSL et aux PEID.



III.

Valeurs et principes

9. Les valeurs et principes énoncés ci-dessous doivent être respectés par tous les acteurs du cycle de vie des systèmes d'IA, en premier lieu, et, s'il y a lieu et selon les besoins, être promus par l'élaboration de nouvelles législations, réglementations et directives commerciales et la modification de celles qui existent déjà. Cette démarche doit respecter le droit international, notamment la Charte des Nations Unies, y compris les obligations des États membres en matière de droits de l'homme, et doit s'inscrire dans le cadre d'objectifs de durabilité sociale, politique, environnementale, éducative, scientifique et économique convenus au niveau international, tels que les objectifs de développement durable (ODD) des Nations Unies.
10. Les valeurs ont un rôle important en tant qu'idéaux motivant l'orientation des mesures politiques et des normes juridiques. Alors que l'ensemble des valeurs énoncées ci-dessous inspirent ainsi des comportements souhaitables et constituent les fondements des principes, les principes quant à eux explicitent les valeurs de manière plus concrète, de façon à faciliter l'application de ces dernières dans les déclarations et actions politiques.
11. Si toutes les valeurs et tous les principes décrits ci-dessous sont souhaitables en soi, dans tout contexte réel, ils peuvent parfois s'opposer entre eux. Quelle que soit la situation donnée, une évaluation du contexte sera nécessaire pour gérer ces oppositions éventuelles, en tenant compte du principe de proportionnalité et en respectant les droits de l'homme et les libertés

fondamentales. Dans tous les cas, toute restriction éventuelle imposée à l'exercice des droits de l'homme et des libertés fondamentales doit reposer sur une base juridique, être raisonnable, nécessaire et proportionnée, et conforme aux obligations qui incombent aux États membres en vertu du droit international. Pour faire des choix judicieux parmi de tels scénarios, il faudra généralement s'engager avec un large éventail de parties prenantes compétentes, en recourant au dialogue social, aux débats sur l'éthique, à la diligence requise et à l'évaluation de l'impact.

12. La fiabilité et l'intégrité du cycle de vie des systèmes d'IA sont essentielles pour faire en sorte que les technologies de l'IA soient mises au service de l'humanité, des individus, des sociétés et de l'environnement et des écosystèmes, et concrétisent les valeurs et principes énoncés dans la présente Recommandation. Les individus doivent avoir de bonnes raisons de ne pas douter de la capacité des systèmes d'IA de produire des avantages sur les plans individuel et collectif lorsque des mesures appropriées sont prises pour atténuer les risques. Une condition essentielle de la fiabilité consiste à soumettre, tout au long de leur cycle de vie, les systèmes d'IA à un suivi approfondi exercé par les parties prenantes compétentes, selon qu'il conviendra. La fiabilité résultant de l'application des principes énoncés dans le présent document, les actions stratégiques proposées dans la Recommandation visent toutes à promouvoir la fiabilité à tous les stades du cycle de vie des systèmes d'IA.

III.1 VALEURS

Respect, protection et promotion des droits de l'homme, des libertés fondamentales et de la dignité humaine

13. La dignité inviolable et intrinsèque de chaque être humain est le fondement du système universel, indivisible, inaliénable, interdépendant et indissociable des droits de l'homme et des libertés fondamentales. Par conséquent, le respect, la protection et la promotion de la dignité humaine et des droits établis par le droit international, notamment le droit international relatif aux droits de l'homme, tiennent une place essentielle tout au long du cycle de vie des systèmes d'IA. La dignité humaine a trait à la reconnaissance de la valeur intrinsèque et égale de chaque être humain, indépendamment de sa race, de sa couleur, de son ascendance, de son genre, de son âge, de sa langue, de sa religion, de ses opinions politiques, de son origine nationale, ethnique ou sociale, de sa condition économique ou sociale de naissance, de son handicap ou de tout autre motif.

14. Aucun être humain ni aucune communauté humaine ne doivent subir de préjudice ou d'assujettissement physique, économique, social, politique, culturel ou mental au cours d'une phase quelconque du cycle de vie des systèmes d'IA. Les systèmes d'IA doivent, tout au long de leur cycle de vie, améliorer la qualité de vie des êtres humains, en laissant aux individus ou groupes le soin de définir la notion de « qualité de vie », tant qu'il n'y a ni violation des droits de l'homme et des libertés fondamentales, ni atteinte à la dignité humaine au sens de cette définition.

15. Les individus peuvent interagir avec les systèmes d'IA tout au long de leur cycle de vie et recevoir leur aide, par exemple pour les soins aux personnes vulnérables ou en situation de vulnérabilité, incluant, mais sans s'y limiter, les enfants, les personnes âgées, les personnes handicapées ou les malades. Dans le cadre de ces interactions, les individus ne doivent jamais être traités comme des objets, leur dignité ne doit pas être mise à mal d'une quelconque autre manière, et il ne doit y avoir ni violation des droits de l'homme et des libertés fondamentales ni atteinte à ces droits et libertés.

- 16.** Les droits de l'homme et les libertés fondamentales doivent être respectés, protégés et promus tout au long du cycle de vie des systèmes d'IA. Les autorités, le secteur privé, la société civile, les organisations internationales, les communautés techniques et les universités doivent respecter les instruments et cadres relatifs aux droits de l'homme lorsqu'ils interviennent dans les processus entourant le cycle de vie des systèmes d'IA. Les nouvelles technologies doivent fournir de nouveaux moyens de promouvoir, de défendre et d'exercer les droits de l'homme, et non de les enfreindre.

Un environnement et des écosystèmes qui prospèrent

- 17.** Tout au long de leur cycle de vie, les systèmes d'IA doivent reconnaître et encourager l'importance d'avoir un environnement et des écosystèmes qui prospèrent, et protéger ces derniers. En outre, l'environnement et les écosystèmes sont une nécessité existentielle pour que l'humanité et les autres êtres vivants puissent profiter des bienfaits liés aux progrès de l'IA.
- 18.** Tous les acteurs impliqués dans le cycle de vie des systèmes d'IA doivent respecter le droit international et les lois, normes et pratiques nationales applicables, telles que le principe de précaution, conçues pour la protection et la restauration de l'environnement et des écosystèmes, et pour le développement durable. Ils devraient réduire l'impact environnemental des systèmes d'IA, ce qui inclut, sans s'y limiter, leur empreinte

carbone, afin de réduire autant que possible les facteurs de risque associés au changement climatique et aux changements environnementaux, et d'empêcher l'exploitation, l'utilisation et la transformation non durables des ressources naturelles, qui contribuent à la détérioration de l'environnement et à la dégradation des écosystèmes.

Assurer la diversité et l'inclusion

- 19.** Il convient d'assurer le respect, la protection et la promotion de la diversité et de l'inclusion tout au long du cycle de vie des systèmes d'IA, en se conformant au droit international, y compris le droit international des droits de l'homme. Cela peut passer par le fait d'encourager la participation active de tous les individus ou groupes indépendamment de leur race, de leur couleur, de leur ascendance, de leur genre, de leur âge, de leur langue, de leur religion, de leurs opinions politiques, de leur origine nationale, ethnique ou sociale, de leur condition économique ou sociale de naissance, de leur handicap ou de tout autre motif.
- 20.** La diversité des choix de style de vie, des croyances, des opinions, des expressions ou des expériences personnelles, ce qui inclut le caractère facultatif de l'utilisation des systèmes d'IA et la coconception de ces architectures, ne doit être restreinte à aucun stade du cycle de vie des systèmes d'IA.
- 21.** Il faudra en outre s'efforcer, notamment par la coopération internationale, de pallier l'absence des infrastructures,



formations et compétences technologiques ainsi que des cadres juridiques nécessaires dont souffrent certaines communautés, en particulier dans les PRITI, les PMA, les PDSL et les PEID, mais jamais d'en profiter.

Vivre dans des sociétés pacifiques, justes et interdépendantes

- 22.** Les acteurs de l'IA doivent favoriser, en y participant, l'édification de sociétés pacifiques et justes fondées sur un avenir interconnecté pour le bien de tous, en respectant les droits de l'homme et les libertés fondamentales. La vie dans des sociétés pacifiques et justes, en tant que valeur, suppose que les systèmes d'IA peuvent contribuer tout au long de leur cycle de vie à l'interdépendance de toutes les créatures vivantes entre elles et avec le milieu naturel.
- 23.** La notion d'interdépendance des êtres humains repose sur le fait que chaque être humain appartient à un

ensemble plus vaste, qui prospère lorsque tous ses membres peuvent s'épanouir. Vivre dans des sociétés pacifiques, justes et interdépendantes nécessite un lien de solidarité organique, immédiat et non calculé, caractérisé par une recherche permanente de relations pacifiques, tendant vers le souci du bien-être des autres et du milieu naturel au sens large du terme.

- 24.** Pour assurer le respect de cette valeur, la paix, l'inclusion, la justice, l'équité et l'interdépendance doivent être promues tout au long du cycle de vie des systèmes d'IA, dans la mesure où les processus du cycle de vie des systèmes d'IA ne devraient pas mettre à l'écart les êtres humains et les communautés, les chosifier ou menacer leur liberté, leur autonomie de décision et leur sécurité, diviser les individus et les groupes et les dresser les uns contre les autres, ou compromettre la coexistence entre les êtres humains, les autres êtres vivants et le milieu naturel.

III.2 PRINCIPES

Principes de proportionnalité et d'innocuité

- 25.** Il faut reconnaître que les technologies de l'IA ne garantissent pas nécessairement, en soi, l'épanouissement des êtres humains ni le maintien d'un environnement et d'écosystèmes prospères. Qui plus est, aucun des processus liés au cycle de vie des systèmes d'IA ne devrait aller au-delà de ce qui est nécessaire pour atteindre des buts ou objectifs légitimes, et ces processus devraient être adaptés au contexte dans lequel ils interviennent. S'il s'avérait qu'un préjudice quelconque puisse affecter les êtres humains, les droits de l'homme et les libertés fondamentales, les communautés et la société en général, ou l'environnement et les écosystèmes, il conviendrait d'assurer la mise en œuvre de procédures d'évaluation des risques et l'adoption de mesures visant à empêcher qu'un tel préjudice ne survienne.
- 26.** La décision de recourir à des systèmes d'IA et le choix de la méthode d'IA devraient être justifiés des manières suivantes : (a) la méthode d'IA retenue devrait être adaptée et proportionnée pour atteindre un objectif légitime ; (b) la méthode d'IA retenue ne devrait pas porter atteinte aux valeurs fondamentales énoncées dans la présente Recommandation – en particulier, son utilisation ne devrait pas constituer une violation ou un abus des droits de l'homme ; et (c) la méthode d'IA retenue devrait être adaptée au contexte et reposer sur des bases scientifiques rigoureuses. Dans les scénarios où les décisions sont considérées comme ayant un impact irréversible ou difficile à renverser, ou qui pourraient impliquer des décisions de vie et de mort, la décision finale devrait être prise par l'homme. En particulier, les systèmes d'IA ne devraient pas être utilisés à des fins de notation sociale ou de surveillance de masse.

Sûreté et sécurité

- 27.** Les préjudices non désirés (risques liés à la sûreté), ainsi que les vulnérabilités aux attaques (risques liés à la sécurité), devraient être évités et il convient de les prendre en compte, de les prévenir et de les éliminer tout au long du cycle de vie des systèmes d'IA afin de garantir la sûreté et la sécurité des personnes, de l'environnement et des écosystèmes. La sûreté et la sécurité de l'IA seront rendues possibles par l'élaboration de cadres d'accès aux données, durables et respectueux de la vie privée, qui favorisent un meilleur entraînement et une meilleure validation des modèles d'IA utilisant des données de qualité.

Équité et non-discrimination

- 28.** Les acteurs de l'IA doivent promouvoir la justice sociale, garantir l'équité et lutter contre les discriminations en tous genres, conformément au droit international. Cela suppose d'adopter une approche inclusive pour s'assurer que les bénéfices des technologies de l'IA sont disponibles et accessibles à tous, en prenant en considération les besoins spécifiques des différents groupes d'âge, des systèmes culturels, des différents groupes linguistiques, des personnes handicapées, des filles et des femmes, ainsi que des personnes défavorisées, marginalisées et vulnérables ou en situation de vulnérabilité. Les États membres devraient s'efforcer de promouvoir un accès inclusif pour tous, y compris pour les communautés locales, à des systèmes d'IA offrant des contenus et des services adaptés localement, et respectant le multilinguisme et la diversité culturelle. Les États membres devraient s'efforcer de réduire les



fractures numériques et de garantir un accès et une participation inclusifs au développement de l'IA. Au niveau national, les États membres devraient promouvoir l'équité entre les zones rurales et urbaines, et entre toutes les personnes indépendamment de leur race, de leur couleur, de leur ascendance, de leur genre, de leur âge, de leur langue, de leur religion, de leurs opinions politiques, de leur origine nationale, ethnique ou sociale, de leur condition économique ou sociale de naissance, de leur handicap ou de tout autre motif en termes d'accès et de participation au cycle de vie des systèmes d'IA. Au niveau international, les pays les plus avancés sur le plan technologique ont un devoir de solidarité envers les pays les moins avancés afin de garantir le partage des bienfaits des technologies de l'IA de telle sorte que l'accès et la participation de ces derniers au cycle de vie des systèmes d'IA contribuent à un ordre mondial plus équitable en matière d'information, de communication, de culture, d'éducation, de recherche et de stabilité socioéconomique et politique.

- 29.** Les acteurs de l'IA doivent faire tout ce qui est raisonnablement possible pour réduire au maximum et éviter de renforcer ou de perpétuer des applications et des résultats discriminatoires ou biaisés, tout au long du cycle de vie des systèmes d'IA, afin de garantir l'équité de ces systèmes. Un recours efficace devrait être disponible contre la discrimination et la détermination algorithmique biaisée.

- 30.** En outre, les fractures numérique et cognitive, au sein des pays et entre eux, doivent être traitées tout au long du cycle de vie des systèmes d'IA, notamment en termes d'accès et de qualité d'accès aux technologies et aux données, conformément aux cadres juridiques nationaux, régionaux et internationaux pertinents, ainsi qu'en termes de connectivité, de connaissances et de compétences et de participation effective des communautés concernées, de sorte que chaque personne soit traitée équitablement.

Durabilité

- 31.** Le développement de sociétés durables repose sur la réalisation d'un ensemble complexe d'objectifs concernant divers aspects humains, sociaux, culturels, économiques et environnementaux. L'arrivée des technologies de l'IA peut soit favoriser les objectifs de durabilité, soit entraver leur réalisation, selon la manière dont elles sont appliquées dans des pays ayant des niveaux de développement différents. L'évaluation continue de l'impact humain, social, culturel, économique et environnemental des technologies de l'IA doit donc être effectuée en tenant pleinement compte des incidences de ces technologies pour la durabilité en tant qu'ensemble d'objectifs en constante évolution dans toute une série de dimensions, tels qu'ils sont actuellement définis dans les objectifs de développement durable (ODD) des Nations Unies..

Droit au respect de la vie privée et protection des données

- 32.** La vie privée, qui constitue un droit essentiel pour la protection de la dignité, de l'autonomie et de la capacité d'action des êtres humains, doit être respectée, protégée et promue tout au long du cycle de vie des systèmes d'IA. Il est important que les données pour les systèmes d'IA soient collectées, utilisées, partagées, archivées et supprimées d'une manière qui soit conforme au droit international et en accord avec les valeurs et principes énoncés dans la présente Recommandation, tout en respectant les cadres juridiques nationaux, régionaux et internationaux pertinents.
- 33.** Des cadres de protection des données et des mécanismes de gouvernance adéquats doivent être mis en place selon une approche multipartite au niveau national ou international, protégés par les systèmes judiciaires et appliqués tout au long du cycle de vie des systèmes d'IA. Les cadres de protection des données et tout mécanisme connexe devraient s'inspirer des principes et des normes internationaux de protection de données en ce qui concerne la collecte, l'utilisation et la divulgation des données personnelles et l'exercice de leurs droits par les personnes concernées, tout en assurant ainsi un objectif légitime et une base juridique valable pour le traitement des données personnelles, y compris le consentement éclairé.
- 34.** Les systèmes algorithmiques nécessitent des évaluations adéquates de l'impact sur la vie privée, qui incluent également des considérations sociétales et éthiques de leur utilisation et une approche innovante consistant à appliquer le respect de la vie privée dès la conception des systèmes. Les acteurs de l'IA doivent assumer la responsabilité de la conception et de la mise en œuvre des systèmes d'IA de manière à garantir que les informations personnelles sont protégées tout au long du cycle de vie du système d'IA.

Surveillance et décision humaines

- 35.** Les États membres devraient veiller à ce qu'il soit toujours possible d'attribuer la responsabilité éthique et juridique de tout stade du cycle de vie des systèmes d'IA à des personnes physiques ou des entités juridiques existantes, y compris dans les cas de recours relatifs à des systèmes d'IA. Ainsi, la surveillance humaine renvoie non seulement à la surveillance humaine individuelle, mais aussi à une surveillance publique inclusive, selon le cas.
- 36.** Les êtres humains choisiront peut-être, dans certains cas, de s'appuyer sur les systèmes d'IA à des fins d'efficacité, mais la décision de céder les fonctions de contrôle dans des contextes limités reste leur prérogative. En effet, les êtres humains peuvent recourir aux systèmes d'IA pour la prise de décisions et l'exécution de tâches, mais un système d'IA ne peut jamais se substituer à la responsabilité ultime des êtres humains et à leur obligation de rendre compte. De manière générale,

les décisions de vie et de mort ne devraient pas être abandonnées à des systèmes d'IA.

Transparence et explicabilité

- 37.** La transparence et l'explicabilité des systèmes d'IA sont souvent des conditions préalables essentielles pour garantir le respect, la protection et la promotion des droits humains, des libertés fondamentales et des principes éthiques. La transparence est nécessaire pour que les réglementations nationales et internationales en matière de responsabilité fonctionnent efficacement. Un manque de transparence peut aussi entraver la possibilité de contester efficacement les décisions fondées sur des résultats produits par les systèmes d'IA et porter ainsi atteinte au droit à un jugement équitable et à un recours effectif, et limite les domaines dans lesquels de tels systèmes peuvent être utilisés légalement.
- 38.** S'il convient de tout mettre en œuvre pour améliorer la transparence et l'explicabilité des systèmes d'IA – y compris ceux qui ont une portée extraterritoriale tout au long de leur cycle de vie – afin de favoriser une gouvernance démocratique, le degré de transparence et d'explicabilité devrait toujours être adapté au contexte et à la portée, car il peut être nécessaire de trouver un équilibre entre la transparence et l'explicabilité et d'autres principes, tels que la protection de la vie privée, la sûreté et la sécurité. Les individus devraient être pleinement informés lorsqu'une décision est fondée sur des algorithmes d'IA ou prise par ceux-ci, notamment lorsqu'elle affecte leur sécurité ou leurs droits humains. Dans ces circonstances, ils devraient avoir la possibilité d'exiger ou de demander des explications à l'acteur de l'IA ou aux institutions du secteur public concernés. En outre, les individus doivent pouvoir connaître les raisons qui ont conduit à une décision se répercutant sur leurs droits et leurs libertés, et avoir la possibilité de soumettre des observations à un membre du personnel de l'entreprise du secteur privé ou de l'institution du secteur public habilitée à réexaminer et corriger la décision. Les acteurs de l'IA devraient informer les utilisateurs lorsqu'un produit ou un service est fourni directement ou par le biais de systèmes d'intelligence artificielle de manière appropriée et en temps utile.
- 39.** D'un point de vue sociotechnique, une plus grande transparence contribue à des sociétés plus pacifiques, plus justes, plus démocratiques et plus inclusives. Elle permet un contrôle public qui peut réduire la corruption et la discrimination, et peut également aider à détecter et à prévenir les impacts négatifs sur les droits de l'homme. La transparence vise à fournir des informations appropriées aux destinataires respectifs afin de permettre la compréhension et de favoriser la confiance. Concernant spécifiquement les systèmes d'IA, la transparence peut permettre aux individus de comprendre comment chaque étape d'un système d'IA est mise en place, en fonction du contexte et du degré de sensibilité du système concerné. Il est possible également de fournir des informations sur les facteurs qui

influencent une prévision ou une décision particulière, et sur l'existence ou non des garanties appropriées (telles que des mesures de sécurité ou d'équité). Dans les cas de menaces sérieuses avec de graves répercussions sur les droits de l'homme, la transparence peut également exiger la communication de codes ou d'ensembles de données.

- 40.** L'explicabilité implique de rendre les résultats des systèmes d'IA intelligibles et de fournir des renseignements à leur sujet. L'explicabilité des systèmes d'IA renvoie également à l'intelligibilité des intrants, des extrants, du fonctionnement des différents modules algorithmiques et de leur contribution aux résultats des systèmes. L'explicabilité est donc étroitement liée à la transparence, puisqu'il convient de rendre les résultats et les sous-processus qui y conduisent intelligibles et traçables, en fonction du contexte. Les acteurs de l'IA devraient s'engager à assurer l'explicabilité des algorithmes mis au point. Dans le cas d'applications de l'IA dont l'impact sur l'utilisateur final n'est pas temporaire, facilement réversible ou à faible risque, il convient de s'assurer qu'une explication valable est fournie avec toute décision ayant entraîné l'action entreprise, afin que les résultats soient considérés comme transparents.
- 41.** La transparence et l'explicabilité sont étroitement liées à des mesures adéquates en matière de responsabilité et d'obligation de rendre compte, ainsi qu'à la fiabilité des systèmes d'IA.

Responsabilité et redevabilité

- 42.** Les acteurs de l'IA et les États membres devraient respecter, protéger et promouvoir les droits de l'homme et les libertés fondamentales, et devraient également favoriser la protection de l'environnement et des écosystèmes, endossant leur responsabilité éthique et juridique respective, conformément au droit national et international, en particulier les obligations des États membres en matière de droits de l'homme, et aux directives éthiques établies tout au long du cycle de vie des systèmes d'IA, y compris en ce qui concerne les acteurs de l'IA sur leur territoire et sous leur contrôle effectifs. La responsabilité éthique des décisions et actions fondées d'une quelconque manière sur un système d'IA devrait toujours incomber en dernier ressort aux acteurs de l'IA selon le rôle qui est le leur dans le cycle de vie du système d'IA.
- 43.** Des mécanismes appropriés de surveillance, d'évaluation de l'impact, de contrôle et de diligence requise, y compris concernant la protection des lanceurs d'alerte, devraient être mis en place pour assurer la redevabilité des systèmes d'IA et de leur impact tout au long de leur cycle de vie. Des dispositifs techniques et institutionnels devraient garantir la vérifiabilité et la traçabilité (du fonctionnement) des systèmes d'IA, en particulier pour résoudre tout conflit lié aux règles et normes relatives aux droits de l'homme et traiter les menaces pesant sur le bien-être de l'environnement et des écosystèmes.

Sensibilisation et éducation

- 44.** La sensibilisation du public et sa compréhension des technologies de l'IA et de la valeur des données devraient être favorisées par une éducation ouverte et accessible, l'engagement civique, l'acquisition de compétences numériques et la formation à l'éthique de l'IA, l'éducation aux médias et à l'information et la formation menée conjointement par les gouvernements, les organisations intergouvernementales, la société civile, les universités, les médias, les dirigeants locaux et le secteur privé, et en tenant compte de la diversité linguistique, sociale et culturelle existante, afin de garantir une participation publique efficace de sorte que tous les membres de la société puissent prendre des décisions éclairées concernant leur utilisation des systèmes d'IA et soient protégés contre toute influence indue.
- 45.** La sensibilisation à l'impact des systèmes d'IA doit inclure un apprentissage sur, par et pour les droits de l'homme et les libertés fondamentales, ce qui signifie que l'approche et la compréhension des systèmes d'IA doivent se baser sur l'impact de ces systèmes sur les droits de l'homme et sur l'accès à ces droits, ainsi que sur l'environnement et les écosystèmes.

Gouvernance et collaboration multipartites et adaptatives

- 46.** Le droit international et la souveraineté nationale doivent être respectés dans l'utilisation des données. Cela signifie que les États, conformément au droit international, peuvent réglementer les données générées sur leur territoire ou transitant par celui-ci, et prendre des mesures en vue d'une réglementation efficace des données, concernant notamment leur protection, qui soit fondée sur le respect du droit à la vie privée conformément au droit international et d'autres règles et normes relatives aux droits de l'homme.
- 47.** La participation des différentes parties prenantes tout au long du cycle de vie des systèmes d'IA est nécessaire pour garantir des approches inclusives de la gouvernance de l'IA, permettant que les bénéfices soient partagés par tous, et de contribuer au développement durable. Les parties prenantes comprennent, sans s'y limiter, les gouvernements, les organisations intergouvernementales, la communauté technique, la société civile, les chercheurs et les universitaires, les médias, les responsables de l'éducation, les décideurs, les entreprises du secteur privé, les institutions de défense des droits de l'homme et les organismes de promotion de l'égalité, les organismes de contrôle des mesures de lutte contre la discrimination et les groupes de jeunes et d'enfants. Il conviendrait d'adopter des normes ouvertes et d'assurer l'interopérabilité pour faciliter la collaboration. Des mesures devraient être adoptées pour tenir compte de l'évolution des technologies et de l'émergence de nouveaux groupes de parties prenantes, ainsi que pour permettre une participation efficace des groupes, communautés et individus marginalisés et, le cas échéant, dans le cas des peuples autochtones, le respect de leur autonomie dans la gestion de leurs données.



IV.

Domaines d'action stratégique

48. Les actions décrites dans les domaines stratégiques suivants traduisent dans les faits les valeurs et les principes énoncés dans la présente Recommandation. La principale action consiste pour les États membres à mettre en place des mesures efficaces, notamment des cadres ou des mécanismes stratégiques, et à veiller à ce que les autres parties prenantes, telles que les entreprises du secteur privé, les établissements universitaires, les instituts de recherche et la société civile, y adhèrent, notamment en encourageant toutes les parties prenantes à élaborer des outils d'évaluation de l'impact sur les droits de l'homme, l'état de droit, la démocratie et l'éthique, ainsi que des outils de diligence requise, conformément aux directives, y compris les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme. Le processus d'élaboration de ces politiques ou mécanismes devrait inclure toutes les parties prenantes et tenir compte de la situation et des priorités de chaque État membre. L'UNESCO peut être un partenaire et aider les États membres dans l'élaboration

ainsi que le suivi et l'évaluation des mécanismes stratégiques.

49. L'UNESCO reconnaît que les États membres se trouveront à différents stades de préparation pour mettre en œuvre la présente Recommandation, des points de vue scientifique, technologique, économique, éducatif, juridique, réglementaire, infrastructurel, sociétal et culturel, entre autres. Il convient de noter que dans le cas présent, « l'état de préparation » est un état dynamique. Afin de permettre la mise en œuvre effective de la présente Recommandation, l'UNESCO : élaborera une méthode d'évaluation de l'état de préparation pour aider les États membres intéressés à identifier leur statut à des moments spécifiques de l'avancement de leur préparation à travers un ensemble de dimensions ; et assurera un soutien aux États membres intéressés pour l'élaboration d'une méthodologie UNESCO concernant l'évaluation de l'impact éthique des technologies de l'IA, le partage de bonnes pratiques, les lignes directrices pour l'évaluation et d'autres mécanismes et travaux analytiques.

DOMAINE STRATÉGIQUE 1 : ÉVALUATIONS DE L'IMPACT ÉTHIQUE

50. Les États membres devraient mettre en place des cadres relatifs aux évaluations de l'impact tels que l'évaluation de l'impact éthique, pour identifier et analyser les avantages et les risques des systèmes d'IA et les questions qu'ils soulèvent, ainsi que des mesures appropriées de prévention, d'atténuation et de suivi des risques, entre autres mécanismes de certification. Ces évaluations de l'impact devraient mettre en évidence les répercussions sur les droits de l'homme et les libertés fondamentales, notamment les droits des personnes marginalisées et vulnérables ou en situation de vulnérabilité, le droit du travail, l'environnement et les écosystèmes ainsi que les incidences éthiques et sociales, et faciliter la participation citoyenne, conformément aux valeurs et principes énoncés dans la présente Recommandation.

51. Les États membres et les entreprises du secteur privé devraient mettre en place des mécanismes de diligence requise et de supervision pour identifier, prévenir et atténuer les risques et rendre compte de la manière dont ils traitent les répercussions des systèmes d'IA sur le respect des droits de l'homme, l'état de droit et les sociétés inclusives. Les États membres devraient également être capables d'évaluer les effets socioéconomiques des systèmes d'IA sur la pauvreté et de s'assurer que le fossé entre les riches et les pauvres, ainsi que la fracture numérique entre les pays et à l'intérieur de ceux-ci, ne s'accroissent pas avec l'adoption massive des technologies de l'IA aujourd'hui et à l'avenir. Pour ce faire, des protocoles de transparence exécutoires devraient notamment être mis en œuvre, correspondant à l'accès à l'information, y compris aux informations d'intérêt public détenues par des entités privées. Les États membres, les entreprises du secteur privé et la société civile devraient étudier les effets sociologiques et psychologiques des

recommandations fondées sur l'IA sur les êtres humains en ce qui concerne leur autonomie de décision. Les systèmes d'IA considérés comme représentant un risque potentiel pour les droits de l'homme devraient être largement testés par les acteurs de l'IA, y compris dans des conditions réelles si nécessaire, dans le cadre de l'évaluation de l'impact éthique, avant de les autoriser sur le marché.

52. Les États membres et les entreprises devraient mettre en œuvre des mesures appropriées pour surveiller toutes les phases du cycle de vie d'un système d'IA dans le cadre de l'évaluation de l'impact éthique, y compris le fonctionnement des algorithmes utilisés pour la prise de décisions, les données, ainsi que les acteurs de l'IA impliqués dans le processus, en particulier dans les services publics et lorsqu'une interaction directe avec l'utilisateur final est nécessaire. Les obligations des États membres en matière de droits de l'homme devraient faire partie des aspects éthiques des évaluations des systèmes d'IA.

53. Les gouvernements devraient adopter un cadre réglementaire qui définisse une procédure permettant en particulier aux autorités publiques de mener à bien des évaluations sur l'impact des systèmes d'IA afin d'anticiper les répercussions, d'atténuer les risques, d'éviter les conséquences préjudiciables, de faciliter la participation des citoyens et de faire face aux défis sociétaux. L'évaluation devrait également établir des mécanismes de supervision adaptés, notamment les principes de vérifiabilité, de traçabilité et d'explicabilité, permettant d'évaluer les algorithmes, les données et les processus de conception, et inclure un examen externe des systèmes d'IA. Les évaluations de l'impact éthique

devraient être transparentes et ouvertes au public, selon le cas. Elles devraient par ailleurs être multidisciplinaires, multipartites, multiculturelles, pluralistes et inclusives. Les

autorités publiques devraient être tenues de surveiller les systèmes d'IA qu'elles mettent en œuvre et/ou déploient, en établissant des mécanismes et des outils appropriés.

DOMAINE STRATÉGIQUE 2 : GOUVERNANCE ET GESTION ÉTHIQUES

- 54.** Les États membres devraient s'assurer que les mécanismes de gouvernance de l'IA sont inclusifs, transparents, multidisciplinaires, multilatéraux (cela inclut la possibilité d'atténuer et de réparer les dommages au-delà des frontières) et multipartites. En particulier, la gouvernance devrait comporter des aspects d'anticipation et des dispositifs effectifs de protection, de suivi des conséquences, de mise en application et de réparation.
- 55.** Les États membres devraient veiller à ce que les préjudices causés par le biais de systèmes d'IA fassent l'objet d'enquêtes et de réparations, en mettant en place des mécanismes d'exécution solides et des mesures correctives afin de s'assurer que les droits de l'homme, les libertés fondamentales et l'état de droit sont respectés dans le monde numérique et dans le monde physique. Ces mécanismes et mesures devraient inclure les mécanismes de réparation fournis par des entreprises des secteurs public et privé. La vérifiabilité et la traçabilité des systèmes d'IA devraient être encouragées à cette fin. En outre, les États membres devraient renforcer leurs capacités institutionnelles à s'acquitter de cet engagement et devraient collaborer avec les chercheurs et les autres parties prenantes pour analyser, prévenir et freiner toute utilisation potentiellement malveillante des systèmes d'IA.
- 56.** Les États membres sont encouragés à élaborer des stratégies nationales et régionales relatives à l'IA et à envisager l'adoption de formes de gouvernance souple, par exemple un mécanisme de certification pour les systèmes d'IA et la reconnaissance mutuelle de leurs certifications, selon la sensibilité du domaine d'application et selon l'impact escompté sur les droits humains, l'environnement et les écosystèmes, ainsi que les autres considérations éthiques énoncées dans la présente Recommandation. Un tel mécanisme pourrait inclure différents niveaux d'audit des systèmes, des données et de la conformité aux principes et règles de procédure en matière d'éthique. Dans le même temps, il ne devrait pas entraver l'innovation ni désavantager les petites et les moyennes entreprises ou les start-up, la société civile, ou encore les organismes scientifiques et de recherche par des lourdeurs administratives. Ce mécanisme devrait en outre inclure un volet consacré au suivi régulier pour garantir la fiabilité du système et le maintien de son intégrité et de sa conformité aux principes éthiques tout au long de sa durée de vie, en exigeant une nouvelle certification si nécessaire.
- 57.** Les États membres et les autorités publiques devraient procéder à une auto-évaluation transparente des systèmes d'IA existants et proposés, qui devrait notamment analyser si l'adoption de l'IA est judicieuse et, le cas échéant, prévoir une évaluation supplémentaire pour identifier la méthode appropriée, ainsi qu'une évaluation visant à déterminer si la méthode adoptée entraînerait des violations des obligations des États membres en matière de droits de l'homme ou des atteintes à celles-ci et, si tel est le cas, d'en interdire l'utilisation.
- 58.** Les États membres devraient encourager les entités publiques, les entreprises du secteur privé et les organisations de la société civile à associer différentes parties prenantes à leur gouvernance en matière d'IA et à envisager d'ajouter une fonction de responsable de l'éthique de l'IA indépendant ou tout autre mécanisme visant à superviser les efforts déployés concernant l'évaluation de l'impact éthique, le contrôle et le suivi continu, ainsi qu'à assurer la conformité éthique des systèmes d'IA. Les États membres, les entreprises du secteur privé et les organisations de la société civile, avec le soutien de l'UNESCO, sont encouragés à créer un réseau de responsables de l'éthique de l'IA indépendants pour appuyer ce mécanisme aux niveaux national, régional et international.
- 59.** Les États membres devraient favoriser la mise en place et l'accessibilité d'un écosystème numérique pour le développement éthique et inclusif des systèmes d'IA à l'échelle nationale, en vue notamment de combler les lacunes en matière d'accès tout au long du cycle de vie des systèmes d'IA, tout en contribuant à la collaboration internationale. Cet écosystème devrait englober notamment des technologies et des infrastructures numériques, ainsi que des mécanismes de partage des connaissances en matière d'IA, le cas échéant.
- 60.** Les États membres devraient établir des mécanismes, en collaboration avec les organisations internationales, les sociétés transnationales, les établissements universitaires et la société civile, afin de garantir la participation active de tous les États membres, notamment les PRITI et en particulier les PMA, les PDSL et les PEID, aux discussions internationales concernant la gouvernance de l'IA. Cela pourrait passer par la mise à disposition de fonds, le fait d'assurer une participation régionale égale, ou tout autre mécanisme. En outre, afin de s'assurer que les forums sur l'IA soient ouverts à tous, les États membres devraient faciliter les déplacements des acteurs de l'IA, notamment ceux originaires des PRITI et en particulier des PMA, des PDSL et des PEID, au sein et en dehors de leur territoire, afin qu'ils puissent participer à ces forums.
- 61.** Les modifications apportées aux législations nationales existantes ou l'élaboration de nouvelles législations nationales concernant les systèmes d'IA doivent

respecter les obligations des États membres en matière de droits de l'homme et promouvoir les droits humains et les libertés fondamentales tout au long de la durée de vie du système d'IA. La promotion des droits de l'homme et des libertés fondamentales devrait également prendre la forme d'initiatives de gouvernance, de modèles de bonnes pratiques de collaboration concernant les systèmes d'IA, ainsi que de directives techniques et méthodologiques élaborées par les autorités nationales et internationales pour s'adapter aux progrès des technologies de l'IA. Différents secteurs, y compris le secteur privé, dans leurs pratiques concernant les systèmes d'IA, doivent respecter, protéger et promouvoir les droits de l'homme et les libertés fondamentales au moyen d'instruments nouveaux et existants, en association avec la présente Recommandation.

62. Les États membres qui se dotent de systèmes d'IA pour les cas d'utilisation sensibles en matière de droits de l'homme, comme l'application des lois, la protection sociale, l'emploi, les médias et les relais d'information, les soins de santé et le système judiciaire indépendant, devraient prévoir des mécanismes permettant d'évaluer l'impact économique et social de ces systèmes, par le biais d'autorités de contrôle appropriées telles que des autorités indépendantes de protection des données, une surveillance sectorielle et des organismes publics chargés de la surveillance.
63. Les États membres devraient renforcer la capacité du pouvoir judiciaire à prendre des décisions relatives aux systèmes d'IA, dans le respect de l'état de droit et conformément aux normes et au droit internationaux, notamment en ce qui concerne l'utilisation des systèmes d'IA dans leurs délibérations, tout en veillant au respect du principe de la surveillance humaine. Dans le cas où les systèmes d'IA sont utilisés par le pouvoir judiciaire, des garanties suffisantes doivent être mises en place pour protéger, entre autres, les droits de l'homme fondamentaux, l'état de droit, l'indépendance judiciaire et le principe de la surveillance humaine, ainsi que pour garantir un développement et une utilisation dignes de confiance, axés sur l'intérêt général et centrés sur l'être humain des systèmes d'IA dans le système judiciaire.
64. Les États membres devraient s'assurer que les gouvernements et les organisations multilatérales jouent un rôle de premier plan pour garantir la sûreté et la sécurité des systèmes d'IA, dans le cadre d'une participation multipartite. En particulier, les États membres, les organisations internationales et les autres organes concernés devraient élaborer des normes internationales qui décriraient des niveaux de sûreté et de transparence mesurables et vérifiables pour pouvoir évaluer les systèmes de manière objective et déterminer les niveaux de conformité. Les États membres et les entreprises devraient en outre apporter un soutien constant à la recherche stratégique sur les risques de sûreté et de sécurité potentiels des technologies de l'IA et devraient encourager la recherche sur la transparence et l'explicabilité, l'inclusion et la maîtrise de ces technologies

en y consacrant des fonds supplémentaires pour différents domaines et à différents niveaux, tels que le langage technique et naturel.

65. Les États membres devraient mettre en œuvre des politiques visant à garantir que les activités des acteurs de l'IA sont conformes aux règles, aux normes et aux principes du droit international des droits de l'homme tout au long de la durée de vie des systèmes d'IA, et qu'elles prennent pleinement en considération les diversités culturelles et sociales actuelles, notamment les coutumes locales et les traditions religieuses, compte dûment tenu de la primauté et de l'universalité des droits de l'homme.
66. Les États membres devraient mettre en place des mécanismes pour exiger des acteurs de l'IA qu'ils révèlent et combattent tout stéréotype présent dans les résultats et les données des systèmes d'IA, volontairement ou par négligence, et qu'ils veillent à ce que les ensembles de données d'entraînement des systèmes d'IA ne favorisent pas les inégalités culturelles, économiques ou sociales, les préjugés, la propagation de la désinformation et de la mésinformation, ou encore les atteintes à la liberté d'expression et à l'accès à l'information. Une attention particulière devrait être portée aux régions où les données sont rares.
67. Les États membres devraient mettre en œuvre des politiques visant à promouvoir et accroître la diversité et l'inclusion au sein des équipes de développement de l'IA et des ensembles de données d'entraînement, de façon que ceux-ci reflètent leur population, ainsi qu'à garantir l'accès équitable aux technologies de l'IA et à leurs avantages, en particulier pour les groupes marginalisés, aussi bien dans les zones rurales que dans les zones urbaines.
68. Les États membres devraient mettre en place, revoir et adapter, le cas échéant, les cadres réglementaires pour établir la redevabilité et la responsabilité des contenus et des résultats des systèmes d'IA aux différentes phases de leur cycle de vie. Les États membres devraient, s'il y a lieu, définir des cadres de responsabilité ou clarifier l'interprétation des cadres existants pour assurer l'attribution de la responsabilité des résultats et du fonctionnement des systèmes d'IA. En outre, lors de l'élaboration de cadres réglementaires, les États membres devraient notamment tenir compte du fait que la responsabilité et la redevabilité devraient toujours incomber en dernier ressort à une personne physique ou morale, et que les systèmes d'IA ne devraient pas eux-mêmes bénéficier d'une personnalité juridique. Pour ce faire, ces cadres réglementaires devraient être compatibles avec le principe de la surveillance humaine et établir une approche globale, axée sur les acteurs de l'IA et les processus technologiques impliqués dans les différentes étapes du cycle de vie des systèmes d'IA.
69. Afin d'établir des normes lorsqu'il n'y en a pas, ou d'adapter les cadres réglementaires existants, les États membres devraient impliquer tous les acteurs de l'IA (chercheurs, représentants de la société civile et des

forces de l'ordre, assureurs, investisseurs, industriels, ingénieurs, avocats et utilisateurs, entre autres). Les normes peuvent évoluer en bonnes pratiques, en lois et en règlements. Les États membres sont également encouragés à utiliser des mécanismes tels que des modèles de politiques et des bacs à sable réglementaires pour accélérer la formulation des lois, règlements et politiques, y compris leur réexamen périodique, et suivre ainsi l'évolution rapide des nouvelles technologies, et pour que les lois et règlements puissent être testés dans un environnement sûr avant d'être officiellement adoptés. Les États membres devraient aider les autorités locales à élaborer des politiques, des règlements et

des lois à l'échelle locale, conformément aux cadres juridiques nationaux et internationaux.

- 70.** Les États membres devraient définir des exigences claires en matière de transparence et d'explicabilité des systèmes d'IA afin d'aider à garantir la crédibilité des systèmes tout au long de leur durée de vie. Ces exigences devraient inclure la conception et la mise en œuvre de mécanismes d'évaluation de l'impact qui prennent en considération la nature du domaine d'application, l'utilisation prévue, le public cible et la faisabilité de chaque système d'IA spécifique.

DOMAINE STRATÉGIQUE 3 : POLITIQUES EN MATIÈRE DE DONNÉES

- 71.** Les États membres devraient s'efforcer d'élaborer des stratégies de gouvernance des données qui garantissent l'évaluation continue de la qualité des données d'entraînement des systèmes d'IA, notamment l'adéquation des processus de collecte et de sélection des données, et qui prévoient des mesures de sécurité et de protection des données appropriées, ainsi que des mécanismes de retour d'information permettant de tirer des enseignements des erreurs et de partager de bonnes pratiques entre tous les acteurs de l'IA.
- 72.** Les États membres devraient mettre en place des garanties appropriées pour protéger le droit au respect

de la vie privée, conformément au droit international, notamment en répondant à des préoccupations telles que la surveillance. Les États membres devraient, entre autres, adopter ou mettre en œuvre des cadres législatifs qui assurent une protection adéquate, conforme au droit international. Les États membres devraient vivement encourager l'ensemble des acteurs de l'IA, y compris les entreprises, à appliquer les normes internationales en vigueur et, en particulier, à conduire les évaluations de l'impact sur la vie privée adéquates, dans le cadre des évaluations de l'impact éthique, qui tiennent compte de l'impact socioéconomique global du traitement prévu des données, ainsi qu'à appliquer le principe du respect



de la vie privée dès la conception de leurs systèmes. Le droit à la vie privée devrait être respecté, protégé et encouragé pendant toute la durée de vie des systèmes d'IA.

73. Les États membres devraient veiller à ce que les personnes conservent des droits sur leurs données personnelles et soient protégées par un cadre qui prévoit notamment : la transparence ; des garanties appropriées pour le traitement des données sensibles ; un niveau approprié de protection des données ; des mécanismes et des systèmes de responsabilisation efficaces et cohérents ; la pleine jouissance des droits des personnes concernées et la possibilité d'accéder à leurs données personnelles dans les systèmes d'IA et de les effacer, sauf dans des circonstances particulières conformes au droit international ; un niveau de protection approprié et pleinement conforme à la législation sur la protection des données lorsque celles-ci sont utilisées à des fins commerciales – par exemple pour permettre la publicité microciblée – ou transférées à l'étranger ; et une surveillance indépendante efficace dans le cadre d'un mécanisme de gouvernance des données qui permette aux individus de garder le contrôle de leurs données personnelles, et renforce les avantages d'une libre circulation des informations au niveau international, y compris l'accès aux données.

74. Les États membres devraient établir leurs politiques en matière de données ou leurs cadres équivalents, ou renforcer les politiques et cadres déjà existants, afin d'assurer une sécurité totale pour les données personnelles et sensibles qui, si elles étaient divulguées, risqueraient de causer des dommages, des blessures ou des difficultés exceptionnelles aux personnes. Il s'agit par exemple des données relatives aux infractions, aux poursuites pénales et aux condamnations, ainsi qu'aux mesures de sécurité qui y sont liées ; des données biométriques, génétiques et médicales ; et des données personnelles concernant par exemple la race, la couleur, l'ascendance, le genre, l'âge, la langue, la religion, les opinions politiques, l'origine nationale, ethnique ou sociale, la condition économique ou sociale de naissance, le handicap, ou toute autre caractéristique.

75. Les États membres devraient promouvoir les données ouvertes. À cet égard, ils devraient envisager de

revoir leurs politiques et leurs cadres réglementaires, notamment en ce qui concerne l'accès à l'information et les principes d'un gouvernement ouvert, pour rendre compte des exigences propres à l'IA et promouvoir des mécanismes, tels que des dépôts en libre accès pour les données et les codes sources publics ou financés par des fonds publics ainsi que des fiducies de données (« data trusts »), afin de favoriser le partage des données de façon sûre, équitable, légale et éthique, entre autres.

76. Les États membres devraient promouvoir et faciliter l'utilisation d'ensembles de données robustes et de qualité pour la formation, le développement et l'exploitation des systèmes d'IA, et faire preuve d'une extrême vigilance dans la supervision de la collecte et de l'exploitation de ces données. Lorsque c'est possible et faisable, cela pourrait supposer d'investir dans la création d'ensembles de données de référence, incluant des ensembles de données ouverts et fiables, diversifiés, établis sur une base légale valable, notamment avec le consentement des personnes concernées, lorsque le consentement est requis par la loi. L'élaboration de normes pour l'annotation des ensembles de données devrait être encouragée, y compris la ventilation des données en fonction du genre et d'autres critères, afin de pouvoir déterminer facilement comment un ensemble de données est constitué et quelles sont ses propriétés.

77. Comme l'indique également le rapport du Groupe de haut niveau du Secrétaire général de l'Organisation des Nations Unies sur la coopération numérique, les États membres devraient, avec le soutien de l'Organisation des Nations Unies et de l'UNESCO, adopter une approche d'« espace commun numérique » des données, le cas échéant, accroître l'interopérabilité des outils et des ensembles de données ainsi que des interfaces des systèmes hébergeant les données, et encourager les entreprises du secteur privé à partager avec toutes les parties prenantes les données qu'elles collectent au profit de la recherche, de l'innovation ou de l'intérêt public, selon les besoins. Ils devraient également promouvoir les efforts publics et privés visant à créer des plates-formes de collaboration pour partager des données de qualité dans des espaces de données fiables et sécurisés.

DOMAINE STRATÉGIQUE 4 : DÉVELOPPEMENT ET COOPÉRATION INTERNATIONALE

78. Les États membres et les sociétés transnationales devraient accorder une attention prioritaire à l'éthique de l'intelligence artificielle en examinant les questions éthiques relatives à l'IA au sein des instances internationales, intergouvernementales et multipartites concernées.

79. Les États membres devraient veiller à ce que l'utilisation de l'IA dans les domaines liés au développement – l'éducation, la science, la culture, la communication et l'information, les soins de santé, l'agriculture et l'approvisionnement alimentaire, l'environnement, la gestion des ressources naturelles et des infrastructures,

la planification et la croissance économiques, entre autres – respecte les valeurs et les principes énoncés dans la présente Recommandation.

- 80.** Les États membres devraient s'employer, au sein des organisations internationales, à mettre en place des cadres de coopération internationale dans le domaine de l'IA au service du développement, notamment en mettant à disposition des compétences techniques, des financements, des données, des connaissances spécialisées, des infrastructures, et en facilitant la collaboration multipartite pour résoudre les problèmes complexes en matière de développement, en particulier pour les PRITI, notamment les PMA, les PDSL et les PEID.
- 81.** Les États membres devraient s'efforcer de promouvoir la collaboration internationale en matière de recherche et d'innovation sur l'IA, notamment au sein des réseaux et des centres de recherche et d'innovation qui favorisent une participation accrue et un plus grand leadership des chercheurs venant des PRITI et d'autres pays, notamment des PMA, des PDSL et des PEID.

- 82.** Les États membres devraient soutenir les recherches sur l'éthique de l'IA en faisant appel aux organisations internationales et aux établissements de recherche, ainsi qu'aux sociétés transnationales, ces travaux pouvant servir de base à l'utilisation éthique des systèmes d'IA par des entités publiques et privées, notamment les recherches sur l'applicabilité de cadres éthiques adaptés selon les cultures et les contextes, et sur les possibilités d'élaborer des solutions réalisables sur le plan technologique en se conformant à ces cadres.
- 83.** Les États membres devraient encourager la coopération et la collaboration internationales dans le domaine de l'IA afin d'éliminer les clivages géotechnologiques. Des échanges et consultations sur les questions d'ordre technologique devraient avoir lieu entre les États membres et leurs populations, entre les secteurs public et privé ainsi qu'entre les pays les plus et les moins avancés sur le plan technologique, dans le plein respect du droit international.

DOMAINE STRATÉGIQUE 5 : ENVIRONNEMENT ET ÉCOSYSTÈMES

- 84.** 84. Les États membres et les entreprises devraient évaluer l'impact environnemental direct et indirect du système d'IA tout au long de son cycle de vie, notamment, sans s'y limiter, son empreinte carbone, sa consommation d'énergie et l'impact environnemental de l'extraction des matières premières nécessaires à la fabrication des technologies de l'IA, et réduire l'impact environnemental des systèmes d'IA et des infrastructures de données. Les États membres devraient garantir que tous les acteurs du domaine de l'IA respectent les législations, les politiques et les pratiques relatives à l'environnement.
- 85.** Les États membres devraient mettre en place, selon les besoins, des mesures d'incitation permettant d'assurer l'élaboration et l'adoption de solutions fondées sur les droits et sur une IA éthique en faveur de la résilience face aux risques de catastrophe ; de la surveillance, de la protection et de la régénération de l'environnement et des écosystèmes ; et de la préservation de la planète. Ces systèmes d'IA devraient reposer, tout au long de leur cycle de vie, sur la participation des communautés locales et autochtones, et favoriser des approches telles que l'économie circulaire et les modes de consommation et de production durables. Les systèmes d'IA peuvent ainsi, selon les besoins et s'il y a lieu, être utilisés pour :
- (a) soutenir la protection, le suivi et la gestion des ressources naturelles ;
 - (b) soutenir la prévision, la prévention, le contrôle et l'atténuation des problèmes liés au climat ;
 - (c) favoriser un écosystème alimentaire plus efficient et plus durable ;
 - (d) contribuer à accélérer l'accès aux énergies durables et leur adoption par le plus grand nombre ;
 - (e) permettre et promouvoir la généralisation d'infrastructures durables, de modèles économiques durables et de financements durables au service du développement durable ;
 - (f) détecter les polluants ou prévoir les niveaux de pollution et aider ainsi les parties prenantes concernées à définir, planifier et mettre en place des interventions ciblées pour prévenir et réduire la pollution et l'exposition à celle-ci.
- 86.** Au moment de sélectionner une méthode d'IA, compte tenu de l'importance potentielle des besoins en données et en ressources de certaines d'entre elles et de l'impact sur l'environnement qui en résulte, les États membres devraient veiller à ce que les acteurs de l'IA, suivant le principe de proportionnalité, privilégient les méthodes d'IA économes en données, en énergie et en ressources. Il convient de fixer des conditions visant à garantir que l'on dispose des éléments probants appropriés pour prouver qu'une application d'IA aura l'effet escompté, ou que cette application est assortie de garanties qui permettent de justifier son utilisation. Si cela n'est pas possible, il convient de privilégier le principe de précaution et, en cas d'impacts négatifs disproportionnés sur l'environnement, l'IA ne devrait pas être utilisée.



DOMAINE STRATÉGIQUE 6 : ÉGALITÉ DES GENRES

87. Les États membres devraient veiller à ce que le potentiel des technologies numériques et de l'intelligence artificielle soit pleinement optimisé afin de contribuer à l'égalité des genres, et à ce qu'il ne soit pas porté atteinte aux droits humains et aux libertés fondamentales des filles et des femmes, ni à leur sécurité et à leur intégrité, à aucun stade du cycle de vie du système d'IA. En outre, les évaluations de l'impact éthique devraient prendre en compte les questions de genre de manière transversale.

88. Les États membres devraient allouer des fonds publics au financement de dispositifs attentifs à l'égalité des genres, veiller à ce que les politiques nationales relatives au numérique prévoient un plan d'action pour l'égalité des genres et élaborer des politiques pertinentes, par exemple sur la préparation à la vie active, en vue de soutenir les filles et les femmes et de faire en sorte que celles-ci ne soient pas exclues de l'économie numérique reposant sur l'IA. Certains investissements spéciaux devraient être envisagés et réalisés afin de proposer des programmes ciblés et un langage respectueux de l'égalité des genres, de façon à élargir les possibilités de participation des filles et des femmes aux sciences, à la technologie, à l'ingénierie et aux mathématiques (STIM),

notamment aux disciplines liées aux technologies de l'information et de la communication (TIC), et à améliorer la préparation, l'employabilité, l'égalité d'évolution de carrières et l'épanouissement professionnel des filles et des femmes.

89. Les États membres devraient faire en sorte qu'il soit tiré parti du potentiel des systèmes d'IA en faveur de la réalisation de l'égalité des genres. Ils devraient s'assurer que ces technologies n'accroissent pas les inégalités entre les genres déjà marquées qui existent dans plusieurs domaines du monde analogique, mais, au contraire, qu'elles les éliminent. Ces inégalités concernent notamment les inégalités salariales entre hommes et femmes ; la représentation inégale dans certaines professions et activités ; la faible représentation aux postes de direction, aux conseils d'administration ou dans les équipes de recherche du domaine de l'IA ; les disparités dans l'accès à l'éducation ; les inégalités en matière d'accessibilité, d'adoption, d'utilisation et de coût abordable du numérique et de l'IA ; et la répartition inégale du travail non rémunéré et des responsabilités en matière de soins dans nos sociétés.

- 90.** Les États membres devraient veiller à ce que les stéréotypes fondés sur le genre et les préjugés discriminatoires ne soient pas transposés dans les systèmes d'IA, mais plutôt repérés et corrigés de manière proactive. Des efforts sont nécessaires pour éviter les effets négatifs cumulés des fractures technologiques dans la réalisation de l'égalité des genres et la prévention de la violence à l'égard des filles et des femmes ainsi que des groupes sous-représentés, prenant notamment la forme du harcèlement, de l'intimidation ou de la traite, y compris en ligne.
- 91.** Les États membres devraient encourager l'esprit d'entreprise, la participation et l'engagement des femmes à tous les stades du cycle de vie d'un système d'IA en proposant et en soutenant des incitations économiques et réglementaires, entre autres mesures d'encouragement et de soutien, ainsi que des politiques visant à assurer une participation équilibrée entre les genres à la recherche sur l'IA au sein des universités, et la représentation des différents genres aux postes de direction ainsi que dans les conseils d'administration et les équipes de recherche des entreprises du secteur numérique ou de l'IA. Les États membres devraient veiller à ce que les fonds publics (en faveur de l'innovation, de la recherche et de la technologie) soient affectés à

des programmes et à des entreprises inclusifs, avec un équilibre clair entre les genres, et à ce que les fonds privés soient encouragés de la même façon suivant les principes de discrimination positive. Des politiques favorisant des environnements exempts de toute forme de harcèlement devraient être élaborées et appliquées, tout en encourageant la diffusion des bonnes pratiques concernant les moyens de promouvoir la diversité tout au long du cycle de vie du système d'IA.

- 92.** Les États membres devraient promouvoir la diversité de genre dans la recherche sur l'IA, dans le milieu universitaire et dans l'industrie, en prenant des mesures qui incitent les filles et les femmes à s'engager dans ce domaine, en mettant en place des mécanismes permettant de lutter contre les stéréotypes sexistes et le harcèlement au sein de la communauté des chercheurs en IA, et en encourageant les entités universitaires et privées à partager leurs meilleures pratiques sur la façon de renforcer la diversité de genre.
- 93.** L'UNESCO peut aider à constituer un recueil des meilleures pratiques visant à encourager la participation des filles, des femmes et des groupes sous-représentés à tous les stades du cycle de vie des systèmes d'IA.

DOMAINE STRATÉGIQUE 7 : CULTURE

- 94.** Les États membres sont encouragés à établir, s'il y a lieu, des systèmes d'IA dans les domaines de la conservation, de l'enrichissement, de la compréhension, de la promotion, de la gestion et de l'accessibilité du patrimoine culturel matériel, documentaire et immatériel, en particulier les langues en péril et les langues et savoirs autochtones, en mettant en place ou en actualisant par exemple les programmes éducatifs concernant l'application de systèmes d'IA dans ces domaines, selon les besoins et suivant une approche participative, à l'intention des institutions et du public.
- 95.** Les États membres sont encouragés à examiner et à traiter les répercussions des systèmes d'IA sur la culture, en particulier celles des applications de traitement du langage naturel, comme la traduction automatique et les assistants vocaux, sur les nuances de l'expression et du langage humains. Ces évaluations devraient permettre de contribuer à la conception et à la mise en œuvre de stratégies visant à tirer parti au maximum des avantages de ces systèmes en réduisant les inégalités culturelles et en améliorant la compréhension humaine, ainsi qu'en luttant contre leurs effets négatifs, tels qu'un recul de l'usage qui pourrait entraîner la disparition de langues en péril, de dialectes locaux et de variations tonales et culturelles associées à l'expression et au langage humains.
- 96.** Les États membres devraient promouvoir l'éducation à l'IA et des formations numériques à l'intention des

artistes et des professionnels de la création afin qu'ils soient capables d'évaluer la pertinence de l'utilisation des technologies de l'IA dans leurs métiers et de contribuer à la conception et la mise en œuvre de technologies de l'IA pertinentes, ces technologies étant actuellement utilisées pour créer, produire, distribuer, diffuser et consommer toute une variété de biens et services culturels, en gardant à l'esprit l'importance de la préservation du patrimoine culturel, de la diversité et de la liberté artistique.

- 97.** Les États membres devraient faire mieux connaître les outils d'IA et promouvoir leur évaluation parmi les industries culturelles locales et les petites et moyennes entreprises travaillant dans le domaine de la culture, afin d'éviter le risque de concentration sur le marché culturel.
- 98.** Les États membres devraient faire appel aux entreprises spécialisées dans les technologies et à d'autres parties prenantes afin de favoriser une offre diversifiée et une pluralité d'accès en matière d'expressions culturelles, en faisant notamment en sorte que la recommandation algorithmique améliore la visibilité et la découvrabilité des contenus locaux.
- 99.** Les États membres devraient encourager les nouvelles recherches à la croisée de l'IA et de la propriété intellectuelle, afin de déterminer par exemple comment protéger par des droits de propriété intellectuelle les œuvres créées au moyen de technologies de l'IA, ou s'il y a lieu le faire. Les États membres devraient également



évaluer les répercussions des technologies de l'IA sur les droits ou les intérêts des titulaires de droits de propriété intellectuelle dont les œuvres sont utilisées pour la recherche, le développement, la formation ou la mise en œuvre d'applications de l'IA.

100. Les États membres devraient encourager les musées, les galeries, les bibliothèques et les archives à l'échelle nationale à utiliser des systèmes d'IA afin de valoriser leurs collections et d'enrichir leurs bibliothèques, leurs bases de données et leurs bases de connaissance, tout en facilitant l'accès à leurs usagers.

DOMAINE STRATÉGIQUE 8 : ÉDUCATION ET RECHERCHE

101. Les États membres devraient collaborer avec des organisations internationales, des établissements d'enseignement et des organismes privés et non gouvernementaux, afin de transmettre au grand public de tous les pays, à tous les niveaux, les connaissances nécessaires à la maîtrise de l'IA, de façon à autonomiser la population et à réduire les fractures numériques et les inégalités d'accès dans le domaine du numérique découlant de l'adoption à grande échelle de systèmes d'IA.

102. Les États membres devraient encourager l'acquisition de « compétences préalables » à l'éducation à l'IA, telles que les compétences de base en lecture, en écriture, en calcul, en programmation et en technologie numérique, l'éducation aux médias et à l'information, ainsi que les compétences relatives à la pensée critique et créative, au travail en équipe et à la communication, les compétences socioémotionnelles et les compétences en matière d'éthique de l'IA, en particulier dans les pays et dans les régions ou zones de pays qui présentent des lacunes notables dans l'enseignement de toutes ces compétences.

103. Les États membres devraient promouvoir des programmes généraux de sensibilisation aux avancées de l'IA, notamment concernant les données et les possibilités et défis découlant des technologies de l'IA, les répercussions des systèmes d'IA sur les droits de l'homme, y compris les droits de l'enfant, et leurs implications. Ces programmes devraient être accessibles aux spécialistes comme aux non-spécialistes.

104. Les États membres devraient encourager les initiatives de recherche portant sur l'utilisation responsable et éthique des technologies de l'IA dans l'enseignement, la formation des enseignants et l'apprentissage à distance, entre autres sujets, afin d'élargir les possibilités et d'atténuer les problèmes et les risques qui existent dans ce domaine. Ces initiatives devraient s'accompagner d'une évaluation adéquate de la qualité de l'éducation et de l'impact sur les élèves et les enseignants de l'utilisation des technologies de l'IA. Les États membres devraient également faire en sorte que les technologies de l'IA autonomisent les élèves et les enseignants et améliorent leur expérience, tout en gardant à l'esprit que les dimensions relationnelles et sociales et la valeur des formes traditionnelles d'enseignement sont

essentielles dans les relations enseignant/élève et élève/élève, et qu'elles devraient être prises en compte lorsque l'on examine l'adoption de technologies de l'IA dans l'éducation. Les systèmes d'IA utilisés dans l'enseignement devraient être soumis à des exigences strictes en matière de suivi, d'évaluation des capacités ou de prédiction des comportements des apprenants. L'IA devrait soutenir le processus d'apprentissage sans réduire les capacités cognitives, ni recueillir de données sensibles, dans le respect des normes pertinentes en matière de protection des données personnelles. Les données communiquées pour acquérir des connaissances qui sont collectées pendant les interactions entre l'apprenant et le système d'IA ne doivent pas être utilisées abusivement, détournées ou exploitées à des fins criminelles, y compris dans un but commercial.

- 105.** Les États membres devraient favoriser la participation, avec un rôle moteur, des filles et des femmes, des personnes d'origines ethniques et culturelles diverses, des personnes handicapées, marginalisées et vulnérables ou en situation de vulnérabilité, des minorités, ainsi

que de tous ceux qui ne profitent pas pleinement des avantages de l'inclusion numérique, aux programmes d'éducation à l'IA à tous les niveaux, ainsi que le suivi et le partage des meilleures pratiques à cet égard avec d'autres États membres.

- 106.** Les États membres devraient mettre au point, conformément à leurs traditions et à leurs programmes d'éducation nationale, des programmes d'enseignement de l'éthique de l'IA pour tous les niveaux, et favoriser la collaboration croisée entre l'enseignement de compétences techniques dans le domaine de l'IA et les aspects humanistes, éthiques et sociaux de l'éducation à l'IA. Des cours en ligne et des ressources numériques concernant l'éducation à l'éthique de l'IA devraient être élaborés dans les langues locales, y compris les langues autochtones, et tenir compte de la diversité des environnements, en veillant notamment à l'accessibilité des formats utilisés pour les personnes handicapées.
- 107.** Les États membres devraient promouvoir et soutenir la recherche sur l'IA, notamment sur l'éthique de l'IA, y compris par exemple en finançant ce type de



recherche ou en prenant des mesures visant à inciter les secteurs public et privé à investir dans ce domaine, en reconnaissant que la recherche contribue de manière significative au développement et à l'amélioration futurs des technologies de l'IA en vue de promouvoir le droit international et les valeurs et principes définis dans la présente Recommandation. Les États membres devraient également promouvoir publiquement les bonnes pratiques des chercheurs et des entreprises qui développent l'IA de manière éthique et coopérer avec eux.

- 108.** Les États membres devraient veiller à ce que les chercheurs en IA possèdent une formation en éthique de la recherche et à ce qu'ils soient tenus de prendre en considération des aspects éthiques dans leurs conceptions, leurs produits et leurs publications, en particulier dans les analyses d'ensembles de données qu'ils utilisent, la façon dont ces ensembles sont annotés, la qualité et la portée des résultats, ainsi que les applications possibles.
- 109.** Les États membres devraient encourager les sociétés du secteur privé à faciliter l'accès de la communauté scientifique à leurs données à des fins de recherche, notamment dans les PRITI, en particulier les PMA, les PDSL et les PEID. Cet accès doit être conforme aux normes pertinentes en matière de respect de la vie privée et de protection des données.

110. Pour garantir une évaluation critique de la recherche sur l'IA et un suivi adéquat des éventuels mauvais usages et effets néfastes, les États membres devraient veiller à ce que toute future avancée concernant les technologies de l'IA soit fondée sur des recherches scientifiques rigoureuses et indépendantes, et promouvoir la recherche interdisciplinaire sur l'IA en faisant appel à des disciplines autres que les sciences, la technologie, l'ingénierie et les mathématiques (STIM), comme les études culturelles, l'enseignement, l'éthique, les relations internationales, le droit, la linguistique, la philosophie, les sciences politiques, la sociologie et la psychologie.

111. Reconnaisant que les technologies de l'IA offrent d'importants moyens de faire progresser la connaissance et la pratique scientifiques, notamment dans les disciplines traditionnellement fondées sur des modèles, les États membres devraient encourager les communautés scientifiques à mesurer les avantages, les limites et les risques que comporte leur utilisation. Il s'agit notamment de veiller à ce que les conclusions tirées d'approches, de modèles et de traitements axés sur les données soient solides et fiables. Par ailleurs, les États membres devraient accueillir avec satisfaction et soutenir l'apport de la communauté scientifique dans la contribution aux politiques et la prise de conscience des atouts et des faiblesses des technologies de l'IA.

DOMAINE STRATÉGIQUE 9 : COMMUNICATION ET INFORMATION

- 112.** Les États membres devraient utiliser les systèmes d'IA pour améliorer l'accès à l'information et au savoir. Ils peuvent notamment mettre en place des initiatives visant à aider les chercheurs, le milieu universitaire, les journalistes, le grand public et les développeurs, à améliorer la liberté d'expression, les libertés académiques et scientifiques ainsi que l'accès à l'information, et à accroître la divulgation proactive des données et informations officielles.
- 113.** Les États membres devraient garantir que les acteurs de l'IA respectent et encouragent la liberté d'expression, ainsi que l'accès à l'information dans le contexte de la création, de la modération et de la conservation automatisées des contenus. Des cadres appropriés, y compris réglementaires, devraient assurer la transparence des opérateurs de communication et d'information en ligne, veiller à ce que les utilisateurs aient accès à des points de vue divers, et prévoir des systèmes de notification immédiate aux utilisateurs des raisons de la suppression ou d'autres traitements des contenus, ainsi que des mécanismes de recours donnant aux utilisateurs la possibilité de demander réparation.

114. Les États membres devraient investir dans les compétences numériques et l'éducation aux médias et à l'information et les promouvoir, afin de renforcer la pensée critique et les compétences indispensables pour comprendre l'utilisation et les implications des systèmes d'IA, en vue d'atténuer et de combattre la désinformation, la mésinformation et le discours de haine. Une meilleure compréhension et une meilleure évaluation des effets positifs et éventuellement indésirables des systèmes de recommandation devraient faire partie de ces efforts.

115. Les États membres devraient créer des environnements favorables aux médias, dans lesquels ces derniers disposeraient des droits et ressources nécessaires pour rendre compte efficacement des avantages et des inconvénients des systèmes d'IA, et encourager les médias à faire un usage éthique de ces systèmes dans leur travail.

DOMAINE STRATÉGIQUE 10 : ÉCONOMIE ET TRAVAIL

116. Les États membres devraient évaluer et traiter l'impact des systèmes d'IA sur le marché du travail ainsi que son incidence sur les besoins en matière d'éducation dans tous les pays, et plus particulièrement dans ceux dont l'économie nécessite une main-d'œuvre importante. Cela peut nécessiter la mise en place, à tous les niveaux d'éducation, d'un éventail plus large de compétences interdisciplinaires « de base » pour donner aux travailleurs actuels et aux nouvelles générations une chance équitable de trouver un emploi sur un marché à l'évolution rapide, et pour garantir leur sensibilisation aux aspects éthiques des systèmes d'IA. En plus des compétences techniques et spécialisées et des tâches requérant peu de qualifications, il convient d'enseigner des compétences comme « apprendre à apprendre », la communication, la réflexion critique, le travail d'équipe, l'empathie et la capacité de transférer ses connaissances d'un domaine à l'autre. Il est essentiel de faire preuve de transparence quant aux compétences recherchées et d'actualiser les programmes d'études en fonction de celles-ci.

117. Les États membres devraient soutenir les accords de coopération entre les autorités, les établissements universitaires, les établissements d'enseignement et de formation professionnels, les entreprises, les organisations de travailleurs et la société civile afin de combler les lacunes en matière de compétences exigées et d'aligner les stratégies et programmes de formation sur les futurs enjeux du travail et sur les besoins des employeurs, y compris les petites et moyennes entreprises. Il convient de promouvoir des approches de l'enseignement et l'apprentissage de l'IA qui soient basées sur des projets, ce qui permet d'établir

des partenariats entre les institutions publiques, les entreprises du secteur privé, les universités et les centres de recherche.

118. Les États membres devraient coopérer avec des entreprises du secteur privé, des organisations de la société civile et d'autres parties prenantes, y compris les travailleurs et les syndicats, pour assurer une transition équitable aux employés menacés. Cela suppose de mettre en place des programmes de perfectionnement et de reconversion, de concevoir des mécanismes efficaces visant à retenir les employés pendant ces périodes de transition, et d'envisager des programmes de « couverture sociale » pour ceux qui ne peuvent pas se reconvertir. Les États membres devraient élaborer et mettre en œuvre des programmes destinés à étudier les défis recensés et à y répondre, notamment le perfectionnement et la reconversion, le renforcement de la protection sociale, la mise en œuvre de politiques et d'interventions sectorielles énergiques, ou encore l'introduction d'avantages fiscaux et de nouvelles modalités d'imposition. Les États membres devraient veiller à ce que les financements publics soient suffisants pour soutenir leurs programmes. Les réglementations pertinentes, telles que les régimes fiscaux, devraient être examinées avec soin et modifiées, si nécessaire, de façon à contrer les conséquences du chômage engendré par l'automatisation basée sur l'IA.

119. Les États membres devraient encourager et aider les chercheurs à se pencher sur l'impact des systèmes d'IA sur l'environnement de travail local afin d'anticiper les tendances et défis à venir. Il conviendrait d'adopter une approche interdisciplinaire et d'analyser les



répercussions des systèmes d'IA sur les secteurs économiques, sociaux et géographiques, ainsi que sur les interactions entre humains et robots et les relations entre êtres humains, dans le but de formuler des conseils sur les pratiques exemplaires en matière de reconversion et de réaffectation.

- 120.** Les États membres devraient prendre les dispositions appropriées pour assurer la compétitivité des marchés et la protection des consommateurs, en envisageant d'éventuels mesures et mécanismes aux niveaux national, régional et international, afin d'empêcher les abus de position dominante sur le marché, y compris les monopoles, en ce qui concerne les systèmes d'IA tout au long de leur cycle de vie, que ces abus portent sur les données, la recherche, la technologie, ou les marchés. Les États membres devraient prévenir

les inégalités qui en résultent, analyser les marchés concernés et promouvoir des marchés concurrentiels. Il convient de prêter dûment attention aux PRITI, et plus particulièrement aux PMA, aux PDSL et aux PEID, qui sont plus exposés et plus vulnérables à la possibilité d'abus de position dominante sur le marché, en raison du manque d'infrastructures, de capacités humaines et de réglementations, entre autres facteurs. Les acteurs de l'IA qui développent des systèmes d'IA dans des pays qui ont établi ou adopté des normes éthiques en matière d'IA devraient également respecter ces normes lorsqu'ils exportent ces produits, développent leurs systèmes d'IA ou les appliquent dans des pays où de telles normes n'existent pas, tout en respectant le droit international et les lois, normes et pratiques nationales applicables de ces pays.

DOMAINE STRATÉGIQUE 11 : SANTÉ ET BIEN-ÊTRE SOCIAL

- 121.** Les États membres devraient s'efforcer d'utiliser des systèmes d'IA efficaces pour améliorer la santé humaine et protéger le droit à la vie, notamment en atténuant les épidémies, tout en développant et maintenant la solidarité internationale en vue de faire face aux risques et incertitudes sanitaires mondiaux. Ils devraient en outre s'assurer que le déploiement des systèmes d'IA dans le domaine des soins de santé se fait dans le respect du droit international ainsi que des obligations qui leur incombent en matière de droits de l'homme. Les États membres devraient veiller à ce que les acteurs impliqués dans les systèmes d'IA relatifs aux soins de santé tiennent compte de l'importance de la relation des patients avec leur famille et le personnel soignant.

- 122.** Les États membres devraient veiller à ce que le développement et le déploiement des systèmes d'IA relatifs à la santé en général et à la santé mentale en particulier – en accordant toute l'attention voulue aux enfants et aux jeunes – soient réglementés, afin de garantir que ces systèmes soient sûrs, efficaces, fiables et éprouvés sur les plans scientifique et médical et qu'ils permettent une innovation et un progrès médical fondés sur des preuves. Par ailleurs, dans le domaine connexe des interventions en cybersanté, les États membres sont vivement encouragés à faire participer activement les patients et leurs représentants à toutes les étapes pertinentes du développement du système.

- 123.** Les États membres devraient accorder une attention particulière à la réglementation des solutions de prévision, de détection et de thérapie médicales contenues dans les applications d'IA, par les moyens suivants :

- (a) assurer une surveillance afin de réduire et d'atténuer les biais ;
- (b) garantir que le professionnel, le patient, le soignant ou l'utilisateur du service est intégré à l'équipe en tant que « spécialiste du domaine » à toutes les étapes pertinentes de l'élaboration des algorithmes ;

- (c) tenir compte du respect de la vie privée, étant donné qu'une surveillance médicale peut être nécessaire et veiller à ce que toutes les obligations nationales et internationales pertinentes en matière de protection des données soient respectées ;
- (d) mettre en place des mécanismes efficaces pour faire en sorte que les personnes dont les données personnelles sont analysées soient informées que leurs données sont utilisées et analysées et qu'elles donnent leur consentement éclairé à ce sujet, sans empêcher l'accès aux soins de santé ;
- (e) garantir que les soins humains et la prise de décisions définitives relatives au diagnostic et au traitement sont systématiquement assurés par des êtres humains, tout en reconnaissant que les systèmes d'IA peuvent les y aider ;
- (f) assurer, si nécessaire, l'examen des systèmes d'IA par un comité de recherche éthique avant l'utilisation clinique.

- 124.** Les États membres devraient mener des recherches sur les effets et la réglementation des dégâts potentiels des systèmes d'IA sur la santé mentale, comme l'augmentation de la dépression, de l'anxiété, de l'isolement social, de la dépendance, du trafic, de la radicalisation, ou encore de la désinformation.

- 125.** Les États membres devraient élaborer des principes directeurs sur les interactions entre humains et robots et leurs conséquences sur les relations entre êtres humains, qui reposeraient sur la recherche et seraient destinés au futur développement des robots, et accorderaient une attention particulière à la santé mentale et physique des êtres humains. Il conviendrait également de prêter une attention particulière aux robots utilisés pour les soins de santé, l'assistance aux personnes âgées et aux personnes handicapées, et dans le domaine de l'éducation, ainsi qu'aux robots destinés aux enfants,



et aux robots à usage ludique, conversationnel et de compagnie destinés aux enfants et aux adultes. En outre, il convient de faire appel aux technologies de l'IA pour améliorer la sécurité et l'ergonomie des robots, y compris dans les environnements de travail associant humains et robots. Une attention particulière devrait être accordée à la possibilité de se servir de l'IA pour manipuler et utiliser à mauvais escient les biais cognitifs humains.

- 126.** Les États membres devraient garantir que les interactions entre humains et robots se font dans le respect des valeurs et principes applicables à tout autre système d'IA, ce qui inclut les droits de l'homme et les libertés fondamentales, la promotion de la diversité et la protection des personnes vulnérables ou en situation de vulnérabilité. Les questions éthiques relatives aux systèmes fondés sur l'IA utilisés dans les neuro-technologies et les interfaces cerveau-machine devraient être prises en considération, afin de préserver la dignité et l'autonomie humaines.
- 127.** Les États membres devraient veiller à ce que les utilisateurs puissent déterminer facilement s'ils interagissent avec un être vivant ou avec un système d'IA imitant des caractéristiques humaines ou animales, et à ce qu'ils puissent efficacement refuser ces interactions et demander une intervention humaine.
- 128.** Les États membres devraient instaurer des politiques visant à sensibiliser à l'anthropomorphisation des

technologies de l'IA et des technologies qui reconnaissent et imitent les émotions humaines, notamment en ce qui concerne les termes utilisés pour les désigner, et évaluer les manifestations, les conséquences éthiques et les possibles limites de ce phénomène, en particulier dans le contexte des interactions entre robots et humains, et plus spécialement lorsqu'il s'agit d'enfants.

- 129.** Les États membres devraient encourager et promouvoir la recherche collaborative sur les effets d'une utilisation à long terme des systèmes d'IA par les individus, en prêtant une attention spéciale aux conséquences psychologiques et cognitives potentielles de ces systèmes sur les enfants et les jeunes. Cette recherche devrait se fonder sur de multiples normes, principes, protocoles, approches disciplinaires, sur une analyse de la modification des comportements et des habitudes, ainsi que sur une évaluation minutieuse des impacts culturels et sociétaux en aval. En outre, les États membres devraient encourager la recherche sur les effets des technologies de l'IA sur les performances des systèmes de santé et les résultats en matière de santé.
- 130.** Les États membres et les parties prenantes devraient établir des mécanismes visant à faire véritablement participer les enfants et les jeunes aux discussions, débats et prises de décisions relatifs aux répercussions des systèmes d'IA sur leur vie et leur avenir.

V.

Suivi et évaluation

131. Les États membres – en fonction de leur situation, de leur mode de gouvernement et de leur Constitution – devraient assurer de manière crédible et transparente le suivi et l'évaluation des politiques, programmes et mécanismes relatifs à l'éthique de l'IA en combinant des approches quantitatives et qualitatives. L'UNESCO peut les aider en :

- (a) élaborant une méthodologie UNESCO d'évaluation de l'impact éthique (EIE) des technologies de l'IA reposant sur des travaux scientifiques rigoureux, et basée sur le droit international des droits de l'homme, des lignes directrices concernant sa mise en œuvre à toutes les étapes du cycle de vie des systèmes d'IA, et des outils de renforcement des capacités destinés à soutenir les initiatives des États membres visant à former des responsables gouvernementaux, les responsables politiques et autres acteurs de l'IA concernés à la méthodologie EIE ;
- (b) mettant au point une méthodologie UNESCO d'évaluation de la préparation pour aider les États membres à déterminer leur état d'avancement, à divers égards, à des moments précis de leur parcours de préparation ;
- (c) concevant une méthodologie UNESCO pour évaluer en amont et en aval l'efficacité et l'efficience des politiques et des mesures d'incitation relatives à l'éthique de l'IA par rapport aux objectifs fixés ;

- (d) renforçant l'analyse fondée sur la recherche et sur des éléments factuels des politiques relatives à l'éthique de l'IA ainsi que l'établissement de rapports à ce sujet ;
- (e) collectant et diffusant des informations sur les progrès et les innovations réalisés, des rapports de recherche, des publications scientifiques, des données et des statistiques concernant l'éthique de l'IA, y compris par le biais d'initiatives déjà en place, afin d'encourager l'échange de pratiques exemplaires et l'apprentissage mutuel et de contribuer à l'application de la présente Recommandation.

132. Les processus de suivi et d'évaluation devraient assurer une large participation de toutes les parties prenantes, notamment, mais pas exclusivement, des personnes vulnérables ou en situation de vulnérabilité. Il conviendrait de garantir une diversité sur les plans social, culturel et du genre, en vue d'améliorer les processus d'apprentissage et de renforcer les liens entre constatations, prises de décisions, transparence et obligation de rendre des comptes concernant les résultats.

133. Dans le but de promouvoir les politiques et pratiques exemplaires en matière d'éthique de l'IA, il conviendrait d'élaborer des outils et indicateurs appropriés pour évaluer leur efficacité et leur efficience par rapport aux normes, priorités et cibles convenues – y compris des

cibles spécifiques pour les personnes appartenant aux populations défavorisées et marginalisées et pour les personnes vulnérables ou en situation de vulnérabilité –, ainsi que les conséquences des systèmes d'IA aux niveaux individuel et sociétal. Le suivi et l'évaluation de l'impact des systèmes d'IA et des politiques et pratiques connexes relatives à l'éthique de l'IA devraient être assurés en permanence et de façon systématique proportionnellement aux risques correspondants. Ce processus devrait s'appuyer sur des cadres convenus au niveau international et s'accompagner d'évaluations des établissements, prestataires et programmes privés et publics, y compris des auto-évaluations, ainsi que des enquêtes de suivi et la mise au point d'une batterie d'indicateurs. La collecte et le traitement des données devraient être menés conformément au droit international, à la législation nationale sur la protection et la confidentialité des données et aux valeurs et principes énoncés dans la présente Recommandation.

134. En particulier, les États membres pourraient envisager la possibilité de mécanismes de suivi et d'évaluation, tels qu'une commission d'éthique, un observatoire de l'éthique de l'IA, un répertoire couvrant le développement éthique et conforme aux droits de l'homme de systèmes d'IA, ou des contributions aux initiatives déjà en place pour renforcer la conformité aux principes éthiques dans l'ensemble des domaines de compétence de l'UNESCO, un mécanisme de partage d'expériences, des bacs à sable réglementaires sur l'IA, et un guide d'évaluation permettant à tous les acteurs de l'IA de déterminer à quel point ils respectent les recommandations stratégiques mentionnées dans le présent document..

VI.

Utilisation de la présente Recommandation

135. Les États membres et toutes les autres parties prenantes identifiées dans la présente Recommandation devraient respecter, promouvoir et protéger les valeurs, principes et normes éthiques relatifs à l'IA qui y sont énoncés, et prendre toutes les mesures en leur pouvoir pour donner effet aux recommandations stratégiques qu'elle contient.

136. Les États membres devraient s'efforcer d'élargir et de compléter leur propre action en ce qui concerne

la présente Recommandation en coopérant avec toutes les organisations nationales et internationales, gouvernementales et non gouvernementales concernées, les sociétés transnationales et les organisations scientifiques dont les activités sont en rapport avec le champ d'application et les objectifs de la présente Recommandation. L'élaboration d'une méthodologie UNESCO d'évaluation de l'impact éthique et la création de commissions nationales d'éthique de l'IA peuvent constituer d'importants outils à cet égard.

VII.

Promotion de la présente Recommandation

137. L'UNESCO a vocation à être la principale institution des Nations Unies chargée de promouvoir et diffuser la présente Recommandation. Elle travaillera par conséquent en collaboration avec d'autres entités compétentes du système onusien, tout en respectant leur mandat et en évitant les chevauchements d'activités.

138. L'UNESCO, y compris certains de ses organes tels que la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST), le Comité international de bioéthique (CIB) et le Comité

intergouvernemental de bioéthique (CIGB), travaillera également en collaboration avec d'autres organisations gouvernementales et non gouvernementales internationales, régionales et sous-régionales.

139. Bien qu'à l'UNESCO, le mandat de promouvoir et protéger relève de la compétence des gouvernements et des organes intergouvernementaux, la société civile jouera un rôle important pour défendre les intérêts du secteur public. L'UNESCO doit par conséquent asseoir et mettre en avant sa légitimité.

VIII.

Dispositions finales

140. La présente Recommandation doit s'entendre comme un tout, et les valeurs et principes fondamentaux comme étant complémentaires et interdépendants.

141. Aucune disposition de la présente Recommandation ne peut être interprétée comme remplaçant, modifiant ou portant atteinte de toute autre manière aux obligations ou aux droits des États en vertu du droit international,

ni comme autorisant un État, tout autre acteur de la vie politique, économique ou sociale, un groupe ou un individu à se livrer à une activité ou à accomplir un acte contraire aux droits de l'homme, aux libertés fondamentales, à la dignité humaine et au souci de l'environnement et des écosystèmes, vivants et non vivants.



unesco

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Secteur des sciences sociales et humaines

7, place de Fontenoy
75352 Paris 07 SP France

 ai-ethics@unesco.org

 on.unesco.org/EthicsAI

Suivez-nous

@UNESCO #AI #HumanAI

