

CONSEIL

Réunion du Conseil au niveau des Ministres, 2-3 mai 2024

RAPPORT SUR LA MISE EN ŒUVRE DE LA RECOMMANDATION DE  
L'OCDE SUR L'INTELLIGENCE ARTIFICIELLE

JT03542431

## *Table des matières*

<b>RÉSUMÉ.....</b>	<b>3</b>
<b>1. Contexte .....</b>	<b>5</b>
<b>2. Méthodologie .....</b>	<b>6</b>
<b>3. Mise en œuvre et maintien de la pertinence.....</b>	<b>7</b>
3.1. Définitions .....	7
3.2. Principes .....	9
<b>4. Réponses des Adhérents au questionnaire .....</b>	<b>33</b>
<b>5. Diffusion .....</b>	<b>37</b>
5.1. Activités des Adhérents .....	37
5.2. Activités du Secrétariat pour diffuser et soutenir la mise en œuvre de la Recommandation .....	37
<b>6. Synthèse et conclusions .....</b>	<b>43</b>
6.1. Mise en œuvre .....	43
6.2. Diffusion.....	43
6.3. Maintien de la pertinence.....	44
<b>Références .....</b>	<b>45</b>

### GRAPHIQUES

Graphique 3.1. Étapes du cycle de vie d'un système d'IA – définition actuelle (deux niveaux)	8
Graphique 3.2. Étapes du cycle de vie d'un système d'IA – proposition de définition actualisée (un niveau)	8
Graphique 4.1. Points de vue des Adhérents sur les cinq principales possibilités offertes par les systèmes d'IA avancés	36
Graphique 4.2. Points de vue des Adhérents sur les principaux risques associés aux systèmes d'IA avancés	36
Graphique 5.1. Travaux de l'OCDE visant à soutenir la mise en œuvre d'une IA digne de confiance et fondée sur des valeurs	39

### TABLEAUX

Tableau 1.1. Les dix Principes de la Recommandation de l'OCDE sur l'intelligence artificielle	6
Tableau 5.1. Orientations de l'OCDE pour aider les Adhérents à mettre en œuvre les cinq recommandations à l'intention des pouvoirs publics	40

## RÉSUMÉ

1. Le présent rapport évalue la mise en œuvre, la diffusion et le maintien de la pertinence de la Recommandation de l'OCDE sur l'intelligence artificielle (IA) [[OCDE/LEGAL/0449](#)] (la « Recommandation » ou la « Recommandation sur l'IA »), première norme intergouvernementale dans le domaine de l'IA. Les Principes de l'OCDE sur l'IA énoncés dans la Recommandation ont servi de base à l'élaboration des Principes du G20 relatifs à l'IA, adoptés par les dirigeants des pays du G20 en juin 2019. Ils visent 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 de l'homme et des valeurs démocratiques. En décembre 2023, outre les 38 Membres de l'OCDE, huit non-Membres (l'Argentine, le Brésil, l'Égypte, Malte, le Pérou, la Roumanie, Singapour et l'Ukraine), avaient adhéré à la Recommandation (les « Adhérents »).
2. Le paysage technologique, économique, géopolitique et de l'action publique a considérablement évolué depuis l'adoption de la Recommandation en 2019, notamment avec la montée en puissance de modèles d'IA pouvant être utilisés à différentes fins, parmi lesquels figurent les modèles de langue et multimodaux qui peuvent produire des contenus nouveaux, transposer du texte en vidéo et en image et interagir avec les gens en langage naturel par le biais d'agents conversationnels. Bien que des réglementations et législations existantes s'appliquent déjà à l'IA, les pays de l'OCDE étudient des propositions concernant des mesures réglementaires variées, notamment des cadres réglementaires propres à l'IA (à la fois transversaux et sectoriels). Dans ce contexte, l'examen de la mise en œuvre de la Recommandation vient à point nommé pour dresser le bilan des évolutions et des tendances observées au cours des cinq dernières années, évaluer le maintien de la pertinence des Principes de l'OCDE sur l'IA et identifier la suite à y donner.
3. Pour assurer le maintien de la pertinence de la Recommandation, la définition d'un « système d'IA » établie par l'OCDE a déjà été révisée en novembre 2023 afin de veiller à ce qu'elle reste exacte d'un point de vue technique et reflète les progrès technologiques à l'œuvre, notamment dans le domaine de l'IA générative.
4. Les conclusions du rapport montrent que la Recommandation est mise en œuvre par les Adhérents. Le principal outil utilisé pour suivre l'avancement de cette mise en œuvre au niveau national est un recensement des politiques nationales en matière d'IA qui est mis à jour deux fois par an et alimente l'Observatoire OCDE des politiques relatives à l'IA. Depuis 2019, les Adhérents proposent des initiatives nationales et internationales visant à suivre les cinq recommandations d'action à l'intention des pouvoirs publics et à traduire en actes les Principes de l'OCDE sur l'IA, fondés sur des valeurs.
5. En 2017, seuls quelques pays disposaient d'une stratégie nationale en matière d'IA. Désormais, l'Observatoire OCDE des politiques relatives à l'IA (OECD.AI) recense plus de 50 stratégies nationales et initiatives gouvernementales traitant des moyens d'orienter le développement et le déploiement d'une IA digne de confiance, 41 Adhérents disposant d'une stratégie nationale en matière d'IA, et trois étant en train d'en élaborer une. Même si les stratégies nationales en matière d'IA de chaque pays sont uniques, elles partagent de nombreux points communs avec les Principes de l'OCDE sur l'IA. Par exemple, la plupart de ces stratégies se concentrent sur la croissance inclusive, le développement durable et le bien-être, les valeurs centrées sur l'humain et l'équité, l'investissement dans la R-D en matière d'IA et le renforcement des capacités humaines.

6. Le nombre de programmes d'action et de pays recensés par l'Observatoire ne cesse d'augmenter : en janvier 2024, la base de données comprenait des informations sur 1 020 programmes d'action dans 70 pays et territoires, dont 850 déclarés par les 46 Adhérents à la Recommandation et l'Union européenne. La richesse des programmes d'action et leur applicabilité directe aux Principes de l'OCDE sur l'IA sont une preuve de l'engagement des Adhérents à mettre en œuvre les Principes et des efforts continus déployés à cette fin, ainsi que du maintien de la pertinence des Principes.

7. Plusieurs pays et territoires font également explicitement référence aux Principes de l'OCDE sur l'IA et aux travaux analytiques ultérieurs de l'OCDE dans leurs lignes directrices, actions législatives ou cadres d'application volontaire nationaux, ou s'en inspirent. C'est le cas notamment des principes intersectoriels d'Israël et du Royaume-Uni, qui reflètent les Principes de l'OCDE sur l'IA. Le Japon en fait mention dans des initiatives publiques majeures, notamment dans le cadre des examens menés par le Conseil stratégique de l'IA. Le cadre de gestion des risques liés à l'IA du National Institute of Standards and Technologies des États-Unis utilise le Cadre de l'OCDE pour la classification des systèmes d'IA, tandis que la législation sur l'IA de l'Union européenne et la convention du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit ont adopté la définition actualisée d'un « système d'IA » incluse dans la Recommandation. En outre, plusieurs Adhérents, dont la Corée, l'Italie, la Lituanie et la Türkiye, ont fait des Principes de l'OCDE sur l'IA les piliers fondamentaux de leurs stratégies et de leurs cadres de gouvernance nationaux en matière d'IA.

8. L'analyse présentée dans ce rapport fait apparaître que les Principes de l'OCDE sur l'IA constituent une référence internationale importante et utile à laquelle les Adhérents ont recours afin d'élaborer leurs politiques nationales en matière d'IA. Ils sont largement diffusés et conservent toute leur pertinence, notamment en tant que cadre solide permettant d'analyser les évolutions récentes telles que celles liées à l'IA générative.

9. La Recommandation, dans sa forme actuelle, reste globalement adaptée à l'objectif poursuivi. Toutefois, son actualisation permettrait de faciliter sa mise en œuvre par les parties prenantes et de refléter les questions qui se font jour et les avancées technologiques, notamment celles relatives à l'IA générative. Il s'agirait plus précisément de : i) refléter l'importance croissante de la lutte contre la mésinformation et la désinformation ainsi que de la sauvegarde de l'intégrité de l'information dans le contexte de l'IA générative ; ii) traiter des utilisations en dehors de la finalité assignée, de l'utilisation abusive intentionnelle ou non intentionnelle ; iii) 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 ; iv) décrire les mécanismes permettant de remédier aux préjudices potentiels ou aux comportements indésirables tout au long du cycle de vie de l'IA ; et v) mettre l'accent sur la conduite responsable des entreprises tout au long du cycle de vie de l'IA, ce qui implique de coopérer avec les fournisseurs de connaissances et de ressources dans le domaine, les utilisateurs de systèmes d'IA et d'autres parties prenantes. En outre, certains intitulés et textes des Principes pourraient être développés, et le texte sur la traçabilité et la gestion du risque pourrait être précisé et déplacé vers le Principe de « responsabilité » qui correspond mieux à ces concepts.

## 1. Contexte

10. L'OCDE a commencé à mener des travaux à visée empirique et stratégique sur l'intelligence artificielle (IA) en 2016 dans le cadre de son Forum de prospective technologique sur l'IA. Elle a ensuite organisé des conférences et réalisé des travaux d'analyse et de mesure qui dressent un tableau général des aspects techniques de l'IA, recensent ses incidences économiques et sociales et identifient les considérations intéressant l'action des pouvoirs publics.

11. 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, la Recommandation sur l'intelligence artificielle (IA) [[OCDE/LEGAL/0449](#)] a été élaborée dans le cadre d'un processus inclusif et participatif, en tenant compte des contributions d'un large éventail de parties prenantes. En particulier, les travaux ont été menés par le Groupe d'experts sur l'intelligence artificielle à l'OCDE (GTGIA), composé de plus de 50 experts de disciplines et de secteurs différents (administrations, industrie, société civile, syndicats, communauté technique et milieux universitaires).

12. Faisant fond sur le document final produit par le GTGIA, le Comité de la politique du numérique (CPN)<sup>1</sup> a préparé un projet de Recommandation en consultation avec d'autres organes concernés au sein de l'OCDE et a approuvé le projet final de Recommandation lors d'une réunion spéciale les 14 et 15 mars 2019. Le 22 mai 2019, la Recommandation a été adoptée par le Conseil de l'OCDE au niveau des ministres [[C/MIN\(2019\)3/FINAL](#) et [C/M\(2019\)10](#), point 102], devenant ainsi la première norme intergouvernementale dans le domaine de l'IA (ci-après, la « Recommandation » ou la « Recommandation sur l'IA »). 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 de l'homme et des valeurs démocratiques. La Recommandation, qui vient compléter les normes existantes de l'OCDE dans des domaines tels que la protection de la vie privée, la gestion du risque de sécurité numérique et la conduite responsable des entreprises, traite des questions 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 (OCDE, 2019<sub>[1]</sub>).

13. La Recommandation énonce cinq principes complémentaires fondés sur des valeurs et cinq recommandations à l'intention des décideurs. Ces cinq valeurs et cinq recommandations sont désignées dans le présent rapport sous le nom « Principes de l'OCDE sur l'IA » (Tableau 1.1). Elle invite également à mettre au point des indicateurs afin de mesurer la recherche et le développement dans le domaine de l'IA, ainsi que son déploiement, et de constituer la base factuelle nécessaire au suivi des progrès dans la mise en application des principes qui y sont exposés. Elle contribue à la normalisation des termes employés par les Adhérents en fournissant des définitions de concepts clés tels qu'un « système d'IA ».

---

<sup>1</sup> Anciennement Comité de la politique de l'économie numérique (CPEN).

**Tableau 1.1. Les dix Principes de la Recommandation de l'OCDE sur l'intelligence artificielle**

Principes d'une approche responsable en appui d'une IA digne de confiance	Politiques nationales et coopération internationale à l'appui d'une IA digne de confiance
1.1 Croissance inclusive, développement durable et bien-être	2.1 Investir dans la recherche et le développement en matière d'IA
1.2 Valeurs centrées sur l'humain et équité	2.2 Favoriser l'instauration d'un écosystème numérique pour l'IA
1.3 Transparence et explicabilité	2.3 Façonner un cadre d'action favorable à l'IA
1.4 Robustesse, sûreté et sécurité	2.4 Renforcer les capacités humaines et préparer la transformation du marché du travail
1.5 Responsabilité	2.5 Favoriser la coopération internationale au service d'une IA digne de confiance

Source : [OECD/LEGAL/0449](https://www.oecd.org/legal/0449).

14. La Recommandation a été révisée par le Conseil le 8 novembre 2023 [C(2023)151 et C/M(2023)14, point 218] pour mettre à jour sa définition d'un « système d'IA », afin d'assurer qu'elle continue d'être techniquement exacte et de refléter les développements technologiques, y compris en ce qui concerne l'IA générative. La mise à jour de la définition visait à : (i) clarifier les objectifs d'un système d'IA (qui peuvent être explicites ou implicites) ; (ii) souligner le rôle des entrées, qui peuvent être fournies par des humains ou des machines ; (iii) clarifier que la Recommandation s'applique aux systèmes d'IA génératifs, qui produisent du « contenu » ; (iv) remplacer le mot « réels » par « physiques » pour plus de clarté et de cohérence avec les autres processus engagés à l'échelle internationale ; et (v) refléter le fait que certains systèmes d'IA peuvent continuer d'évoluer après leur conception et leur mise en œuvre. En outre, le CPN, sur proposition du GTGIA, a approuvé un Mémoire explicatif destiné à accompagner la définition actualisée d'un « système d'IA » le 15 décembre 2023 (OCDE, 2024<sup>[2]</sup>).

15. La Recommandation est ouverte à l'adhésion de non-Membres. En décembre 2023, outre les 38 Membres de l'OCDE, huit non-Membres, (l'Argentine, le Brésil, l'Égypte, Malte, le Pérou, la Roumanie, Singapour et l'Ukraine), avaient adhéré à la Recommandation.

16. En adoptant la Recommandation, le Conseil a chargé le CPN « d'assurer le suivi, en consultation avec les autres Comités concernés, de la mise en œuvre de la présente Recommandation et d'en faire rapport au Conseil au plus tard dans les cinq ans suivant son adoption ».

17. Le présent rapport évalue la mise en œuvre, la diffusion et le maintien de la pertinence de la Recommandation cinq ans après son adoption. Pendant cette période, l'IA a connu des avancées considérables, tant sur le plan de la technologie que de l'action des pouvoirs publics. Dans ce contexte, le rapport au Conseil de l'OCDE sur la mise en œuvre, la diffusion et le maintien de la pertinence de la Recommandation vient à point nommé pour dresser le bilan des évolutions et des tendances observées au cours des cinq dernières années et apprécier leurs effets sur la pertinence de la Recommandation ainsi que la suite à y donner.

## 2. Méthodologie

18. L'OCDE a commencé à suivre la mise en œuvre des politiques en matière d'IA en 2020 en les recensant et en publiant ces données au fur et à mesure sur l'Observatoire OCDE des politiques relatives à l'IA. Les points de contact nationaux dans les pays et

territoires couverts (actuellement au nombre de 70) sont invités deux fois par an à rendre compte de l'évolution de leurs politiques nationales et régionales en matière d'IA, selon un questionnaire structuré, organisé en programmes d'action, instruments d'action et autres champs spécifiques, tels que description, budget et groupes cibles. Ces informations alimentent la base de données de l'Observatoire OCDE des politiques relatives à l'IA (OCDE, 2024<sup>[3]</sup>). Cette dernière est une plateforme en ligne par laquelle les utilisateurs peuvent explorer les programmes d'action par pays, par type et catégorie d'instrument d'action ou par groupe cible, entre autres.

19. Depuis sa création, la base de données de l'Observatoire OCDE des politiques relatives à l'IA a permis de suivre l'évolution des politiques et de faire le point sur les efforts déployés par les pays et les organisations internationales pour mettre en œuvre les Principes de l'OCDE sur l'IA. Le nombre de programmes d'action et de pays couverts n'a cessé d'augmenter depuis 2020. En janvier 2024, la base de données comprenait des informations sur 1 020 programmes d'action dans 70 pays et territoires, dont 850 déclarés par les 46 Adhérents à la Recommandation et l'Union européenne.

20. Deux rapports sur l'état d'avancement de la mise en œuvre des Principes de l'OCDE sur l'IA ont également été approuvés par le CPN, respectivement en 2021 [[DSTI/CDEP\(2020\)15/REV1](#)] (le « Rapport 2021 ») et en 2023 [[DSTI/CDEP/AIGO\(2023\)5/REV2](#)] (le « Rapport 2023 »). Le Rapport 2021 (OCDE, 2021<sup>[4]</sup>) fait le point sur les efforts déployés par les Adhérents pour élaborer des politiques en matière d'IA et établir des mécanismes de gouvernance pour les initiatives nationales dans le domaine l'IA. Il présente également de nombreux exemples d'action publique des Adhérents qui mettent en œuvre les cinq recommandations à l'intention des pouvoirs publics. Le Rapport 2023 (OCDE, 2023<sup>[5]</sup>) rend compte des développements survenus dans les stratégies nationales des Adhérents en matière d'IA et met l'accent sur certains cadres réglementaires propres à l'IA en cours d'élaboration partout dans le monde. En outre, il illustre les efforts déployés par différents Adhérents pour mettre en œuvre les Principes de l'OCDE sur l'IA.

21. Pour compléter ces informations et rendre compte de façon détaillée et approfondie de la mise en œuvre, de la diffusion et du maintien de la pertinence de la Recommandation, un questionnaire a été distribué aux délégués auprès du GTGIA en décembre 2023 afin d'éclairer certains aspects du rapport au Conseil, à savoir les difficultés rencontrées dans le cadre de la mise en œuvre, du maintien de la pertinence et de la diffusion des Principes de l'OCDE sur l'IA au niveau national. Ce questionnaire comprenait également des questions sur les principales possibilités et les principaux risques associés aux modèles d'IA avancés, tirées du questionnaire distribué aux membres du G7 à l'appui du Processus du G7, dit d'Hiroshima, sur l'IA générative.

### 3. Mise en œuvre et maintien de la pertinence

#### 3.1. Définitions

##### 3.1.1. Système d'IA

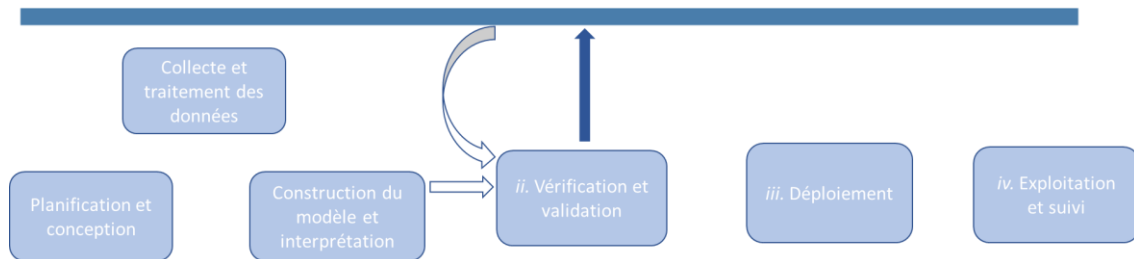
22. Pour assurer le maintien de la pertinence de la définition de l'OCDE d'un système d'IA, la Recommandation a été révisée par le Conseil le 8 novembre 2023 ([[C\(2023\)151](#)] et [[C/M\(2023\)14](#), point 218]) pour mettre à jour cette définition, afin d'assurer qu'elle continue d'être techniquement exacte et reflète les développements technologiques, y compris en ce qui concerne l'IA générative. En outre, le CPN a

approuvé et déclassifié un Mémoire explicatif destiné à accompagner la définition actualisée d'un « système d'IA » le 15 décembre 2023 (OCDE, 2024<sup>[2]</sup>).

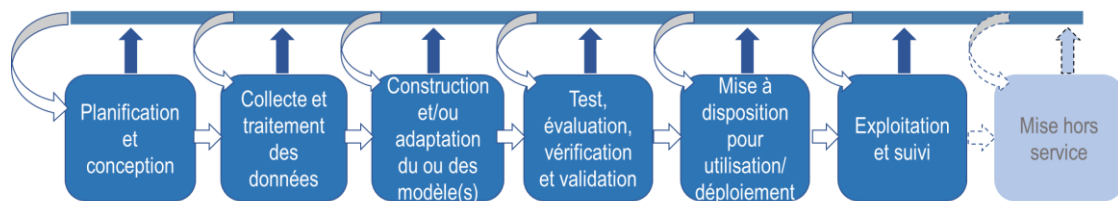
### 3.1.2. Cycle de vie d'un système d'IA

23. La définition du cycle de vie d'un système d'IA demeure exacte et pertinente. Toutefois, tant l'OCDE – dans ses travaux sur le Cadre de l'OCDE pour la classification des systèmes d'IA (OCDE, 2022<sup>[6]</sup>) et dans ses travaux sur les risques et les responsabilités (OCDE, 2023<sup>[7]</sup>) – que ses membres, tels que le National Institute of Standards and Technology (NIST) dans son cadre de gestion des risques liés à l'IA (National Institute of Standards and Technology - US Department of Commerce, 2023<sup>[8]</sup>), utilisent une définition simplifiée à un niveau (Graphique 3.2), au lieu de deux (Graphique 3.1). Cette définition est considérée comme étant identique sur le fond, mais plus claire pour décrire les différentes étapes de la conception et du déploiement d'un système d'IA.

**Graphique 3.1. Étapes du cycle de vie d'un système d'IA – définition actuelle (deux niveaux)**



**Graphique 3.2. Étapes du cycle de vie d'un système d'IA – proposition de définition actualisée (un niveau)**



24. **La définition du cycle de vie d'un système d'IA gagnerait à être simplifiée afin de décrire de manière précise la façon dont les systèmes d'IA sont conçus et exploités.**

### 3.1.3. Connaissances en matière d'IA, acteurs de l'IA et parties prenantes

25. Les définitions des expressions « acteurs de l'IA » et « parties prenantes » sont toujours adaptées au contexte et elles servent ou ont servi de base à d'autres définitions similaires aux niveaux national et régional.

26. **La définition de l'expression « connaissances en matière d'IA » gagnerait à être complétée de sorte à inclure les connaissances sur la gestion des risques à l'ensemble des compétences et des ressources nécessaires pour comprendre le cycle de vie d'un système d'IA et y participer.**

27. Les sections ci-dessous abordent tour à tour chaque Principe, examinant leur mise en œuvre, le maintien de leur pertinence ainsi que l'éventuelle nécessité d'une mise à jour.

## 3.2. Principes

### 3.2.1. *Stratégies nationales en matière d'IA et modèles de gouvernance*

28. En 2017, seuls quelques pays disposaient d'une stratégie nationale en matière d'IA. En janvier 2024, l'Observatoire OCDE des politiques relatives à l'IA recensait plus de 850 programmes d'action en faveur de l'IA déclarés par les Adhérents, dont 41 stratégies nationales, c'est-à-dire des initiatives gouvernementales visant à orienter le développement et le déploiement d'une IA digne de confiance.

29. La stratégie nationale en matière d'IA de chaque pays présente des spécificités et aborde différents aspects de l'action des pouvoirs publics dans ce domaine. Pourtant, elles partagent toutes des points communs avec les recommandations à l'intention des pouvoirs publics énoncées dans les Principes de l'OCDE sur l'IA. En effet, elles sont souvent axées sur la croissance inclusive, le développement durable et le bien-être, les valeurs centrées sur l'humain et l'équité. La plupart des stratégies comprennent également des piliers correspondant aux cinq recommandations d'action, l'investissement dans la recherche et le développement en matière d'IA et le renforcement des capacités humaines figurant parmi les principales priorités des décideurs. Plusieurs Adhérents, dont la Corée, l'Italie, la Lituanie et la Türkiye, ont indiqué que leurs stratégies nationales en matière d'IA intègrent les Principes de l'OCDE sur l'IA ou s'en inspirent.

30. Les Adhérents à la Recommandation utilisent différents types de modèles de gouvernance pour mettre en œuvre leurs politiques nationales en matière d'IA. Certains ont établi des organes de coordination publics pour l'IA. Le Royaume-Uni a créé le Government Office for AI, une unité au sein du ministère de la Science, de l'Innovation et de la Technologie, et les États-Unis ont mis en place le National Artificial Intelligence Initiative Office (NAIIO), intégré au Bureau de la politique scientifique et technologique de la Maison Blanche (White House Office of Science and Technology Policy). D'autres Adhérents s'appuient sur les ministères existants et ont établi des comités interministériels et multipartites chargés de superviser l'élaboration et la mise en œuvre des stratégies en matière d'IA. Le Comité de gouvernance de la stratégie brésilienne en matière d'IA et le Conseil national égyptien pour l'IA sont des exemples de cette approche.

31. Les Adhérents mettent également en place des groupes multipartites d'experts en IA chargés de donner des avis et faire rapport à propos des possibilités, des risques et des enjeux actuels et futurs liés à l'utilisation de l'IA. Le Comité national pilote d'éthique du numérique en France et le Conseil consultatif en matière d'intelligence artificielle au Canada en sont des exemples pertinents.

32. Des Adhérents comme le Canada ont mis en place des observatoires de l'IA dédiés afin de suivre et d'évaluer le déploiement de leur stratégie dans le domaine. Bien que les mécanismes de suivi et d'évaluation soient encore rares, il est probable que d'autres Adhérents en établissent au fur et à mesure que les stratégies nationales en matière d'IA passent aux dernières phases de leur mise en œuvre.

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

#### *Croissance inclusive, développement durable et bien-être (Principe 1.1)*

*« Les parties prenantes devraient adopter de manière proactive une approche responsable en soutien d'une IA digne de confiance afin de tendre vers des résultats bénéfiques pour les individus et la planète, tels que le renforcement des capacités humaines et le renforcement de la créativité humaine, 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 développement durable et le bien-être. »*

33. Ce Principe invite les parties prenantes à orienter le développement, le déploiement et l'utilisation de l'IA de manière à en tirer des avantages pour la société dans son ensemble, en responsabilisant tous ses membres, en atténuant les inégalités et en réduisant au minimum les dommages causés à l'environnement.

34. L'IA a le potentiel d'augmenter la productivité des pays, de conduire à la croissance économique et de contribuer au développement durable, y compris à la durabilité environnementale. La plupart des pays ont pris conscience de ces possibilités et encouragent, par le biais de diverses initiatives, la recherche et le développement en matière d'IA ainsi que le développement de l'infrastructure, des capacités et des outils associés. Cependant, les systèmes d'IA peuvent également perpétuer les inégalités existantes et avoir un effet préjudiciable sur les populations vulnérables et sous-représentées, telles que les minorités ethniques, les femmes, les enfants, les personnes âgées et les personnes moins instruites ou peu qualifiées.

35. En outre, les ressources de calcul nécessaires à la formation et à l'utilisation des systèmes d'IA peuvent avoir une empreinte environnementale significative en raison de la consommation d'énergie et d'eau, de l'émission de gaz à effet de serre (GES) ainsi que des éléments à prendre en compte pour la fin du cycle de vie (OCDE, 2022<sup>[9]</sup>). Le Principe souligne l'importance d'harmoniser les progrès réalisés dans le domaine de l'IA avec les objectifs de durabilité environnementale. La lutte contre la dégradation de l'environnement et le changement climatique nécessite d'intégrer la durabilité dans le développement et le déploiement de l'IA. La forte consommation énergétique, l'augmentation des émissions, la consommation et la dégradation des ressources, ainsi qu'une croissance rapide et incontrôlée mettent en évidence la nécessité critique pour les initiatives dans le domaine de l'IA d'être respectueuses de l'environnement. Par conséquent, le Principe doit être compris comme approuvant le soutien aux architectures d'IA économes en énergie et encourage l'utilisation de l'IA pour la protection de l'environnement, le suivi et la facilitation du passage vers des sources d'énergie renouvelables. Il est essentiel de donner la priorité à la création et à l'application de technologies d'IA qui soutiennent et accélèrent la transition environnementale afin de garantir la contribution positive de l'IA à la santé écologique mondiale et aux objectifs de développement durable.

36. En janvier 2024, la base de données des politiques nationales en matière d'IA comptait plus de 300 initiatives en rapport avec le Principe 1.1 sur la croissance inclusive, le développement durable et le bien-être. La plupart des stratégies nationales globales et des cadres éthiques en matière d'IA ou des lignes directrices pour la mise en œuvre de l'IA font référence à ces thèmes. Des Adhérents ont notamment lancé des programmes d'action pour s'assurer que les groupes vulnérables de la population sont impliqués dans le développement des systèmes d'IA et en bénéficient, aussi bien par le

biais d'initiatives ciblées ou dans le cadre de l'élaboration de politiques. D'autres financent et promeuvent des projets qui utilisent l'IA pour relever des défis environnementaux. D'autres initiatives, par exemple dans les secteurs de la création et des soins de santé, illustrent la manière dont l'IA peut améliorer le bien-être des personnes.

- **Croissance inclusive** : on peut citer entre autres le programme français « IA Booster » qui vise à soutenir la numérisation des petites et moyennes entreprises (PME) grâce à des solutions d'IA, et le programme « Women in Data Science » de l'Alan Turing Institute au Royaume-Uni, qui se concentre sur la réduction des inégalités entre les genres en analysant la participation, en examinant les cultures en milieu de travail et en promouvant une conception de l'IA inclusive pour les deux sexes.
- **Consultations des citoyens** : plusieurs Adhérents impliquent de façon proactive les citoyens dans la conception de leurs politiques en matière d'IA. Par exemple, en Autriche, plus de 160 experts et organisations de la société civile ont œuvré à l'élaboration de la stratégie nationale en matière d'IA. Au Canada, un groupe de travail sur la sensibilisation du public fait participer la population à des discussions sur l'IA par le biais d'ateliers virtuels. L'Écosse (Royaume-Uni) a organisé des ateliers de cocréation dans le domaine de l'IA ouverts au public et a élaboré des principes de conception sur la manière dont les citoyens devraient participer à l'avenir à la prise de décision en matière d'IA. De même, au Chili, le processus de participation sur l'IA recueille les opinions et les préoccupations des citoyens et des organisations concernant l'utilisation et le développement de l'IA. Au Mexique, l'Alliance nationale pour l'intelligence artificielle encourage un dialogue ouvert et inclusif sur l'IA et ses répercussions, en faisant participer de multiples parties prenantes afin de démocratiser le débat sur le sujet. Au Royaume-Uni, l'enquête en ligne sur l'écosystème de l'IA, réalisée en juin 2021 par l'institut Alan Turing, a recueilli plus de 400 réponses reflétant les points de vue des acteurs impliqués dans l'écosystème de l'IA.
- **Des résultats bénéfiques améliorés par une collaboration multipartite** : le Forum IA Québec au Canada collabore avec diverses parties prenantes afin de faire de l'IA un levier de développement économique et social. En Colombie, les organes de coordination pour la mise en œuvre des politiques en matière d'IA harmonisent les politiques en matière d'IA entre les entités publiques nationales et locales et fournissent des orientations aux entités publiques, au secteur privé, aux milieux universitaires et aux pouvoirs publics nationaux. L'initiative allemande « Civic Coding » est un effort de collaboration entre plusieurs ministères visant à renforcer les compétences en matière d'IA et à promouvoir l'utilisation de l'IA pour le bien commun. Depuis février 2022, la Corée a mis en place un forum multipartite chargé de réfléchir aux préoccupations éthiques que suscitent les progrès réalisés par les technologies d'IA et de parvenir à un consensus sur la manière d'instaurer la confiance dans l'IA. Au sein du forum, trois comités d'experts (éthique, technologie et éducation) travaillent à la recherche d'un consensus.
- **Développement durable** : plusieurs Adhérents ont lancé des initiatives visant à promouvoir l'utilisation de l'IA en faveur de la durabilité environnementale. Le projet allemand « AI Lighthouses for the Environment », par exemple, soutient des projets d'IA visant à relever des défis environnementaux ; le Portugal utilise l'IA pour lutter contre la pêche illégale et améliorer la gestion des déchets afin de préserver l'environnement, et le projet « Destination Terre » de l'Union européenne utilise l'IA pour créer un jumeau numérique de la planète Terre afin d'effectuer des simulations et d'aider à se préparer aux catastrophes naturelles et à s'adapter au changement climatique.

37. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 1.1 sur la croissance inclusive, le développement durable et le bien-être. Le Principe reste pertinent, mais pourrait bénéficier d'un ajout mineur faisant spécifiquement référence à la durabilité environnementale, dont l'importance s'est considérablement accrue au cours des cinq dernières années.**

*Valeurs centrées sur l'homme et équité (Principe 1.2)*

*« Les acteurs de l'IA devraient respecter l'état de droit, les droits de l'homme et les valeurs démocratiques tout au long du cycle de vie des systèmes d'IA. Ces droits et valeurs comprennent la liberté, la dignité et l'autonomie, la protection de la vie privée et des données, la non-discrimination et l'égalité, la diversité, l'équité, la justice sociale, ainsi que les droits des travailleurs reconnus à l'échelle internationale. Pour ce faire, les acteurs de l'IA devraient instituer des garanties et des mécanismes, tels que l'attribution de la capacité de décision finale à l'homme, qui soient adaptés au contexte et à l'état de l'art. »*

38. Certaines utilisations des systèmes d'IA ont des répercussions sur les droits de l'homme, notamment le risque que les valeurs centrées sur l'humain soient délibérément ou accidentellement violées. Pour faire face à ces risques, les Adhérents ont publié des lignes directrices ou des initiatives principalement non contraignantes visant à réduire les biais liés à l'IA et à favoriser l'alignement des valeurs en promouvant les droits de l'homme et les valeurs centrées sur l'humain. Quelques études d'impact sur les droits humains (EIDH) et labels de qualité ont également été développés.

- **Initiatives visant à favoriser l'alignement des valeurs en promouvant les droits de l'homme et les valeurs centrées sur l'homme et en rendant obligatoires les études d'impact sur les droits de l'homme (EIDH) :** la France, en collaboration avec le Forum économique mondial, a établi un cadre d'action pour répondre aux préoccupations en matière de droits de l'homme découlant de l'utilisation de la reconnaissance faciale par l'IA. En décembre 2020, la Corée a élaboré des lignes directrices nationales sur l'éthique dans le domaine de l'IA, en vue de réaliser une IA centrée sur l'homme pour tous les membres de la société à tous les stades de son développement et de son utilisation. Aux États-Unis, le département d'État a publié des orientations pratiques en matière de droits de l'homme à l'intention des entreprises américaines qui proposent des produits ou des services impliquant des capacités de surveillance. Une évaluation d'impact algorithmique et des droits fondamentaux est désormais obligatoire en cas d'utilisation d'algorithmes par les autorités publiques néerlandaises.
- **Protection de la vie privée :** l'IA, et en particulier l'IA générative, fait peser des menaces considérables sur la vie privée. Les Adhérents mettent en œuvre diverses politiques pour protéger la vie privée dans l'IA, notamment par le biais de législations existantes sur la protection des données et de la vie privée, d'orientations sur la protection des données dans les applications d'IA (Corée et Mexique), de bacs à sable réglementaires destinés à promouvoir le développement d'une utilisation de solutions d'IA respectueuse de la vie privée (OCDE, 2023<sup>[10]</sup>), et de la promotion de technologies protectrices de la vie privée afin de prévenir les atteintes à la vie privée lors de la conception ou de l'exploitation de technologies et de services d'IA (par exemple, Estonie, États-Unis, Royaume-Uni, Türkiye) (OCDE, 2023<sup>[11]</sup>).
- **Labels de qualité et certifications promouvant les valeurs centrées sur l'humain :** quelques Adhérents ont développé ou sont en train de développer des labels de qualité

et des certifications confirmant qu'un outil d'IA est éthique et centré sur l'humain. Parmi les exemples à ce jour le label pour une IA digne de confiance en Türkiye.

- **Initiatives visant à réduire les biais de l'IA :** plusieurs Adhérents ont lancé des initiatives destinées à identifier et à rectifier les biais et les problèmes d'éthique dans les systèmes d'IA. Il s'agit notamment de l'atelier Bigscience en France, qui analyse les biais et les problèmes d'éthique dans les modèles de langage et propose des mesures et des outils pour les évaluer et les atténuer ; d'une enquête sur les causes de la discrimination dans l'IA menée par le ministère néerlandais de l'Intérieur et des Relations au sein du royaume ; et du document « Review into Bias in Algorithmic Decision-Making » publié par le Centre for Data Ethics and Innovation (CDEI) du Royaume-Uni en 2020. Aux États-Unis, l'Equal Employment Opportunity Commission a lancé une « AI and Algorithmic Fairness Initiative » à l'échelle de l'agence pour veiller à ce que l'utilisation de l'IA soit conforme aux lois américaines en matière de lutte contre la discrimination. En outre, la Commission fédérale du commerce (FTC) a désigné l'IA comme une priorité pour l'action coercitive et réglementaire dans les enquêtes portant sur des pratiques ou des actes déloyaux, trompeurs, anticoncurrentiels, collusoires, coercitifs, prédateurs, d'exploitation et d'exclusion liés aux algorithmes et à la biométrie. Certains pays soumettent également les outils de recrutement à des exigences en matière d'audit de l'IA (OCDE, 2023<sup>[12]</sup>).
- **Valeurs démocratiques :** les Adhérents et les organisations internationales proposent des législations visant à garantir que les systèmes d'IA respectent les valeurs démocratiques, par exemple la proposition de loi canadienne sur l'intelligence artificielle et les données (LIAD), la législation sur l'IA de l'Union européenne et la convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit comprennent des dispositions sur la transparence et la responsabilité (voir Principe 2.3).

39. Ce Principe reste pertinent, mais pourrait gagner à être actualisé afin de refléter les préoccupations du point de vue de l'action des pouvoirs publics suscitées par les évolutions récentes survenues dans le domaine de l'IA générative.

40. L'intitulé de ce Principe pourrait mieux refléter le texte, en rappelant la nécessité de respecter l'État de droit, les droits de l'homme et les valeurs démocratiques. Les préoccupations relatives aux atteintes à la vie privée figuraient parmi les principaux risques jugés prioritaires par l'OCDE dans le contexte de l'IA générative (Lorenz, Perset et Berryhill, 2023<sup>[13]</sup>). Si le texte du Principe aborde déjà la question de la protection de la vie privée, il serait utile que son intitulé y fasse explicitement mention, au même titre que l'équité, afin d'assurer la cohérence avec les domaines prioritaires d'action identifiés par l'OCDE et les membres du G7 et par l'UE dans sa proposition de législation sur l'IA, ainsi qu'avec les risques prioritaires abordés dans les documents d'orientation pratiques (notamment dans le cadre de gestion des risques liés à l'IA du NIST aux États-Unis).

41. Les évolutions récentes survenues dans le domaine de l'IA générative suscitent une grande inquiétude quant au risque d'abus lié à la création et à la diffusion de contenus générés par ordinateur, de plus en plus réalistes et convaincants. Les conséquences pourraient s'étendre à la diffusion de mésinformations et de désinformations, à la perpétuation de la discrimination, à la distorsion du discours public et des marchés, à l'incitation à la violence et aux menaces connexes pour les processus démocratiques et les droits de l'homme. Étant donné que les pouvoirs publics reconnaissent l'effet transformateur de l'IA générative et s'emploient activement à relever les défis qu'il pose, les Principes pourraient explicitement souligner la nécessité

de lutter contre la mésinformation et la désinformation, en accord avec les domaines prioritaires d'action en matière d'IA générative identifiés par les Adhérents (Graphique 3.4) et le G7 (OCDE, 2023<sup>[14]</sup>).

42. En conséquence de la sophistication croissante des technologies d'IA, leur utilisation en dehors de la finalité assignée, ou une utilisation abusive, qu'elle soit intentionnelle ou non, est un facteur de risque. En raison de leurs capacités d'adaptabilité et d'apprentissage, les systèmes d'IA peuvent être réaffectés à des tâches ou à des applications non prévues dans leur conception initiale, ce qui peut avoir des conséquences imprévues. L'utilisation abusive intentionnelle de l'IA est également très préoccupante, en particulier dans des scénarios où des acteurs malveillants cherchent à exploiter les systèmes d'IA à des fins telles que des cyberattaques, la mésinformation ou la désinformation. En outre, une utilisation abusive non intentionnelle de l'IA peut découler d'une erreur humaine, de défaillances techniques ou de circonstances imprévues. À la lumière de ces risques, le Principe pourrait appeler les acteurs de l'IA à mettre en œuvre des mécanismes et des garanties solides, notamment la capacité d'une surveillance et d'une intervention humaines tout au long du cycle de vie de l'IA.

43. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 1.2 sur les valeurs centrées sur l'homme et l'équité. Bien que le Principe reste pertinent, des clarifications et des ajouts seraient appropriés pour refléter les défis importants qui sont apparus au cours des cinq dernières années. Il s'agit notamment de mentionner explicitement la protection de la vie privée dans l'intitulé et d'ajouter dans le texte la nécessité de lutter contre la mésinformation et la désinformation. Enfin, le texte pourrait appeler les acteurs de l'IA à mettre en œuvre des mécanismes pour gérer les risques découlant d'une utilisation en dehors de la finalité assignée et d'une utilisation abusive, qu'elle soit intentionnelle ou non, des systèmes d'IA.**

#### *Transparence et explicabilité (Principe 1.3)*

*« 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 : de favoriser une compréhension générale des systèmes d'IA, d'informer les parties prenantes de leurs interactions avec les systèmes d'IA, y compris dans la sphère professionnelle, de permettre aux personnes concernées par un système d'IA d'en appréhender le résultat, et, de permettre aux personnes subissant les effets néfastes d'un système d'IA de contester les résultats sur la base d'informations claires et facilement compréhensibles sur les facteurs, et sur la logique ayant servi à la formulation de prévisions, recommandations ou décisions. »*

44. La plupart des stratégies nationales en matière d'IA, des cadres d'éthique et des principes généraux pour la mise en œuvre de l'IA citent la transparence et l'explicabilité parmi les principales caractéristiques d'un système d'IA digne de confiance. La transparence et l'explicabilité figurent également en bonne place dans plusieurs lignes directrices non contraignantes pour une mise en œuvre éthique de l'IA. Cependant, bien que l'importance de ces concepts soit largement reconnue, leurs multiples nuances les rendent complexes à mettre en œuvre. La transparence de l'IA implique de communiquer clairement aux utilisateurs qu'ils ont affaire à un système d'IA, de leur permettre d'interpréter ses résultats et, dans certains cas, d'expliquer la logique étayant la prise de décision.

45. Les Adhérents garantissent la transparence de l'IA de différentes façons, allant de l'élaboration de lignes directrices pour la mise en œuvre de l'IA à l'instauration

d'organes de surveillance. Des dispositions relatives à la transparence sont établies dans la législation existante (par exemple, la législation sur la protection des données et de la vie privée, la législation sur la protection des consommateurs) et sont également incluses aux propositions de réglementations propres à l'IA, avec plusieurs dispositions spécifiques applicables à la sphère professionnelle. Dans le secteur public, les Adhérents améliorent la transparence entourant l'utilisation de l'IA pour les services publics grâce à des registres d'IA. Exemples d'initiatives visant à promouvoir la transparence et l'explicabilité :

- **Initiatives exigeant la divulgation d'informations en cas d'utilisation de systèmes d'IA :** le Japon a introduit des exigences de transparence dans la loi sur la transparence des plateformes numériques en exigeant des fournisseurs de plateformes numériques désignés qu'ils garantissent la transparence et l'égalité de traitement dans leurs transactions avec des utilisateurs professionnels. La réglementation de l'UE sur l'intelligence artificielle (ci-après « législation de l'UE sur l'IA ») (qui en est aux derniers stades du processus législatif au moment de la rédaction du présent rapport) prévoit des obligations de transparence pour les systèmes d'IA à haut risque, ainsi que pour ceux destinés à interagir avec des personnes physiques et ceux utilisés pour la reconnaissance des émotions, pour la catégorisation biométrique ou pour générer ou manipuler des contenus image, audio ou vidéo, et pour les systèmes d'IA à usage général.
  - **Dans le secteur public :** la loi française de 2016 pour une République numérique impose la transparence des algorithmes utilisés par les pouvoirs publics. La Finlande et les Pays-Bas ont lancé des registres d'IA en libre accès grâce auxquels il est possible de savoir de quelle façon les algorithmes sont utilisés dans leur capitale respective (Helsinki et Amsterdam). La norme britannique d'enregistrement de la transparence algorithmique (2022) comprend des lignes directrices complètes à l'intention du secteur public, y compris du gouvernement, sur la divulgation d'informations dans le cadre de l'utilisation d'outils algorithmiques.
  - **Dans la sphère professionnelle :** la France, l'Allemagne et l'Italie, entre autres, ont adopté des lois qui incluent l'obligation d'obtenir l'accord préalable des représentants des travailleurs en cas d'utilisation de technologies numériques, dont l'IA, pour surveiller les travailleurs. La loi visant à œuvrer pour les travailleurs de la province canadienne de l'Ontario exige des employeurs qu'ils informent leurs employés de leur politique de surveillance électronique. De même, plusieurs États fédéraux des États-Unis ont adopté des lois obligeant les employeurs à informer leurs employés en cas de surveillance électronique.
- **Initiatives concernant la fourniture d'informations sur le fonctionnement de l'IA :** la proposition de loi canadienne sur l'intelligence artificielle et les données (LIAD) introduit des exigences visant à promouvoir la transparence sur l'utilisation de l'IA. Le décret-loi royal espagnol 9/2021 (la « Loi Rider ») rend la transparence obligatoire pour les systèmes d'IA qui influencent les conditions de travail ou le statut de l'emploi, ou prennent des décisions les concernant. La législation de l'UE sur l'IA prescrit des exigences en matière d'information qui permettraient aux utilisateurs d'interpréter les résultats du système et de les utiliser de manière appropriée.
- **Initiatives concernant la fourniture d'informations sur les facteurs et les processus de décision et créant des voies de recours :** au Canada, la loi sur la protection de la vie privée des consommateurs comporte des exigences en matière de transparence. De plus, la directive sur la prise de décision automatisée fixe un large éventail de mesures

obligatoires pour garantir l'utilisation responsable de l'IA par les institutions fédérales du pays. Le cadre juridique mexicain en matière de protection des données à caractère personnel contient également des dispositions applicables aux décisions automatisées. La loi norvégienne sur l'administration publique dispose que toute décision prise par le secteur public et concernant une personne en particulier doit être expliquée afin de garantir la responsabilité et de rendre possible une procédure de plainte ou de recours. Le règlement général sur la protection des données de l'UE établit un « droit à l'explication » dans son article 22.

**46. Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 1.3 sur la transparence et l'explicabilité. Le Principe reste pertinent, mais pourrait bénéficier de quelques clarifications quant aux informations que les acteurs de l'IA devraient fournir en relation avec les systèmes d'IA en général, ainsi qu'aux personnes concernées par un système d'IA et à celles affectées de façon préjudiciable par les résultats d'un système d'IA.**

*Robustesse, sûreté et sécurité (Principe 1.4)*

*« 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 un risque de sécurité démesuré.*

*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 produits par lesdits systèmes d'IA et le traitement des demandes d'information, compte tenu du contexte et de l'état de l'art de la technologie.*

*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, afin de gérer les risques y afférents, notamment ceux liés au respect de la vie privée, à la sécurité numérique, à la sûreté et aux biais. »*

47. Les questions de robustesse, de sûreté et de sécurité de l'IA sont interdépendantes. Par exemple, la sécurité numérique peut affecter la sûreté des produits connectés tels que les automobiles et les appareils ménagers quand les risques ne sont pas gérés de manière appropriée. Elles sont donc analysées ensemble. Toutefois, les Adhérents peuvent les mettre en œuvre de différentes manières, qui ne s'excluent pas mutuellement.

48. Les Adhérents s'appuient sur des lignes directrices, des cadres d'éthique, des évaluations d'impact, de nouvelles législations, des modifications de législations existantes et d'autres instruments pour mettre en œuvre le Principe 1.4. En voici quelques exemples :

- **Évaluations d'impact algorithmique** : la directive canadienne sur la prise de décision automatisée exige que les institutions fédérales qui prévoient d'utiliser un système automatisé, y compris un système s'appuyant sur l'IA, pour prendre ou soutenir des décisions administratives, réalisent et publient une évaluation d'impact algorithmique (EAI) avant la mise en service du système. Les principes et le guide d'analyse d'impact du Mexique pour le développement et l'utilisation de systèmes basés sur l'IA dans l'administration publique fédérale sont conçus pour évaluer les implications sociétales et éthiques des systèmes d'IA qu'elle développe.

- **Initiatives concernant la conservation d'enregistrements des caractéristiques des données à des fins de traçabilité :** en Türkiye, des efforts sont en cours pour créer un dictionnaire national des données, destiné à compiler un inventaire national de données et à établir des processus de gestion et de contrôle par le biais d'une architecture nationale d'intégration des données. Aux États-Unis, le mémorandum présidentiel sur la sécurité nationale (NSPM) « Protecting the United States Advantage in AI and Related Critical Technologies » vise à améliorer l'accès à des ressources informatiques, des données et des modèles fédéraux de haute qualité et entièrement traçables.
- **Lois et règlements prévenant les risques déraisonnables pour la sécurité des systèmes d'IA – conduite autonome :** en Autriche, en Allemagne, au Danemark, au Japon, en Lituanie et au Royaume-Uni, une nouvelle législation a été adoptée (ou la législation existante a été modifiée) pour encadrer l'utilisation des voitures autonomes sur les routes nationales de chacun de ces pays.

49. Les travaux sur la responsabilité ont mis en évidence la nécessité d'une traçabilité et d'une approche systématique de la gestion du risque à chaque phase du cycle de vie de l'IA et sur une base continue. Bien que faisant initialement partie du Principe intitulé « Robustesse, sûreté et sécurité », il est désormais clair que la traçabilité et la gestion du risque s'appliquent à tous les Principes fondés sur des valeurs et à toutes les phases du cycle de vie, et devraient donc être intégrées au Principe intitulé « Responsabilité ». Les deux thèmes sont actuellement couverts aux points (b) et (c) du Principe 1.4. (Robustesse, sûreté et sécurité), et trouveraient mieux leur place sous le Principe 1.5 sur la responsabilité.

50. Comme souligné dans l'examen du Principe 1.2, l'essor de l'IA générative suscite des préoccupations quant au risque de création et d'utilisation abusive de contenus générés par ordinateur, ainsi que de propagation de la désinformation et de la désinformation, ce qui pourrait nuire aux processus démocratiques. Cette situation souligne la nécessité de soutenir les efforts visant à protéger l'intégrité de l'information, car elle est essentielle pour garantir que les systèmes d'IA n'affectent pas négativement la capacité des individus à obtenir des sources d'information précises, fondées sur des données empiriques et plurielles. Respecter l'intégrité de l'information est essentiel pour préserver la liberté d'expression, qui englobe celle de chercher, de recevoir et de communiquer des éléments d'information et des idées (OCDE, 2024<sub>[15]</sub>).

51. En outre, à la lumière des débats portant sur le risque que les humains perdent le contrôle des systèmes d'IA – un thème important du Sommet sur la sécurité de l'intelligence artificielle organisé par le Royaume-Uni en novembre 2023 (AI Safety Summit, 2023<sub>[16]</sub>) et des prochains sommets prévus en Corée et en France –, le Principe pourrait inclure une référence à la nécessité de disposer de mécanismes garantissant que les systèmes d'IA susceptibles de causer un préjudice ou présentant un comportement inapproprié peuvent être neutralisés, réparés et/ou mis hors service selon les besoins.

52. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 1.4 sur la robustesse, la sûreté et la sécurité. Si la pertinence du Principe demeure, il pourrait être intéressant de souligner explicitement la nécessité de mettre en place des mécanismes permettant de renforcer l'intégrité de l'information et de retirer les systèmes d'IA qui risquent de causer un préjudice ou présentent un comportement inapproprié. En outre, le texte qui figure actuellement sous le Principe 1.4. (Robustesse, sûreté et sécurité) et concerne la traçabilité et une approche systématique de la gestion du risque trouverait mieux sa place sous le Principe 1.5 (Responsabilité).**

### *Responsabilité (Principe 1.5)*

*« 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. »*

53. La responsabilité implique que les organisations ou les individus qui conçoivent, développent, exploitent ou déploient des systèmes d'IA, en garantissent le bon fonctionnement et en sont tenus responsables, tout au long du cycle de vie, en fonction de leur rôle et des cadres réglementaires applicables, et qu'ils démontrent s'acquitter de leur obligation par leurs actes et leurs processus de prise de décision. Dans le cas d'un résultat négatif, elle implique également de prendre des mesures pour garantir un meilleur résultat à l'avenir.

54. La demande d'outils et de processus permettant de documenter les décisions du système d'IA et de garantir la responsabilité est en hausse, tant dans le secteur public que dans le secteur privé. Ce domaine englobe d'importantes initiatives de normalisation de l'IA menées par des organisations telles que l'Organisation internationale de normalisation (ISO), l'Institut des ingénieurs électriciens et électroniciens (IEEE), l'Union internationale des télécommunications (UIT), le National Institute of Standards and Technology aux États-Unis (NIST), l'Institut européen des normes de télécommunications (ETSI), l'Internet Engineering Task Force (IETF) et le Comité européen de normalisation électrotechnique (CEN-CENELEC). Ces initiatives portent sur différents axes, notamment la conception de l'IA (intégration des problématiques de confiance dès la phase de conception, par exemple), les analyses d'impact, les évaluations de conformité et les cadres de gestion des risques pour l'IA. À cela s'ajoutent des initiatives gouvernementales et intergouvernementales comme la législation de l'UE sur l'IA, l'AI Standards Hub au Royaume-Uni, l'Alliance européenne de l'IA, le Comité sur l'intelligence artificielle du Conseil de l'Europe et le Conseil du commerce et des technologies UE-États-Unis. Des systèmes de certification font également partie de ce paysage (OCDE, 2023<sup>[7]</sup>).

55. Les Adhérents ont élaboré des lignes directrices pour l'utilisation ou la mise en œuvre de l'IA dans plusieurs secteurs (administration publique, soins de santé, conduite autonome), ces initiatives pouvant être considérées comme des codes de conduite éthique. La proposition de réglementation propre à l'IA prévoit la documentation du bon fonctionnement des systèmes d'IA tout au long de leur cycle de vie. Enfin, les Adhérents et l'UE ont mis en place des organismes de surveillance indépendants pour l'IA et les algorithmes. En voici quelques exemples :

- **Législation imposant la documentation du bon fonctionnement des systèmes d'IA tout au long de leur cycle de vie** : au Canada, la proposition de loi sur l'intelligence artificielle et les données (LIAD) garantit la responsabilité par la documentation proactive des politiques, des procédures et des mesures mises en œuvre ainsi que des moyens de satisfaire aux exigences de conception et de développement. Dans l'UE, la législation sur l'IA rend obligatoire la documentation technique.
- Codes de conduite éthique et outils techniques pratiques : la France, le Portugal, la Norvège et le Royaume-Uni ont, entre autres, publié des orientations, y compris des exigences de transparence, pour le développement et l'utilisation de l'IA dans le secteur public. La France et Singapour ont également établi des codes de conduite éthique dans le secteur des soins de santé. En février 2022, la Corée a élaboré une « liste de contrôle d'auto-évaluation » comme moyen pratique de mettre en œuvre les lignes directrices nationales sur l'éthique en matière d'IA, des lignes directrices plus détaillées (exemples spécifiques aux différents secteurs et situations) ayant été publiées en avril 2023. Au

Mexique, l'Institut national pour la transparence, l'accès à l'information et la protection des données à caractère personnel (Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales, INAI) a élaboré conjointement avec le Réseau ibéro-américain de protection des données (RIPD) des recommandations générales pour le traitement des données à caractère personnel dans l'intelligence artificielle (Ibero-American Data Protection Network, 2019<sub>[17]</sub>) et des lignes directrices spécifiques pour le respect des principes et des droits qui régissent la protection des données à caractère personnel dans les projets d'intelligence artificielle (Ibero-American Data Protection Network, 2019<sub>[18]</sub>), qui énoncent des orientations à l'intention des développeurs et des fabricants travaillant dans le domaine de l'IA sur les exigences réglementaires applicables au traitement des données à caractère personnel. Au Royaume-Uni, le Centre for Data Ethics and Innovation (CDEI) a développé un portefeuille de cas d'utilisation et un référentiel consultable en ligne d'outils d'assurance dans l'IA (CDEI, 2024<sub>[19]</sub>). Singapour a lancé I.A. Verify, un cadre de test pour la gouvernance de l'IA qui aide les entreprises en matière de transparence.

- **Organes de surveillance indépendants chargés de la supervision de l'IA** : les agences de contrôle et les organes de surveillance jouent un rôle de plus en plus important en garantissant une utilisation responsable de l'IA. En décembre 2022, l'Espagne a annoncé la création de l'Agence espagnole de supervision de l'IA, chargée de promouvoir une IA responsable, durable et digne de confiance, et de favoriser la collaboration et la coordination avec d'autres autorités nationales et supranationales ayant la même fonction. Les statuts de l'Agence ont été approuvés en août 2023 et son organe directeur a été désigné en décembre 2023. Aux Pays-Bas, l'unité de supervision des algorithmes, au sein de l'Autorité néerlandaise de protection des données, a pour objectif de renforcer la supervision des algorithmes qui traitent des données à caractère personnel.

56. Étant donné que l'IA pénètre tous les secteurs de l'économie et de la société, et conformément aux réflexions en cours du groupe d'experts de l'OCDE sur les risques et les responsabilités en matière d'IA, il serait important de mentionner dans ce Principe la nécessité pour les acteurs de l'IA d'adopter une conduite responsable afin de gérer les risques liés aux systèmes d'IA. À l'avenir, les acteurs de l'IA devront coopérer de plus en plus entre eux, mais aussi avec des acteurs en amont et en aval, c'est-à-dire des fournisseurs de connaissances et de ressources en matière d'IA, des utilisateurs de systèmes d'IA et d'autres parties prenantes. Ces dernières peuvent être des fournisseurs de matériel, des investisseurs, des détenteurs de droits et des instituts de recherche, par exemple.

57. D'autres révisions pourraient préciser que les risques de biais à gérer ne se réfèrent pas à l'ensemble des biais, mais spécifiquement à ceux qui sont « préjudiciables », par exemple les biais qui conduisent à l'injustice, à la discrimination ou qui renforcent les stéréotypes et les inégalités. Le texte pourrait mentionner explicitement les risques de violation des droits de propriété intellectuelle, une préoccupation du point de vue de l'action des pouvoirs publics qui s'est considérablement accrue avec l'essor de l'IA générative, ainsi que les droits des travailleurs

58. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 1.5 sur la responsabilité. Bien que le Principe reste pertinent, le texte qui figure actuellement sous le Principe 1.4. (Robustesse, sûreté et sécurité) et concerne la traçabilité et une approche systématique de la gestion du risque trouverait mieux sa place sous le Principe 1.5 (Responsabilité). En outre, il serait important de mentionner la nécessité pour les acteurs de l'IA d'adopter une**

**conduite responsable afin de gérer les risques liés aux systèmes d'IA. D'autres révisions pourraient préciser que les risques de biais à gérer ne se réfèrent pas à l'ensemble des biais, mais spécifiquement à ceux qui sont « préjudiciables », et mentionner explicitement les risques liés à l'atteinte aux droits de travail et de propriété intellectuelle.**

### ***3.2.3. Section 2 : Politiques nationales et coopération internationale à l'appui d'une IA digne de confiance***

#### *Investir dans la recherche et le développement en matière d'IA (Principe 2.1)*

*« a) Les pouvoirs publics devraient envisager des investissements publics à long terme et encourager les investissements privés dans la recherche et le développement, notamment interdisciplinaire, afin de stimuler l'innovation dans une IA digne de confiance, relevant des 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 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 sans partialité induite et renforcer l'interopérabilité et l'utilisation de normes. »*

59. De nombreux Adhérents ont reconnu l'importance des politiques de soutien à la R-D en matière d'IA et y donnent suite en lançant des initiatives visant à intensifier les efforts dans ce domaine. À ce jour, les Adhérents ont fait état d'environ 290 programmes d'action se rapportant à ce Principe. La plupart des stratégies nationales en matière d'IA mettent l'accent sur la R-D et en font un domaine d'action prioritaire. Les Adhérents disposent de programmes de financement dédiés et utilisent différents instruments pour leur mise en œuvre. L'une des grandes tendances est la création d'instituts et de centres nationaux de recherche sur l'IA. D'autres actions comprennent la mise en place de politiques, de plans, de programmes et de fonds axés sur la R-D en matière d'IA afin de soutenir la diffusion de l'IA au sein des entreprises et du secteur public et de consolider les réseaux de recherche dans le domaine.

- **Financement public à l'appui de la R-D en matière d'IA** : bien que disponibles pour certains pays, les informations sur la R-D en matière d'IA financée par le secteur public restent imprécises et difficiles à comparer. Par exemple, la France a alloué par le biais de son Programme national de recherche en IA 445 millions EUR à la recherche en IA entre 2018 et 2022. Aux États-Unis, un financement de 1,8 milliard USD a été demandé pour la R-D en IA non militaire en 2023 (National Council of Science and Technology, 2022<sup>[20]</sup>). Plus récemment, la Fondation nationale pour la science (National Science Foundation, NSF) et le Bureau de la politique scientifique et technologique (Office of Science and Technology Policy, OSTP) ont proposé d'établir une ressource nationale de recherche en IA (National AI Research Resource, NAIRR) dotée d'un financement estimé à 2,6 milliards USD pour les années 2023-2029. L'Union européenne a affecté 1 milliard EUR par an à l'IA, notamment dans le cadre du programme Horizon Europe et du programme pour une Europe numérique (Commission européenne, 2023<sup>[21]</sup>).
- **Instituts, centres et réseaux nationaux de recherche en IA** : plusieurs pays ont créé des instituts et des centres nationaux de recherche en IA, en chargeant des institutions ou des organisations spécialisées de promouvoir la recherche, le développement et l'innovation dans ce domaine. Ces centres ont généralement pour mission de faire

progresser les technologies d'IA, de favoriser la collaboration entre le monde universitaire, l'industrie et les pouvoirs publics, et de contribuer à l'écosystème national de l'IA au sens large. Parmi ces centres, on peut citer les trois instituts nationaux pour l'IA du Canada (Amii, l'Institut Vecteur et Mila) ; en France, l'Institut national de recherche en sciences et technologies du numérique (INRIA), le centre IA pour la science et Science pour l'IA (AISSAI) au sein du Centre national de la recherche scientifique (CNRS), LaborIA, et un réseau de quatre instituts interdisciplinaires pour l'IA (3AI) établi par la stratégie nationale française en matière d'IA ; les six centres d'excellence allemands en matière d'IA (dont le Centre allemand de recherche en intelligence artificielle « DFKI ») et les quatre centres de services d'IA allemands, le centre national australien de l'IA et le centre coréen de données de recherche de l'AI Innovation Hub. L'Italie a établi en 2022 « FAIR » (Future Artificial Intelligence Research), un réseau de recherche national qui regroupe des instituts de recherche, des universités et des entreprises et cherche à faire progresser la recherche et l'innovation dans le domaine de l'IA. Aux États-Unis, le NSF finance divers instituts dans tout le pays, soutenant des projets interdisciplinaires portant sur des questions fondamentales de la recherche en IA. En outre, les Centres de recherche appliquée en IA, le Centre d'innovation appliquée et le Centre national pour l'innovation et l'IA ont été créés respectivement au Brésil, en Égypte et au Pérou.

- **Investissements dans les ensembles de données en libre accès :** les données de l'indice des données Ouvertes, Utiles et Réutilisables 2023 de l'OCDE montrent que « les pays de l'OCDE ont amélioré la qualité des données publiques ouvertes, une ressource importante compte tenu des progrès récents de l'IA » (OCDE, 2023<sup>[22]</sup>). Plusieurs Adhérents ont lancé des initiatives à l'appui des ensembles de données en libre accès destinés à la recherche et au développement, tout en garantissant la protection des données et de la vie privée. Parmi les exemples, on peut citer le Portail du gouvernement ouvert canadien, qui permet d'accéder à un large éventail d'ensembles de données provenant d'agences et de ministères fédéraux, la plateforme Etalab en France et le portail néerlandais de données ouvertes aux Pays-Bas. L'Open Data Institute (ODI) du Royaume-Uni encourage l'utilisation de données ouvertes et fournit des orientations sur la protection des données et de la vie privée, et Data.gov aux États-Unis permet d'accéder à une vaste gamme d'ensembles de données dans différents secteurs.

60. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 2.1 sur l'investissement dans la recherche et le développement en matière d'IA. Le Principe reste pertinent. Il pourrait toutefois bénéficier de quelques ajouts pour refléter l'importance d'investir dans la science ouverte et les outils en libre accès, qui ont tous deux été essentiels au développement de l'IA à ce jour. En outre, le texte pourrait être révisé pour être cohérent avec la clarification du Principe 1.5 concernant les biais « préjudiciables ».**

*Favoriser l'instauration d'un écosystème numérique pour l'IA (Principe 2.2)*

*« Les pouvoirs publics devraient favoriser le développement et l'accessibilité d'un écosystème numérique à l'appui d'une IA digne de confiance. Cet écosystème se compose notamment des technologies et infrastructures numériques 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 fiduciaires de données (« data trusts »), pour favoriser le partage des données de façon sûre, équitable, légale et éthique. »*

61. La capacité à amorcer une transformation fondée sur l'IA dépend de la disponibilité des données, des infrastructures et des logiciels nécessaires pour entraîner

et utiliser les modèles d'IA à grande échelle. Favoriser l'instauration d'un écosystème numérique pour l'IA représente donc un élément essentiel des efforts déployés par les pays pour progresser dans leur adoption de l'IA. À ce jour, les Adhérents ont fait état de 400 programmes d'action en rapport avec le Principe 2.2 visant à favoriser l'instauration d'un écosystème numérique pour l'IA. Il s'agit aussi bien de documents stratégiques de haut niveau, comme les stratégies nationales en matière d'IA, les stratégies en matière de données et les feuilles de route technologiques, que d'actions spécifiques, comme les mesures en matière d'IA prises en réponse à la pandémie de COVID-19.

62. Les grandes tendances observées dans ce domaine sont l'accroissement de la capacité de calcul et l'amélioration de l'accès aux infrastructures, les politiques publiques en faveur de l'ouverture et du partage des données et les investissements dans les technologies de traitement automatique du langage naturel.

- **Initiatives visant à accroître la capacité de calcul et l'accès à l'infrastructure :** les chercheurs et les industries ont besoin d'un accès aux technologies d'IA et aux capacités de calcul pour pouvoir développer des solutions innovantes. La moitié des Adhérents et l'Union européenne ont fait état à ce jour d'instruments d'action portant sur l'infrastructure de calcul et de recherche en matière d'IA. Il s'agit par exemple de politiques visant à soutenir les capacités de calcul haute performance (dont la République tchèque, la Finlande, la France, le Japon, le Portugal, la Slovaquie, la Turquie, le Royaume-Uni), notamment pour améliorer l'accès des chercheurs et des jeunes pousses dans le domaine de l'IA (Canada, Israël, États-Unis et Union européenne), ou pour mettre en commun des ressources afin de renforcer la capacité de calcul (Union européenne) ou des investissements afin d'améliorer la compétitivité de l'infonuagique (Corée et Union européenne) et de renforcer l'offre nationale de puces destinées à l'IA (Corée et États-Unis). À ce jour, rares sont les Adhérents à avoir réalisé une cartographie nationale de leurs capacités de calcul et besoins connexes. On peut citer l'évaluation des besoins de l'infrastructure de recherche numérique du Canada de 2020 (Alliance de recherche numérique du Canada, 2020<sup>[23]</sup>) et l'examen des besoins de l'infrastructure de recherche numérique du Royaume-Uni en 2023 (UK DSIT, 2023<sup>[24]</sup>).
- **Politiques en matière de données :** l'IA a besoin d'une grande quantité de données pour reconnaître les formes, apprendre, faire des prédictions ou prendre des décisions précises. Il est important de lier les politiques en matière de données aux politiques en matière d'IA, car cela aide à préserver l'éthique, la protection et la sécurité des données face aux problèmes de confidentialité, tout en garantissant le fonctionnement efficace de l'IA. Alors que l'indice relatif à l'administration numérique 2023 de l'OCDE montre que « seuls 59 % des pays de l'OCDE disposent d'une stratégie en matière de données ou d'un instrument similaire pour le secteur public » (OCDE, 2024<sup>[25]</sup>), plusieurs Adhérents ont commencé à établir un lien entre leurs politiques en matière d'accès aux données et de partage des données et leurs politiques en matière d'IA. Parmi les exemples de ces initiatives, citons des stratégies visant à améliorer l'accès aux données et leur partage (Corée, Suède), à fournir un accès aux informations et aux données scientifiques (République tchèque), y compris dans des secteurs spécifiques (par exemple la santé en Norvège), à créer des plateformes (Belgique, France, Hongrie, Norvège) ou des infrastructures et des plateformes de données (Pérou) pour centraliser les données du secteur public et partager les données publiques ouvertes (OCDE, 2023<sup>[26]</sup>), ou encore à fournir une infrastructure de données accessibles à l'appui du développement de systèmes d'IA (Colombie). D'autres Adhérents, comme le Royaume-Uni, ont créé des fiduciaires de données pour permettre à des individus ou à des organisations de gérer, partager et administrer collectivement des données dans un but

spécifique. Plusieurs Adhérents et l'UE établissent également des espaces de données pour stocker, gérer, traiter et rendre accessibles les données (par exemple, l'Union européenne), y compris dans des secteurs spécifiques (par exemple, la santé en France et la santé et mobilité en Allemagne).

- **Investissements dans les technologies de traitement automatique du langage naturel** : plusieurs Adhérents ont lancé des programmes d'action visant à promouvoir les modèles de langage dans leurs langues nationales. Parmi les exemples de ces initiatives, citons « Danish Gigaword Project » au Danemark, « Estonian Language Technology 2018-2027 » en Estonie, « Pour des IA Francophones » en France, les projets d'Israël visant à créer des ensembles de données, des modèles et des outils pour le traitement automatique du langage naturel en hébreu et en arabe, « Global Communication Plan 2025 » au Japon, « National Initiative for Language Technologies » en Corée, « Norwegian Language Bank » en Norvège et « National Plan for the Advancement of Language Technologies » en Espagne.

63. Au vu de l'importance croissante de la puissance de calcul pour le développement et l'adoption de l'IA, ce Principe est de plus en plus pertinent, et les Adhérents devraient continuer à travailler pour assurer la disponibilité d'une infrastructure de calcul avancée, et pour mettre la capacité de calcul à la disposition des chercheurs et des jeunes pousses. Ces efforts devraient s'appuyer sur des évaluations nationales des capacités de calcul et des besoins connexes, en mettant l'accent sur la capacité, l'efficacité et la résilience (OCDE, 2023<sup>[27]</sup>). Sur cette base, les Adhérents pourraient élaborer des plans nationaux en matière de capacité de calcul pour l'IA ou intégrer de tels plans à leurs stratégies nationales en matière d'IA.

64. Les politiques visant à favoriser la capacité de calcul devraient également prendre en considération les incidences de l'infrastructure de calcul sur l'environnement et la durabilité. Il est important de noter que, parallèlement, le Comité de la politique du numérique (CPN) examine la Recommandation de l'OCDE de 2010 sur les technologies de l'information et des communications et l'environnement [[OCDE/LEGAL/0380](#)], afin de clarifier le paysage de l'action publique dans le domaine du numérique à cet égard.

65. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 2.2 visant à favoriser l'instauration d'un écosystème numérique pour l'IA. Bien que le Principe reste pertinent, il pourrait mieux refléter dans son intitulé la nature habilitante de l'écosystème numérique et la nécessité de favoriser l'instauration d'un écosystème numérique inclusif, c'est-à-dire d'assurer l'accès aux ressources de l'IA à un ensemble diversifié d'utilisateurs. La formulation pourrait être clarifiée en mentionnant explicitement les éléments de l'écosystème numérique qui devraient être abordés sous cet intitulé, notamment les données, les technologies d'IA, l'infrastructure de calcul et de connectivité, et le partage des connaissances. Enfin, le texte devrait également faire référence aux caractéristiques qu'un écosystème d'IA devrait incarner, à savoir l'inclusivité, le dynamisme, la durabilité et l'interopérabilité. Ceci signifie que les gouvernements devraient favoriser un écosystème qui encourage l'accès de diverses parties prenantes et leur contribution, réagit à bref délai à l'évolution rapide des technologies, donne la priorité au bien-être sociétal, environnemental et économique à long terme - notamment en minimisant son empreinte carbone et sa consommation de ressources -, et facilite une intégration sans heurts des systèmes d'IA dans les différents domaines.**

*Façonner un cadre d'action favorable à l'IA (Principe 2.3)*

- a) *Les pouvoirs publics devraient promouvoir l'instauration d'un cadre d'action favorable qui soutienne une transition souple du stade de recherche et développement à celui de déploiement 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 monter en puissance, selon les besoins.*
- 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.*

66. Des pays réfléchissent aux façons de garantir une IA digne de confiance et d'atténuer les risques associés au développement et au déploiement des systèmes d'IA. Outre un examen de l'application de la législation actuelle à l'IA et de la nécessité de l'adapter à cette fin, les nouvelles actions réglementaires pour une IA digne de confiance comprennent : i) l'établissement de cadres et de principes éthiques, ii) l'étude d'approches basées sur des normes contraignantes, iii) l'appui aux efforts en matière de normalisation internationale et de droit international.

- **Lignes directrices pour une IA digne de confiance** : de nombreux Adhérents ont élaboré des lignes directrices afin de tendre vers une IA digne de confiance qui sont dans le droit fil des Principes de l'OCDE sur l'IA et fournissent des normes pour l'utilisation éthique de l'IA et sa gouvernance. Selon le cas, elles s'adressent aux décideurs, aux entreprises, aux instituts de recherche et à d'autres acteurs de l'IA. Parmi les exemples, on peut citer le cadre AI Ethics Framework en Australie, l'outil d'auto-évaluation en ligne destiné à promouvoir une IA digne de confiance et spécifiquement adapté au secteur public en Belgique, le cadre d'éthique pour l'IA en Colombie, les lignes directrices pour le développement d'une IA digne de confiance en Corée, les lignes directrices éthiques en matière d'IA en Hongrie, les lignes directrices relatives à la R-D dans le domaine de l'IA et les Principes pour une IA centrée sur l'homme au Japon, les lignes directrices pour la Confédération « Intelligence artificielle » en Suisse, les principes d'éthique pour le développement de l'IA en Argentine et la charte pour une IA responsable en Égypte.
- **Environnements contrôlés pour l'expérimentation réglementaire** : un nombre croissant d'Adhérents utilisent des bacs à sable réglementaires pour l'IA, c'est-à-dire des espaces dans lesquels les pouvoirs publics permettent aux entreprises de tester des produits et services innovants qui remettent en cause les cadres juridiques existants (OCDE, 2023<sub>[10]</sub>). Parmi les exemples, on peut citer la stratégie de bac à sable réglementaire de l'Allemagne (2019) (Gouvernement fédéral de l'Allemagne, 2022<sub>[28]</sub>), et le bac à sable réglementaire de l'autorité norvégienne de protection des données (*Datatilsynet*) (2020). La promotion de l'expérimentation réglementaire est également l'un des outils essentiels prévus par la stratégie nationale israélienne en matière d'IA pour garantir un déploiement sûr et innovant de l'IA. L'Espagne, en collaboration avec la Commission européenne, a créé un bac à sable réglementaire pour l'IA en 2022. Il s'agissait du premier programme pilote permettant de tester la future proposition de législation de l'UE sur l'IA avec des applications d'IA réelles, afin d'évaluer la façon dont la réglementation et les applications réagissent, et de suggérer des modifications ou des lignes directrices explicatives (OCDE, 2023<sub>[10]</sub>).
- **Nouvelles réglementations propres à l'IA** : des dispositions s'appliquent déjà aux systèmes d'IA dans différents domaines de la législation. Toutefois, ces dernières

années, des pays ont commencé à transposer les Principes de l'OCDE sur l'IA en cadres législatifs et réglementaires contraignants propres à l'IA qui traitent des incidences ou des systèmes à haut risque de l'IA, même si la façon de procéder varie considérablement d'un pays à l'autre et d'un territoire à l'autre.

- Certains Adhérents et organisations internationales abordent la réglementation de l'IA de manière intersectorielle en élaborant un cadre réglementaire propre à l'IA applicable à tous les secteurs. Le Canada, le Brésil, l'Union européenne et le Conseil de l'Europe ont proposé de réglementer les systèmes d'IA dans tous les domaines et toutes les applications, en créant un cadre réglementaire propre à l'IA applicable à tous les secteurs.
  - Le Canada a proposé un cadre réglementaire complet au niveau fédéral, la loi sur la mise en œuvre de la Charte numérique, qui comprend la loi sur l'intelligence artificielle et les données (Parlement du Canada, 2022<sup>[29]</sup>). Cette dernière viserait à garantir la conception, le développement et le déploiement sûrs et responsables de systèmes d'IA qui respectent les valeurs des Canadiens. Elle établit une approche fondée sur les impacts qui cherche à atténuer les risques de préjudice et de biais associés aux systèmes d'IA à « incidence élevée » (Fasken, 2022<sup>[30]</sup>). Le document complémentaire à la proposition de loi sur l'intelligence artificielle et les données (Gouvernement du Canada, 2023<sup>[31]</sup>) indique que cette dernière est alignée sur les Principes de l'OCDE sur l'IA, la proposition de législation de l'UE sur l'IA et le cadre de gestion des risques du National Institute of Standards and Technology (NIST) des États-Unis.
  - Le Brésil a proposé le projet de loi n° 2338/2023 (la « loi brésilienne sur l'IA »), dont les thèmes clés sont notamment : une approche axée sur les droits de l'homme ; une approche fondée sur les risques et une classification des risques des systèmes d'IA ; la mise en place d'une autorité de surveillance ; des règles en matière de responsabilité civile ; l'encouragement de l'innovation par la promotion de bacs à sable réglementaires (OECD.AI, 2023<sup>[32]</sup>).
  - L'Union européenne en est aux dernières étapes du processus législatif en vue de l'adoption de sa législation sur l'IA. Cette dernière suit une approche fondée sur les risques et présente un cadre juridique uniforme et horizontal pour l'IA. Elle s'appuie sur la définition d'un système d'IA de l'OCDE, actualisée par le Conseil de l'OCDE en novembre 2023. Elle introduit une classification des systèmes d'IA en fonction des niveaux de risque qu'ils représentent pour la santé, la sécurité et les droits fondamentaux, notamment la démocratie, l'État de droit et la protection de l'environnement. Les systèmes d'IA « à haut risque » doivent faire l'objet d'une évaluation de la conformité avant leur mise sur le marché, et doivent respecter des dispositions de surveillance par la suite. Les systèmes d'IA présentant des risques « inacceptables » sont interdits.
  - Le Comité sur l'intelligence artificielle du Conseil de l'Europe a finalisé une Convention sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit (Conseil de l'Europe, 2024<sup>[33]</sup>) à la mi-mars 2024. La Convention sera envoyée au Comité des ministres pour adoption et ouverte à la signature à un stade ultérieur. Une fois adoptée, elle devrait devenir un instrument juridique international contraignant pour ses parties. Elle reprend la définition révisée de l'OCDE d'un système d'IA, actualisée par le Conseil de l'OCDE en novembre 2023, ainsi que la définition révisée proposée du cycle de vie d'un système d'IA.

- D'autres Adhérents envisagent des approches plus sectorielles et élaborent des réglementations par secteur ou par domaine.
  - S'inspirant des Principes de l'OCDE sur l'IA, Israël a adopté dans son livre blanc intitulé « Draft Policy White Paper for Regulation and Ethics in the Field of AI » des principes éthiques non contraignants à prendre en compte lors du développement, de l'utilisation et de la réglementation de l'IA, et préconise des efforts réglementaires sectoriels fondés sur des approches de l'évaluation et de la gestion des risques, plutôt qu'une réglementation sectorielle globale (Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal counsel and legislative affairs, 2022<sup>[34]</sup>).
  - Dans son document intitulé « AI regulation: A pro-innovation approach » (Gouvernement du Royaume-Uni, 2023<sup>[35]</sup>), le Royaume-Uni établit des principes intersectoriels non contraignants – basés sur les Principes de l'OCDE sur l'IA –, laissant aux régulateurs la tâche de les mettre en œuvre, de les réglementer et de les faire respecter dans leurs secteurs et domaines respectifs. Israël procède de façon similaire.
  - Aux États-Unis, le document intitulé « Blueprint for an AI Bill of Rights » (Gouvernement des États-Unis, 2022<sup>[36]</sup>) cherche à soutenir l'élaboration de politiques et de pratiques destinées à protéger les droits civils et à promouvoir les valeurs démocratiques dans le développement, le déploiement et la gouvernance des systèmes d'IA (Gouvernement des États-Unis, 2022<sup>[36]</sup>). Il établit cinq principes non contraignants visant à atténuer les risques pour les droits civils et les valeurs démocratiques que pose l'utilisation de systèmes automatisés dans les différents secteurs. En 2023, le Président a signé un décret pour le développement et l'utilisation d'une intelligence artificielle sûre, sécurisée et digne de confiance, qui prévoit l'établissement de normes de sûreté et de sécurité de l'IA, avec diverses exigences applicables aux équipes de développement d'IA, aux instances de l'administration publique, et à d'autres acteurs aux fins de la protection de la vie privée, des droits civils, des consommateurs et des travailleurs, notamment (The White House, 2023<sup>[37]</sup> ; The White House, 2023<sup>[38]</sup>). Le cadre de gestion des risques liés à l'IA, publié par le National Institute of Standards and Technology (NIST) (National Institute of Standards and Technology - US Department of Commerce, 2023<sup>[8]</sup>), préconise d'adopter une approche basée sur la préservation des droits en rapport avec les lignes directrices techniques et les normes relatives aux caractéristiques de fiabilité et à une pratique et une utilisation responsables, et s'appuie sur les Principes de l'OCDE sur l'IA et le Cadre de l'OCDE pour la classification des systèmes d'IA (OCDE, 2022<sup>[6]</sup>) pour proposer un cadre permettant de cartographier, de mesurer et de gérer les risques de l'IA.

67. La Recommandation souligne spécifiquement que « certains cadres juridiques, réglementaires et politiques nationaux 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 ». Cette nécessité de veiller au maintien de la pertinence des approches réglementaires et de gouvernance et de les adapter le cas échéant est également un principe clé de la Recommandation en faveur d'une gouvernance réglementaire agile permettant de mettre l'innovation à profit (OCDE, 2021<sup>[39]</sup>) et des Principes de bonne pratique pour l'examen de la réglementation existante (OCDE, 2021<sup>[40]</sup>).

68. L'IA, en particulier générative, soulève des questions concernant les droits de propriété intellectuelle : contenus non soumis à licence dans les données d'entraînement ; risques de voir les créations d'une IA porter atteinte à un droit d'auteur, une marque ou un brevet ; et interrogations concernant la propriété des travaux générés par IA. Les Adhérents procèdent à des évaluations et élaborent des approches concernant le droit d'auteur et l'IA. En août 2023, par exemple, le United States Copyright Office a publié un avis d'enquête dans le registre fédéral sur le droit d'auteur et l'IA (US Copyright Office, 2023<sup>[41]</sup>), demandant des informations et des points de vue sur les questions de droit d'auteur soulevées par les progrès de l'IA générative. Le décret américain (The White House, 2023<sup>[38]</sup>) reconnaît les questions importantes que pose l'IA en matière de protection de la propriété intellectuelle. Il charge le Trademark and Patent Office (USPTO) de publier des orientations à l'intention des examinateurs et des demandeurs de brevets concernant l'intersection entre IA et la propriété intellectuelle, et adresse au président des recommandations sur le droit d'auteur et l'IA. La législation de l'UE sur l'IA impose aux fournisseurs de modèles d'IA à usage général d'établir et de mettre à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour la formation de leur modèle.

69. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 2.3 visant à façonner un cadre d'action favorable à l'IA. Ce Principe reste pertinent. Notamment, de nouvelles réglementations propres à l'IA sont alignées sur les Principes de l'OCDE sur l'IA, ou sur les travaux de l'OCDE basés sur les Principes (le Cadre de l'OCDE pour la classification des systèmes d'IA, (OCDE, 2022<sup>[6]</sup>)) ou y font explicitement référence. Le Principe pourrait éventuellement bénéficier d'une clarification de son intitulé, car la formulation « un cadre d'action favorable » est très large et pourrait en conséquence se référer à de nombreux domaines d'action non liés à l'IA. En outre, l'intitulé pourrait explicitement mentionner la gouvernance de l'IA, ainsi que l'importance de veiller à ce que les différents cadres de gouvernance soient « interopérables », c'est-à-dire compatibles et capables de répondre à des défis communs. Cette interopérabilité des cadres de gouvernance diffère de l'interopérabilité technologique et se trouve au cœur du mandat de l'OCDE. En tant que telle, elle peut apporter des éclaircissements utiles. En outre, si l'expérimentation réglementaire est l'une des options permettant de mettre en œuvre une réglementation agile promouvant l'innovation, d'autres solutions ont également été identifiées dans les travaux de l'OCDE (OCDE, 2021<sup>[39]</sup>), et pourraient être mentionnées dans le texte.**

*Renforcer les capacités humaines et préparer la transformation du marché du travail (Principe 2.4)*

*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 et d'interagir efficacement avec les systèmes d'IA 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 d'IA, notamment par le biais de programmes de formation tout au long de la vie active, du soutien aux personnes affectées par les suppressions de postes 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, favoriser l'entrepreneuriat et la productivité, et veiller à ce que les avantages de l'IA soient partagés largement et équitablement.*

70. L'IA modifie déjà la nature de nombreux aspects de notre vie à mesure qu'elle se diffuse dans les différents secteurs, en particulier sur le plan du travail, de l'emploi et de la sphère professionnelle. Si, à ce jour, elle a surtout affecté la qualité des emplois plutôt que leur quantité (OCDE, 2023<sup>[12]</sup>), certains signes semblent cependant indiquer que les marchés du travail pourraient bientôt connaître des bouleversements importants dont les effets seraient tant positifs que négatifs. Elle peut être bénéfique pour l'emploi en engendrant une demande pour de nouvelles tâches et des compétences complémentaires, entraînant, de fait, la création de nouveaux emplois pour lesquels la main-d'œuvre humaine dispose d'un avantage comparatif. Des travaux de recherche récents montrent que l'IA générative peut améliorer les performances des travailleurs moins qualifiés (OCDE, 2023<sup>[12]</sup>). Dans le même temps, ses progrès ont mis en lumière les possibles conséquences de l'IA sur les marchés du travail.

71. Les développements récents survenus dans le domaine de l'IA générative, notamment GPT-4, ont élargi le champ des tâches que les systèmes d'IA peuvent accomplir, augmentant ainsi leur influence potentielle sur le marché de l'emploi. Si les professions peu qualifiées ont jusqu'à présent été les plus exposées à l'automatisation (Lassébie et Quintini, 2022<sup>[42]</sup>), les progrès récents de l'IA générative ont conduit à remettre en question les hypothèses traditionnelles concernant les types de professions susceptibles d'être touchées par l'automatisation. Les travaux de recherche sur l'IA générative basée sur le langage révèlent que 32,8 % des emplois de la Classification internationale type des professions (CITP) pourraient être affectés à grande échelle par les modèles d'IA générative, 36,5 % pourraient l'être partiellement et seulement 30,7 % ne le seraient pas (Zarifhonarvar, 2023<sup>[43]</sup>). Cela oblige les organisations à s'adapter à l'IA générative et à soutenir leur main-d'œuvre, et les décideurs à orienter les évolutions et les transitions du marché du travail. Toutefois, la recherche sur les effets de l'IA générative sur le marché du travail est relativement récente et d'autres travaux soumis à un examen par les pairs sont nécessaires pour tirer des conclusions plus définitives.

72. Les pays reconnaissent que pour à la fois gérer une transition équitable du marché du travail et montrer la voie en matière de recherche, de développement et d'adoption de l'IA, il est nécessaire de mettre en œuvre des politiques de développement des compétences en matière d'IA et d'attirer des talents. La plupart des stratégies nationales en matière d'IA comprennent un pilier sur l'éducation à l'IA et l'amélioration des compétences dans le domaine, et à ce jour, les Adhérents ont fait état de plus de 200 programmes d'action en rapport avec le Principe 2.4. Les Adhérents ont principalement mis en place des initiatives visant à permettre à la main-d'œuvre d'acquérir les compétences requises pour l'IA par le biais de programmes d'enseignement formel et d'initiatives de formation tout au long de la vie. Nombre de ces programmes sont axés sur le développement des talents parmi les personnes qui mettront au point des systèmes d'IA, qu'ils soient fondamentaux ou destinés à être utilisés dans des secteurs/domaines particuliers. Des Adhérents ont également lancé des initiatives visant à attirer et à retenir les talents spécialisés dans l'IA.

73. Si les pays de l'OCDE ont pour la plupart pris des initiatives en faveur du développement des compétences numériques, ils devront également mettre en place des programmes de formation qui sont destinés spécifiquement à préparer les salariés à travailler avec l'IA et ciblent de façon appropriée les personnes peu qualifiées. Il est essentiel de veiller à ce que les travailleurs de nombreux secteurs et industries possèdent les compétences nécessaires pour appliquer l'IA dans leur propre secteur, tout comme il

est essentiel de développer des talents parmi les personnes qui mettront au point les systèmes d'IA.

74. Les initiatives visant à surveiller l'incidence de l'IA sur le marché du travail et à accompagner les transitions sur le marché du travail semblent limitées à ce jour. Comme examiné dans le programme *IA dans le travail, l'innovation, la productivité et les compétences* (AI-WIPS) – un programme de l'OCDE soutenu par l'Allemagne qui analyse les répercussions de l'IA sur le marché du travail, les compétences et la politique sociale (OECD.AI, 2024<sup>[44]</sup>) –, il est important de continuer à développer des initiatives destinées à surveiller l'incidence de l'IA sur les marchés du travail et à identifier les mesures fondées sur des données empiriques que les pouvoirs publics ont prises.

75. Dans de nombreux pays de l'OCDE, les systèmes d'IA utilisés dans la sphère professionnelle sont tenus de respecter des textes législatifs ou réglementaires, notamment une législation anti-discrimination, une réglementation relative à la santé et à la sécurité au travail, une réglementation relative au respect de la vie privée des employés et la liberté d'association (OCDE, 2023<sup>[45]</sup>). Les nouvelles législations propres à l'IA, notamment la législation de l'UE sur l'IA, ont d'importantes implications pour l'IA au travail. La législation de l'UE sur l'IA classe certains systèmes d'IA utilisés à des fins de recrutement et de prise de décision dans des relations de travail contractuelles comme étant à « haut risque », les soumettant ainsi à des exigences légales en matière de gestion des risques, de qualité et de gouvernance des données, de conservation de la documentation et des enregistrements, de transparence et d'information des utilisateurs, de surveillance humaine, de robustesse, de précision et de sécurité. En février 2024, l'Espagne et les États-Unis ont signé une déclaration commune sur les « Algorithmic Bias in the World of Work » appelant la communauté internationale à gérer de manière appropriée les risques néfastes pour les droits sociaux et des travailleurs posés par certains systèmes automatisés (Department of Labour, 2024<sup>[46]</sup>).

- **Programmes d'enseignement formel pour les STIM, l'IA et les domaines liés à l'IA :** plusieurs Adhérents et l'UE soutiennent le développement de talents spécialisés dans l'IA par le biais d'un financement ciblé des étudiants dans des programmes d'enseignement supérieur. Parmi les exemples, on peut citer le programme australien Next Generation AI Graduates, un programme de bourses de doctorat cofinancé par l'industrie, les bourses financées par le gouvernement israélien pour soutenir les étudiants en IA en master, en doctorat et en post-doc, le programme national de doctorat en IA de l'Italie et les 2 500 cours de conversion de niveau master du Royaume-Uni pour les candidats issus de milieux proches ou non des STIM, qui encouragent également une plus grande diversité dans les carrières liées à l'IA. Le programme pour une Europe numérique de l'Union européenne finance également des actions visant à renforcer les compétences numériques avancées en Europe, y compris dans le domaine de l'IA. La Corée a élaboré trois types de manuels d'éthique de l'IA à l'intention des élèves des écoles primaires et secondaires, ainsi que trois types de manuels d'enseignement de l'éthique de l'IA à l'intention des enseignants.
- **Formation et apprentissage tout au long de la vie dans le domaine de l'IA et programmes connexes :** parmi les exemples d'initiatives visant à dispenser aux professionnels une formation alignée sur les besoins du marché du travail, on peut citer le programme Digital Talent au Chili, la stratégie globale de développement de la main-d'œuvre numérique en Corée, le guide pratique sur la fourniture de données pour favoriser le développement de ressources humaines expertes en IA et en science des données au Japon, le programme AI Competence for Sweden, la collaboration entre dix universités suédoises pour développer des cours destinés aux professionnels qui peuvent contribuer au développement de la Suède dans le domaine de l'IA, et la désignation

Chartered AI Engineer de Singapour, un programme de qualification professionnelle par l'AI Professionals Association pour reconnaître et attribuer des titres aux professionnels qui travaillent dans des rôles d'ingénierie liés à l'IA.

- **Initiatives pour retenir et attirer les talents spécialisés dans l'IA :** le Canada a été l'un des premiers pays à adopter une stratégie axée sur les compétences : sa Stratégie pancanadienne en matière d'intelligence artificielle vise à attirer et à retenir les talents universitaires de premier plan. De même, attirer et former de nouveaux professeurs dans le domaine de l'IA est une initiative clé de la stratégie nationale allemande en matière d'IA.
- **Initiatives visant à favoriser la collaboration avec les parties prenantes afin de garantir une transition équitable pour les travailleurs.** Les pays élaborent des mesures visant à favoriser ou à renforcer la collaboration avec les parties prenantes, y compris les partenaires sociaux, par le biais du dialogue social. En Espagne, la Charte des droits numériques précise que « les représentants des travailleurs doivent être informés des changements technologiques qui ont lieu dans l'entreprise et participer à la prise de décision concernant la transformation numérique et les conséquences qu'elle peut avoir sur le travail ». Aux États-Unis, le décret du Président (The White House, 2023<sup>[38]</sup>) précise que « l'IA créant de nouveaux emplois et de nouvelles industries, tous les travailleurs doivent participer aux discussions, notamment par le biais de la négociation collective, pour s'assurer qu'ils bénéficient de ces opportunités » et que « les prochaines étapes du développement de l'IA devraient s'appuyer sur les points de vue des travailleurs, des syndicats, des éducateurs et des employeurs pour soutenir des utilisations responsables de l'IA qui améliorent la vie des travailleurs, augmentent positivement le travail humain et aident tout le monde à profiter en toute sécurité des gains et des opportunités qu'apporte l'innovation technologique ».
- **Surveiller les répercussions de l'IA sur le marché du travail :** aux États-Unis, l'Office of the White House a créé l'American Workforce Policy Advisory Board dans le cadre d'une initiative nationale visant à combler le déficit de compétences qui se creuse en partie en raison de l'essor de l'automatisation et du besoin croissant de compétences en technologie de pointe. Le décret du Président (The White House, 2023<sup>[38]</sup>) appelle à prendre des mesures afin de protéger les droits et la sécurité des travailleurs, soulignant la nécessité d'adapter l'éducation et la formation professionnelle afin de soutenir une main-d'œuvre diverse et d'aider à accéder aux possibilités offertes par l'IA. Le « Guide to Job Redesign in the Age of AI de Singapour » est un document qui aide les organisations et les employés à comprendre comment les emplois existants peuvent être redéfinis pour exploiter le potentiel de l'IA et augmenter la valeur de leur travail.

76. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 2.4 sur le renforcement des capacités humaines et la préparation de la transformation du marché du travail. Toutefois, il reste encore beaucoup à faire pour contrôler l'impact de l'IA sur le marché du travail, en examinant la quantité, la qualité et l'inclusivité des emplois, et en mettant en œuvre une réglementation spécifique au marché du travail afin de promouvoir une utilisation digne de confiance de l'IA dans la sphère professionnelle. Le Principe reste pertinent et pourrait bénéficier de deux ajouts. Il conviendrait de mentionner la protection sociale comme l'un des moyens de soutenir les personnes affectées par les changements, et le potentiel de l'IA pour améliorer la qualité des services publics.**

*Favoriser la coopération internationale au service d'une IA digne de confiance (Principe 2.5)*

*a) Les pouvoirs publics, y compris ceux des pays en voie de développement, en association avec les parties prenantes, devraient coopérer activement afin de faire progresser 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.*

77. Les Adhérents coopèrent de plus en plus au niveau international afin de promouvoir l'utilisation bénéfique de l'IA et de relever les défis qu'elle pose. Pour ce faire, ils ont recours à différents types d'initiatives, notamment :

- **Collaboration internationale pour la recherche en IA :** le Partenariat mondial sur l'intelligence artificielle (PMIA) est une initiative internationale et multipartite fondée conjointement par le Canada et la France. Lancé en juin 2020, il mène à bien des travaux de recherche de pointe et des projets pilotes sur les priorités de l'IA afin de promouvoir son développement et son utilisation responsables. Parmi les exemples de collaboration transfrontière en matière de recherche au niveau régional, citons la déclaration sur l'IA dans la région nordique et balte, le cadre de l'accord de partenariat économique entre le Japon et Singapour (JSEPA) et le dialogue quadrilatéral sur la sécurité (QUAD), un forum stratégique informel réunissant les États-Unis, l'Inde, l'Australie et le Japon.
- **Coopération internationale et multipartite sur l'IA :** en mai 2023, sous la présidence japonaise du G7, les dirigeants du G7 ont établi le Processus du G7, dit d'Hiroshima, sur l'IA générative afin d'examiner les possibilités et enjeux associés à l'IA générative (OCDE, 2023<sup>[47]</sup>). En décembre 2023, les dirigeants du G7 ont approuvé le Cadre d'action exhaustif relatif au processus d'Hiroshima sur l'IA, qui comprend des principes directeurs (MIC, 2023<sup>[48]</sup>) et un code de conduite (MIC, 2023<sup>[49]</sup>) visant à promouvoir le développement sûr, sécurisé et digne de confiance des systèmes d'IA avancés. En novembre 2023, le Royaume-Uni a tenu le Sommet sur la sécurité de l'intelligence artificielle, dont deux des résultats sont la création d'un Institut sur la sécurité de l'IA (AI Safety Institute), pour conduire des recherches avancées en sécurité de l'IA, possiblement en collaboration avec les établissements analogues actuellement créés dans d'autres pays, et la commande d'un rapport sur l'état de l'art en IA (AI Safety Summit, 2023<sup>[16]</sup>). L'Envoyé pour les technologies des Nations unies (ONU) a établi un « organe consultatif sur l'IA » qui a remis son rapport intermédiaire en décembre 2023 (UN Advisory Board on AI, 2023<sup>[50]</sup>) et qui est chargé de remettre un rapport final présentant ses recommandations lors du « Sommet de l'avenir », un événement majeur prévu en septembre 2024. En Corée, le forum sur l'éthique de l'IA s'appuie sur une

approche multipartite pour diffuser le système d'éthique de l'IA. Les États-Unis et la Commission européenne ont créé ensemble un Conseil du commerce et des technologies pour soutenir le renforcement des relations transatlantiques, y compris dans le domaine de l'IA. Le secteur privé s'associe également à d'autres parties prenantes. Le Partenariat sur l'intelligence artificielle est une collaboration entre de grandes entreprises technologiques telles qu'Amazon, Google et Microsoft, des organisations de la société civile et des organisations à but non lucratif telles que l'American Civil Liberties Union (ACLU) et l'Electronic Frontier Foundation (EFF), des institutions universitaires et des médias.

- **Accords commerciaux incluant l'IA :** en 2020, le Chili, la Nouvelle-Zélande et Singapour ont signé l'Accord de partenariat sur l'économie numérique (DEPA), qui vise notamment à promouvoir l'utilisation sûre et responsable des technologies d'IA. En 2023, la Corée a accepté de rejoindre le DEPA à partir de 2030. Également en 2023, l'Australie et Singapour, s'appuyant sur leur accord commercial préexistant, ont signé l'Accord sur l'économie numérique Singapour-Australie (SADEA) par lequel les parties ont convenu de faire progresser leur coopération en matière d'IA. En outre, l'Accord de partenariat numérique Corée-Singapour (KSDPA), qui comprend une référence à l'IA et un protocole d'accord visant à renforcer la collaboration pratique et à promouvoir le développement et l'utilisation responsables de l'IA, est entré en vigueur en janvier 2023.
- **Coopération pour le renforcement des capacités en IA dans les pays en développement :** la coopération internationale est nécessaire pour éviter l'émergence d'une « fracture de l'IA » et garantir que ses avantages soient répartis plus équitablement entre les différentes régions et collectivités. « AI for All - FAIR Forward » (2019-2023) est une initiative stratégique lancée par le ministère fédéral allemand de la Coopération économique et du Développement et axée sur l'application et le développement ouverts et durables de l'IA. Le Centre Asie-Pacifique de formation aux technologies de l'information et de la communication pour le développement (CAPFTIC) propose aux pays en développement de la région Asie-Pacifique une formation aux TIC, un partage des connaissances et des consultations stratégiques. Le Programme des Nations unies pour le développement (PNUD) a mis au point iVerify, un outil de vérification automatisée des faits à code source ouvert qui est utilisé pour repérer les fausses informations et prévenir et atténuer leur diffusion. Cet outil est actuellement utilisé en Zambie, au Kenya et au Honduras et son déploiement est prévu au Libéria. L'initiative « Harnessing AI for Development », préparée avec le soutien du Partenariat pour le développement numérique, est un travail en cours au sein de la Digital Development Global Practice de la Banque mondiale.

78. **Les données indiquent que les Adhérents ont progressé dans la mise en œuvre du Principe 2.5 visant à favoriser la coopération internationale au service d'une IA digne de confiance. Ce Principe gagne en importance à mesure que les programmes d'action en faveur de l'IA se multiplient dans le monde entier. Les difficultés et les possibilités que représente l'IA ont souvent une dimension mondiale, comme le démontrent les nombreuses initiatives lancées en 2023 pour favoriser le dialogue sur l'IA au niveau international. Les considérations géopolitiques jouent un rôle de plus en plus important, parallèlement au rôle sociétal croissant de l'IA, et soulignent encore davantage l'importance de ce Principe. Le Principe pourrait bénéficier d'une révision mineure du texte en anglais afin de remplacer le terme « metrics » par « indicators » pour plus de clarté (la traduction en français, « indicateurs », ne serait pas modifiée).**

#### 4. Réponses des Adhérents au questionnaire

79. La plupart des Adhérents ont indiqué retirer des effets bénéfiques de leur adhésion à la Recommandation. Quelques réponses neutres ont également été enregistrées. Les Adhérents ont indiqué que les Principes de l'IA constituent une base solide pour l'élaboration de politiques nationales en matière d'IA, qu'ils fournissent des orientations pour l'établissement du cadre d'action, qu'ils favorisent l'alignement international en matière de gouvernance de l'IA et qu'ils facilitent la collaboration entre les parties prenantes.

- **Alignement international** : plusieurs Adhérents ont souligné l'avantage de s'appuyer sur les Principes pour travailler à l'alignement international de la gouvernance de l'IA. Ils estiment qu'ils constituent une base universellement reconnue pour orienter les politiques en matière d'IA, favoriser l'interopérabilité et promouvoir la coopération mondiale dans le domaine.
- **Orientations pratiques** : les Adhérents ont noté que les Principes fournissent des orientations essentielles pour l'élaboration de politiques. Ils aident à recenser les considérations relatives à l'action publique de haut niveau, à allouer les ressources de manière efficace et à orienter le lancement de nouvelles initiatives. Plusieurs Adhérents ont indiqué qu'ils ont servi de base solide à l'élaboration de leur stratégie nationale en matière d'IA.
- **Collaboration intersectorielle** : les Adhérents ont noté que les Principes facilitent la collaboration intersectorielle au niveau national, car ils soulignent la nécessité d'impliquer les milieux universitaires, l'industrie et la société civile. Ils ont noté que cette démarche multipartite enrichit les stratégies d'IA, car elle tient compte de divers points de vue.

80. Toutefois, certains Adhérents ont déclaré avoir rencontré des problèmes lors de la mise en œuvre des Principes de l'OCDE sur l'IA, le principal étant la difficulté de transposer les Principes fondés sur des valeurs en mesures pratiques et réalisables, car ils sont de haut niveau et de vaste portée. Le manque de connaissances et de capacités techniques nécessaires à la compréhension de l'IA, ainsi que l'insuffisance de ressources consacrées aux investissements et à l'élaboration de politiques ont également été perçus comme des obstacles à la mise en œuvre des Principes au niveau national.

81. Les Adhérents ont suggéré que l'OCDE pourrait entreprendre les activités supplémentaires suivantes afin de les aider à mettre en œuvre les Principes :

- **Lignes directrices et orientations pratiques** : certains Adhérents ont suggéré de fournir des lignes directrices détaillées et des explications pratiques pour faciliter la mise en œuvre des Principes de haut niveau, comme les travaux de l'OCDE sur la classification des systèmes d'IA ou l'élaboration d'un catalogue d'outils et d'indicateurs pour une IA digne de confiance. En outre, plusieurs Adhérents ont souligné l'importance de transposer les Principes en normes techniques et en processus de certification.
- **Partage des connaissances et renforcement des capacités** : plusieurs Adhérents ont souligné l'importance du partage des connaissances et du renforcement des capacités afin de soutenir les décideurs et les parties prenantes dans les pays adhérents. Il a notamment été suggéré de créer des programmes de formation et des ateliers pratiques à l'intention des décideurs. Les Adhérents ont reconnu l'importance de l'Observatoire OCDE des politiques relatives à l'IA en tant que plateforme de partage de l'information et des connaissances, et ont encouragé la mise en valeur des pratiques exemplaires par

le biais d'études de cas, d'analyses comparatives et d'exemples de mise en œuvre réussie des Principes de l'OCDE sur l'IA.

- **Mécanismes de suivi et de rapport** : certains Adhérents ont recommandé d'établir des mécanismes de suivi et de rapport sur les progrès de la mise en œuvre des Principes de l'OCDE sur l'IA, y compris des analyses comparatives et des études de cas.
- **Assistance technique et services de conseil** : il a également été suggéré de fournir une assistance technique et des services de conseil aux pays, comme une aide à l'élaboration de politiques en matière d'IA et à l'examen des politiques en vigueur et l'offre de conseils d'experts.
- **Collaboration et partenariats** : les Adhérents ont souligné l'importance de favoriser la collaboration internationale, les partenariats public-privé et les collaborations entre les administrations publiques, les milieux universitaires et la société civile afin de mettre en commun les ressources et l'expertise de tous ces acteurs. Ils ont également souligné la nécessité de disposer de mécanismes pour reconnaître les efforts d'engagement et la conformité ainsi que pour promouvoir l'implication du secteur.

82. Les Adhérents ont convenu que les Principes de l'OCDE sur l'IA restent pertinents pour guider le développement de l'IA, encourager l'innovation et renforcer la confiance du public. Ils considèrent que les Principes de l'OCDE sur l'IA sont fondamentaux pour promouvoir les valeurs centrées sur l'humain, la transparence, la responsabilité et l'équité, prévenir les biais, promouvoir l'inclusion et veiller à ce que l'IA respecte les droits de l'homme et les valeurs démocratiques. Il a également été indiqué que les Principes fournissent un cadre d'éthique qui soutient la croissance inclusive, la collaboration internationale et l'établissement de normes communes dans le paysage mondial de l'IA.

83. Certains Adhérents ont souligné la nécessité d'une législation nationale ou régionale complémentaire et d'une collaboration entre les initiatives internationales pour traiter efficacement les risques liés à l'IA. Les Adhérents ont également suggéré d'adapter les Principes de l'OCDE sur l'IA aux différents types d'IA et d'élaborer des lignes directrices spécifiques pour les différentes parties prenantes, telles que les utilisateurs, les développeurs et les secteurs. D'autres Adhérents ont formulé des recommandations visant à renforcer la pertinence des Principes de l'OCDE sur l'IA, telles que l'évaluation de l'applicabilité des principes aux modèles d'avant-garde et aux modèles de fondation, et l'importance de procéder à des examens réguliers afin de maintenir l'adaptabilité des Principes dans un paysage de l'IA en rapide mutation.

84. Les Adhérents ont également été interrogés sur les possibilités et les risques clés associés aux systèmes d'IA avancés, tels que les modèles de fondation et l'IA générative, ainsi que sur le rôle que peuvent jouer les Principes afin de tirer profit des avantages et d'atténuer les risques.

85. Les Adhérents considèrent les gains de productivité comme la plus grande opportunité apportée par les modèles de fondation et l'IA générative, suivie de près par l'amélioration des soins de santé et la promotion de l'innovation et de l'entrepreneuriat (Graphique 4.1). Ils ont également souligné la possibilité pour ces modèles de contribuer à la lutte contre la criminalité, notamment la cybercriminalité. La plupart des Adhérents ont estimé que la Recommandation sur l'IA pouvait contribuer à tirer parti des possibilités offertes par les systèmes d'IA avancés, certaines délégations adoptant une position neutre à cet égard. Toutefois, ils ont tous convenu que compléter les Principes de l'OCDE sur l'IA par des recommandations plus spécifiques et plus concrètes, en

particulier dans des domaines émergents tels que l'IA générative, leur permettrait de mieux tirer parti des possibilités offertes par l'IA.

86. Les menaces pour la cybersécurité et la désinformation sont considérées par la plupart des Adhérents comme les principaux risques posés par les systèmes d'IA avancés (Graphique 4.2). La plupart des Adhérents considèrent également que les risques majeurs concernent la sécurité ainsi que l'exacerbation des biais et de la discrimination. Trois Adhérents ont mentionné d'autres risques dans leur réponse : menaces pour la sécurité, perturbations de la cohésion sociale et polarisation, et menaces pour les droits de l'homme et les processus démocratiques.

87. Si la capacité de la Recommandation à traiter les risques liés à l'IA est généralement bien perçue, les réponses des Adhérents varient toutefois à cet égard. Plusieurs Adhérents ont souligné que si les Principes de l'OCDE sur l'IA sont utiles pour traiter les risques associés à l'IA, ils peuvent ne pas être suffisants en soi, des mesures supplémentaires, telles que la réglementation, étant nécessaires pour atténuer efficacement ces risques. Un Adhérent a souligné que si les Principes de l'OCDE sur l'IA sont utiles pour définir des objectifs, des travaux supplémentaires pourraient être nécessaires pour évaluer leur efficacité, notamment en ce qui concerne les risques posés par les systèmes d'IA d'avant-garde, et il a insisté sur la nécessité de disposer de techniques et de normes d'assurance supplémentaires pour garantir une mise en œuvre réussie de ces Principes. Un autre Adhérent a également souligné l'importance de disposer de lignes directrices détaillées pour traiter les risques liés à la vie privée, à la sécurité, à l'équité, à la diversité et à la non-discrimination. Enfin, l'importance de la prise en compte des risques tout au long du cycle de vie des systèmes d'IA et la nécessité d'une approche globale pour garantir une gestion adéquate des risques et des répercussions ont été mentionnées.

88. La « gouvernance » des systèmes d'IA avancés a été considérée par beaucoup comme la priorité la plus « importante » et la plus « urgente » pour les pouvoirs publics. Outre la « gouvernance », l'« utilisation responsable » des technologies d'IA générative a également été considérée comme la priorité la plus « importante » pour les pouvoirs publics, suivie par la « sécurité des personnes », l'« alignement des objectifs des systèmes d'IA sur les valeurs humaines » et la « gouvernance des données ». La promotion de la responsabilité et la protection de la vie privée ont également été considérées comme des questions urgentes à traiter.

89. Dans l'ensemble, les réponses des Adhérents suggèrent que la Recommandation est considérée comme un outil précieux pour traiter les priorités importantes liées aux systèmes d'IA avancés, mais la façon dont les différents Adhérents ont perçu son efficacité à cet égard varie. Tout en soulignant l'importance de la Recommandation de l'OCDE en tant qu'orientation pour relever les défis cruciaux liés à l'IA, les Adhérents ont également mis l'accent sur la nécessité de disposer de cadres de gouvernance pratiques et de Principes adaptés aux questions qui se font jour, comme la désinformation. Ils ont souligné le caractère général des Principes de l'OCDE sur l'IA et la nécessité d'une orientation plus détaillée pour l'élaboration de cadres réglementaires et d'options non réglementaires adaptés à des cas spécifiques.

### Graphique 4.1. Points de vue des Adhérents sur les cinq principales possibilités offertes par les systèmes d'IA avancés

Nombre d'Adhérents ayant opté pour chaque possibilité (cinq choix possibles) parmi une liste préétablie



Note : le chiffre indiqué est la somme des réponses données par les dix-sept participants à la question : « Quelles sont, pour votre pays ou votre région, les cinq principales possibilités offertes par l'IA générative en vue de la réalisation des objectifs nationaux et régionaux ? (Merci d'indiquer cinq choix) »

### Graphique 4.2. Points de vue des Adhérents sur les principaux risques associés aux systèmes d'IA avancés

Nombre d'Adhérents ayant opté pour chaque risque (cinq choix possibles) parmi une liste préétablie



Note : le chiffre indiqué est la somme des réponses données par les sept participants à la question : « Quels sont, pour votre pays ou votre région, les cinq risques principaux de l'IA pour la réalisation des objectifs nationaux et régionaux ? (Merci d'indiquer cinq choix) »

## 5. Diffusion

### 5.1. Activités des Adhérents

90. Plusieurs Adhérents ont fait état de mesures concrètes prises pour diffuser les Principes de l'OCDE sur l'IA au niveau national et les intégrer dans la formulation et la mise en œuvre de stratégies et de cadres de gouvernance en rapport avec l'IA. Ces initiatives vont de la diffusion des Principes au sein des milieux universitaires, de la communauté technique et des pouvoirs publics à leur intégration explicite dans des normes d'éthique et des stratégies nationales en matière d'IA. Cependant, les Adhérents n'ont pas tous explicitement mentionné d'exemples de diffusion, ce qui suggère que des efforts supplémentaires pourraient être nécessaires pour promouvoir la sensibilisation et la compréhension des Principes de l'OCDE sur l'IA à différents niveaux.

91. Plusieurs Adhérents ont fait des Principes de l'OCDE sur l'IA les piliers fondamentaux de leurs stratégies nationales en matière d'IA et de leurs cadres de gouvernance. Il s'agit notamment de l'Italie, dont la stratégie nationale en matière d'IA et le programme stratégique pour l'IA (2022-2024) mettent en œuvre les Principes de l'OCDE sur l'IA, et de la Corée, qui a intégré les Principes de l'OCDE sur l'IA dans sa stratégie nationale en matière d'IA, ses normes d'éthique en matière d'IA et sa stratégie de mise en œuvre pour une IA digne de confiance. De même, la Türkiye a intégré les Principes de l'OCDE sur l'IA dans sa stratégie nationale en matière d'IA et a entamé des discussions dans le cadre de réunions techniques et de haut niveau. La Lituanie a également inclus certains Principes de l'OCDE sur l'IA dans sa stratégie en matière d'IA, et le Japon y a fait référence dans des initiatives publiques majeures, notamment pour soutenir les débats au sein du Conseil pour des principes sociaux d'une IA centrée sur l'humain et du Conseil stratégique de l'IA, et pour élaborer un résumé provisoire des questions liées à l'IA et un projet de lignes directrices sur l'IA à l'intention des entreprises. Le Danemark et l'Irlande ont intégré les Principes de l'OCDE sur l'IA dans leurs lignes directrices à destination du secteur public, soulignant leur importance pour une adoption responsable de l'IA dans les administrations. Le Mexique s'est référé aux Principes de l'OCDE sur l'IA dans ses recommandations relatives au traitement des données à caractère personnel issues de l'utilisation de l'IA. Enfin, le Livre blanc du Royaume-Uni sur la gouvernance de l'IA s'aligne directement sur les Principes de l'OCDE sur l'IA.

### 5.2. Activités du Secrétariat pour diffuser et soutenir la mise en œuvre de la Recommandation

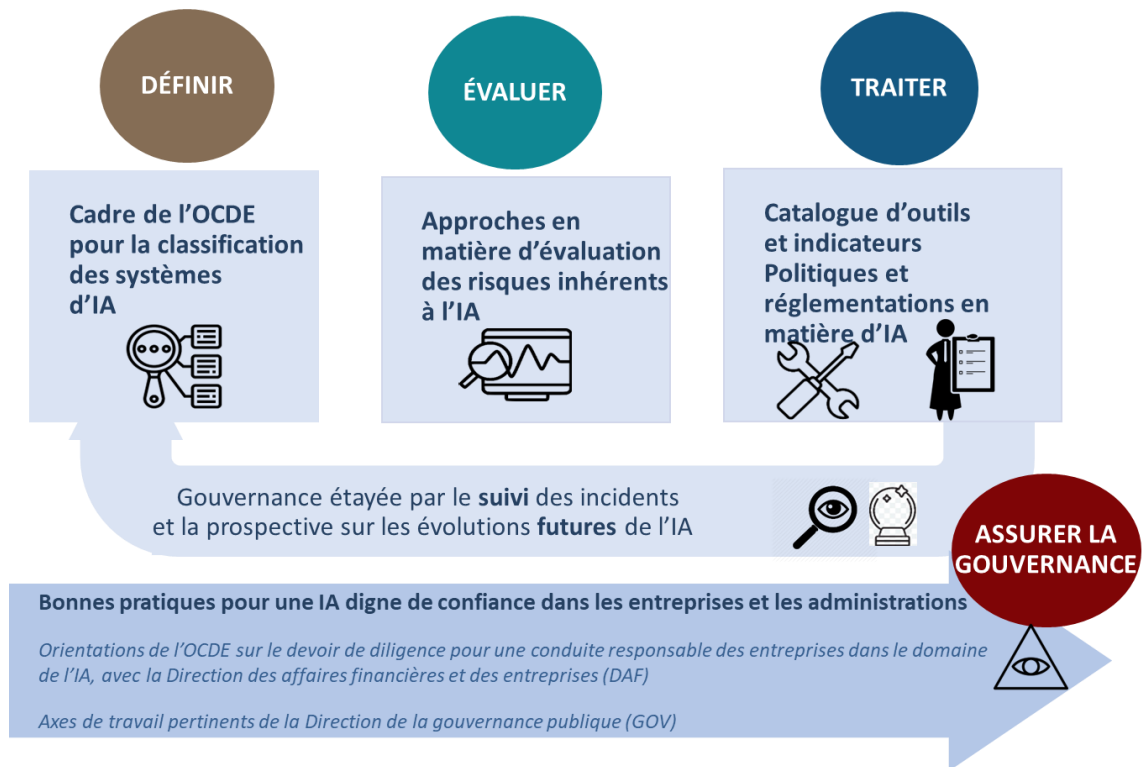
92. Le Secrétariat de l'OCDE collabore avec d'autres organisations intergouvernementales et régionales par le biais de la coalition globalpolicy.AI – établie par l'OCDE et sept autres organisations intergouvernementales – qui comprend une plateforme en ligne, et qui convoque régulièrement des discussions internationales sur la politique en matière d'IA et y participe. Grâce à cette collaboration, les organisations intergouvernementales partenaires aident les décideurs à s'orienter parmi les différentes initiatives internationales, se tiennent mutuellement informées de leurs activités respectives en matière d'IA, s'efforcent d'assurer l'interopérabilité de leurs travaux et travaillent ensemble pour faire progresser une IA digne de confiance là où leurs mandats se recoupent.

93. L'OCDE aide à la mise en œuvre de la Recommandation en constituant une base factuelle solide, en associant à ses travaux des experts de renommée mondiale et en faisant progresser les stratégies, les cadres et les outils. Le Secrétariat de l'OCDE mène ce travail dans le cadre du programme de travail du Groupe de travail sur la gouvernance de l'intelligence artificielle (GTGIA) du Comité de la politique du numérique.

- **L’Observatoire OCDE des politiques relatives à l’IA** : afin de soutenir la mise en œuvre pratique des Principes de l’OCDE sur l’IA, de fournir un forum inclusif pour l’échange d’informations sur les politiques en matière d’IA et les activités connexes, et de favoriser un dialogue multipartite et interdisciplinaire, l’OCDE a lancé l’Observatoire des politiques relatives à l’IA (OECD.AI) en février 2020. Cet observatoire réunit des ressources provenant de l’ensemble de l’OCDE, de ses partenaires et de groupes de parties prenantes, dans des domaines tels que la concurrence, l’innovation, les échanges, la santé et les compétences, des pages étant dédiées aux travaux analytiques connexes, aux actualités et aux visualisations de données. Preuve de l’intérêt croissant suscité par l’IA et du rôle de premier plan de l’OCDE au niveau mondial, le trafic sur la plateforme OECD.AI a progressé de 350 % au cours de l’année écoulée, et compte plus de 50 000 utilisateurs mensuels provenant de 100 pays à la mi-2023.
- **Nouvelles approches pour la mesure des tendances et évolutions propres à l’IA** : l’OCDE améliore également ses approches en ce qui concerne la mesure du développement et de l’adoption de l’IA dans le but de fournir aux pouvoirs publics les données actualisées nécessaires à l’élaboration de politiques en la matière. L’OCDE produit en continu plusieurs [mesures de l’IA](#) comme références et outils de recherche aux côtés de [données en temps réel](#) présentant les dernières actualités en matière d’IA et les principales tendances dans des domaines tels que : la [démographie](#) et la [fracture entre les sexes](#), la [recherche](#), l’[investissement](#), les [emplois et compétences](#), les brevets, le [développement de logiciels](#), l’[éducation](#) et [plus encore](#). D’autres travaux en cours s’appuient sur différentes sources de microdonnées pour étudier plus en profondeur les liens entre l’utilisation de l’IA par les entreprises et la productivité, et s’intéressent au rôle que jouent les actifs complémentaires, à l’instar du capital humain.
- **L’indice IA de l’OCDE** : sur la base des paramètres déjà disponibles sur OECD.AI (tendances et données, et base de données des politiques nationales en matière d’IA), l’OCDE élabore un cadre de mesure complet pour l’IA.
- **L’AI Incidents Monitor** : pour comprendre les incidents réels qui résultent de l’utilisation de l’IA, l’OCDE commence à surveiller les incidents d’IA en temps réel grâce à un nouvel outil de suivi dénommé [AI Incidents Monitor \(AIM\)](#).

94. Au cours des dernières années, l’OCDE a fourni des orientations pratiques sur la mise en œuvre des Principes de l’OCDE sur l’IA, par le biais des travaux analytiques qu’il a menés dans plusieurs domaines, et en collaboration avec d’autres services de l’OCDE. Graphique 5.1 résume le soutien de l’OCDE à la mise en œuvre des Principes de l’OCDE sur l’IA fondés sur des valeurs, à savoir contribuer à la gestion efficace des risques liés à l’IA. Tableau 5.1 fournit une vue d’ensemble des activités visant à aider les Adhérents à mettre en œuvre les cinq recommandations d’action.

**Graphique 5.1. Travaux de l'OCDE visant à soutenir la mise en œuvre d'une IA digne de confiance et fondée sur des valeurs**



**Tableau 5.1. Orientations de l'OCDE pour aider les Adhérents à mettre en œuvre les cinq recommandations à l'intention des pouvoirs publics**

Cinq recommandations à l'intention des pouvoirs publics	Sélection de projets
Investir dans la recherche et le développement en matière d'IA	Travaux sur les futurs de l'IA (menés en collaboration avec le Comité de la politique scientifique et technologique [CSTP] et l'unité de prospective du Secrétaire général) Événements et analyses concernant les systèmes énergétiques intelligents (avec le Groupe de travail sur les politiques d'infrastructure et de services de communication [GTPISC] et l'Agence internationale de l'énergie [AIE])
Favoriser l'instauration d'un écosystème numérique pour l'IA	Rapport intitulé « Measuring the environmental impacts of artificial intelligence compute and applications » (OCDE, 2022 <sup>[51]</sup> ) Rapport intitulé « Blueprint for building national compute capacity for artificial intelligence report » (OCDE, 2023 <sup>[27]</sup> ) Rapport intitulé « AI language models: Technological, socio-economic and policy considerations » (OCDE, 2023 <sup>[52]</sup> ) Rapport intitulé « Initial policy considerations for generative artificial intelligence » (Lorenz, Perset et Berryhill, 2023 <sup>[13]</sup> ).
Façonner un cadre d'action favorable à l'IA	Rapport sur les bacs à sable réglementaires (OCDE, 2023 <sup>[5]</sup> ) Rapports sur l'état d'avancement de la mise en œuvre (OCDE, 2021 <sup>[4]</sup> ), (OCDE, 2023 <sup>[5]</sup> ) Examens des stratégies nationales en matière d'IA (Allemagne et Égypte, à paraître)
Renforcer les capacités humaines et préparer la transformation du marché du travail	Programme sur l'IA dans le travail, l'innovation, la productivité et les compétences – OCDE.AI/WIPS (avec la Direction de l'emploi, du travail et des affaires sociales de l'OCDE [ELS] et la Direction de l'éducation de l'OCDE [EDU])
Favoriser la coopération internationale au service d'une IA digne de confiance	Globalpolicy.ai (avec 8 organisations intergouvernementales dont la CE, le Conseil de l'Europe, l'UNESCO, la Banque interaméricaine de développement)

### 5.2.1. Le réseau d'experts de l'IA à l'OCDE

95. Le réseau d'experts de l'IA à l'OCDE (ONE AI) est un groupe d'experts informel multipartite et multidisciplinaire qui soutient le GTGIA. Il comprend plus de 400 experts issus de tous les groupes de parties prenantes organisés en six groupes de travail informels, qui se réunissent tous les mois ou tous les deux mois afin de fournir des informations et des conseils au GTGIA sur les principaux domaines prioritaires d'action, notamment :

- **Risque et responsabilité** : l'OCDE collabore avec des organisations partenaires, des décideurs et des experts, pour identifier des indicateurs communs en vue d'évaluer les risques et les répercussions de l'IA afin de tendre vers une IA digne de confiance. L'objectif est de contribuer à la mise en œuvre de systèmes d'IA dignes de confiance, efficaces et responsables, en promouvant une cohérence globale.
- **Incidents liés à l'IA** : l'IA offre des avantages considérables, mais présente également des risques. Certains de ces risques se matérialisent déjà sous la forme de préjudices pour les individus et la société, tels que les biais et les discriminations, la polarisation des opinions, les atteintes à la vie privée et les problèmes de sécurité et de sûreté. Ces préjudices sont généralement désignés par l'expression « incident lié à l'IA » dont la définition est en cours d'élaboration. Le suivi des incidents liés à l'IA nécessite une cohérence et une interopérabilité globales dans la notification des incidents, de sorte que les opérateurs de systèmes d'IA et les décideurs puissent tirer des enseignements des risques et des incidents ayant affecté d'autres acteurs à l'échelle internationale. Le groupe d'experts auprès de l'OCDE travaille à l'élaboration d'un cadre de notification

des incidents liés à l'IA comprenant des définitions. Parallèlement, l'OCDE a mis au point un outil de suivi dénommé AI Incidents Monitor (AIM), qui suit les incidents réels liés à l'IA en temps réel et permet de vérifier que le cadre de notification et la définition fonctionnent dans la pratique ([OECD.AI/incidents](#)). Les incidents liés à l'IA rapportés dans les médias internationaux sont utilisés comme point de départ, de nombreux autres incidents n'étant pas rendus publics.

- **Calcul et climat** : à côté des données et des algorithmes, la capacité de calcul dédiée à l'IA (« capacité de calcul pour l'IA ») est un déterminant essentiel de l'IA ainsi que de la croissance économique et de la compétitivité qui en découlent. Les décideurs qui souhaitent élaborer des politiques en matière d'IA efficaces et faire des choix intelligents en matière d'investissement dans l'IA se doivent de comprendre leur capacité de calcul pour l'IA nationale. Le groupe d'experts OECD.AI sur la capacité de calcul pour l'IA et le climat aide le Secrétariat de l'OCDE à établir un cadre de base pour comprendre, mesurer et comparer les capacités de calcul pour l'IA nationales par pays et par région.
- **Les futurs de l'IA** : l'OCDE continue de suivre de près et d'analyser en détail les progrès rapides accomplis dans le domaine et leurs conséquences pour l'action publique. À mesure que l'IA évolue vers l'IA générative et au-delà, les décideurs doivent être en mesure d'anticiper et d'exploiter ces avancées technologiques, de doter les individus des compétences nécessaires et d'atténuer les risques associés. En avril 2023, l'OCDE a créé, dans le cadre du réseau ONE AI, un Groupe d'experts sur les futurs de l'IA, qui a pour mission d'étudier la diffusion et l'impact des systèmes d'IA générative ainsi que les conséquences de leur utilisation, notamment des aspects tels que l'innovation, la productivité, l'inclusion, l'emploi et l'épanouissement personnel, et l'éducation, dans le but de doter les administrations des connaissances et des outils nécessaires pour adapter l'action publique plus rapidement en tant que de besoin.
- **Indice de l'OCDE sur l'IA** : faisant fond sur la richesse des indicateurs sur les tendances et les politiques en matière d'IA présents dans l'Observatoire OCDE des politiques relatives à l'IA, le groupe d'experts nouvellement créé (septembre 2023) œuvre pour élaborer un cadre de mesure complet et synthétique sur l'intelligence artificielle digne de confiance. Ce travail est mené en collaboration avec le Groupe de travail sur l'économie, la mesure et l'analyse du numérique (GTEMAN) et la Saïd Business School de l'Université d'Oxford.
- **AI, données et vie privée** : établi début 2024, le Groupe d'experts sur l'IA, les données et la vie privée aide l'OCDE à créer des liens entre la communauté de l'IA et celle de la protection de la vie privée et des données, à explorer les possibilités et les enjeux en matière d'action publique, et à examiner les synergies possibles entre les principaux cadres existants en matière d'IA et de protection de la vie privée en s'appuyant sur les Principes de l'OCDE en matière d'IA et de protection de la vie privée. Les travaux sont menés en coopération avec le Groupe de travail sur la gouvernance des données et la vie privée (unité Gouvernance des données et vie privée).

### 5.2.2. Travaux analytiques de l'OCDE

96. Le Secrétariat de l'OCDE a mené un large éventail de travaux analytiques fondés sur les Principes de l'OCDE sur l'IA énoncés dans la Recommandation et qui contribuent à diffuser et à soutenir leur mise en œuvre :

- **Le Cadre de l'OCDE pour la classification des systèmes d'IA** (OCDE, 2022<sup>[6]</sup>) jette les bases d'une compréhension commune du fonctionnement des systèmes d'IA de sorte

qu'il soit plus facile d'adapter l'évaluation des risques au contexte (par exemple la différence entre un système d'IA utilisé pour la traduction et un système utilisé pour détecter des maladies). Le cadre, aujourd'hui largement utilisé, a été élaboré par ONE AI en s'appuyant sur une contribution internationale par le biais d'une consultation publique. On le retrouve notamment dans le projet de cadre de gestion des risques du National Institute of Standards and Technology mis au point par les États-Unis, ainsi que dans les travaux du Parlement européen, du Conseil de l'Europe et de la Commission européenne en vue de l'adoption de la législation de l'UE sur l'IA.

- **Le rapport intitulé « Advancing Accountability in AI : Governing and managing risks through the lifecycle for trustworthy AI »** (OCDE, 2023<sup>[7]</sup>) fournit un cadre d'interopérabilité à haut niveau et établit une cartographie des normes, cadres et lignes directrices existants et en cours d'élaboration en matière d'IA en fonction de leurs caractéristiques fondamentales. L'OCDE prévoit de réaliser une analyse des lacunes et de développer un outil interactif en ligne pour aider les organisations et les parties prenantes à comparer les cadres et à s'y orienter parmi les approches existantes pour identifier, évaluer, traiter et gérer les risques liés à l'IA.
- **Le catalogue d'outils et de mesures de l'OCDE pour une IA digne de confiance** (OECD.AI/tools) : le catalogue constitue une base de données actualisée à laquelle des organisations du monde entier contribuent et qu'elles mettent à jour régulièrement. Avec actuellement plus de 700 outils (dont 34 concernent spécifiquement l'IA générative), il s'agit d'un guichet unique permettant de rechercher et de partager des outils et des méthodes – techniques, normatives et éducatives – visant à rendre l'IA digne de confiance. Les outils sont classés en fonction de leurs objectifs, qui reflètent les Principes de l'IA de l'OCDE fondés sur des valeurs. Le catalogue est notamment cité dans la feuille de route conjointe du Conseil du commerce et des technologies UE-États-Unis sur les outils d'évaluation et de mesure pour une IA digne de confiance et la gestion des risques.

### 5.2.3. Analyse des progrès technologiques de l'IA : l'IA générative

97. Le paysage technologique et juridique de l'IA a connu une évolution rapide depuis l'adoption des Principes de l'OCDE sur l'IA en 2019.

98. L'évolution la plus notable est la montée en puissance des modèles d'IA « de fondation » ou « à usage général », notamment les grands modèles de langue qui peuvent produire des contenus inédits, transposer du texte en vidéo et en image, et offrir des agents conversationnels avancés accessibles d'une simple pression sur un écran. Vers la fin de 2022 et en 2023, l'IA générative a occupé le devant de la scène dans les débats publics, universitaires et politiques, à la suite du lancement de l'agent conversationnel ChatGPT et de l'adoption rapide d'outils d'IA générative qui produisent de nouveaux textes, images et contenus audio ou vidéo.

99. En avril 2023, l'OCDE était l'une des premières organisations à publier une analyse concernant les évolutions d'un domaine clé de l'IA générative, à savoir les grands modèles de langue, avec un examen des avantages et des risques liés à leur utilisation, ainsi qu'un exposé des considérations relatives à l'action publique par le prisme de la Recommandation de l'OCDE (OCDE, 2023<sup>[52]</sup>). Elle a également soutenu le « processus d'Hiroshima sur l'IA générative » du G7, notamment en dressant un bilan des principaux risques, possibilités et approches en matière d'IA générative dans les pays du G7 et l'UE (OCDE, 2023<sup>[47]</sup>). En septembre 2023, l'OCDE a lancé une série de documents de travail qui suivent les progrès de l'IA générative, avec une première publication axée sur les considérations technologiques et socio-économiques intéressant l'action des pouvoirs

publics des modèles de langage de l'IA, notamment sur le travail, les biais, la désinformation et la propriété intellectuelle (OCDE, 2023<sup>[53]</sup>).

100. Les trois rapports susmentionnés analysent l'IA générative par le prisme des Principes de l'OCDE sur l'IA, qui se sont révélés être un cadre pertinent et complet pour comprendre les possibilités et les risques associés aux progrès technologiques et les enseignements à en tirer sur le plan de l'action des pouvoirs publics, et pour identifier des mesures d'atténuation potentielles.

## 6. Synthèse et conclusions

### 6.1. Mise en œuvre

101. Le rapport constate que la Recommandation sur l'IA est mise en œuvre par les Adhérents. Depuis 2019, les Adhérents ont établi des initiatives nationales et internationales visant à suivre les cinq recommandations d'action à l'intention des pouvoirs publics et à traduire en actes les Principes fondés sur des valeurs. Les dix Principes de l'OCDE sur l'IA énoncés dans la Recommandation ont été utilisés pour orienter les initiatives nationales, notamment les stratégies nationales et les cadres réglementaires et de gouvernance en matière d'IA des Adhérents, pour encadrer les travaux analytiques menés par l'OCDE et les Adhérents, et pour analyser les progrès technologiques dans le domaine.

102. Les Adhérents ont indiqué que les Principes de l'OCDE sur l'IA constituent une base solide pour l'élaboration de politiques nationales en matière d'IA, qu'ils fournissent des orientations pour l'établissement du cadre d'action, qu'ils favorisent l'alignement international en matière de gouvernance de l'IA et qu'ils facilitent la collaboration entre les parties prenantes. Toutefois, certains Adhérents ont rencontré des problèmes lors de la mise en œuvre des Principes fondés sur les valeurs, le principal étant la difficulté de les transposer en mesures pratiques et réalisables, car ils sont de haut niveau et de vaste portée. Les connaissances et les capacités techniques limitées pour comprendre l'IA, ainsi que les contraintes de ressources qui pèsent sur l'élaboration des politiques et les investissements sont également perçues comme des obstacles à la mise en œuvre des Principes de l'OCDE sur l'IA au niveau national.

103. Les Adhérents ont demandé à l'OCDE de continuer à fournir des lignes directrices et des orientations pratiques sur la mise en œuvre. Il s'agit notamment de partager les pratiques exemplaires, comme des orientations sur le devoir de diligence concernant la conduite responsable des entreprises dans le domaine de l'IA, des évaluations comparatives internationales et le renforcement des capacités, et d'apporter une assistance technique aux pays. Les Adhérents ont également souligné la nécessité de disposer d'orientations plus détaillées pour la mise en œuvre des Principes de l'OCDE sur l'IA dans les différents secteurs, et notamment d'un soutien à l'élaboration de cadres réglementaires et d'options non réglementaires adaptés à des cas spécifiques et à des secteurs particuliers tels que les soins de santé ou la finance, ou pour contribuer à relever les défis environnementaux.

### 6.2. Diffusion

104. Plusieurs Adhérents ont déclaré avoir pris des initiatives concrètes pour diffuser les Principes de l'OCDE sur l'IA au niveau national et les intégrer dans la formulation et la mise en œuvre des stratégies et des cadres de gouvernance en matière d'IA. Ces initiatives vont de leur diffusion au sein des milieux universitaires, de la communauté technique et des pouvoirs publics à leur intégration explicite dans des normes d'éthique et des stratégies nationales en matière d'IA.

### 6.3. Maintien de la pertinence

105. Les Adhérents considèrent que les Principes de l'OCDE sur l'IA restent pertinents pour guider le développement de l'IA, encourager l'innovation et renforcer la confiance du public. Les Principes de l'OCDE sur l'IA sont également largement considérés comme étant des lignes directrices précieuses pour promouvoir un développement de l'IA responsable et conforme aux droits humains et aux valeurs démocratiques. Ils sont perçus comme étant fondamentaux afin de promouvoir les valeurs centrées sur l'humain, la transparence, la responsabilité et l'équité, prévenir les biais, promouvoir l'inclusion et veiller à ce que l'IA respecte les droits de l'homme et les valeurs démocratiques. Ils fournissent également un cadre d'éthique qui soutient la croissance inclusive, la collaboration internationale et l'établissement de normes communes dans le paysage mondial de l'IA. Certains Adhérents ont souligné la nécessité d'une législation nationale ou régionale complémentaire et d'une collaboration avec d'autres initiatives internationales pour gérer efficacement les risques liés à l'IA.

106. En ce qui concerne les systèmes d'IA avancés, les Adhérents considèrent les Principes de l'OCDE sur l'IA comme un outil précieux pour aborder les domaines prioritaires d'action importants liés à l'IA, mais avec des divergences quant à leur efficacité perçue. Tout en soulignant l'importance de la Recommandation en tant qu'orientation pour relever les défis cruciaux liés à l'IA, les Adhérents ont également mis l'accent sur la nécessité de disposer de cadres de gouvernance pratiques et de Principes adaptés pour traiter les préoccupations qui se font jour, comme la désinformation. Les Adhérents ont estimé que compléter les Principes de l'OCDE sur l'IA par des recommandations plus spécifiques et plus concrètes, en particulier dans des domaines émergents tels que l'IA générative, pourrait améliorer leur efficacité à tirer parti des possibilités offertes par l'IA.

107. Les Principes de l'OCDE sur l'IA constituent une référence internationale importante et utile à laquelle les Adhérents ont recours afin d'élaborer leurs politiques nationales en matière d'IA. Ils bénéficient d'une large diffusion et conservent toute leur pertinence, offrant notamment un cadre solide à l'analyse des évolutions récentes telles que celles liées à l'IA générative.

108. La Recommandation, sous sa forme actuelle, reste globalement pertinente. Il est néanmoins possible de l'actualiser afin d'aider les parties prenantes à la mettre en œuvre et de tenir compte des problématiques et des avancées technologiques nouvelles, notamment dans le domaine de l'IA générative. Il s'agirait plus précisément de : i) refléter l'importance croissante de la lutte contre la mésinformation et la désinformation ainsi que de la sauvegarde de l'intégrité de l'information dans le contexte de l'IA générative ; ii) traiter des utilisations en dehors de la finalité assignée, de l'utilisation abusive intentionnelle ou non intentionnelle ; iii) 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 ; iv) décrire les mécanismes permettant de remédier aux préjudices potentiels ou aux comportements indésirables tout au long du cycle de vie de l'IA ; et v) mettre l'accent sur la conduite responsable des entreprises tout au long du cycle de vie de l'IA, ce qui implique de coopérer avec les fournisseurs de connaissances et de ressources dans le domaine, les utilisateurs de systèmes d'IA et d'autres parties prenantes. De plus, les titres et le texte de certains Principes de l'OCDE sur l'IA pourraient être étoffés, et les passages sur la traçabilité et la gestion du risque pourraient être développés et déplacés vers le Principe de responsabilité, dont ils semblent davantage relever.

## Références

- AI Safety Summit (2023), *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>. [16]
- Alliance de recherche numérique du Canada (2020), *Évaluation des besoins de l'infrastructure de recherche numérique au Canada*, <https://alliancecan.ca/fr/initiatives/evaluation-des-besoins-de-linfrastructure-de-recherche-numerique-au-canada>. [23]
- CDEI (2024), *Portfolio of AI Assurance Technique*, <https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-technique>. [19]
- Commission européenne (2023), *Une approche européenne de l'intelligence artificielle*, <https://digital-strategy.ec.europa.eu/fr/policies/european-approach-artificial-intelligence>. [21]
- Conseil de l'Europe (2024), *Comité sur l'intelligence artificielle*, <https://www.coe.int/fr/web/artificial-intelligence/cai>. [33]
- Department of Labour (2024), *United States/Spain Joint Statement Algorithmic Bias in the World of Work*, [https://www.dol.gov/sites/dolgov/files/ILAB/USA-SPAIN\\_JointStatementOnAlgorithmicBias.pdf](https://www.dol.gov/sites/dolgov/files/ILAB/USA-SPAIN_JointStatementOnAlgorithmicBias.pdf). [46]
- Fasken (2022), *La réglementation sur l'intelligence artificielle au Canada et à l'étranger : comparaison entre le projet de loi C-27 édictant la LIAD et la législation sur l'IA de l'Union européenne*, <https://www.fasken.com/fr/knowledge/2022/10/18-the-regulation-of-artificial-intelligence-in-canada-and-abroad> (consulté le 29 mars 2023). [30]
- Gouvernement des États-Unis (2022), *Blueprint for an AI Bill of Rights*, <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> (consulté le 29 mars 2023). [36]
- Gouvernement du Canada (2023), *La Loi sur l'intelligence artificielle et les données (LIAD) – document complémentaire*, <https://ised-isde.canada.ca/site/innover-meilleur-canada/fr/loi-lintelligence-artificielle-donnees-liad-document-complementaire> (consulté le 29 mars 2023). [31]
- Gouvernement du Royaume-Uni (2023), *AI regulation: a pro-innovation approach*, <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach> (consulté le 29 mars 2023). [35]
- Gouvernement fédéral de l'Allemagne (2022), *Regulatory Sandboxes – Testing Environments for Innovation and Regulation*, <https://www.bmwk.de/Redaktion/EN/Dossier/regulatory-sandboxes.html> (consulté le 29 mars 2023). [28]
- Ibero-American Data Protection Network (2019), *General Recommendations for the Processing of Personal Data in Artificial Intelligence*, <https://www.redipd.org/sites/default/files/2020-02/guide-general-recommendations-processing-personal-data-ai.pdf>. [17]
- Ibero-American Data Protection Network (2019), *Specific Guidelines for Compliance with the Principles and Rights that Govern the Protection of Personal Data in Artificial Intelligence Projects*, <https://www.redipd.org/sites/default/files/2020-02/guide-specific-guidelines-ai-projects.pdf>. [18]

- INAI (2022), *Recomendaciones para el tratamiento de datos personales derivado del uso de la Inteligencia Artificial*, <https://home.inai.org.mx/wp-content/documentos/DocumentosSectorPublico/RecomendacionesPDP-IA.pdf>. [54]
- Lassébie, J. et G. Quintini (2022), « What skills and abilities can automation technologies replicate and what does it mean for workers? : New evidence », *Documents de travail de l'OCDE sur les questions sociales, l'emploi et les migrations*, n° 282, Éditions OCDE, Paris, <https://doi.org/10.1787/646aad77-en>. [42]
- Lorenz, P., K. Perset et J. Berryhill (2023), « Initial policy considerations for generative artificial intelligence », *OECD Artificial Intelligence Papers*, n° 1, Éditions OCDE, Paris, <https://doi.org/10.1787/fae2d1e6-en>. [13]
- MIC (2023), *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*, [https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05\\_en.pdf](https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05_en.pdf). [49]
- MIC (2023), *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System*, [https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document04\\_en.pdf](https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document04_en.pdf). [48]
- Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal counsel and legislative affairs (2022), *Principles for Policy, Regulation and Ethics in AI: technical paper (annex to policy paper)*, <https://www.gov.il/BlobFolder/rfp/061122/he/professional-letter.pdf>. [34]
- National Council of Science and Technology (2022), *The Networking & Information Technology R&D Program and the National Artificial Intelligence Initiative Office Supplement to the President's FY 2023 Budget*, <https://www.nitrd.gov/pubs/FY2023-NITRD-NAIIO-Supplement.pdf>. [20]
- National Institute of Standards and Technology - US Department of Commerce (2023), *AI Risk Management Framework*, <https://www.nist.gov/itl/ai-risk-management-framework> (consulté le 28 mars 2023). [8]
- OCDE (2024), *2023 OECD Digital Government Index: Results and key findings*, Éditions OCDE, Paris. [25]
- OCDE (2024), « Explanatory memorandum on the updated OECD definition of an AI system », *OECD Artificial Intelligence Papers*, n° 8, Éditions OCDE, Paris, <https://doi.org/10.1787/623da898-en>. [2]
- OCDE (2024), *Les faits sans le faux : Lutter contre la désinformation, renforcer l'intégrité de l'information*, Éditions OCDE, <https://doi.org/10.1787/4078bb32-fr>. [15]
- OCDE (2024), *OECD database of national AI policies & strategies*, <https://oecd.ai/fr/dashboards/overview>. [3]
- OCDE (2023), *2023 OECD Open, Useful and Re-usable data (OURdata) Index: Results and key findings*, Éditions OCDE, Paris. [22]
- OCDE (2023), *A blueprint for building national compute capacity for artificial intelligence*, Éditions OCDE, <https://doi.org/10.1787/876367e3-en>. [27]
- OCDE (2023), *Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI*, Éditions OCDE, <https://doi.org/10.1787/2448f04b-en>. [7]

- OCDE (2023), « AI language models: Technological, socio-economic and policy considerations », *Documents de travail de l'OCDE sur l'économie numérique* n° 352, <https://doi.org/10.1787/13d38f92-en>. [52]
- OCDE (2023), « AI language models: Technological, socio-economic and policy considerations », *Documents de travail de l'OCDE sur l'économie numérique* n° 352, <https://doi.org/10.1787/13d38f92-en>. [53]
- OCDE (2023), *Digital Government Review of Latin America and the Caribbean: Building Inclusive and Responsive Public Services*, <https://www.oecd.org/publications/digital-government-review-of-latin-america-and-the-caribbean-29f32e64-en.htm> (consulté le 5 décembre 2023). [26]
- OCDE (2023), *Emerging privacy-enhancing technologies: Current regulatory and policy approaches*, Éditions OCDE, <https://doi.org/10.1787/bf121be4-en>. [11]
- OCDE (2023), *Perspectives de l'emploi de l'OCDE 2023 : Intelligence artificielle et marché du travail*, Éditions OCDE, Paris, <https://doi.org/10.1787/aae5dba0-fr>. [12]
- OCDE (2023), *Perspectives de l'emploi de l'OCDE 2023 : Intelligence artificielle et marché du travail*, Éditions OCDE, Paris, <https://doi.org/10.1787/aae5dba0-fr>. [45]
- OCDE (2023), *Processus du G7, dit d'Hiroshima, sur l'intelligence artificielle générative : Vers une vision commune à l'échelle du G7 sur l'IA générative*, Éditions OCDE, <https://doi.org/10.1787/678bee9f-fr>. [14]
- OCDE (2023), *Processus du G7, dit d'Hiroshima, sur l'intelligence artificielle générative : Vers une vision commune à l'échelle du G7 sur l'IA générative*, Éditions OCDE, Paris, <https://doi.org/10.1787/678bee9f-fr>. [47]
- OCDE (2023), « Regulatory sandboxes in artificial intelligence », vol. n° 356, <http://10.1787/8f80a0e6-en>. [10]
- OCDE (2023), *State of implementation of the OECD AI Principles four years on*. [5]
- OCDE (2022), « Measuring the environmental impacts of artificial intelligence compute and applications : The AI footprint », *Documents de travail de l'OCDE sur l'économie numérique*, n° 341, Éditions OCDE, Paris, <https://doi.org/10.1787/7babf571-en>. [9]
- OCDE (2022), *Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint*, Éditions OCDE, <https://doi.org/10.1787/7babf571-en>. [51]
- OCDE (2022), *OECD Framework for the Classification of AI systems*, Éditions OCDE, <https://doi.org/10.1787/cb6d9eca-en>. [6]
- OCDE (2021), *Examiner la réglementation existante*, Principes de bonne pratique de l'OCDE en matière de politique réglementaire, Éditions OCDE, Paris, <https://doi.org/10.1787/eb656b8d-fr>. [40]
- OCDE (2021), « Recommandation du Conseil en faveur d'une gouvernance réglementaire agile permettant de mettre l'innovation à profit », *OECD/LEGAL/0464*, <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0464>. [39]
- OCDE (2021), *State of Implementation of the OECD AI Principles: Insights from National AI Policies*, <https://doi.org/10.1787/1cd40c44-en>. [4]

- OCDE (2019), *Recommandation du Conseil sur l'intelligence artificielle*, OECD/LEGAL/0449, <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. [1]
- OECD.AI (2024), *Programme OECD.AI sur le travail, l'innovation, la productivité et les compétences*, <https://oecd.ai/fr/work-innovation-productivity-skills>. [44]
- OECD.AI (2023), *Politiques et stratégies nationales en matière d'IA*, <https://oecd.ai/fr/dashboards/overview> (consulté le 2023). [32]
- Parlement du Canada (2022), *Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*, <https://www.parl.ca/legisinfo/fr/projet-de-loi/44-1/c-27> (consulté le 28 mars 2023). [29]
- The White House (2023), *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. [38]
- The White House (2023), *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-united-states-bans-imports-of-russian-oil-liquefied-natural-gas-and-coal/>. [37]
- UK DSIT (2023), *Independent Review of The Future of Compute: Final report and recommendations*, Department for Science, Innovation & Technology. [24]
- UN Advisory Board on AI (2023), *Governing AI for Humanity - Interim Report*, [https://www.un.org/sites/un2.un.org/files/ai\\_advisory\\_body\\_interim\\_report.pdf](https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf). [50]
- US Copyright Office (2023), *Copyright Office Issues Notice of Inquiry on Copyright and Artificial Intelligence*, [https://www.copyright.gov/newsnet/2023/1017.html?utm\\_campaign=subscriptioncenter&utm\\_content=&utm\\_medium=email&utm\\_name=&utm\\_source=govdelivery&utm\\_term=](https://www.copyright.gov/newsnet/2023/1017.html?utm_campaign=subscriptioncenter&utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term=). [41]
- Zarifhonarvar, A. (2023), « Economics of ChatGPT: A Labour Market View on the Occupational Impact of Artificial Intelligence », <https://ssrn.com/abstract=4350925>. [43]