

Note aux facilitateurs: Ce document est un condensé du texte original de projet de recommandation de l'UNESCO sur l'éthique de l'IA, pour lequel l'ensemble de ce travail de délibération est conçu. Il s'accompagne d'un index permettant de lier chacun des points ci-dessous avec les enjeux éthiques identifiés

L'UNESCO est impliquée dans la question de l'éthique de l'IA car, de part sa Constitution, elle a pour mission "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"

La présente Recommandation **(2)** accorde une attention particulière aux domaines centraux de l'UNESCO: **(2a)** éducation, **(2b)** science, **(2c)** culture, **(2d)** communication et information - chacun étant directement ou indirectement lié à L'IA à bien des égards. Cette recommandation est **(3)** adressée aux États, aux organisations publiques et privées et aux individus qui participent à **la recherche, à la conception, au développement, au déploiement ou à l'utilisation (RCDDU¹)** de l'IA.

La recommandation vise **(4)** la formulation de valeurs éthiques, principes, recommandations politiques pour le RDDDU de l'IA, pour le bien de l'humanité. Ce document **(6)** s'adresse principalement aux décideurs mais peut aussi servir de cadre favorisant une approche multipartite, aux acteurs non étatiques.

VALEURS

-Dignité humaine: la RCDDU de l'IA devrait **(9)** respecter et préserver la dignité humaine et cette valeur **(10)** devrait être respectée par tous les acteurs impliqués

-Droits de l'homme et libertés fondamentales: la RCDDU d'IA doit être **(11)** cohérent et conforme au droit, aux principes et aux normes internationaux relatifs aux droits humain

-Ne laisser personne de côté: La RCDDU de l'IA doit se faire **(12)** d'une manière qui respecte tous les groupes d'humanité et stimule la créativité. **(13)** Elle doit être compatible avec l'autonomisation de tous les humains

-Vivre en harmonie: la RCDDU de l'IA devrait **(14)** reconnaître l'interdépendance de tous les humains (appartenant à un ensemble), **(15)** devrait éviter les conflits, la violence, et ne devrait pas séparer, objectiver ou compromettre la sécurité des êtres humains, diviser et retourner les groupes les uns contre les autres

-Crédibilité: l'IA devrait être crédible et **(16)** digne de confiance et la RCDDU devrait inspirer, au lieu de porter atteinte à, la confiance entre les personnes et en l'IA. La confiance doit **(17)** être conquise pour chaque type d'utilisation

¹ l'expression "recherche, à la conception, au développement, au déploiement ou à l'utilisation" apparaît 23 fois. A des fins de lisibilité, elle est remplacée par l'abréviation "**RCDDU**" dans ce résumé.

-Protection de l'environnement: la RCDDU de l'IA devrait **(18)** reconnaître promotion du bien-être environnemental et suivre les lois environnementales pour minimiser les facteurs de risque liés au changement climatique, empêcher l'exploitation et l'épuisement des ressources naturelles contribuant à la détérioration de l'environnement et devrait être mis à profit **(19)** pour fournir des solutions pour protéger l'environnement.

PRINCIPES

-Au service de l'humanité et de son épanouissement: la RCDDU de l'IA doit **(23)** avoir pour but l'épanouissement de l'homme et de leur l'environnement et l'amélioration de la qualité de vie. La RCDDU of AI peut être utilisé **(24)** dans les interactions impliquant des personnes vulnérables et ne doit jamais objectiver les humains ou porter atteinte à la dignité humaine, ou violer ou abuser des droits de l'homme.

-Proportionnalité: la RCDDU de AI **(25)** ne doit pas aller au-delà de ce qui est nécessaire pour atteindre des objectifs légitimes et **(26)** la méthode choisie doit être **(26a)** souhaitable et proportionnelle à l'objectif donné **(26b)** ne devrait pas porter d'atteinte excessive aux valeurs fondamentales énoncées dans la Recommandation **(26c)** adaptée au contexte.

-Surveillance et décision humaines: il faut toujours être en mesure **(27)** d'attribuer la responsabilité éthique et juridique de la RCDDU de l'IA à une personne physique ou à une entité juridique. **(28)** Le contrôle partagé avec l'IA doit être pour des raisons d'efficacité; cette décision dans des contextes limités appartient aux humains. La RCDDU de l'IA vise à aider les humains à prendre des décisions et à agir, jamais à remplacer la responsabilité humaine.

-Durabilité: la RCDDU de l'IA devrait **(29)** promouvoir la réalisation de la durabilité liée aux cadres mondialement acceptés tels que les objectifs de développement durable.

-Diversité et inclusion: la RCDDU de l'IA devrait **(30)** respecter et favoriser la diversité et l'inclusion au minimum conformément au droit international des droits de l'homme.

-Vie privée: l'IA devrait **(31)** respecter, protéger et promouvoir la vie privée. Des mécanismes adéquats de gouvernance des données doivent être en place tout le long du cycle de vie de l'IA.

-Sensibilisation et éducation: **(32)** La sensibilisation du public et sa compréhension de l'IA et de la valeur des données devraient être encouragées afin que les citoyens puissent prendre des décisions éclairées sur leur utilisation de l'IA.

-Gouvernance multipartite et adaptative: la gouvernance de l'IA devrait être **(33)** sensible aux changements technologiques et aux modèles commerciaux associés. La gouvernance **(34)** devrait envisager une gamme de réponses allant de la gouvernance souple à travers des processus d'autorégulation et de certification à une gouvernance stricte avec les lois nationales et, si possible et nécessaire, les instruments internationaux.

-Équité: les acteurs de l'IA devraient **(35)** respecter la justice, l'équité pour minimiser et éviter de renforcer ou de perpétuer les biais sociotechniques, notamment les préjugés raciaux, ethniques, de genre, d'âge et de culture.

-Transparence et explicabilité: **(36)** Le niveau de transparence et d'explicabilité doit toujours être approprié au contexte d'utilisation, **(37)** transparence = permettre aux gens de comprendre la RCDDU de l'IA de manière appropriée au contexte d'utilisation et à la sensibilité du système d'IA. **(38)** Explicabilité = rendre intelligible et à fournir un aperçu des résultats de l'IA.

-Sûreté et sécurité: la RCDDU de l'IA devrait **(39)** éviter les préjudices involontaires (risques liés à la sécurité) et les vulnérabilités aux attaques (fonctions de sécurité). Les gouvernements doivent **(40)** jouer un rôle de premier plan pour assurer la sûreté et la sécurité de l'IA. La recherche sur les risques potentiels de sécurité/sûreté associés doit être continue

-Responsabilité et redevabilité: les acteurs de l'IA devraient (41) assumer la responsabilité morale et juridique et (42) des mécanismes appropriés devraient être développés pour garantir la responsabilité, l'auditabilité et la traçabilité de l'IA.

DOMAINES D'ACTION STRATÉGIQUE POUR LES ÉTATS MEMBRES

PREMIER OBJECTIF DES ACTIONS : UNE GESTION ÉTHIQUE

-Action stratégique 1: Promouvoir la diversité et l'inclusion: (44) Les États membres devraient travailler activement aux discussions sur l'IA, (45) devraient exiger des acteurs de l'IA qu'ils combattent les stéréotypes culturels et sociaux présents dans l'IA volontairement ou par négligence, devraient (46) veiller à ce que les acteurs de l'IA fassent preuve de considération et de respect pour les diversités culturelles et sociales, devraient (47) combler les lacunes en matière de diversité, (48) intégrer l'éthique de l'IA au centre des préoccupations dans les discussions sur l'éthique liée à l'IA

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

-Action stratégique 2: Faire face à l'évolution du marché du travail: les États membres devraient (50) évaluer et traiter l'impact de l'IA sur les marchés du travail et ses implications pour les exigences en matière d'éducation; "Apprendre à apprendre", devrait être enseigné parallèlement aux compétences techniques, (51) assurer une transition équitable pour les employés à risque, (52) analyser l'impact de l'IA sur le marché du travail local, (53) élaborer des politiques de main-d'œuvre ciblant les femmes et les populations sous-représentées

-Action stratégique 3: Faire face aux répercussions économiques et sociales de l'IA: (54) Empêcher la monopolisation de l'IA que les monopoles concernent les données, la recherche, la technologie, le marché ou d'autres domaines et les inégalités qui en résultent, (55) fournir au public une connaissance adéquate de l'IA, (56) établir des mécanismes de suivi et d'évaluation des initiatives et des politiques liées à l'éthique de l'IA, (57) envisager un mécanisme de certification pour les systèmes d'IA similaire à ceux utilisés pour les dispositifs médicaux, (58) encourager le secteur privé à impliquer différentes parties prenantes dans la gouvernance de l'IA et envisager d'ajouter une fonction de responsable de l'éthique de l'IA pour superviser l'évaluation de l'impact, (59) développer des stratégies de gouvernance des données équilibrant les métadonnées et la confidentialité des utilisateurs.

-Action stratégique 4: Les répercussions sur la culture et l'environnement: (60) Intégrer l'IA, le cas échéant, dans le patrimoine culturel, (61) examiner l'impact de l'IA sur le langage humain, (62) étudier les effets de l'interaction des personnes avec l'IA, (63) promouvoir l'éducation et la sensibilisation à l'IA pour les créatifs, (64) les industries culturelles et les startups, (65) évaluer et réduire l'impact environnemental de l'IA.

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

-Action stratégique 5: Promouvoir l'éducation et la sensibilisation à l'éthique de l'IA: encourager (67) l'intégration de l'éthique de l'IA et promouvoir la collaboration entre les compétences techniques et les sciences humaines, (68) promouvoir la sensibilisation à l'IA et l'accès aux connaissances sur les opportunités et les défis apportés par l'IA, (69) introduire de la flexibilité dans l'enseignement supérieur compte tenu des innovations en matière d'IA, (70) promouvoir des programmes de sensibilisation à l'IA, (71) encourager les initiatives de recherche sur l'utilisation de l'IA dans l'enseignement. Soutenir la collaboration entre le milieu universitaire et l'industrie, (72) soutenir la collaboration dans la recherche (73) promouvoir la participation des femmes et de toutes sortes de minorités.

-Action stratégique 6: Promouvoir la recherche en éthique de l'IA: (74) Promouvoir la recherche en IA, en veillant à ce que (75) les chercheurs en IA soient formés en éthique, (76) faciliter l'accès aux

données pour la recherche pour la communauté scientifique, et non au détriment de la vie privée des citoyens, (77) promouvoir la diversité des genres dans la recherche sur l'IA dans les universités et l'industrie, lutter contre les stéréotypes et le harcèlement, (78) promouvoir une recherche interdisciplinaire et critique sur l'IA en incluant des disciplines autres que les STEM (Science Technologie Ingénierie Mathématiques).

QUATRIÈME OBJECTIF DES ACTIONS: LE DÉVELOPPEMENT ET LA COOPÉRATION INTERNATIONALE

-Action stratégique 7: Promouvoir une utilisation éthique de l'IA dans le domaine du développement: (80) encourager l'utilisation éthique de l'IA dans les domaines du développement, (81) s'efforcer de fournir des plateformes de coopération internationale en matière d'IA pour le développement, (82) travailler à promouvoir des collaborations internationales dans la Recherche sur l'IA

-Action stratégique 8: Promouvoir la coopération internationale en matière d'éthique de l'IA: (83) Travailler par le biais d'organisations internationales et d'institutions de recherche pour mener des recherches sur l'éthique de l'IA, en veillant à ce que les algorithmes et les données utilisés dans les domaines de l'IA soient appliqués de manière égale et équitable, et (84) en encourageant la coopération internationale dans Développement de l'IA pour éliminer les clivages géotechniques.

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

-Action stratégique 9: Mettre en place des mécanismes de gouvernance pour l'éthique de l'IA: (86) s'assurer que tout mécanisme de gouvernance de l'IA soit (86a) inclusif, (86b) transparent, (86c) multidisciplinaire, (86d) multilatéral, et (87) favoriser le développement et l'accès à un écosystème numérique pour une IA éthique, (88) encourager l'élaboration et l'utilisation de lignes directrices sur l'IA, (89) considérant l'élaboration d'un cadre juridique international pour la coopération internationale

-Action stratégique 10: garantir la fiabilité de l'IA: (90) mettre en œuvre des mesures appropriées pour surveiller toutes les phases d'un cycle de vie de l'IA, et (91) définir des exigences de transparence et d'explicabilité de l'IA en fonction (91a) du domaine d'application, (91b) du public cible et (91c) de la faisabilité; les états devraient (92) encourager la recherche sur la transparence et l'explicabilité, (93) élaborer des normes internationales pour mesurables de transparence pour une évaluation objective des niveaux de conformité

Action stratégique 11: Garantir la responsabilité, la redevabilité et le respect de la vie privée: (94) adapter les cadres juridiques pour assurer la responsabilisation et la responsabilité du contenu et des résultats de l'IA pour attribuer la responsabilité des décisions et du comportement de l'IA à une personne physique ou morale; aucune responsabilité ni personnalité juridique ne devrait être accordée à l'IA, (95) introduire des évaluations d'impact pour identifier et évaluer les avantages et les risques de l'IA, ainsi que des mécanismes de surveillance appropriés, y compris l'auditabilité, la traçabilité et l'explicabilité, ainsi que des examens externes et multidisciplinaires des AI, (96) impliquer tous les acteurs de l'écosystème de l'IA dans l'établissement de normes là où elles n'existent pas, (97) assurer que les dommages causés aux utilisateurs par l'IA peuvent être examinés, punis et réparés, (98) appliquer les garanties appropriées du droit fondamental des individus à la vie privée, (99) veiller à ce que les individus puissent superviser l'utilisation de leurs informations / données privées, (100) assurer une sécurité accrue pour les données ou données personnellement identifiables, (101) adopter une approche commune des données pour promouvoir l'interopérabilité des ensembles de données et faire preuve de vigilance pour superviser leur collecte et leur utilisation

Crédits : Document réalisé par les chercheur.e.s d'Algora Lab et de Mila-Institut Québécois d'intelligence artificielle (Juillet 2020)