Recommandation du Conseil sur l'intelligence artificielle



Instruments juridiques de l'OCDE

OCDE
DES POLITIQUES MEILLEURES POUR UNE VIE MEILLEURE

Ce document est publié sous la responsabilité du Secrétaire général de l'OCDE. Il reproduit un instrument juridique de l'OCDE et peut contenir des informations complémentaires. Les opinions ou arguments exprimés dans ces informations complémentaires ne reflètent pas nécessairement les vues officielles des pays Membres de l'OCDE.

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Pour accéder aux textes officiels à jour des instruments juridiques de l'OCDE, ainsi qu'aux informations s'y rapportant, veuillez consulter le Recueil des instruments juridiques de l'OCDE http://legalinstruments.oecd.org.

Merci de citer cet ouvrage comme suit :
OCDE, *Recommandation du Conseil sur l'intelligence artificielle*, OECD/LEGAL/0449

Collection : Instruments juridiques de l'OCDE

Crédits photo : © kras99/Shutterstock.com

© OECD 2025

Ce document est mis à disposition à titre gratuit. Il peut être reproduit et distribué gratuitement sans autorisation préalable à condition qu'il ne soit modifié d'aucune façon. Il ne peut être vendu.

Ce document est disponible dans les deux langues officielles de l'OCDE (anglais et français). Il peut être traduit dans d'autres langues à condition que la traduction comporte la mention "traduction non officielle" et qu'elle inclut l'avertissement suivant : *"Cette traduction a été préparée par [NOM DE L'AUTEUR DE LA TRADUCTION] à des fins d'information seulement et son exactitude ne peut être garantie par l'OCDE. Les seules versions officielles sont les textes anglais et français disponibles sur le site Internet de l'OCDE http://legalinstruments.oecd.org"*

Informations Générales

La Recommandation sur l'intelligence artificielle (ci-après dénommée la « Recommandation »), première norme intergouvernementale sur l'intelligence artificielle (IA), a été adoptée par le Conseil de l'OCDE réuni au niveau des Ministres le 22 mai 2019, sur proposition du Comité de la politique du numérique (CPN, anciennement Comité de la politique de l'économie numérique, CPEN). Elle vise à stimuler l'innovation et renforcer la confiance dans l'IA en promouvant une approche responsable au service d'une IA digne de confiance, tout en garantissant le respect des droits humains et des valeurs démocratiques. En juin 2019, à l'occasion du Sommet d'Osaka, les chefs d'État et de gouvernement des pays du G20 ont salué les Principes du G20 relatifs à l'IA, inspirés de la Recommandation de l'OCDE.

La Recommandation a été révisée par le Conseil de l'OCDE le 8 novembre 2023 afin de mettre à jour sa définition d'un « système d'IA », de sorte qu'elle continue d'être techniquement exacte et de refléter les évolutions technologiques, notamment pour ce qui est de l'IA générative. À la lumière du rapport de 2024 au Conseil sur sa mise en œuvre, sa diffusion et le maintien de sa pertinence, la Recommandation a été révisée par le Conseil de l'OCDE réuni au niveau des Ministres, le 3 mai 2024, dans le but de tenir compte des évolutions des technologies et de l'action des pouvoirs publics, en particulier dans le domaine de l'IA générative, et de faciliter la mise en œuvre de l'instrument.

Travaux de l'OCDE sur l'intelligence artificielle

L'intelligence artificielle (IA) est une technologie générique qui promet d'améliorer le bien-être des individus, de contribuer à une activité économique mondiale dynamique et durable, de stimuler l'innovation et la productivité, et d'aider à affronter les grands défis planétaires. Elle est déployée dans de nombreux secteurs, de la production à l'éducation, en passant par la finance, les transports, la santé, ou encore la sécurité.

Si elle présente des avantages, l'IA ne va pas sans poser aux sociétés et aux économies un certain nombre de défis, notamment en termes de mutations économiques et d'inégalités, de concurrence, de transitions sur les marchés du travail et de conséquences sur la démocratie et les droits humains.

L'OCDE a réalisé des travaux à visée empirique et stratégique sur l'IA afin de nourrir les débats d'orientation menés depuis 2016, notamment à l'occasion du Forum de prospective technologique sur l'IA organisé cette année-là, puis d'une conférence internationale sur le thème *AI: Intelligent Machines, Smart Policies*, en 2017. L'Organisation a par ailleurs effectué des travaux d'analyse et de mesure qui dressent un tableau général des aspects techniques de l'IA, recensent les incidences économiques et sociales des technologies connexes et de leurs applications, identifient les grandes considérations intéressant l'action des pouvoirs publics et décrivent les initiatives menées par les gouvernements et les autres parties prenantes dans le domaine de l'IA, aux niveaux tant national qu'international.

Ces travaux ont mis au jour la nécessité de bâtir un cadre d'action stable à l'échelle internationale afin de favoriser la confiance dans l'IA et son adoption au sein de la société. Dans ce contexte, le Conseil de l'OCDE a adopté, sur proposition du CPN, une Recommandation visant à promouvoir une approche d'une IA digne de confiance qui soit centrée sur l'humain, favorise la recherche, préserve les incitations économiques en faveur de l'innovation et vaille pour l'ensemble des parties prenantes.

Recours à un processus inclusif et participatif pour l'élaboration de la Recommandation

L'élaboration de la Recommandation s'est faite dans le cadre d'un processus participatif, en tenant compte, à chaque étape, des contributions d'un large éventail de sources. En mai 2018, le CPN est convenu de constituer un groupe d'experts chargé d'identifier des principes susceptibles de favoriser la confiance dans l'IA et son adoption, dans l'optique de l'élaboration d'un projet de Recommandation courant 2019. De là est né le Groupe informel d'experts de l'IA à l'OCDE, composé de plus de 50 experts de disciplines et de secteurs différents (administrations, industrie, société civile, syndicats, communauté technique et milieux universitaires) – pour consulter la liste complète, voir http://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf. Entre septembre 2018 et février 2019, le groupe d'experts s'est réuni à quatre reprises. Les travaux ont bénéficié de la diligence, de l'engagement et des contributions substantielles des experts membres du groupe, ainsi que de leur expérience multipartite et pluridisciplinaire.

Faisant fond sur le document final produit par le groupe informel, le CPN a préparé un projet de Recommandation en concertation avec d'autres organes concernés au sein de l'OCDE et l'a approuvé lors d'une réunion spéciale organisée les 14 et 15 mars 2019. Le Conseil de l'OCDE a adopté la Recommandation à sa réunion au niveau des Ministres, les 22 et 23 mai 2019.

Champ d'application de la Recommandation

La Recommandation, qui vient compléter les normes de l'OCDE s'appliquant d'ores et déjà à l'IA — à l'instar de celles relatives à la protection de la vie privée et des données, à la gestion du risque de sécurité numérique et à la conduite responsable des entreprises —, traite des questions de fond propres à l'IA et a vocation à définir une norme susceptible d'être mise en œuvre et suffisamment souple pour résister à l'épreuve du temps, dans un domaine en rapide mutation. Elle énonce cinq principes de haut niveau fondés sur des valeurs et cinq recommandations à suivre dans le cadre des politiques nationales et de la coopération internationale. Elle cherche également à établir une compréhension commune des expressions clés utilisées dans ses dispositions, telles que « système d'IA », « cycle de vie d'un système d'IA » et « acteurs de l'IA ».

La Recommandation s'articule autour de deux grandes sections :

1. **Principes d'une approche responsable à l'appui d'une IA digne de confiance** : la première section énonce cinq principes complémentaires intéressant l'ensemble des parties prenantes : i) croissance inclusive, développement durable et bienêtre ; ii) respect de l'état de droit, des droits humains et des valeurs démocratiques, y compris de l'équité et de la vie privée ; iii) transparence et explicabilité ; iv) robustesse, sûreté et sécurité ; et v) responsabilité. On y appelle en outre les acteurs de l'IA à promouvoir et mettre en œuvre ces principes, selon leurs rôles respectifs.
2. **Politiques nationales et coopération internationale à l'appui d'une IA digne de confiance** : dans la droite lignée des cinq principes précités, la deuxième section énonce cinq recommandations à l'intention des Membres et des non-Membres ayant adhéré à la Recommandation (ci-après dénommés les « Adhérents »), recommandations qu'ils sont invités à mettre en œuvre dans le cadre de leurs politiques nationales et de la coopération internationale : i) investir dans la recherche et le développement en matière d'IA ; ii) favoriser l'instauration d'un écosystème inclusif propice à l'IA ; iii) façonner un cadre d'action et de gouvernance interopérable favorable à l'IA ; iv) renforcer les capacités humaines et préparer la transformation du marché du travail ; et v) favoriser la coopération internationale au service d'une IA digne de confiance.

Révisions de la Recommandation en 2023 et 2024

En 2023, il a semblé opportun, pour assurer le maintien de la pertinence de la Recommandation, de mettre à jour la définition d'un « système d'IA » et le CPN a approuvé une proposition de version révisée de la définition lors d'une session qu'il a tenue conjointement avec le Groupe de travail sur la gouvernance de l'intelligence artificielle (GTGIA), le 16 octobre 2023. Le Conseil de l'OCDE a adopté la définition révisée d'un « système d'IA » à sa réunion du 8 novembre 2023. Les modifications apportées à la définition visaient à :

- clarifier les objectifs d'un système d'IA (qui peuvent être explicites ou implicites) ;
- souligner le rôle des entrées, qui peuvent être fournies par des humains ou des machines ;
- préciser que la Recommandation s'applique aux systèmes d'IA générative, qui produisent des

« contenus » ;
- remplacer le terme « réels » par « physiques » pour plus de clarté et de cohérence avec les autres processus engagés à l'échelle internationale ;
- refléter le fait que certains systèmes d'IA peuvent continuer d'évoluer après leur conception et leur déploiement.

Conformément aux conclusions du rapport au Conseil de 2024, la Recommandation a fait l'objet d'une nouvelle révision adoptée lors de la Réunion du Conseil au niveau des Ministres de 2024, afin d'assurer le maintien de sa pertinence et de faciliter sa mise en œuvre dans les cinq ans suivant son adoption. Les modifications visaient en particulier à :

- refléter l'importance croissante de la lutte contre la mésinformation et la désinformation, et de la préservation de l'intégrité de l'information dans le contexte de l'IA générative ;
- traiter des risques découlant d'utilisations à des fins autres que celles prévues ou d'une mauvaise utilisation délibérée ou involontaire ;
- préciser les informations que les acteurs de l'IA sont tenus de fournir concernant les systèmes d'IA pour assurer la transparence et une divulgation responsable ;
- aborder les problématiques de sécurité, en incitant à prévoir des mécanismes garantissant que, dans l'éventualité où des systèmes d'IA risqueraient de causer des préjudices injustifiés ou présenteraient un comportement indésirable, ils puissent être neutralisés, réparés et/ou mis hors service en toute sécurité, par une intervention humaine ;
- mettre l'accent sur l'importance d'assurer la conduite responsable des entreprises, tout au long du cycle de vie des systèmes d'IA, notamment par la coopération avec les pourvoyeurs de connaissances et de ressources liées à l'IA, les utilisateurs des systèmes d'IA et les autres parties prenantes ;
- souligner la nécessité pour les pays et territoires de travailler de concert afin de promouvoir l'instauration de cadres d'action et de gouvernance interopérables dans le domaine de l'IA, face à la multiplication des mesures prises en la matière par les pouvoirs publics ; et
- ajouter une référence explicite à la durabilité environnementale, dont l'importance s'est considérablement accrue depuis l'adoption de la Recommandation en 2019.

De plus, les titres de plusieurs principes et recommandations ont été étoffés pour plus de clarté, et les passages sur la traçabilité et la gestion du risque ont été développés et déplacés vers le principe de responsabilité, dont ils semblent davantage relever.

Pour plus d'informations, veuillez consulter : oecd.ai.
Contact : ai@oecd.org.

Mise en œuvre

Dans la Recommandation, le Conseil charge le CPN de lui faire rapport sur la mise en œuvre, la diffusion et le maintien de la pertinence de l'instrument cinq ans après son adoption, puis régulièrement par la suite.

Rapport au Conseil de 2024

Le CPN, par l'intermédiaire du GTGIA, a rédigé un rapport à l'intention du Conseil sur la mise en œuvre, la diffusion et le maintien de la pertinence de la Recommandation cinq ans après son adoption, et a proposé d'apporter des modifications à la lumière des conclusions qui y sont énoncées.

Il ressort du rapport de 2024 que la Recommandation constitue une référence internationale importante et utile dans le cadre de l'élaboration des politiques nationales en matière d'IA. Elle est mise en œuvre par les Adhérents, bénéficie d'une large diffusion et conserve toute sa pertinence, offrant notamment un cadre solide à l'analyse des évolutions technologiques telles que celles liées à l'IA générative.

En revanche, il montre que des modifications s'avèrent nécessaires afin de clarifier la teneur de certaines de ses dispositions, de faciliter la mise en œuvre, de renforcer la pertinence et de veiller à ce qu'elle tienne compte des progrès technologiques importants à l'œuvre, notamment dans le domaine de l'IA générative.

Suite des travaux à l'appui de la mise en œuvre de la Recommandation

En plus de faire rapport au Conseil sur la mise en œuvre de la Recommandation, le CPN est également chargé de poursuivre ses activités sur l'IA, en s'appuyant sur la Recommandation et en prenant en considération les travaux menés au sein d'autres instances internationales, telles que l'UNESCO, l'Union européenne, le Conseil de l'Europe et l'initiative visant à créer un regroupement international d'experts en intelligence artificielle (voir https://pm.gc.ca/fra/nouvelles/2018/12/06/mandat-groupe-international-dexperts-intelligence-artificielle et https://www.gouvernement.fr/en/france-and-canada-create-new-expert-international-%20panel-on-artificial-intelligence).

Dans le cadre de l'appui à la mise en œuvre de la Recommandation, le Conseil a chargé le CPN d'élaborer des orientations pratiques à cet effet, de faire office de forum d'échange d'informations sur les politiques et les activités ayant trait à l'IA, et de favoriser un dialogue multipartite et interdisciplinaire sur le sujet.

Pour offrir un forum inclusif d'échange d'informations sur les politiques et les activités ayant trait à l'IA et favoriser le dialogue multipartite et interdisciplinaire, l'OCDE a créé : i) l'Observatoire des politiques relatives à l'IA (OECD.AI) et ii) le Réseau informel d'experts de l'IA à l'OCDE (ONE AI), en février 2020.

OECD.AI est une plateforme inclusive dédiée aux politiques publiques en matière d'IA et destinée à aider les pays à encourager, accompagner et suivre le développement responsable de systèmes d'intelligence artificielle dignes de confiance dans l'intérêt de la société. L'Observatoire allie les ressources mobilisées à l'échelle de l'OCDE et celles des partenaires issus de tous les groupes de parties prenantes, afin de fournir une analyse pluridisciplinaire, fondée sur des données probantes, des politiques liées à l'IA. Il intègre une base de données dynamique des stratégies, politiques et initiatives en matière d'IA, que les pays et d'autres parties prenantes peuvent partager et mettre à jour ; ils peuvent en outre comparer leurs éléments clés de manière interactive. La base de données est enrichie en permanence avec les indicateurs, mesures, politiques et bonnes pratiques en matière d'IA, ce qui donne lieu à de nouvelles mises à jour des orientations pratiques à l'appui de la mise en œuvre.

Le Réseau d'experts de l'IA à l'OCDE (ONE AI) est un groupe informel d'experts de l'IA issus d'administrations, d'entreprises, du monde universitaire et de la société civile qui fournit à l'OCDE une expertise et des conseils stratégiques propres à l'IA. Le réseau offre un espace permettant à la communauté internationale de l'IA de mener des discussions approfondies sur les possibilités et les défis communs que l'IA fait naître du point de vue de l'action publique.

En 2022, face à l'importance et au périmètre croissants de ses travaux sur l'IA, le CPN a créé un organe dédié : le Groupe de travail sur la gouvernance de l'intelligence artificielle (GTGIA). Il a pour mission de soutenir la mise en œuvre de la Recommandation de l'OCDE sur l'intelligence artificielle, de rédiger des orientations pratiques à l'appui de la mise en œuvre, de faire office de forum et de fournir une plateforme en ligne pour l'échange d'informations sur les politiques et les activités liées à l'IA, par le biais de l'Observatoire OCDE des politiques relatives à l'IA (OECD.AI), et de favoriser le dialogue multipartite et interdisciplinaire dans le cadre des groupes du Réseau d'experts OECD.AI.

LE CONSEIL,

CONSIDÉRANT l'article 5 b) de la Convention relative à l'Organisation de coopération et de développement économiques en date du 14 décembre 1960 ;

CONSIDÉRANT les normes élaborées par l'OCDE dans les domaines de la protection de la vie privée, de la sécurité numérique, de la protection des consommateurs et de la conduite responsable des entreprises ;

CONSIDÉRANT les Objectifs de développement durable énoncés dans le Programme de développement durable à l'horizon 2030 adopté par l'Assemblée générale des Nations Unies (A/RES/70/1), ainsi que la Déclaration universelle des droits de l'homme de 1948 ;

CONSIDÉRANT les travaux importants menés sur l'intelligence artificielle (ci-après dénommée « IA ») dans le cadre d'autres instances internationales gouvernementales et non gouvernementales ;

RECONNAISSANT que l'IA a des incidences mondiales généralisées et profondes qui transforment les sociétés, les secteurs économiques et le monde du travail, une tendance qui devrait s'accentuer à l'avenir ;

RECONNAISSANT que l'IA promet d'améliorer la prospérité et le bien-être des individus, de contribuer à une activité économique mondiale dynamique et durable, de stimuler l'innovation et la productivité, et d'aider à affronter les grands défis planétaires ;

RECONNAISSANT que, dans le même temps, ces transformations pourraient avoir des effets disparates d'une société et d'une économie à l'autre et en leur sein, notamment en termes de mutations économiques, de concurrence, de transitions sur les marchés du travail, d'inégalités, et de conséquences sur la démocratie et les droits de l'homme, la protection de la vie privée et la confidentialité des données, et la sécurité numérique ;

RECONNAISSANT que la confiance est un déterminant essentiel de la transformation numérique ; que, bien qu'il soit difficile de prévoir la nature des futures applications de l'IA et leurs incidences, la confiance dans la fiabilité des systèmes d'IA est un facteur clé de la diffusion et de l'adoption de l'IA ; et qu'un débat public bien informé de l'ensemble de la société est nécessaire afin de concrétiser le potentiel de cette technologie tout en limitant les risques qui y sont associés ;

SOULIGNANT que certains cadres juridiques, réglementaires et politiques nationaux et internationaux existants s'appliquent d'ores et déjà à l'IA, y compris ceux qui ont trait aux droits de l'homme, à la protection des consommateurs et des données à caractère personnel, aux droits de propriété intellectuelle, à la conduite responsable des entreprises et à la concurrence, tout en sachant qu'il pourrait être nécessaire d'évaluer la pertinence de certains cadres et de concevoir de nouvelles approches ;

RECONNAISSANT que compte tenu de l'évolution et de la mise en œuvre rapides de l'IA, il est nécessaire de bâtir un cadre d'action stable, qui privilégie une IA digne de confiance centrée sur l'humain, favorise la recherche, préserve les incitations économiques à innover et s'applique à l'ensemble des parties prenantes selon leurs responsabilités et le contexte ;

CONSIDÉRANT qu'il est essentiel de saisir les opportunités offertes et de relever les défis posés par les applications de l'IA et de permettre aux parties prenantes de s'investir pour favoriser l'adoption d'une IA digne de confiance dans la société et faire de la fiabilité de l'IA un atout dans le marché mondial.

Sur proposition du Comité de la politique du numérique :

I. **CONVIENT** qu'aux fins de la présente Recommandation, les termes suivants s'entendent comme suit :

- *Système d'IA* : Un système d'intelligence artificielle est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats

en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influer sur des environnements physiques ou virtuels. Différents systèmes d'IA présentent des degrés variables d'autonomie et d'adaptabilité après déploiement.

- *Cycle de vie d'un système d'IA* : Le cycle de vie d'un système d'IA comporte généralement plusieurs phases, à savoir : la phase de planification et de conception ; la phase de collecte et de traitement des données ; la phase de construction du ou des modèle(s) et/ou d'adaptation du ou des modèle(s) existant(s) à des tâches spécifiques ; la phase de test, d'évaluation, de vérification et de validation ; la phase de mise à disposition pour utilisation/déploiement ; la phase d'exploitation et de suivi ; et la phase de mise hors service. Ces phases présentent souvent un caractère itératif et ne suivent pas nécessairement un ordre séquentiel. La décision de mettre un terme à l'utilisation d'un système d'IA peut intervenir à n'importe quel stade de la phase d'exploitation et de suivi.
- *Acteurs de l'IA* : Les acteurs de l'IA sont les parties jouant un rôle actif dans le cycle de vie d'un système d'IA, y compris les organisations et les individus qui déploient ou exploitent l'IA.
- *Connaissances en matière d'IA* : Les connaissances en matière d'IA désignent les compétences et les ressources, à l'instar des données, du code, des algorithmes, des modèles, de la recherche, du savoir-faire, des programmes de formation, de la gouvernance, des processus et des meilleures pratiques nécessaires pour comprendre le cycle de vie des systèmes d'IA et y prendre part, y compris pour ce qui est de la gestion des risques.
- *Parties prenantes* : On entend par parties prenantes l'ensemble des organisations et des individus intervenant dans les systèmes d'IA ou concernés par ces systèmes, que ce soit directement ou indirectement. Les acteurs de l'IA constituent un sous-ensemble des parties prenantes.

Section 1 : Principes d'une approche responsable en appui d'une IA digne de confiance

II. **RECOMMANDE** que les Membres et non-Membres adhérant à la présente Recommandation (ci-après dénommés les « Adhérents ») promeuvent et mettent en œuvre les principes d'une approche responsable en appui d'une IA digne de confiance suivants, qui intéressent l'ensemble des parties prenantes.

III. **APPELLE** tous les acteurs de l'IA à promouvoir et mettre en œuvre, selon leurs rôles respectifs, les principes suivants pour une approche responsable en appui d'une IA digne de confiance.

IV. **SOULIGNE** que les principes suivants sont complémentaires et devraient être considérés comme un tout.

1.1. Croissance inclusive, développement durable et bien-être

Les parties prenantes devraient adopter de manière proactive une approche responsable en appui d'une IA digne de confiance afin de tendre vers des résultats bénéfiques pour les individus et la planète, tels que l'augmentation des capacités humaines et le développement de la créativité, l'inclusion des populations sous-représentées, la réduction des inégalités économiques, sociales, entre les sexes et autres, et la protection des milieux naturels, favorisant ainsi la croissance inclusive, le bien-être, le développement durable et la durabilité environnementale.

1.2. Respect de l'état de droit, des droits humains et des valeurs démocratiques, y compris de l'équité et de la vie privée

a) Les acteurs de l'IA devraient respecter l'état de droit, les droits humains, les valeurs démocratiques et les valeurs centrées sur l'humain, tout au long du cycle de vie des systèmes d'IA. Cela comprend la non-discrimination et l'égalité, la liberté, la dignité, l'autonomie des individus, la protection de la vie privée et des données, la diversité, l'équité, la justice sociale, ainsi que les droits des travailleurs reconnus à l'échelle internationale. Cela recouvre aussi la

lutte contre la mésinformation et la désinformation amplifiées par l'IA, dans le respect de la liberté d'expression et des autres droits et libertés protégés par le droit international en vigueur.

b) Pour ce faire, les acteurs de l'IA devraient instituer des garanties et des mécanismes, tels que la capacité d'intervention et de surveillance humaines, y compris pour faire face aux risques découlant d'utilisations à des fins autres que celles prévues ou d'une mauvaise utilisation délibérée ou involontaire, d'une manière qui soit adaptée au contexte et à l'état de l'art.

1.3. Transparence et explicabilité

Les acteurs de l'IA devraient s'engager à assurer la transparence et une divulgation responsable des informations liées aux systèmes d'IA. À cet effet, ils devraient fournir des informations pertinentes, adaptées au contexte et à l'état de l'art, afin :

i. de favoriser une compréhension générale des systèmes d'IA, y compris de leurs capacités et de leurs limites,

ii. d'informer les parties prenantes de leurs interactions avec les systèmes d'IA, y compris dans la sphère professionnelle,

iii. de donner, lorsque cela est réalisable et utile, des informations claires et facilement compréhensibles sur les sources des données/entrées, les facteurs, les processus et/ou la logique ayant abouti à la prévision, au contenu, à la recommandation ou à la décision, pour permettre aux personnes concernées par un système d'IA d'en appréhender le résultat, et

iv. de fournir des informations permettant aux personnes qui subissent les effets néfastes d'un système d'IA d'en contester les résultats.

1.4. Robustesse, sûreté et sécurité

a) Les systèmes d'IA devraient être robustes, sûrs et sécurisés tout au long de leur cycle de vie, de sorte que, dans des conditions d'utilisation normales ou prévisibles, ou en cas d'utilisation abusive ou de conditions défavorables, ils soient à même de fonctionner convenablement, et ne fassent pas peser des risques de sûreté et/ou de sécurité démesurés.

b) Des mécanismes devraient être en place, le cas échéant, afin de garantir que, dans l'éventualité où des systèmes d'IA risqueraient de causer des préjudices injustifiés ou présenteraient un comportement indésirable, ils puissent être neutralisés, réparés et/ou mis hors service en toute sécurité, en tant que de besoin.

c) Des mécanismes devraient également être en place, lorsque cela est techniquement réalisable, pour renforcer l'intégrité de l'information, tout en veillant au respect de la liberté d'expression.

1.5. Responsabilité

a) Les acteurs de l'IA devraient être responsables du bon fonctionnement des systèmes d'IA et du respect des principes exposés ci-dessus, selon leurs rôles, le contexte et l'état de l'art.

b) Pour ce faire, les acteurs de l'IA devraient veiller à la traçabilité, notamment pour ce qui est des ensembles de données, des processus et des décisions prises au cours du cycle de vie des systèmes d'IA, afin de permettre l'analyse des résultats issus desdits systèmes d'IA et le traitement des demandes d'information, compte tenu du contexte et de l'état de l'art.

c) Les acteurs de l'IA devraient, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, et assurer la conduite responsable des entreprises afin de gérer les risques y afférents, notamment, en tant que de besoin, par la coopération entre les différents acteurs de l'IA, les pourvoyeurs de connaissances et de ressources liées à l'IA, les utilisateurs de systèmes d'IA et les autres parties prenantes. Ces risques sont ceux liés aux biais

préjudiciables, au non-respect des droits humains tels que la sûreté, la sécurité et la protection de la vie privée, et au non-respect des droits des travailleurs et des droits de propriété intellectuelle.

Section 2 : Politiques nationales et coopération internationale en appui d'une IA digne de confiance

V. RECOMMANDE que les Adhérents mettent en œuvre les recommandations suivantes, en respectant les principes énoncés à la section 1, dans le cadre de leurs politiques nationales et de la coopération internationale, en prêtant une attention particulière aux petites et moyennes entreprises (PME).

2.1. Investir dans la recherche et le développement en matière d'IA

a) Les pouvoirs publics devraient envisager des investissements publics à long terme et encourager les investissements privés dans la recherche et le développement et dans la science ouverte, notamment les efforts interdisciplinaires, afin de stimuler l'innovation dans une IA digne de confiance, axée sur les défis techniques importants, ainsi que sur les implications sociales, juridiques et éthiques et les politiques liées à l'IA.

b) Les pouvoirs publics devraient par ailleurs envisager des investissements publics et encourager les investissements privés dans des outils en open source et des ensembles de données en libre accès qui soient représentatifs et qui garantissent la protection de la vie privée et des données, pour soutenir un environnement de recherche et développement en matière d'IA qui soit exempt de biais préjudiciables, et renforcer l'interopérabilité et l'utilisation de normes.

2.2. Favoriser l'instauration d'un écosystème inclusif propice à l'IA

Les pouvoirs publics devraient favoriser le développement et l'accessibilité d'un écosystème numérique inclusif, dynamique, durable et interopérable en appui d'une IA digne de confiance. Cet écosystème se compose, entre autres, des données, des technologies d'IA, des infrastructures de calcul et de connectivité et des mécanismes de partage des connaissances en matière d'IA, en fonction des besoins. À cet égard, les pouvoirs publics devraient envisager de promouvoir des mécanismes, tels que les fiducies de données (« data trusts »), pour favoriser le partage des données de façon sûre, équitable, légale et éthique.

2.3. Façonner un cadre d'action et de gouvernance interopérable favorable à l'IA

a) Les pouvoirs publics devraient promouvoir l'instauration d'un cadre d'action souple qui soutienne la transition du stade de recherche et développement à celui de déploiement et d'exploitation de systèmes d'IA dignes de confiance. À cette fin, ils devraient envisager le recours à l'expérimentation, afin de fournir un environnement contrôlé dans lequel les systèmes d'IA peuvent être testés et déployés à une plus grande échelle, selon les besoins. Ils devraient en outre adopter des approches axées sur les résultats qui offrent de la souplesse dans la réalisation des objectifs de gouvernance et favoriser la coopération entre les pays et territoires et en leur sein, afin de promouvoir la mise en place de cadres de gouvernance et d'action interopérables, en tant que de besoin.

b) Les pouvoirs publics devraient examiner et adapter, selon les besoins, leur cadres politiques et réglementaires et leurs mécanismes d'évaluation applicables aux systèmes d'IA, afin d'encourager l'innovation et la concurrence dans le développement d'une IA digne de confiance.

2.4. Renforcer les capacités humaines et préparer la transformation du marché du travail

a) Les pouvoirs publics devraient travailler en étroite collaboration avec les parties prenantes en vue de préparer la transformation du monde du travail et de la société. Ils devraient donner aux personnes les moyens d'utiliser efficacement les systèmes d'IA et d'interagir avec eux, au travers de leurs différentes applications, notamment en les dotant des compétences nécessaires.

b) Les pouvoirs publics devraient prendre des mesures, y compris en recourant au dialogue social, pour assurer une transition équitable des travailleurs au fur et à mesure du déploiement de l'IA, en particulier par le biais de programmes de formation tout au long de la vie active, du soutien aux personnes affectées par les suppressions de postes, au moyen notamment de la protection sociale, et de l'accès aux nouvelles opportunités sur le marché du travail.

c) Les pouvoirs publics devraient par ailleurs travailler en étroite collaboration avec les parties prenantes pour promouvoir l'utilisation responsable de l'IA au travail, renforcer la sécurité des travailleurs et la qualité des emplois et des services publics, favoriser l'entrepreneuriat et la productivité, et veiller à ce que les avantages de l'IA soient partagés largement et équitablement.

2.5. Favoriser la coopération internationale au service d'une IA digne de confiance

a) Les pouvoirs publics, y compris ceux des pays en développement, en association avec les parties prenantes, devraient coopérer activement afin de faire avancer la mise en œuvre de ces principes et de progresser dans une approche responsable en appui d'une IA digne de confiance.

b) Les pouvoirs publics devraient travailler de concert, au sein de l'OCDE et d'autres instances mondiales et régionales pertinentes, en vue de favoriser le partage des connaissances en matière d'IA, selon les besoins. Ils devraient encourager les initiatives multipartites internationales, intersectorielles et ouvertes afin de bâtir une expertise à long terme en matière d'IA.

c) Les pouvoirs publics devraient encourager l'élaboration de normes techniques internationales multipartites fondées sur la recherche de consensus, au service d'une IA interopérable et digne de confiance.

d) Les pouvoirs publics devraient en outre encourager la mise au point et l'utilisation, pour leurs propres besoins, d'indicateurs comparables au plan international, afin de mesurer la recherche et le développement dans le domaine de l'IA et le déploiement de l'IA, et de constituer la base factuelle nécessaire au suivi des progrès quant à la mise en œuvre des principes exposés dans les présentes.

VI. **INVITE** le Secrétaire général et les Adhérents à diffuser la présente Recommandation.

VII. **INVITE** les non-Adhérents à tenir dûment compte de la présente Recommandation et à y adhérer.

VIII. **CHARGE** le Comité de la politique du numérique, par l'intermédiaire du Groupe de travail sur la gouvernance de l'intelligence artificielle, de :

a) poursuivre ses travaux importants sur l'intelligence artificielle en s'appuyant sur la présente Recommandation et en prenant en considération les travaux menés au sein d'autres instances internationales, et de poursuivre la mise au point du cadre de mesure à l'appui de politiques en matière d'IA fondées sur des données empiriques ;

b) continuer d'élaborer et d'enrichir les orientations pratiques sur la mise en œuvre de la présente Recommandation, afin de répondre aux évolutions à l'œuvre et aux nouvelles priorités d'action ;

c) faire office de forum d'échange d'informations sur les politiques et les activités ayant trait à l'IA, notamment sur les retours d'expériences de mise en œuvre de la présente Recommandation, et de favoriser un dialogue multipartite et interdisciplinaire afin de promouvoir la confiance dans l'IA et son adoption ; et

d) faire rapport au Conseil, en concertation avec les autres comités concernés, sur la mise en œuvre, la diffusion et le maintien de la pertinence de la présente Recommandation au plus tard cinq ans après sa révision, puis au moins tous les dix ans.

À propos de l'OCDE

L'OCDE est un forum unique en son genre où les gouvernements œuvrent ensemble pour relever les défis économiques, sociaux et environnementaux que pose la mondialisation. L'OCDE est aussi à l'avant- garde des efforts entrepris pour comprendre les évolutions du monde actuel et les préoccupations qu'elles font naître. Elle aide les gouvernements à faire face à des situations nouvelles en examinant des thèmes tels que le gouvernement d'entreprise, l'économie de l'information et les défis posés par le vieillissement de la population. L'Organisation offre aux gouvernements un cadre leur permettant de comparer leurs expériences en matière de politiques, de chercher des réponses à des problèmes communs, d'identifier les bonnes pratiques et de travailler à la coordination des politiques nationales et internationales.

Les pays Membres de l'OCDE sont : l'Allemagne, l'Australie, l'Autriche, la Belgique, le Canada, le Chili, la Colombie, la Corée, le Costa Rica, le Danemark, l'Espagne, l'Estonie, les États Unis, la Finlande, la France, la Grèce, la Hongrie, l'Irlande, l'Islande, Israël, l'Italie, le Japon, la Lettonie, la Lituanie, le Luxembourg, le Mexique, la Norvège, la Nouvelle Zélande, les Pays Bas, la Pologne, le Portugal, la République slovaque, la République tchèque, le Royaume-Uni, la Slovénie, la Suède, la Suisse et la Türkiye. L'Union européenne participe aux travaux de l'OCDE.

Instruments juridiques de l'OCDE

Environ 460 instruments juridiques de substance ont été développés dans le cadre de l'OCDE depuis sa création en 1961. Ces instruments comprennent les Actes de l'OCDE (les Décisions et Recommandations adoptées par le Conseil de l'OCDE conformément à la Convention relative à l'OCDE) et d'autres instruments juridiques développés dans le cadre de l'OCDE (notamment les Déclarations et les accords internationaux).

L'ensemble des instruments juridiques de substance de l'OCDE, qu'ils soient en vigueur ou abrogés, est répertorié dans le Recueil des instruments juridiques de l'OCDE. Ils sont présentés selon cinq catégories :

- Les **Décisions** sont adoptées par le Conseil et sont juridiquement contraignantes pour tous les Membres, à l'exception de ceux qui se sont abstenus au moment de leur adoption. Elles définissent des droits et des obligations spécifiques et peuvent prévoir des mécanismes de suivi de la mise en œuvre**.**
- Les **Recommandations** sont adoptées par le Conseil et n'ont pas une portée juridique obligatoire. Elles représentent un engagement politique vis-à-vis des principes qu'elles contiennent, il est attendu que les Adhérents feront tout leur possible pour les mettre en œuvre.
- Les **Documents finaux de substance** sont adoptés individuellement par les Adhérents indiqués plutôt que par un organe de l'OCDE et sont le résultat d'une réunion ministérielle, à haut niveau ou autre, tenue dans le cadre de de l'Organisation. Ils énoncent habituellement des principes généraux ou des objectifs à long terme et ont un caractère solennel.
- Les **accords internationaux** sont négociés et conclus dans le cadre de l'Organisation. Ils sont juridiquement contraignants pour les parties.
- **Arrangement, accord/arrangement et autres** : plusieurs autres types d'instruments juridiques de substance ont été développés dans le cadre de l'OCDE au fil du temps, comme l'Arrangement sur les crédits à l'exportation bénéficiant d'un soutien public, l'Arrangement international sur les Principes à suivre dans les transports maritimes et les Recommandations du Comité d'aide au développement (CAD).