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

Groupe d'experts ad hoc (GEAH) pour l'élaboration d'un projet de recommandation sur l'éthique de l'intelligence artificielle

Distribution limitée

SHS/BIO/AHEG-AI/2020/4
Paris, le 7 mai 2020
Original anglais

DOCUMENT FINAL

PREMIÈRE VERSION DU PROJET DE RECOMMANDATION SUR L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE

Conformément à la décision adoptée par la Conférence générale de l'UNESCO à sa 40e session (résolution 40 C/37), la Directrice générale a constitué, en mars 2020, un Groupe d'experts ad hoc (GEAH) chargé d'élaborer un projet de recommandation sur l'éthique de l'intelligence artificielle.

Compte tenu de la situation difficile liée à la pandémie de COVID-19, le GEAH a conduit ses travaux à distance de fin mars à début mai 2020, aboutissant à la production d'une première version du projet de recommandation sur l'éthique de l'intelligence artificielle, contenue dans le présent document.

Il est à noter que le GEAH continuera d'affiner cette première version jusqu'au début de septembre 2020, en tenant compte des avis reçus au cours du processus de consultation multipartite qui se tiendra de juin à juillet 2020.

Le présent document ne prétend pas être exhaustif et ne représente pas nécessairement les opinions des États membres de l'UNESCO.

PREMIÈRE VERSION DU PROJET DE RECOMMANDATION SUR L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE

PRÉAMBULE

La Conférence générale de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO), réunie à Paris à l'occasion de sa xx session,

Rappelant que l'UNESCO se propose, aux termes de son Acte constitutif, de contribuer à élever les défenses de la paix dans l'esprit des femmes et des hommes et de resserrer, par l'éducation, la science, 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,

Réfléchissant à l'influence profonde que l'intelligence artificielle (IA) pourrait avoir sur les sociétés, les écosystèmes et la vie humaine, y compris l'esprit humain, en raison notamment des nouvelles façons dont elle agit sur la pensée et la prise de décision des êtres humains et dont elle retentit sur l'éducation, la science, la culture et la communication et l'information,

Considérant que les systèmes d'IA peuvent rendre de grands services à l'humanité, mais qu'ils soulèvent également des préoccupations éthiques de fond, à l'égard, par exemple, des préjugés qu'ils sont susceptibles de comporter et d'accentuer, lesquels pourraient entraîner inégalité et exclusion et menacer la diversité culturelle et sociale ainsi que l'égalité des genres ; la nécessité d'assurer la transparence et l'intelligibilité du fonctionnement des algorithmes et des données à partir desquelles ils ont été formés ; et leurs éventuelles conséquences sur la vie privée, la liberté d'expression, les processus sociaux, économiques et politiques et l'environnement,

Ayant à l'esprit que le développement de l'IA pourrait creuser les écarts et les inégalités qui existent dans le monde et que personne ne devrait être laissé de côté contre son gré, qu'il s'agisse de profiter des avantages de l'IA ou de se prémunir contre ses conséquences négatives, tout en reconnaissant les différences de situation qui prévalent entre les pays,

Consciente que les pays à revenu faible et intermédiaire, y compris, sans s'y limiter, ceux d'Afrique, d'Amérique latine et des Caraïbes et d'Asie centrale, ainsi que les petits États insulaires en développement (PEID), connaissent une accélération de l'utilisation des technologies de l'information et de l'IA, et que l'économie numérique représente pour les sociétés créatives des défis et des possibilités immenses, ce qui exige de prendre en compte les cultures, valeurs et connaissances endogènes pour développer l'économie de ces pays,

Reconnaissant que l'IA pourrait avoir des effets bénéfiques sur l'environnement par les fonctions qu'elle assure dans la recherche écologique et climatologique, la gestion des risques de catastrophe et l'agriculture, mais que la matérialisation de ces avantages exige d'assurer un accès équitable à la technologie et de mettre en balance les avantages potentiels avec l'impact environnemental du cycle complet de production de l'IA et des technologies de l'information,

Notant que la prise en compte des risques et des préoccupations éthiques ne devrait pas freiner l'innovation, mais plutôt encourager de nouvelles pratiques de recherche et d'innovation responsables plaçant les valeurs morales et la réflexion éthique au fondement de la recherche, de la conception, du développement, du déploiement et de l'utilisation de l'IA,

Rappelant 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,

Convaincue que l'instrument normatif ici présenté, fondé sur une approche normative globale et centré sur la dignité humaine et les droits de l'homme, y compris la diversité, l'interdépendance, l'inclusion et l'équité, peut donner une orientation responsable à la recherche, à la conception, au développement, au déploiement et à l'utilisation de l'IA,

Constatant que le cadre normatif applicable à l'IA et à ses implications sociales se situe au croisement de l'éthique, des droits de l'homme, des cadres juridiques internationaux et nationaux, de la liberté de recherche et d'innovation et du bien-être de l'humanité,

Reconnaissant que les valeurs et principes éthiques ne sont pas nécessairement des normes juridiques par nature, mais qu'ils peuvent profondément influencer la définition et l'application des mesures politiques et des normes juridiques, en fournissant des orientations lorsque le champ d'application des normes n'est pas clairement circonscrit, ou lorsque ces normes ne sont pas encore établies en raison de la rapidité du progrès technologique et de la relative lenteur des réponses politiques,

Convaincue que des normes éthiques mondialement reconnues peuvent jouer un rôle utile pour harmoniser les normes juridiques relatives à l'IA dans le monde, ainsi que pour assurer l'application responsable du droit international en vigueur, sous réserve que cette application soit conforme aux cadres éthiques et ne cause aucun dommage à l'échelle locale,

Reconnaissant la Déclaration universelle des droits de l'homme (1948), notamment l'article 27 qui énonce le droit de participer au progrès scientifique et aux bienfaits qui en résultent ; les instruments internationaux relatifs aux droits de l'homme, dont la Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes (1979), la Convention des Nations Unies relative aux droits de l'enfant (1989), et la Convention des Nations Unies relative aux droits des personnes handicapées (2006) ; et la Convention de l'UNESCO sur la protection et la promotion de la diversité des expressions culturelles (2005),

Prenant acte de la Déclaration de l'UNESCO sur les responsabilités des générations présentes envers les générations futures (1997) ; de la Déclaration des Nations Unies sur les droits des peuples autochtones (2007) ; du Rapport du Secrétaire général de l'ONU de 2011 sur la suite donnée à la deuxième Assemblée mondiale sur le vieillissement (A/66/173), qui traite de la situation des droits des personnes âgées ; du Rapport du Représentant spécial du Secrétaire général de l'ONU chargé de la question des droits de l'homme et des sociétés transnationales et autres entreprises de 2011 (A/HRC/17/31), qui présente les « Principes directeurs relatifs aux entreprises et aux droits de l'homme : mise en œuvre du cadre de référence « protéger, respecter et réparer » des Nations Unies » ; 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) adoptée le 26 septembre 2019 ; de la Recommandation de l'UNESCO concernant la science et les chercheurs scientifiques (2017) ; des indicateurs de l'UNESCO sur l'universalité de l'Internet (2019), notamment les principes ROAM ; du rapport du Groupe de haut niveau du Secrétaire général de l'ONU sur la coopération numérique, intitulé « L'ère de l'interdépendance numérique » (2019) ; et des résultats et rapports des sommets mondiaux de l'Union internationale des télécommunications sur l'intelligence artificielle au service de l'intérêt général,

Prenant acte également des cadres relatifs à l'éthique de l'IA établis par d'autres organisations intergouvernementales, tels que les instruments des droits de l'homme et autres textes juridiques pertinents adoptés par le Conseil de l'Europe, y compris les travaux de son Comité ad hoc sur l'intelligence artificielle (CAHAI) ; les travaux de l'Union européenne traitant

de l'IA et ceux du Groupe d'experts de haut niveau de la Commission européenne sur l'IA, notamment les Lignes directrices en matière d'éthique pour une IA digne de confiance ; les travaux du Groupe d'experts de l'Organisation de coopération et de développement économiques (OCDE) sur l'IA (AIGO) et la Recommandation du Conseil de l'OCDE sur l'IA ; les Principes sur l'IA du G20, inspirés des travaux de l'OCDE et présentés dans la Déclaration ministérielle du G20 sur le commerce et l'économie numérique ; la Vision commune de Charlevoix sur l'avenir de l'IA adoptée par le G7 ; les travaux du Groupe de travail de l'Union africaine sur l'IA ; et les travaux du Groupe de travail de la Ligue des États arabes sur l'IA,

Soulignant qu'il est nécessaire de prêter une attention particulière aux pays à revenu faible et intermédiaire, y compris, sans s'y limiter, ceux d'Afrique, d'Amérique latine et des Caraïbes et d'Asie centrale, ainsi qu'aux PEID en raison de leur sous-représentation 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 et éthique, des systèmes de valeurs et des exigences d'équité mondiale,

Consciente qu'il existe de nombreux cadres éthiques et réglementaires relatifs à l'IA à l'échelle nationale,

Consciente également qu'il existe de nombreuses initiatives et cadres relatifs à l'IA émanant du secteur privé, des organisations professionnelles et des organisations non gouvernementales, notamment l'Initiative mondiale de l'Institut des ingénieurs électriciens et électroniciens (IEEE) sur l'éthique des systèmes autonomes et intelligents et ses travaux sur une conception conforme à l'éthique ; le livre blanc du Forum économique mondial sur une gouvernance multipartite mondiale des technologies ; les « 10 grands principes pour une intelligence artificielle éthique » de l'UNI Global Union ; la Déclaration de Montréal pour un développement responsable de l'IA ; les principes pour une intelligence artificielle harmonieuse ; et les principes du Partenariat sur l'IA,

Convaincue que l'IA peut être porteuse d'importants avantages, mais que leur matérialisation pourrait s'accompagner d'une dette d'innovation, d'un accès asymétrique aux connaissances, de limitations du droit à l'information, d'écarts en matière de capacité de créativité, de cycles de développement et de capacités humaines et institutionnelles, de restrictions de l'accès à l'innovation technologique, et d'un manque d'infrastructures et de cadres réglementaires adéquats concernant les données,

Reconnaissant qu'une concurrence économique s'exerce au sein des pays et entre eux, ainsi qu'entre entreprises multinationales, ce qui pourrait conduire à orienter les stratégies et cadres réglementaires relatifs à l'IA vers des intérêts nationaux et commerciaux, alors qu'une coopération mondiale est nécessaire pour aborder les défis posés par l'IA dans des cultures et systèmes éthiques divers et interdépendants et pour réduire les risques d'utilisation abusive,

Tenant pleinement compte du fait que le développement rapide des systèmes d'IA se heurte à des obstacles liés à la compréhension et la mise en œuvre de l'IA, qui découlent de la diversité des orientations éthiques et des cultures du monde, du manque de souplesse de la législation concernant la technologie et la société de l'information, et du risque que l'IA perturbe les normes et valeurs éthiques locales et régionales,

1. ***Adopte*** la présente Recommandation sur l'éthique de l'intelligence artificielle ;

2. ***Recommande*** aux États membres d'appliquer 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 ;

3. ***Recommande également*** aux États membres de porter la présente Recommandation à la connaissance des autorités, organismes et institutions des secteurs public, commercial et non commercial engagés dans des activités de recherche, de conception, de développement, de déploiement et d'utilisation de systèmes d'IA.

I. CHAMP D'APPLICATION

1. La présente Recommandation traite des questions éthiques soulevées par l'intelligence artificielle. Elle aborde l'éthique de l'IA en tant que cadre global de valeurs, de principes et d'actions interdépendants de nature à orienter les sociétés tout au long du cycle de vie des systèmes d'IA, en fixant la dignité et le bien-être humains pour repères, afin d'apporter des réponses responsables aux conséquences connues et inconnues des interactions des systèmes d'IA avec les êtres humains et leur environnement. Le cycle de vie d'un système d'IA englobe les processus de recherche, de conception, de développement, de déploiement et d'utilisation de ce système ; l'utilisation d'un système d'IA peut être entendue comme la maintenance, le fonctionnement, la fin d'usage et le démantèlement de ce système. Le présent instrument 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 et font l'objet d'un vaste consensus international. Aux fins de la présente Recommandation, les systèmes d'IA peuvent être envisagés comme des systèmes technologiques capables de traiter l'information par un processus s'apparentant à un comportement intelligent, et comportant généralement des fonctions d'apprentissage, de perception, d'anticipation, de planification ou de contrôle. La Recommandation aborde les systèmes d'IA selon les axes suivants :

(a) tout d'abord, les systèmes d'IA intègrent des modèles et des algorithmes qui génèrent une capacité d'apprentissage et d'exécution de tâches cognitives, par exemple formuler des recommandations et prendre des décisions dans des environnements réels et virtuels. Les systèmes d'IA sont conçus pour fonctionner à 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 approches et technologies, 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, la recherche et l'optimisation ;

   (iii) des systèmes cyberphysiques, y compris l'Internet des objets et la robotique, 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) ensuite, outre les questions éthiques semblables à celles posées par toute technologie, les systèmes d'IA soulèvent des interrogations nouvelles. Certaines ont trait à leur capacité 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. En poussant plus loin le raisonnement, les systèmes d'IA pourraient à long terme disputer aux êtres humains le sentiment d'expérience et de conscience qui leur est propre, ce qui susciterait de nouvelles inquiétudes quant à l'autonomie, la valeur et la dignité humaines, mais tel n'est pas encore le cas ;

(c) enfin, si les interrogations éthiques relatives à l'IA portent généralement sur les conséquences concrètes des systèmes d'IA sur les sociétés et les personnes humaines, elles peuvent également concerner les interactions entre ces systèmes et les humains et leur retentissement sur notre conception de l'être humain et de la technologie. Cette Recommandation reconnaît que les deux catégories de questions sont intimement liées et doivent faire partie de toute approche éthique de l'IA.

2. La présente Recommandation prête une attention particulière aux implications éthiques plus larges de l'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) les systèmes d'IA entretiennent de multiples liens avec l'éducation : ils mettent en question son rôle dans la société par leurs conséquences sur le marché de l'emploi et l'employabilité ; peuvent avoir une incidence sur les pratiques éducatives ; et rendent nécessaire l'intégration de la sensibilisation aux implications sociétales et éthiques de l'IA dans la formation des ingénieurs en IA et en informatique ;

(b) dans tous les domaines des sciences et des sciences sociales et humaines, l'IA influence nos conceptions de la compréhension et de l'explication scientifiques, ainsi que la manière dont nous appliquons la connaissance scientifique pour étayer la prise de décision ;

(c) l'IA a des conséquences sur l'identité et la diversité culturelles. Elle peut avoir des effets positifs sur les industries culturelles et créatives, mais pourrait également aboutir à une concentration accrue de l'offre, des données et des revenus de la culture entre les mains d'un petit nombre d'acteurs, avec des répercussions potentiellement négatives sur la diversité des expressions culturelles et l'égalité ;

(d) dans le domaine de la communication et de l'information, la traduction automatisée des langues est appelée à jouer un rôle croissant. Cela pourrait avoir d'importantes conséquences sur les langues et l'expression humaine, dans toutes les dimensions de la vie, ce qui oblige à adopter une attitude prudente vis-à-vis des langues humaines et de leur diversité. En outre, l'IA met en question les pratiques journalistiques et le rôle social des journalistes, des agents des médias et des producteurs de médias sociaux qui participent à des activités journalistiques, et intervient aussi bien dans la diffusion des fausses informations ou des erreurs d'interprétation que dans leur détection.

3. La présente Recommandation s'adresse aux États. Elle permet aussi, dans la mesure appropriée et pertinente, de guider les décisions ou pratiques des individus, des groupes, des communautés, des institutions et des sociétés, publiques ou privées, et en particulier des acteurs de l'IA, qui s'entendent comme ceux qui jouent un rôle actif dans le cycle de vie des systèmes d'IA, y compris les organisations et individus qui participent à la recherche, à la conception, au développement, au déploiement ou à l'utilisation de l'IA.

II. BUTS ET OBJECTIFS

4. La présente Recommandation a pour objet de formuler des valeurs et des principes éthiques ainsi que des recommandations concrètes concernant la recherche, la conception, le développement, le déploiement et l'utilisation de l'IA, en vue de mettre les systèmes d'IA au service de l'humanité, des individus, des sociétés et de l'environnement.

5. La complexité des questions éthiques qui entourent l'IA appelle des réponses elles aussi complexes, nécessitant une coopération des multiples parties prenantes aux différents niveaux et dans les différents secteurs des communautés internationales, régionales et nationales.

6. Bien que cette Recommandation s'adresse principalement aux responsables politiques des États membres ou non membres de l'UNESCO, elle entend également offrir aux organisations internationales, sociétés nationales et transnationales, ingénieurs et chercheurs, notamment en sciences humaines, naturelles et sociales, organisations non gouvernementales, organisations religieuses, ainsi qu'à la société civile, un cadre favorisant une approche multipartite fondée sur un système d'éthique mondialement reconnu, qui permettra aux parties prenantes de collaborer et d'assumer leur responsabilité commune par le biais d'un dialogue interculturel mondial.

III. VALEURS ET PRINCIPES

7. Les valeurs et principes énoncés dans la présente Recommandation ne sont pas nécessairement des normes juridiques par nature, comme indiqué dans le préambule. Leur rôle est important pour orienter les mesures politiques et les normes juridiques, puisque les valeurs recouvrent les attentes internationalement convenues concernant ce qui est positif et ce qu'il faut préserver. En tant que telles, les valeurs sous-tendent les principes.

8. Les valeurs inspirent ainsi des comportements moraux positifs conformes à la vision que la communauté internationale a de ces comportements, et constituent les fondements des principes. Les principes, quant à eux, explicitent les valeurs de manière plus concrète, de façon à faciliter l'actualisation de ces dernières dans les déclarations et les actions politiques.

III.1 VALEURS

Dignité humaine

9. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient respecter et préserver la dignité humaine. La dignité de chaque être humain fait partie des valeurs qui sont au fondement de tous les droits de l'homme et de toutes les libertés fondamentales, et tient une place essentielle dans le développement et l'adaptation des systèmes d'IA. La dignité humaine a trait à la reconnaissance de la valeur intrinsèque de chaque être humain, et n'est donc pas liée à son origine nationale, son statut juridique, sa situation socioéconomique, son genre et son orientation sexuelle, sa religion, son origine ethnique, son idéologie politique ou ses opinions.

10. Cette valeur devrait, en premier lieu, être respectée par tous les acteurs participant à la recherche, à la conception, au développement, au déploiement et à l'utilisation des systèmes d'IA ; et, en second lieu, être promue par de nouvelles mesures législatives, des initiatives de gouvernance, des modèles collaboratifs positifs de développement et d'utilisation de l'IA, ou des directives techniques et méthodologiques élaborées par les autorités nationales et internationales pour s'adapter aux progrès des technologies d'IA.

Droits de l'homme et libertés fondamentales

11. Dans le contexte de l'IA, le respect, la protection et la promotion des droits de l'homme et des libertés fondamentales signifient, en tant que valeur, que la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient être compatibles et en conformité avec les règles, principes et normes du droit international des droits de l'homme.

Ne laisser personne de côté

12. Il est essentiel de veiller à ce que la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA se déroulent dans le respect de l'humanité dans son ensemble et favorisent la créativité dans toute sa diversité. La discrimination et les préjugés, les fractures numérique et cognitive et les inégalités dans le monde doivent être traitées tout au long du cycle de vie des systèmes d'IA.

13. Par conséquent, la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA doivent être compatibles avec l'autonomisation de tous les êtres humains, en tenant compte des besoins spécifiques des différents groupes d'âge, des systèmes culturels, des personnes en situation de handicap, des femmes et des filles, et des populations défavorisées, marginalisées et vulnérables ; et ne devraient pas servir à restreindre les choix de style de vie ou le champ des expériences personnelles, ce qui inclut le caractère facultatif de l'utilisation des systèmes d'IA. En outre, il faudra s'efforcer de pallier l'absence des infrastructures, formations et compétences technologiques ainsi que des cadres juridiques nécessaires, en particulier dans les pays à revenu faible et intermédiaire.

Vivre en harmonie

14. La vie en harmonie, en tant que valeur, suppose que la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA tiennent compte de l'interdépendance de tous les êtres humains. La notion d'interdépendance repose sur le fait que chaque être humain appartient à un ensemble plus vaste, qui s'affaiblit lorsque d'autres êtres humains se trouvent diminués d'une quelconque façon.

15. Pour assurer le respect de cette valeur, la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient éviter le conflit et la violence, ne pas mettre à l'écart, chosifier les êtres humains ou menacer leur sécurité, ne pas diviser et dresser les uns contre les autres les individus et les groupes, et ne pas compromettre la coexistence harmonieuse entre les êtres humains et l'environnement naturel, car cela nuirait à l'humanité tout entière. L'objet de cette valeur est de souligner le rôle moteur que les acteurs de l'IA devraient jouer pour atteindre l'objectif de la vie en harmonie, qui est d'assurer un avenir bénéfique pour tous.

Crédibilité

16. Les systèmes d'IA devraient être crédibles. La crédibilité est un concept sociotechnique en vertu duquel la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient établir une confiance entre les personnes et dans les systèmes d'IA, plutôt que trahir cette confiance.

17. La confiance est à conquérir pour chaque type d'utilisation et constitue, plus généralement, un indicateur du degré d'acceptation sociale des systèmes d'IA. Les personnes devraient donc avoir de bonnes raisons de ne pas douter des avantages des technologies de l'IA lorsque des mesures appropriées sont prises pour atténuer les risques.

Protection de l'environnement

18. La protection de l'environnement, en tant que valeur, suppose de veiller à ce que la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA prennent en compte la promotion du bien-être environnemental. Tous les acteurs participant au cycle de vie des systèmes d'IA devraient respecter l'ensemble des lois nationales et internationales pertinentes en matière de protection de l'environnement et de développement durable, afin de réduire autant que possible les facteurs de risque associés au changement

climatique, y compris les émissions de carbone des systèmes d'IA, et d'empêcher l'exploitation et l'épuisement des ressources naturelles, qui contribuent à la dégradation de l'environnement.

19. En parallèle, les systèmes d'IA devraient être mis à profit pour proposer des solutions permettant de protéger l'environnement et de préserver la planète, en soutenant des approches telles que l'économie circulaire.

III.2 PRINCIPES

20. Étant donné que tout système d'IA comporte des caractéristiques circonstancielles essentielles et en évolution subordonnées aux facteurs humain et technologique, les principes sont répartis en deux groupes.

21. Le premier groupe rassemble les principes qui reflètent les caractéristiques associées à l'interface entre êtres humains et technologies, c'est-à-dire les interactions entre personnes humaines et systèmes d'IA. Il est à noter que la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA ont une double incidence sur les agents humains : ils étendent la marge d'autonomie et de prise de décision des machines, et ont des effets positifs et négatifs sur la qualité des agents humains.

22. Le deuxième groupe réunit les principes qui reflètent les propriétés des systèmes d'IA pertinentes pour faire en sorte que la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA se déroulent conformément aux attentes internationalement reconnues en matière de comportement éthique.

GROUPE 1

Au service de l'humanité et de son épanouissement

23. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient avoir pour but l'épanouissement des êtres humains et de leur environnement. Tout au long de leur cycle de vie, les systèmes d'IA devraient améliorer la qualité de vie et favoriser l'exercice de tous les droits de l'homme de chaque être humain, en laissant aux individus ou groupes le soin de définir la notion de « qualité de vie », tant qu'aucune personne humaine ne subit d'atteinte physique ou mentale ou ne voit sa dignité rabaissée par cette définition.

24. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA pourraient viser à faciliter la communication avec des personnes vulnérables, y compris, sans s'y limiter, les enfants et les personnes âgées ou malades, mais ne devraient jamais chosifier les êtres humains, porter atteinte à leur dignité ou violer leurs droits.

Proportionnalité

25. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA ne devraient pas aller au-delà de ce qui est nécessaire pour atteindre des buts ou objectifs légitimes, et devraient être adaptés au contexte dans lequel ils interviennent.

26. Le choix de la méthode d'IA devrait être justifié des manières suivantes : (a) la méthode d'IA retenue devrait être souhaitable et adéquate pour atteindre un objectif donné ; (b) la méthode d'IA retenue ne devrait pas porter d'atteinte excessive aux valeurs fondamentales énoncées dans la présente Recommandation ; (c) la méthode d'IA retenue devrait être adaptée au contexte.

Surveillance et décision humaines

27. Il devrait toujours être possible d'attribuer la responsabilité éthique et juridique de la recherche, de la conception, du développement, du déploiement et de l'utilisation des systèmes d'IA à une personne physique ou une entité juridique existante. Ainsi, la surveillance humaine renvoie non seulement à la surveillance humaine individuelle, mais aussi à la surveillance publique.

28. Les êtres humains devront peut-être, dans certains cas, partager les fonctions de contrôle avec les systèmes d'IA à des fins d'efficacité, mais la décision de céder ces fonctions dans des contextes limités reste leur prérogative. En effet, la recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient avoir pour but d'assister les êtres humains dans la prise de décision et l'exécution de tâches, mais jamais de se substituer à leur responsabilité ultime.

Durabilité

29. Dans le contexte de la promotion de l'édification de sociétés durables, les acteurs de l'IA devraient respecter les dimensions sociales, économiques et environnementales du développement durable de l'humanité dans son ensemble et de l'environnement. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient chercher à favoriser la durabilité visée par les cadres internationalement reconnus, tels que les objectifs de développement durable.

Diversité et inclusion

30. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient respecter et favoriser la diversité et l'inclusion en se conformant, à tout le moins, aux règles, principes et normes du droit international des droits de l'homme, notamment en ce qui concerne la diversité et l'inclusion démographiques, culturelles et sociales.

Vie privée

31. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient respecter, protéger et favoriser la vie privée, qui constitue un droit essentiel pour la protection de la dignité humaine et des agents humains. Des mécanismes appropriés de gouvernance des données devraient être en place tout au long du cycle de vie des systèmes d'IA, notamment pour la collecte et le contrôle de l'utilisation des données, sous la forme d'une procédure de consentement éclairé, d'autorisations et de divulgations relatives à l'application et l'utilisation des données, garantissant les droits des personnes à l'égard de leurs données et leur accès à celles-ci.

Sensibilisation et éducation

32. La sensibilisation du public et sa compréhension des technologies d'IA et de la valeur des données devraient être favorisées par le biais de l'éducation, de campagnes publiques et de formations, afin d'assurer une participation publique effective et de permettre aux citoyens de prendre des décisions éclairées concernant leur utilisation des systèmes d'IA. Les enfants devraient être protégés des dangers raisonnablement prévisibles associés aux systèmes d'IA, avoir accès à ces systèmes à travers l'éducation et la formation, et ne pas voir leurs capacités affaiblies par leurs interactions avec les systèmes d'IA.

Gouvernance multipartite et adaptative

33. La gouvernance de l'IA doit prendre en compte les évolutions des technologies et des modèles commerciaux associés, être inclusive (en assurant la participation de multiples parties prenantes), être potentiellement répartie à différents niveaux, et garantir, grâce à une approche systémique interdomaine, des réponses de gouvernance adaptées à l'objectif visé.

34. La gouvernance devrait envisager un éventail de réponses varié, allant de la gouvernance « douce », qui passe par l'autorégulation et des processus de certification, à la gouvernance « dure », qui se traduit par des lois nationales et, si cela est possible et nécessaire, des instruments internationaux. Pour éviter toute conséquence négative et tout préjudice involontaire, la gouvernance devrait comporter des aspects d'anticipation, de protection, de suivi des conséquences, de mise en application et de réparation.

GROUPE 2

Équité

35. Tout au long du cycle de vie des systèmes d'IA, les acteurs de l'IA devraient respecter l'équité et l'inclusion, et faire tout leur possible pour réduire au maximum les préjugés sociotechniques et éviter de les renforcer ou de les perpétuer, notamment les préjugés culturels et ceux liés à la race, à l'origine ethnique, au genre et à l'âge.

Transparence et explicabilité

36. Si, en principe, il convient de tout mettre en œuvre pour améliorer la transparence et l'explicabilité des systèmes d'IA afin de gagner la confiance des êtres humains, le degré de transparence et d'explicabilité devrait toujours être adapté au contexte d'utilisation, car de nombreux arbitrages sont effectués entre la transparence et l'explicabilité et d'autres principes, tels que la sûreté et la sécurité.

37. La transparence suppose de permettre aux personnes de comprendre les modalités de la recherche, de la conception, du développement, du déploiement et de l'utilisation des systèmes d'IA, en fonction du contexte d'utilisation 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, mais cela n'inclut généralement pas la communication de codes ou d'ensembles de données spécifiques. En ce sens, la transparence est un enjeu sociotechnique, le but étant d'établir la confiance des êtres humains dans les systèmes d'IA.

38. L'explicabilité implique de rendre les résultats des systèmes d'IA intelligibles et de fournir des renseignements à leur sujet. L'explicabilité des modèles d'IA renvoie également à l'intelligibilité des intrants, des extrants, du comportement des différents modules algorithmiques et de leur contribution aux résultats des modèles. L'explicabilité est donc étroitement liée à la transparence, puisque les résultats et les sous-processus qui y conduisent doivent être intelligibles et traçables, en fonction du contexte d'utilisation.

Sûreté et sécurité

39. La recherche, la conception, le développement, le déploiement et l'utilisation des systèmes d'IA devraient éviter les préjudices involontaires (risques liés à la sûreté) et les vulnérabilités aux attaques (fonctions de sécurité), afin de garantir la sûreté et la sécurité tout au long du cycle de vie des systèmes d'IA.

40. Les gouvernements devraient jouer un rôle de premier plan pour garantir la sûreté et la sécurité des systèmes d'IA, notamment en établissant des normes nationales et internationales conformes aux règles, normes et principes applicables du droit international des droits de l'homme. Pour éviter de causer des dommages catastrophiques, un soutien constant devrait être apporté à la recherche stratégique sur les risques de sûreté et de sécurité potentiels liés aux différentes approches utilisées pour bâtir une IA durable.

Responsabilité et redevabilité

41. Les acteurs de l'IA devraient endosser une responsabilité morale et juridique, conformément au droit international des droits de l'homme en vigueur et aux directives éthiques établies tout au long du cycle de vie des systèmes d'IA. La responsabilité 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.

42. Des mécanismes appropriés devraient être mis en place pour assurer la redevabilité des systèmes d'IA et de leurs résultats. Des dispositifs techniques et institutionnels devraient être envisagés pour garantir la vérifiabilité et la traçabilité (du fonctionnement) des systèmes d'IA.

IV. DOMAINES D'ACTION STRATÉGIQUE

PREMIER OBJECTIF DES ACTIONS : UNE GESTION ÉTHIQUE

43. Garantir l'alignement de la recherche en matière d'IA ainsi que de la conception, du développement, du déploiement et de l'utilisation de l'IA sur les valeurs éthiques fondamentales que sont les droits de l'homme et les principes de diversité et d'inclusion, entre autres.

Action stratégique 1 : Promouvoir la diversité et l'inclusion

44. Les États membres devraient coopérer avec les organisations internationales pour garantir la participation active aux discussions internationales concernant l'IA de tous les États membres, en particulier des PRITI. Cela peut passer par la mise à disposition de fonds, le fait d'assurer une participation régionale égale, ou tout autre mécanisme.

45. Les États membres devraient exiger des acteurs de l'IA qu'ils révèlent et combattent tout stéréotype culturel et social présent dans le fonctionnement des systèmes d'IA volontairement ou par négligence, et qu'ils veillent à ce que les ensembles de données d'entraînement ne favorisent pas les inégalités culturelles et sociales. Des mécanismes devraient être adoptés pour permettre aux utilisateurs finaux de signaler de tels inégalités, biais et stéréotypes.

46. Les États membres devraient veiller à ce que les acteurs de l'IA prennent en considération et respectent les diversités culturelles et sociales actuelles, notamment les coutumes locales et les traditions religieuses, lors de leurs recherches et lors de la conception, du développement, du déploiement et de l'utilisation des systèmes d'IA, tout assurant la conformité de ces systèmes avec les normes et règles internationales en matière de droits de l'homme.

47. Les États membres devraient s'efforcer de combler les lacunes en matière de diversité actuellement constatées dans le développement des systèmes d'IA, notamment en matière de diversité des ensembles de données d'entraînement et des acteurs de l'IA eux-mêmes. Les États membres devraient coopérer avec tous les secteurs, les organisations internationales et régionales et d'autres entités pour donner aux femmes et aux filles les moyens de participer à toutes les étapes du cycle de vie d'un système d'IA en proposant des mesures incitatives, l'accès à des mentors et à des modèles, ainsi qu'une protection contre le harcèlement. Ils

devraient également s'efforcer de rendre le domaine de l'IA plus accessible aux personnes d'origines ethniques variées et aux personnes handicapées. En outre, il convient de promouvoir l'égalité d'accès aux avantages des systèmes d'IA, en particulier pour les groupes marginalisés.

48. Les États membres devraient coopérer avec les organisations internationales pour placer l'éthique de l'IA au centre des préoccupations en intégrant des discussions sur des questions éthiques en lien avec l'IA dans les forums internationaux, intergouvernementaux et multipartites pertinents.

DEUXIÈME OBJECTIF DES ACTIONS : L'ÉVALUATION DE L'IMPACT

49. Développer les capacités d'observation et d'anticipation pour réagir en temps voulu aux conséquences négatives ou autres conséquences involontaires découlant des systèmes d'IA.

Action stratégique 2 : Faire face à l'évolution du marché du travail

50. Les États membres devraient s'employer à évaluer et traiter l'impact de l'IA sur le marché du travail, ainsi que son incidence sur les besoins en matière d'éducation. Cela peut nécessiter la mise en place d'un éventail plus large de « compétences de base » à tous les niveaux d'éducation pour donner 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 de l'IA. En plus des compétences techniques et spécialisées, il convient d'enseigner des compétences comme « apprendre à apprendre », la communication, 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 scolaires en fonction de celles-ci.

51. Les États membres devraient coopérer avec des entités privées, des ONG et d'autres parties prenantes pour assurer une transition équitable aux employés menacés. Cela suppose de mettre en place des programmes de perfectionnement et de reconversion, de trouver des moyens créatifs de retenir les employés pendant ces périodes de transition, et d'envisager des programmes de « couverture sociale » pour ceux qui ne peuvent pas se reconvertir.

52. Les États membres devraient encourager les chercheurs à analyser l'impact de l'IA sur le marché du travail local afin d'anticiper les tendances et défis à venir. Ces études devraient mettre en évidence les secteurs économiques, sociaux et géographiques qui seront le plus touchés par l'introduction massive de l'IA.

53. Les États membres devraient élaborer des politiques relatives à la population active visant à soutenir les femmes et les populations sous-représentées pour que personne ne soit exclu de l'économie numérique basée sur l'IA. Il convient d'envisager, et de mettre en œuvre si possible, des investissements spéciaux en faveur de programmes ciblés destinés à améliorer la préparation, l'employabilité, l'avancement professionnel et l'épanouissement professionnel des femmes et des populations sous-représentées.

Action stratégique 3 : Faire face aux répercussions économiques et sociales de l'IA

54. Les États membres devraient concevoir des mécanismes destinés à empêcher la monopolisation de l'IA et les inégalités qui en résulteraient, que les monopoles concernent les données, la recherche, la technologie, le marché ou d'autres domaines.

55. Les États membres devraient coopérer avec des organisations internationales, des entités privées et non gouvernementales pour fournir les connaissances adéquates en matière

d'IA à la population, en particulier dans les PRITI, afin de réduire la fracture numérique et les inégalités d'accès au numérique découlant de l'adoption à grande échelle de systèmes d'IA.

56. Les États membres devraient mettre en place des mécanismes de suivi et d'évaluation pour les initiatives et les politiques liées à l'éthique de l'IA. Ces mécanismes peuvent être : un répertoire recensant les initiatives en matière de conformité éthique dans les domaines de compétence de l'UNESCO, un mécanisme de partage d'expériences permettant aux États membres de demander à d'autres États membres leur avis sur leurs politiques et initiatives, ou un guide permettant aux développeurs de systèmes d'IA d'évaluer le respect des recommandations stratégiques mentionnées dans le présent document.

57. Les États membres sont encouragés à envisager un mécanisme de certification pour les systèmes d'IA semblables à ceux utilisés pour les dispositifs médicaux. Cela peut inclure différentes catégories de certification selon la sensibilité du domaine d'application et selon l'impact escompté sur les vies humaines, l'environnement, les considérations éthiques telles que l'égalité, la diversité et les valeurs culturelles, entre autres. Un tel mécanisme pourrait inclure différents niveaux d'audit des systèmes, des données et de la conformité éthique. Dans le même temps, il ne doit pas entraver l'innovation ni désavantager les petites entreprises ou les startups en exigeant de grandes quantités de documents administratifs. Ce mécanisme inclurait en outre un volet consacré à un suivi régulier pour garantir la fiabilité et le maintien de l'intégrité et de la conformité du système d'IA tout au long de sa durée de vie, en exigeant une nouvelle certification si nécessaire.

58. Les États membres devraient encourager les sociétés privées 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 ou tout autre mécanisme visant à superviser les efforts déployés concernant l'évaluation de l'impact, le contrôle et le suivi continu, ainsi qu'à assurer la conformité éthique des systèmes d'IA.

59. 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. Il est fondamental de trouver un équilibre entre les métadonnées et la protection de la vie privée des utilisateurs lors de l'élaboration de ces stratégies.

Action stratégique 4 : Les répercussions sur la culture et l'environnement

60. Les États membres sont encouragés à mettre en place, lorsque cela se justifie, des systèmes d'IA dans les domaines de la préservation, de l'enrichissement et de la compréhension du patrimoine culturel, tant matériel qu'immatériel, dont les langues rares, par exemple en instaurant ou en actualisant des programmes éducatifs concernant l'application des systèmes d'IA dans ces domaines, à l'intention des institutions et du public.

61. Les États membres sont encouragés à examiner et à traiter les répercussions des systèmes d'IA, en particulier des applications de traitement du langage naturel telles que la traduction automatique et les assistants vocaux, sur les nuances du langage humain. Cet examen peut inclure la maximisation des bienfaits de ces systèmes par la réduction des écarts culturels et l'amélioration de la compréhension humaine, ainsi que les incidences négatives telles que la réduction de la présence des langues rares, des dialectes locaux, et des variations tonales et culturelles associées à l'expression et au langage humains.

62. 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. Cette recherche devrait se fonder sur de multiples normes, principes, protocoles, approches disciplinaires, sur une analyse de la modification des habitudes, ainsi que sur une évaluation minutieuse des impacts culturels et sociétaux en aval.

63. Les États membres devraient promouvoir l'éducation à l'IA pour les artistes et les professionnels de la création de sorte de déterminer la pertinence d'une application de l'IA dans leur profession, l'IA étant actuellement utilisée pour créer, produire, distribuer et diffuser une grande variété de biens et services culturels, en gardant à l'esprit l'importance de préserver le patrimoine et la diversité culturels.

64. Les États membres devraient sensibiliser les industries culturelles locales et les startups travaillant dans le domaine de la culture et promouvoir l'évaluation des outils d'IA auprès d'elles afin d'éviter le risque d'une plus grande concentration sur le marché culturel.

65. Les États membres devraient s'efforcer d'évaluer et de réduire l'impact environnemental des systèmes d'IA, ce qui inclut, sans s'y limiter, leur empreinte carbone. Ils devraient en outre mettre en place des mesures incitatives pour favoriser les solutions environnementales basées sur une IA éthique et faciliter leur adoption dans différents contextes. L'IA peut ainsi être utilisée pour :

(a) accélérer la protection, le suivi et la gestion des ressources naturelles ;

(b) soutenir la prévention, le contrôle et la gestion des problèmes liés au climat ;

(c) soutenir un écosystème alimentaire plus efficace et durable ;

(d) accélérer l'accès à l'énergie verte et son adoption massive.

TROISIÈME OBJECTIF DES ACTIONS : LE RENFORCEMENT DES CAPACITÉS EN MATIÈRE D'ÉTHIQUE DE L'IA

66. Développer les capacités humaines et institutionnelles pour permettre une évaluation de l'impact, une supervision et une gouvernance éthiques.

Action stratégique 5 : Promouvoir l'éducation et la sensibilisation à l'éthique de l'IA

67. Les États membres devraient encourager l'inclusion de l'éthique de l'IA dans les programmes de tous les niveaux scolaires et universitaires conformément à leurs programmes d'éducation nationale et à leurs traditions, et promouvoir une collaboration croisée entre les domaines techniques et les sciences humaines et sociales. Les cours en ligne et les ressources numériques devraient être élaborés dans les langues locales et dans des formats accessibles aux personnes handicapées.

68. Les États membres devraient promouvoir l'acquisition de « compétences préalables » à l'éducation à l'IA, telles que les compétences de base en lecture, en écriture, en calcul, et des compétences en programmation, en particulier dans les pays où il existe des lacunes notables dans l'enseignement de ces compétences.

69. Les États membres devraient assouplir les programmes universitaires et faciliter leur mise à jour, compte tenu du rythme accéléré des innovations concernant les systèmes d'IA. Par ailleurs, la mise en cohérence des formations en ligne et continues et le cumul des diplômes devraient être envisagés pour faciliter la souplesse et l'actualisation des programmes d'enseignement.

70. Les États membres devraient promouvoir des programmes généraux de sensibilisation à l'IA, ainsi que l'accès de tous aux connaissances relatives aux possibilités et aux défis découlant de l'IA. Ces connaissances devraient être accessibles aux groupes techniques et non techniques et cibler particulièrement les populations sous-représentées.

71. Les États membres devraient encourager les initiatives de recherche sur l'utilisation de l'IA dans l'enseignement, la formation des enseignants et l'apprentissage en ligne, entre autres sujets, de sorte d'accroître les possibilités et d'atténuer les difficultés et les risques associés à ces technologies. Cela devrait toujours aller de pair avec une évaluation adéquate de l'impact de la qualité de l'éducation et de l'impact sur les élèves et les enseignants de l'utilisation de l'IA, et garantir que l'IA confère plus d'autonomie à ces deux groupes et améliore leur expérience.

72. Les États membres devraient soutenir les accords de coopération entre les établissements universitaires et les entreprises pour combler les lacunes en matière de compétences exigées, et promouvoir la collaboration entre les secteurs d'activité, les universités, la société civile et les autorités en vue de l'alignement des stratégies et programmes de formation fournis par les établissements d'enseignement sur les besoins des employeurs. Il convient de promouvoir des approches de l'apprentissage de l'IA qui soient basées sur des projets, ce qui permet d'établir des partenariats entre les entreprises, les universités et les centres de recherche.

73. Les États membres devraient particulièrement promouvoir la participation des femmes, des personnes d'origines raciales et culturelles variées et des personnes handicapées aux programmes d'éducation à l'IA, de l'école primaire à l'enseignement supérieur, ainsi que promouvoir le suivi et le partage de bonnes pratiques avec d'autres États membres.

Action stratégique 6 : Promouvoir la recherche sur l'éthique de l'IA

74. Les États membres devraient promouvoir la recherche sur l'éthique de l'IA, soit par des investissements directs, soit par l'élaboration de mesures visant à inciter les secteurs public et privé à investir dans ce domaine.

75. Les États membres devraient veiller à ce que les chercheurs en IA soient formés à l'éthique de la recherche et leur demander de tenir compte de considérations éthiques lors de l'organisation de leurs recherches et dans les produits finaux, en particulier les analyses des ensembles de données qu'ils utilisent, la façon dont ils sont annotés et la qualité et la portée des résultats.

76. Lorsque cela est possible, les États membres et les sociétés privées devraient faciliter l'accès de la communauté scientifique nationale aux données à des fins de recherche afin de renforcer les capacités de cette communauté, en particulier dans les pays en développement. Cet accès ne doit pas se faire au détriment de la vie privée des citoyens.

77. Les États membres devraient promouvoir la diversité de genre dans la recherche sur l'IA dans le milieu universitaire et dans l'industrie en proposant des incitations aux femmes pour qu'elles s'engagent 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 bonnes pratiques sur la façon de promouvoir la diversité.

78. Les États membres et les organismes de financement devraient promouvoir des recherches interdisciplinaires sur l'IA par l'intégration de disciplines autres que les sciences, la technologie, l'ingénierie et les mathématiques (STEM), telles que le droit, les relations internationales, les sciences politiques, l'éducation, la philosophie, la culture et les études

linguistiques, afin de garantir une approche critique de la recherche sur l'IA et un suivi adéquat des utilisations abusives ou effets préjudiciables possibles.

QUATRIÈME OBJECTIF DES ACTIONS : LE DÉVELOPPEMENT ET LA COOPÉRATION INTERNATIONALE

79. Assurer une approche coopérative et éthique de l'utilisation de l'IA dans les applications au service du développement, cette technologie offrant une formidable occasion d'accélérer les efforts en faveur du développement.

Action stratégique 7 : Promouvoir une utilisation éthique de l'IA dans le domaine du développement

80. Les États membres devraient encourager une utilisation éthique de l'IA dans les domaines du développement tels que les soins de santé, l'agriculture/l'approvisionnement alimentaire, l'éducation, la culture, l'environnement, la gestion de l'eau, la gestion des infrastructures, la planification et la croissance économiques, entre autres.

81. Les États membres et les organisations internationales devraient s'efforcer de mettre en place des cadres de coopération internationale dans le domaine de l'IA au service du développement, notamment en fournissant des compétences techniques, des financements, des données, des connaissances spécialisées, des infrastructures, et en facilitant l'organisation d'ateliers entre experts techniques et commerciaux pour remédier aux problèmes complexes liés au développement, en particulier pour les PRITI et les PMA.

82. Les États membres devraient s'employer à promouvoir les collaborations internationales en matière de recherche sur l'IA, y compris les centres de recherche et les réseaux qui encouragent une plus grande participation des chercheurs venant des PRITI et d'autres zones géographiques émergentes.

Action stratégique 8 : Promouvoir la coopération internationale en matière d'éthique de l'IA

83. Les États membres devraient coopérer dans le cadre d'organisations internationales et d'instituts de recherche pour mener des recherches sur l'éthique de l'IA. Les entités publiques et privées devraient veiller à ce que les algorithmes et les données utilisés dans un grand nombre de domaines de l'IA – de la surveillance policière et de la justice pénale à l'emploi, la santé et l'éducation – soient appliqués de manière égale et équitable, notamment en examinant quels types d'égalité et d'équité conviennent à différents contextes et cultures, et en étudiant comment les associer à des solutions techniquement réalisables.

84. Les États membres devraient encourager la coopération internationale en matière de développement et de déploiement de l'IA pour éliminer les clivages géotechnologiques. Cela nécessite un effort multipartite aux niveaux national, régional et international. Des échanges/consultations technologiques devraient être organisés entre les États membres et leur population, entre les secteurs public et privé, au sein des États membres et entre eux.

CINQUIÈME OBJECTIF DES ACTIONS : UNE GOUVERNANCE POUR L'ÉTHIQUE DE L'IA

85. Promouvoir et guider l'inclusion de considérations éthiques dans la gouvernance des systèmes d'IA.

Action stratégique 9 : Mettre en place des mécanismes de gouvernance pour l'éthique de l'IA

86. Les États membres devraient s'assurer que tout mécanisme de gouvernance de l'IA est :

(a) inclusif : invite et encourage la participation de représentants des communautés autochtones, des femmes, des jeunes et des personnes âgées, des personnes handicapées, et d'autres groupes minoritaires et sous-représentés ;

(b) transparent : accepte la supervision de structures nationales compétentes ou de tiers de confiance. Pour les médias, il pourrait s'agir d'une équipe spéciale transsectorielle qui vérifierait les sources ; pour les entreprises spécialisées dans la technologie, il pourrait s'agir d'audits externes des processus de conception, de déploiement et d'audit interne ; pour les États membres, il pourrait s'agir d'examens menés par des instances relatives aux droits de l'homme ;

(c) multidisciplinaire : toute question devrait être envisagée de manière globale et non seulement du point de vue technologique ;

(d) multilatéral : des accords internationaux devraient être établis pour atténuer et réparer tout préjudice pouvant apparaître dans un pays, qui serait causé par une entreprise ou un utilisateur basé dans un autre pays. Cela n'empêche pas les différents pays et régions d'élaborer leurs propres règles adaptées à leur culture.

87. Les États membres devraient favoriser le développement et l'accessibilité d'un écosystème numérique à l'appui d'une IA éthique. Cet écosystème se composerait notamment des technologies et infrastructures numériques et des mécanismes de partage des connaissances en matière d'IA, le cas échéant. À cet égard, les États membres 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 fiducies de données (« data trusts »), afin de favoriser le partage des données de façon sûre, équitable, légale et éthique, entre autres.

88. Les États membres devraient encourager l'élaboration et l'utilisation de lignes directrices comparables en matière d'IA, incluant des aspects éthiques aux niveaux mondial et régional, et réunir les données nécessaires pour pouvoir évaluer, suivre et contrôler les progrès accomplis dans la mise en œuvre éthique des systèmes d'IA.

89. Les États membres devraient envisager d'élaborer et de mettre en œuvre un cadre juridique international pour encourager la coopération internationale entre les États et d'autres parties prenantes.

Action stratégique 10 : Garantir la fiabilité des systèmes d'IA

90. Les États membres et les sociétés privées devraient mettre en œuvre des mesures appropriées pour surveiller toutes les phases du cycle de vie d'un système d'IA, y compris le comportement des algorithmes en charge de la prise de décision, 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.

91. Les États membres devraient s'efforcer de définir des exigences claires en matière de transparence et d'explicabilité des systèmes d'IA :

(a) Domaine d'application : le besoin de transparence et d'explicabilité peut être plus important dans certains secteurs que dans d'autres (maintien de l'ordre, sécurité, éducation ou soins de santé) ;

(b) Public cible : le niveau d'information concernant les algorithmes et les résultats d'un système d'IA et la forme de l'explication requise peuvent varier en fonction de la personne qui demande l'explication (utilisateurs, spécialistes du domaine, développeurs, etc.) ;

(c) Faisabilité : de nombreux algorithmes d'IA ne sont pas encore explicables ; pour d'autres, l'explicabilité ajoute une importante surcharge de mise en œuvre. Tant que l'explicabilité complète avec un impact minimal sur la fonctionnalité n'est techniquement pas possible, il y aura un compromis entre la précision/qualité d'un système et son niveau d'explicabilité.

92. Les États membres devraient encourager la recherche sur la transparence et l'explicabilité en y consacrant des fonds supplémentaires pour différents domaines et à différents niveaux (technique, langage naturel, etc.).

93. Les États membres et les organisations internationales devraient envisager d'élaborer des normes internationales qui décriraient des niveaux de transparence mesurables et vérifiables pour pouvoir évaluer les systèmes de manière objective et déterminer les niveaux de conformité.

Action stratégique 11 : Garantir la responsabilité, la redevabilité et le respect de la vie privée

94. Les États membres devraient revoir et adapter, le cas échéant, les cadres réglementaires et juridiques 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 gouvernements devraient définir des cadres de responsabilité ou clarifier l'interprétation des cadres existants pour qu'il soit possible d'attribuer la responsabilité des décisions et du comportement des systèmes d'IA. Lors de l'élaboration de cadres réglementaires, les gouvernements devraient notamment tenir compte du fait que la responsabilité et la redevabilité devraient toujours incomber à une personne physique ou morale ; la responsabilité ne devrait pas être déléguée à un système d'IA, ni une personnalité juridique donnée à un système d'IA.

95. Les États membres sont encouragés à mettre en place des études d'impact pour identifier et analyser les avantages et les risques des systèmes d'IA, ainsi que des mesures de prévention, d'atténuation et de suivi des risques. L'évaluation des risques devrait mettre en évidence les répercussions sur les droits de l'homme et l'environnement, ainsi que les incidences éthiques et sociales conformément aux principes énoncés dans la présente Recommandation. Les gouvernements devraient adopter un cadre réglementaire qui définisse une procédure permettant aux autorités publiques de mener à bien des études sur l'impact des systèmes d'IA qu'elles ont acquis, développés et/ou déployés 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. Dans le cadre de l'étude d'impact, les autorités publiques devraient être tenues de procéder à une auto-évaluation des systèmes d'IA existants et proposés, qui devrait notamment analyser si l'utilisation de systèmes d'IA dans un domaine particulier du secteur public est appropriée et indiquer la méthode appropriée. L'étude doit 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, ainsi qu'inclure un examen externe des systèmes d'IA. Enfin, une telle étude devrait être multidisciplinaire, multipartite, multiculturelle, pluraliste et inclusive.

96. Les États membres devraient impliquer tous les acteurs de l'écosystème d'IA (représentants de la société civile, forces de l'ordre, assureurs, investisseurs, industriels, ingénieurs, avocats et utilisateurs, entre autres) dans un processus visant à établir des normes lorsqu'il n'y en a pas. Les normes peuvent évoluer en bonnes pratiques et en lois. Les États

membres sont également encouragés à utiliser des mécanismes tels que des bacs à sable réglementaires pour accélérer la formulation de lois et politiques et suivre ainsi l'évolution rapide des nouvelles technologies, et pour que les lois puissent être testées dans un environnement sûr avant d'être officiellement adoptées.

97. Les États membres devraient veiller à ce que les préjudices causés à des utilisateurs par le biais de systèmes d'IA puissent faire l'objet d'enquêtes, de sanctions et de réparations, notamment en encourageant les entreprises du secteur privé à mettre en place des mécanismes de réparation. La vérifiabilité et la traçabilité des systèmes d'IA, en particulier de ceux qui sont autonomes, devraient être encouragées à cette fin.

98. Les États membres devraient appliquer les garanties appropriées concernant le droit fondamental des individus au respect de la vie privée, notamment par l'adoption ou la mise en œuvre de cadres législatifs qui assurent une protection appropriée, conforme au droit international. En l'absence d'une telle législation, les États membres devraient vivement encourager l'ensemble des acteurs de l'IA, y compris les sociétés privées qui développent et exploitent des systèmes d'IA, à appliquer le principe du respect de la vie privée dès la conception de leurs systèmes.

99. Les États membres devraient veiller à ce que les individus puissent superviser l'utilisation qui est faite de leurs informations/données privées, et en particulier à ce qu'ils conservent le droit d'accéder à leurs données personnelles et le « droit à l'oubli numérique ».

100. Les États membres devraient assurer une sécurité accrue pour les données permettant d'identifier une personne ou les données qui, si elles étaient divulguées, risqueraient de causer des dommages, des blessures ou des difficultés exceptionnelles à une personne. 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 ; des données personnelles concernant l'origine « raciale » ou ethnique, les opinions politiques, l'appartenance à un syndicat, les croyances religieuses ou autres, la santé ou la vie sexuelle.

101. Les États membres devraient chercher à adopter une approche commune des données qui leur permettraient de promouvoir l'interopérabilité des ensembles de données, de garantir leur fiabilité et de faire preuve d'une extrême vigilance dans la supervision de leur collecte et de leur utilisation. 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 avec le consentement des personnes concernées, lorsque le consentement est requis par la loi, ainsi que d'encourager des pratiques éthiques en matière d'IA, notamment grâce au partage de données de qualité dans un espace commun de données fiables et sécurisées.

V. SUIVI ET ÉVALUATION

102. Les États membres – en fonction de leur situation, de leur mode de gouvernement et de leur Constitution – devraient assurer le suivi et l'évaluation des politiques, programmes et mécanismes relatifs à l'éthique de l'IA en combinant, selon les cas, des approches quantitatives et qualitatives. Les États membres sont invités à envisager :

(a) de mettre en place des mécanismes de recherche adaptés pour mesurer l'efficacité et l'efficience des politiques et des mesures incitatives relatives à l'éthique de l'IA à l'aune des objectifs définis ;

(b) de recueillir et de diffuser – avec l'appui de l'UNESCO et des communautés internationales de l'éthique de l'IA – des données sur les progrès accomplis, des bonnes pratiques, des innovations et des rapports de recherche relatifs à l'éthique de l'IA et à ses retombées.

103. Les mécanismes possibles de suivi et d'évaluation peuvent inclure un observatoire de l'IA portant sur la conformité éthique dans l'ensemble des domaines de compétence de l'UNESCO, un mécanisme de partage d'expériences permettant aux États membres de réagir à leurs initiatives respectives, et un dispositif de « mesure de la conformité » permettant aux développeurs des systèmes d'IA de déterminer à quel point ils respectent les recommandations stratégiques mentionnées dans le présent document.

104. Il conviendrait d'élaborer des outils et indicateurs appropriés pour mesurer l'efficacité et l'efficience des politiques relatives à l'éthique de l'IA par rapport aux normes, priorités et cibles convenues, y compris des cibles spécifiques pour les groupes défavorisés et vulnérables. Cela pourrait nécessiter notamment des évaluations des établissements, prestataires et programmes publics et privés, 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 à la législation sur la protection des données.

105. Les processus de suivi et d'évaluation devraient assurer une large participation des parties prenantes concernées, notamment, mais pas exclusivement, des personnes des différents groupes d'âge, des personnes handicapées, des femmes et des filles, des populations défavorisées, marginalisées et vulnérables, ainsi que garantir le respect de la diversité sociale et culturelle, dans le but d'améliorer les processus d'apprentissage et de renforcer les liens entre constatations, prise de décision, transparence et redevabilité des résultats.

VI. UTILISATION ET MISE EN ŒUVRE DE LA PRÉSENTE RECOMMANDATION

106. 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, dont les activités sont en rapport avec le champ d'application et les objectifs de la présente Recommandation.

107. Les États membres et les parties prenantes identifiées dans la présente Recommandation devraient prendre toutes les mesures en leur pouvoir pour faire appliquer les dispositions énoncées ci-dessus afin de donner effet aux valeurs, actions et principes fondamentaux contenus dans la présente Recommandation.

VII. PROMOTION DE LA PRÉSENTE RECOMMANDATION

108. L'UNESCO a vocation à être la principale institution des Nations Unies chargée de promouvoir et de diffuser la présente Recommandation et, par conséquent, devrait travailler en collaboration avec d'autres entités du système des Nations Unies, notamment le Groupe de haut niveau du Secrétaire général de l'ONU sur la coopération numérique, la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST), le Comité international de bioéthique (CIB), le Comité intergouvernemental de bioéthique (CIGB), l'Union internationale des télécommunications (UIT) et d'autres entités compétentes des Nations Unies concernées par l'éthique de l'IA.

109. L'UNESCO devrait également travailler en collaboration avec d'autres organisations internationales, notamment l'Union africaine (UA), l'Association des nations de l'Asie du Sud-Est (ASEAN), le Conseil de l'Europe (COE), l'Union économique eurasiatique (UEE), l'Union européenne (UE), l'Organisation de coopération et de développement économiques (OCDE) et l'Organisation pour la sécurité et la coopération en Europe (OSCE), ou encore l'Institute of Electrical and Electronic Engineers (IEEE) et l'Organisation internationale de normalisation (ISO).

VIII. DISPOSITIONS FINALES

110. La présente Recommandation doit s'entendre comme un tout, et les valeurs et principes fondamentaux comme étant complémentaires et interdépendants. Chaque principe doit être considéré dans le contexte des valeurs fondamentales.

111. Aucune disposition de la présente Recommandation ne peut être interprétée comme autorisant un État, tout autre acteur de la vie 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 la vie sur Terre et au-delà.