Rapport d'étape

Gouverner l'intelligence artificielle au bénéfice de l'humanité

Décembre 2023

www.un.org/en/ai-advisory-body (non disponible en français)

Table des matières

		Page
	Abréviations	3
I.	Introduction	4
II.	Un manque de gouvernance mondiale	8
III.	Des débouchés et des catalyseurs	8
	A. Principaux éléments permettant de mettre l'intelligence artificielle au service de l'humanité	10
	B. La gouvernance, un facteur essentiel	11
IV.	Des risques et des difficultés	12
	A. Risques posés par l'intelligence artificielle	12
	B. Difficultés à résoudre	15
V.	Une gouvernance internationale de l'intelligence artificielle	16
	A. Situation en matière de gouvernance de l'intelligence artificielle	16
	B. Vers des principes et des fonctions de gouvernance internationale de l'intelligence artificielle	18
	C. Recommandations préliminaires	19
	1. Principes directeurs	19
	a) Principe directeur nº 1 : l'intelligence artificielle doit être gouvernée de manière inclusive, par toutes et par tous et dans l'intérêt commun	19
	b) Principe directeur nº 2 : l'intelligence artificielle doit être gouvernée dans l'intérêt général	19

c) Principe directeur nº 3 : la gouvernance de l'intelligence artificielle doit être articulée avec la gouvernance des données et la promotion des biens communs en matière de données 20

d) Principe directeur nº 4 : la gouvernance de l'intelligence artificielle doit être universelle, en réseau et ancrée dans une collaboration multipartite adaptative 20

e) Principe directeur nº 5 : la gouvernance de l'intelligence artificielle devrait prendre ses racines dans la Charte des Nations Unies, le droit international des droits de l'homme et d'autres engagements convenus au niveau international, tels que les objectifs de développement durable. 21

2. Fonctions institutionnelles 21

a) Fonction institutionnelle nº 1 : évaluer régulièrement les orientations et les implications futures de l'intelligence artificielle. 22

b) Fonction institutionnelle nº 2 : renforcer l'interopérabilité des efforts de gouvernance et leur ancrage dans les normes internationales grâce à un cadre mondial de gouvernance de l'intelligence artificielle 23

c) Fonction institutionnelle nº 3 : élaborer et harmoniser des normes et des cadres de gestion de la sécurité et des risques 24

d) Fonction institutionnelle nº 4 : faciliter le développement, le déploiement et l'utilisation de l'intelligence artificielle dans l'intérêt de l'économie et de la société grâce à une coopération internationale multipartite 24

e) Fonction institutionnelle nº 5 : promouvoir la collaboration internationale en matière de développement des talents, d'accès aux infrastructures de calcul, de constitution de divers jeux de données de haute qualité, de partage responsable de modèles libres de droits et de biens collectifs utilisant l'intelligence artificielle aux fins de la réalisation des objectifs de développement durable 25

f) Fonction institutionnelle nº 6 : surveiller les risques, signaler les incidents et coordonner les interventions d'urgence 25

g) Fonction institutionnelle nº 7 : conformité et responsabilité fondées sur des normes 26

VI. Conclusions 28

VII. Les prochaines étapes 29

Annexes

I. À propos de l'Organe consultatif de haut niveau sur l'intelligence artificielle 32

II. Membres de l'Organe consultatif de haut niveau sur l'intelligence artificielle 33

III. Mandat de l'Organe consultatif de haut niveau sur l'intelligence artificielle. 34

IV. Groupes de travail et thèmes transversaux 35

Abréviations

AI	Intelligence artificielle
AIEA	Agence internationale de l’énergie atomique
CERN	Organisation européenne pour la recherche nucléaire
GIEC	Groupe d’experts intergouvernemental sur l’évolution du climat
OACI	Organisation de l’aviation civile internationale
OIT	Organisation internationale du Travail
OMI	Organisation maritime internationale
UNESCO	Organisation des Nations Unies pour l’éducation, la science et la culture
UTI	Union internationale des télécommunications

I. Introduction

1. L'intelligence artificielle (IA) a de plus en plus d'incidences sur la vie des personnes[1]. Elle existe depuis des années, mais des capacités autrefois difficilement imaginables sont apparues à un rythme rapide et sans précédent. Elle a aujourd'hui un potentiel extraordinaire qui pourrait être mis au service du bien : elle pourrait conduire à des découvertes scientifiques qui repoussent les limites de la connaissance humaine et à des outils qui optimisent les ressources limitées et aident les personnes dans leurs tâches quotidiennes. Elle pourrait aussi changer la donne dans la transition vers un avenir plus vert ou aider les pays en développement à transformer la santé publique et à résoudre plus rapidement les problèmes liés à l'accès à l'éducation, même dans les zones les plus reculées. Les pays développés à la population vieillissante pourraient même l'utiliser pour remédier aux pénuries de main-d'œuvre.

2. Néanmoins, l'IA comporte aussi des risques. Elle pourrait renforcer les préjugés ou élargir la surveillance ; l'automatisation de la prise de décision pourrait brouiller la chaîne de responsabilité hiérarchique parmi les fonctionnaires, et la désinformation renforcée par l'IA menace le processus même qui régit leur élection. La vitesse, l'autonomie et l'opacité de ces systèmes remettent en question les modèles traditionnels de réglementation, alors même que des systèmes toujours plus puissants sont développés, déployés et utilisés.

3. Les chances à saisir et les risques que présente l'IA pour l'humanité et la société sont évidents et ont suscité l'intérêt du public : ils se manifestent également au niveau mondial, où on observe des tensions géostratégiques liées à l'accès aux données, à la puissance de calcul et aux talents qui alimentent l'IA, et des débats sur une nouvelle course aux armements en lien avec l'IA. De plus, les bienfaits et les risques associés à l'IA ne sont pas équitablement répartis. Même si l'humanité ne devait exploiter que les aspects positifs de celle-ci, le risque que ceux-ci ne bénéficient qu'au « club des riches » est réel. En effet, les bienfaits associés à l'IA profitent essentiellement à une poignée d'États, d'entreprises et de personnes à l'heure actuelle.

4. Il est nécessaire d'instaurer une gouvernance pour cette technologie, non seulement pour relever les problèmes et les risques qu'elle crée, mais aussi pour que son potentiel soit exploité de manière à ne laisser personne de côté. La contribution de l'IA à la réalisation des objectifs de développement durable constituera l'un des principaux critères d'évaluation du succès de cette technologie. On trouvera dans l'encadré 1 ci-dessous une présentation de son potentiel en matière de lutte contre les changements climatiques et leurs conséquences (objectif 13).

Encadré 1

Au point de rencontre crucial entre les changements climatiques et les possibilités offertes par l'intelligence artificielle : étude de cas

Les changements climatiques constituent un problème mondial et universel : y répondre de façon collective nécessite une transformation numérique durable, de nouvelles infrastructures bien conçues et la capacité de prendre des décisions précises à grande échelle. Les approches fondées sur l'IA sont particulièrement bien adaptées à ce problème, car elles créent de nouvelles capacités en tenant compte, entre autres, des évolutions majeures en matière d'apprentissage automatique, de grands modèles de langage et d'analyse de données de haute qualité.

[1] Voir la définition de l'intelligence artificielle de l'Observatoire OCDE des politiques de l'IA, à l'adresse https://oecd.ai/en/wonk/ai-system-definition-update (non disponible en français).

Les informations qui décrivent des phénomènes déconnectés et disparates (imagerie géospatiale, capteurs distribués, surveillance en temps réel et données sur les effets des changements climatiques hyperlocaux transmises par les citoyens et citoyennes) peuvent être utilisées pour créer une nouvelle compréhension des éléments, des conséquences et des systèmes complexes qui déterminent les répercussions climatiques. En parallèle des systèmes prédictifs capables de transformer les données en informations et de traduire ces informations en actes, les outils utilisant l'IA peuvent contribuer à l'élaboration de nouvelles stratégies et de nouveaux investissements visant à réduire les émissions, à influencer les nouveaux investissements du secteur privé dans des projets à zéro émission nette, à protéger la biodiversité et à renforcer la résilience sociale à grande échelle. Ces mesures concernent non seulement l'objectif 13 relatif à l'action climatique, mais aussi d'autres objectifs.

On trouvera ci-dessous une liste non exhaustive des domaines dans lesquels l'IA a de bonnes chances de contribuer à la lutte contre les changements climatiques :

- Attribuer la responsabilité de l'action climatique aux institutions de gouvernance nationales et infranationales en créant de nouvelles ressources prédictives très granulaires pour l'investissement dans le domaine du climat. Par exemple, des cartes thermiques en temps réel des inondations urbaines liées aux tempêtes pourraient contribuer à améliorer les infrastructures hyperlocales des égouts et des réseaux d'évacuation des eaux.
- Construire des systèmes publics de données et d'intelligence artificielle libres de droits afin que les rapports sur les émissions nettes publiés par le secteur privé cessent de jouer un rôle statique (conformité) et deviennent une base de données en temps réel accessible au public, ce qui augmenterait la confiance, la transparence et la responsabilité en ce qui concerne les engagements publics.
- Utiliser des modélisations climatiques de pointe associées à des informations sur la mobilité urbaine et les comportements pour créer de nouveaux dispositifs d'alerte rapide, ce qui améliorerait l'efficacité des secours et de la reprise après un conflit ou une catastrophe.
- Mettre au point des solutions utilisant l'IA qui seraient fondées sur des preuves, aux fins de leur utilisation dans des systèmes ouverts et d'autres technologies d'élimination du carbone dans lesquels des intervalles d'incertitude élevés peuvent limiter les investissements essentiels à un stade précoce. Les techniques de modélisation de pointe peuvent réduire le coût de la recherche scientifique et faciliter le prototypage rapide de nouvelles solutions.

Toutefois, des obstacles structurels subsistent et empêchent ces technologies d'atteindre l'échelle requise pour répondre à l'ampleur de la crise climatique et aux divers besoins des parties prenantes majeures qui participent à la lutte contre les changements climatiques, notamment les entreprises, les gouvernements, les activistes et la société civile. Les risques systémiques, tels que les biais algorithmiques, les biais liés au contexte de transfert, les biais d'interprétation et les préjudices liés à la

représentation et à l'attribution, doivent être pris en compte. Pour surmonter ces obstacles, on peut, par exemple :

- Améliorer l'explicabilité des modèles et la confiance afin d'accroître l'intégration des informations produites par l'IA dans la prise de décisions essentielles en matière de climat.
- Veiller à ce que les modèles d'IA soient entraînés sur des jeux de données diversifiés et réellement représentatifs qui intègrent des données commercialement viables collectées par des entités à but lucratif et des données qui « comblent les lacunes », financées par des ressources à but non lucratif, philanthropiques et gouvernementales, et qui complètent les connaissances locales et tacites.
- Permettre aux populations vulnérables face aux changements climatiques d'accéder à des prévisions générées par l'IA qui, en d'autres circonstances, ne seraient communiquées qu'à des entreprises privées.
- Réduire le coût du calcul et de l'expertise en apprentissage automatique afin que les organisations à but non lucratif et la société civile puissent créer et maintenir des produits liés à l'IA qui soient gratuits et libres de droits.
- Avoir raison des activités cloisonnées menées par plusieurs organisations qui élaborent des solutions exclusives ou détiennent des données exclusives afin de rivaliser pour obtenir des investissements privés ou philanthropiques.
- Fournir un financement pour donner plus d'ampleur aux solutions susmentionnées.

Les relations transversales existant entre l'IA et l'expérience des changements climatiques en première ligne sont essentielles pour que les approches transformatrices susmentionnées puissent voir le jour. Même lorsqu'elles sont dotées de puissance de calcul, de données et d'effectifs, les solutions techniques en silo rencontrent des problèmes importants pour ce qui est de l'adoption et de la distribution lorsqu'elles ne reflètent pas l'expérience vécue par les populations et les décideurs locaux.

Pour chacune des possibilités décrites ci-dessus, les parties prenantes non techniques doivent participer à la conception, à l'élaboration, à l'exécution et à l'intégration du projet en apportant des contributions essentielles dès le départ. Les facilitateurs ont donc besoin d'une démarche fondée sur les valeurs qui donne la priorité aux intérêts de la population, d'une combinaison d'expertise technique et d'expertise axée sur les problèmes et d'une prise en compte globale du développement de la nouvelle IA. Il faut aussi garder à l'esprit les effets négatifs potentiels que peut avoir l'IA sur les changements climatiques en raison de la consommation d'énergie et d'eau.

5. L'Organe consultatif de haut niveau sur l'intelligence artificielle[2] a été créé pour analyser la gouvernance internationale de l'intelligence artificielle et formuler des recommandations. Dans le cadre de ce mandat, il se demande non seulement comment gouverner l'IA aujourd'hui, mais aussi comment préparer les institutions de

[2] On trouvera davantage d'informations sur l'Organe consultatif de haut niveau à l'adresse www.un.org/en/ai-advisory-body (non disponible en français).

gouvernance à un environnement dans lequel le rythme des changements ne fera que s'accélérer. La gouvernance de l'IA doit donc à la fois tenir compte des qualités de la technologie elle-même (c'est-à-dire être agile, en réseau et flexible) et de l'évolution rapide de ses utilisations, et être responsabilisante et inclusive, au bénéfice de toute l'humanité.

6. Les travaux de l'Organe consultatif ne se déroulent pas dans un vide normatif ou institutionnel. Les activités de l'Organisation des Nations Unies sont guidées par des règles et des principes que tous les États Membres s'engagent à respecter. Ces normes et valeurs partagées et codifiées constituent la base de tous les travaux menés par l'Organisation, et la gouvernance de l'IA ne fait pas exception. Les principes et normes établis du droit international, dont les obligations résultant de la Charte des Nations Unies et de la Déclaration universelle des droits de l'homme, ainsi que d'autres aspects du droit international, tels que le droit de l'environnement et le droit humanitaire international, s'appliquent à l'IA. Les institutions créées à l'appui d'objectifs multilatéraux allant de la paix et de la sécurité au développement durable ont un rôle à jouer pour saisir les possibilités qu'offre l'IA, tout en se protégeant des risques.

7. L'Organe consultatif partage néanmoins le sentiment d'urgence associé aux initiatives de gouvernance complémentaires relatives à l'IA, y compris celles menées par les États et les processus régionaux et intergouvernementaux, tels que l'Union européenne, le Groupe des Sept, le Groupe des Vingt, l'Organisation de coopération et de développement économiques et l'UNESCO. Un engagement plus inclusif est toutefois nécessaire, car de nombreuses populations – en particulier dans le monde du Sud ou dans la majorité mondiale – ont été largement absentes de ces débats, en dépit des effets potentiels sur leur vie. Il faut adopter une approche plus cohérente, plus inclusive, plus participative et mieux coordonnée, qui implique diverses populations à travers le monde, en particulier celles du Sud ou de la majorité mondiale.

8. L'Organisation des Nations Unies ne détient pas de panacée en ce qui concerne la question de la gouvernance de l'IA. Toutefois, la légitimité unique de l'Organisation, qui est la seule instance mondiale composée de tous les pays du monde, fondée sur les principes universellement reconnus de la Charte, et sa volonté d'accepter sans réserve la diversité de l'ensemble de la population mondiale en font un nœud central aux fins du partage des connaissances, de l'adoption de normes et de principes et de la garantie d'une bonne gouvernance et d'une responsabilité effective. Dans le système des Nations Unies, les projets relatifs au pacte numérique mondial et au Sommet de l'avenir, qui se tiendra en septembre 2024, ouvrent la voie à des mesures opportunes.

9. L'Organe consultatif réunit des personnes représentant un large éventail de domaines et sa composition tient compte de critères de diversité géographique, de genre et d'âge. Il s'appuie sur l'expertise des gouvernements, de la société civile, du secteur privé et du monde universitaire. Des discussions intenses et de grande envergure ont permis de dégager un large consensus sur l'absence d'une gouvernance mondiale en matière d'IA et sur le rôle que l'Organisation des Nations Unies a à jouer à cet égard.

10. Dans le présent rapport, l'Organe consultatif commence par recenser les possibilités et les catalyseurs qui peuvent contribuer à exploiter les bienfaits potentiels de l'IA pour l'humanité. Il met ensuite en évidence les risques et les problèmes que l'IA présente aujourd'hui et aussi loin qu'on puisse prévoir. Il poursuit en expliquant pourquoi des principes clairs, ainsi que des fonctions et des dispositions institutionnelles nouvelles, sont nécessaires pour combler le déficit de gouvernance mondiale. Il conclut par des recommandations préliminaires et la présentation des prochaines étapes, qui seront développées dans son rapport final, prévu pour août 2024.

11. L'Organe consultatif a confiance dans les orientations générales choisies, mais il ne fait pas cavalier seul pour autant. Il attend avec intérêt de procéder à une large consultation sur les prochaines étapes afin que davantage de voix et d'opinion soient prises en compte et que l'IA serve le bien commun.

II. Un manque de gouvernance mondiale

12. L'IA est en train de transformer notre monde, mais son développement et ses retombées sont actuellement concentrés entre les mains d'un petit nombre d'acteurs du secteur privé situés dans un nombre d'États encore plus restreint. Les dommages associés sont aussi répartis de manière inégale. Une gouvernance mondiale, à laquelle tous les États Membres participent de façon égale, est nécessaire pour que les ressources soient accessibles, que les mécanismes de représentation et de contrôle soient largement représentatifs, qu'il y ait une obligation de rendre des comptes en cas de dommage et que la concurrence géopolitique n'entraîne pas une utilisation irresponsable ou n'entrave pas une gouvernance responsable.

13. L'Organisation des Nations Unies est au centre de l'ordre international reposant sur des règles. Sa légitimité tient au fait qu'il s'agit d'un véritable forum mondial fondé sur le droit international au service de la paix et de la sécurité, des droits humains et du développement durable. L'Organisation offre donc un socle institutionnel et normatif pour une action collective en matière de gouvernance de l'IA. Outre les considérations relatives à l'équité, à l'accès et à la prévention des dommages, la nature même de la technologie nécessite une approche globale, les systèmes d'IA étant transfrontaliers de par leur structure, leur fonctionnement et leur application et utilisés par un large éventail d'acteurs.

14. Les initiatives d'autorégulation, les législations nationales et régionales et les travaux des forums multilatéraux viennent compléter ce puzzle. Des lacunes subsistent cependant et le défi à relever est clair : il faut mettre en place un cadre de gouvernance mondial pour réglementer cet ensemble de technologies qui se développent rapidement ainsi que leur utilisation par différents acteurs, qu'il s'agisse de développeurs ou d'utilisateurs. L'Organisation des Nations Unies est particulièrement bien placée pour relever les défis et réaliser les promesses que présente l'IA, par la transformation d'une mosaïque d'initiatives évolutives en un ensemble caractérisé par la cohérence et l'interopérabilité, fondé sur les valeurs universelles adoptées par ses États Membres et adaptable à tous les contextes.

15. Dans les chapitres suivants, l'Organe consultatif décrit les rôles qu'une institution ou un réseau d'institutions ancrées dans le cadre universel de l'Organisation des Nations Unies pourraient jouer pour accroître les bienfaits et atténuer les risques créés par l'IA, ainsi que les principes et les fonctions qui permettraient d'atteindre au mieux ces objectifs.

III. Des débouchés et des catalyseurs

16. L'IA a le potentiel de transformer l'accès à la connaissance et d'accroître l'efficacité au niveau mondial : une nouvelle génération d'innovateurs est en train de repousser les frontières de la science et de l'ingénierie de l'IA. Celle-ci accroît la productivité et l'innovation dans des secteurs allant des soins de santé à l'agriculture, dans les économies avancées comme dans les économies en développement (voir encadré 2). Dans le même temps, des questions se posent quant aux instruments à utiliser pour que les bienfaits soient répartis équitablement et en toute sécurité dans l'ensemble de l'humanité et que les effets perturbateurs, notamment sur l'emploi, soient pris en compte et limités. Une question importante pour les décideurs politiques

est de savoir comment développer des écosystèmes d'IA performants dans le monde entier, tout en responsabilisant les acteurs établis et émergents.

Encadré 2
Exemples de possibilités offertes par l'intelligence artificielle

Systèmes d'assistance aux personnes

L'IA peut aider les personnes à accomplir leurs tâches quotidiennes et leurs efforts les plus ambitieux, qu'ils soient créatifs ou productifs. Les systèmes d'assistance aux personnes comprennent des outils d'accessibilité et des améliorations en matière de formation. Des applications ont été développées pour servir d'assistants virtuels aux personnes ayant une vision ou une élocution limitées, et répondent ainsi à des besoins d'accessibilité ignorés ou négligés jusqu'alors. La traduction assistée par l'intelligence artificielle, qui couvre désormais plus de 100 langues, peut favoriser l'accès aux services en ligne et hors ligne, ainsi que la compréhension et la communication interculturelles. Une nouvelle génération d'applications de tutorat promet d'élargir l'accès à un enseignement de qualité dans le monde entier.

Perspectives sectorielles

L'IA aura des répercussions plus importantes dans certains secteurs que dans d'autres. Les domaines les plus prometteurs sont l'agriculture et la sécurité alimentaire, la santé, l'éducation, la protection de l'environnement, la résilience aux catastrophes naturelles et la lutte contre les changements climatiques. L'IA a ainsi été utilisée pour créer des dispositifs d'alerte rapide pour les inondations (qui couvrent maintenant plus de 80 pays), les feux incontrôlés et l'insécurité alimentaire, mais aussi pour surveiller les espèces menacées d'extinction (dauphins, baleines, etc.) et pour optimiser les pratiques agricoles. Il existe une myriade de possibilités dans chaque domaine.

L'IA élargit l'accès à des soins de qualité, par exemple dans le domaine des soins de santé maternelle en Afrique subsaharienne. Elle peut aussi contribuer à lutter contre les problèmes environnementaux, à rendre l'éducation plus accessible, à réduire la pauvreté et la faim et à améliorer la sécurité dans les villes.

Perspectives scientifiques

L'IA transforme la manière dont la recherche scientifique est menée et repousse les frontières du progrès scientifique, notamment en accélérant la recherche moléculaire et génomique. Les systèmes d'IA sont particulièrement prometteurs en ce qui concerne l'accélération du travail des scientifiques dans de nombreuses disciplines et le changement de paradigme potentiel dans la pratique de la science, par exemple en aidant à explorer de nouveaux espaces de découverte et en automatisant l'expérimentation à grande échelle. Par exemple, les outils alimentés par l'IA qui prédisent les structures des protéines sont utilisés par plus d'un million de chercheurs pour découvrir des médicaments et faire progresser la compréhension de maladies telles que la tuberculose et de nombreuses maladies auparavant négligées. Dans le domaine des soins de santé, l'IA alimente des outils de diagnostic qui aident les médecins à détecter plus rapidement différents types de cancers et de maladies oculaires et à sauver

des vies. Dans le secteur de l'énergie, l'IA participe à l'optimisation des systèmes énergétiques et à la transition vers des énergies renouvelables. Elle a, par exemple, été utilisée pour augmenter la valeur de l'énergie éolienne, contrôler les plasmas des tokamaks dans la fusion nucléaire et permettre la séquestration du carbone. L'Organisation des Nations Unies peut encourager les progrès de la science fondée sur l'IA en attirant l'attention sur les problèmes qui doivent être résolus pour le bien de toutes et de tous.

Perspectives pour le secteur public

Un élément essentiel est que l'IA peut favoriser les progrès dans des domaines dans lesquels les forces du marché seules ont traditionnellement échoué, qu'il s'agisse des prévisions de phénomènes météorologiques extrêmes, de la surveillance de la biodiversité, de l'élargissement des possibilités d'éducation, de l'accès à des soins de santé de qualité ou de l'optimisation des systèmes énergétiques. Les gouvernements et le secteur public peuvent améliorer les services aux citoyens et renforcer la prestation de services aux populations vulnérables en mettant l'IA au service du bien social.

Comment l'Organisation des Nations Unies peut exploiter l'intelligence artificielle

Enfin, l'utilisation de l'IA pourrait contribuer à accélérer la réalisation des objectifs de développement durable et renforcer le rôle et l'efficacité de l'Organisation des Nations Unies dans la promotion du développement durable, des droits humains, de la paix et de la sécurité. L'Organisation pourrait par exemple l'utiliser pour suivre l'évolution des situations de crise telles que les atteintes aux droits humains dans le monde entier ou pour mesurer les progrès accomplis dans la réalisation des objectifs de développement durable. Le fait que l'IA puisse contribuer aux efforts déployés pour atteindre un grand nombre des 17 objectifs a été mis en exergue, tout comme les obstacles importants qui empêchent d'exploiter pleinement son potentiel à cette fin. L'Organisation des Nations Unies et d'autres organisations internationales ont commencé à élaborer des cas d'utilisation de l'IA et des exemples prometteurs dans des domaines tels que la prévision de l'insécurité alimentaire, la gestion des interventions de secours et les prévisions météorologiques.

A. Principaux éléments permettant de mettre l'intelligence artificielle au service de l'humanité

17. Le développement de l'IA repose désormais sur les données, la puissance de calcul et le talent, parfois complétés par un travail manuel d'annotation des données aux fins de l'apprentissage automatique. À l'heure actuelle, seuls les États Membres disposant de ressources importantes et les géants du secteur des technologies ont accès aux trois premiers éléments, ce qui crée une concentration de l'influence. À la pénurie mondiale de matériel essentiel, tel que les processeurs graphiques, s'ajoute une pénurie de talents techniques de haut niveau dans le domaine de l'IA. Il a été suggéré que le développement de modèles ouverts pourrait changer cette dynamique, même si les effets et la sécurité des modèles ouverts font encore l'objet d'analyses et de débats.

18. L'IA et les possibilités qu'elle recèle arrivent dans un contexte difficile, en particulier pour le monde du Sud, la « fracture de l'IA » s'inscrivant dans une fracture numérique et de développement bien plus large. Selon les estimations de l'UIT pour 2023, plus de 2,6 milliards de personnes n'ont toujours pas accès à Internet. Les socles de l'économie numérique (accès à large bande, appareils et données abordables, habileté numérique, approvisionnement en électricité fiable et abordable) font défaut. La marge de manœuvre budgétaire est limitée et l'environnement international des flux commerciaux et des flux d'investissement n'est pas facile. Il sera crucial d'investir dans des infrastructures de base telles que la large bande et l'électricité, sans lesquelles la capacité de participer au développement et à l'utilisation de l'IA sera gravement limitée. Même en dehors du monde du Sud, les efforts visant à tirer parti de l'IA nécessiteront le développement d'écosystèmes d'IA locaux, la capacité d'entraîner des modèles locaux à partir de données locales et l'adaptation de modèles développés ailleurs aux situations et aux objectifs locaux.

19. L'accès et les bienfaits doivent être indissociables. Les entrepreneurs situés dans les régions qui accusent un retard en matière de capacité d'IA ont besoin et méritent de pouvoir créer leurs propres solutions. Il faut donc investir au niveau national dans les talents, les données et les ressources de calcul, ainsi que dans les capacités nationales en matière de réglementation et d'achats. Ces efforts nationaux doivent être complétés par une assistance et une coopération internationales entre les gouvernements, mais aussi entre les acteurs du secteur privé. Réunir des scientifiques pour résoudre des problèmes sociétaux pourrait constituer l'un des principaux moyens de mettre le potentiel de l'IA au service de l'humanité. Les solutions libres de droits et le partage de données et de modèles pourraient jouer un rôle important dans la diffusion de ces bienfaits et le développement de chaînes de valeur bénéfiques en matière de données et d'IA, au-delà des frontières.

20. Les facilitateurs, ou « voies communes », pour le développement, le déploiement et l'utilisation de l'IA devraient être contrebalancés par des « voies de sécurité » afin de prendre en charge les effets sur les sociétés et les populations. Déterminer dans quelle mesure les efforts de gouvernance de l'IA ont fait progresser l'humain plutôt que de le remplacer ou de l'aliéner constituera un critère décisif. Une partie du développement de l'IA repose sur une main-d'œuvre bon marché et exploitable dans le monde du Sud, tandis que, dans le monde du Nord, des questions se posent quant à la valorisation de l'expression artistique, de la propriété intellectuelle et de la dignité du travail humain. Un accès équitable à ces technologies et les compétences nécessaires pour les utiliser pleinement sont indispensables si nous voulons éviter les « fractures de l'IA » au sein des nations et entre elles.

B. La gouvernance, un facteur essentiel

21. L'IA peut et doit être déployée à l'appui de la réalisation des objectifs de développement durable. Toutefois, cela ne peut et ne doit pas dépendre uniquement des pratiques actuelles du marché ou de la bienveillance d'une poignée d'entreprises du secteur des technologies. Tout dispositif de gouvernance devrait prévoir des mesures d'incitation à l'échelle mondiale pour promouvoir la réalisation de ces objectifs plus vastes et plus largement représentatifs et contribuer à détecter et à prendre en compte les compromis.

22. À cet égard, les comparaisons avec d'autres secteurs peuvent apporter des enseignements. Des mécanismes tels que Gavi, l'Alliance du Vaccin, peuvent fournir des exemples à court terme montrant comment faire en sorte que les bienfaits de l'IA sont partagés. Des référentiels de modèles d'IA adaptables à différents contextes

pourraient servir d'équivalent aux médicaments génériques, afin d'élargir l'accès sans favoriser la concentration ou la consolidation de l'IA.

23. Certaines de ces aspirations bénéfiques pour la société peuvent être réalisées grâce aux progrès de la recherche sur l'IA elle-même ; d'autres peuvent être abordées en tirant parti de nouveaux mécanismes de marché pour uniformiser les règles du jeu ou en incitant les acteurs à atteindre toutes les populations et à permettre à toutes et à tous d'accéder à ces bienfaits. Mais beaucoup resteront lettre morte. Pour que l'IA soit mise au service du bien commun et que ses bienfaits soient répartis équitablement, il faudra que les gouvernements et les organisations intergouvernementales agissent, notamment en promouvant des méthodes innovantes pour encourager la participation du secteur privé, des universités et de la société civile. Une solution plus durable consisterait à permettre un accès fédéré aux éléments fondamentaux des données, à la puissance de calcul et aux talents qui alimentent l'IA, ainsi qu'à l'infrastructure numérique et à l'électricité, le cas échéant. À cet égard, le CERN, qui exploite le plus grand laboratoire de physique des particules au monde, et d'autres collaborations scientifiques internationales similaires peuvent offrir des enseignements utiles. Un « CERN distribué » réimaginé pour l'IA et organisé en réseau entre différents États et régions pourrait élargir les possibilités de participation. Le Laboratoire européen de biologie moléculaire et ITER, le réacteur thermonucléaire expérimental international, sont d'autres exemples de science ouverte applicables à l'IA.

IV. Des risques et des difficultés

24. Il faut non seulement offrir un accès équitable aux possibilités créées par l'IA, mais aussi redoubler d'efforts pour faire face aux dangers connus, inconnus et encore impossibles à connaître. Aujourd'hui, des systèmes de plus en plus puissants sont déployés et utilisés en l'absence de nouvelles réglementations, l'objectif étant d'apporter des avantages et de gagner de l'argent. Les systèmes d'IA peuvent créer de la discrimination sur la base de la race ou du genre. L'utilisation généralisée des systèmes actuels peut menacer la diversité linguistique. De nouvelles méthodes de désinformation et de manipulation menacent les processus politiques, y compris les processus démocratiques. Les utilisateurs malveillants et bienveillants de l'IA jouent au chat et à la souris dans le contexte de la cybersécurité et de la cyberdéfense.

A. Risques posés par l'intelligence artificielle

25. L'Organe consultatif a examiné les risques liés à l'IA d'abord sous l'angle des caractéristiques techniques de celle-ci, puis sous celui de l'utilisation inappropriée, en tenant compte du double usage, et de considérations plus larges relatives à l'interaction homme-machine. Enfin, l'Organe consultatif a examiné les risques sous l'angle de la vulnérabilité.

26. Certains risques découlent des limites techniques des systèmes d'IA ; il s'agit, entre autres, des préjugés nuisibles et de divers dangers liés à l'information, tels que le manque de précision et les « hallucinations » ou confabulations, qui sont des problèmes connus de l'IA générative.

27. D'autres risques relèvent davantage des humains que de l'IA. Les hypertrucages (*deepfakes*) et les campagnes d'information hostiles ne sont que le dernier exemple en date de technologies déployées à des fins malveillantes. Ils peuvent poser des risques graves pour la confiance de la société et le débat démocratique.

28. Une autre catégorie de risques est liée à l'interaction homme-machine. Au niveau des personnes, il s'agit notamment d'une confiance excessive dans les

systèmes d'IA (biais d'automatisation) et d'une perte potentielle de compétences au fil du temps. Au niveau des sociétés, elle comprend les effets sur les marchés du travail si de larges pans de la main-d'œuvre sont déplacés ou sur la créativité si les droits de propriété intellectuelle ne sont pas protégés. On ne peut pas non plus exclure des changements dans la manière dont nous nous comportons les uns envers les autres en tant qu'humains, étant donné que de plus en plus d'interactions se font par l'intermédiaire de l'IA. Ces changements peuvent avoir des conséquences imprévisibles sur la vie de famille et le bien-être physique et émotionnel.

29. Une autre catégorie de risques concerne les questions de sécurité de plus grande ampleur. Par exemple, le débat se poursuit sur les lignes rouges qui pourraient être fixées, notamment dans le contexte des systèmes d'armes létaux autonomes ou de la militarisation plus large de l'IA. Il existe des preuves crédibles attestant de l'utilisation croissante de systèmes dotés d'intelligences artificielles et de fonctions autonomes sur les champs de bataille. Il se pourrait qu'une nouvelle course aux armements soit en route : elle aurait des conséquences sur la stabilité mondiale et sur la définition de ce qui constitue le seuil d'un conflit armé. Le fait que des machines puissent cibler et blesser des êtres humains de manière autonome est l'une de ces « lignes rouges » qui ne doivent pas être franchies. Dans de nombreux pays, l'utilisation de l'IA par les services chargés de l'application de la loi, en particulier la surveillance biométrique en temps réel, est considérée comme un risque inacceptable, portant atteinte au droit à la vie privée. L'idée qu'une IA puisse devenir incontrôlable ou incontrôlée suscite aussi des inquiétudes, telle que la possibilité qu'elle menace l'humanité (même s'il y a des débats sur la question de savoir s'il faut évaluer de telles menaces et comment le faire).

30. Dresser une liste exhaustive et définitive des risques liés à l'IA est une perte de temps. Compte tenu de l'omniprésence et de l'évolution rapide de l'IA et de son utilisation, l'Organe consultatif estime qu'il est plus utile d'examiner les risques du point de vue des populations vulnérables et des biens communs. En suivant cette approche, il a tenté une première catégorisation (voir encadré 3), qui sera développée dans un cadre d'évaluation des risques, sur la base des efforts existants. Les risques évolueront au fur et à mesure des progrès de la technologie, de son adoption et de son utilisation. Il faut donc surveiller les risques au moyen de la science interdisciplinaire et d'approches fondées sur des données probantes. Des cadres de gestion des risques adaptables, pouvant être ajustés en fonction de l'expérience des différentes régions à différents moments, seraient aussi nécessaires. L'Organisation des Nations Unies peut offrir un espace d'une grande utilité pour un tel apprentissage mutuel et une adaptation agile.

Encadré 3
Catégorisation des risques liés à l'intelligence artificielle en fonction de la vulnérabilité existante ou potentielle

- Personnes
  - Dignité humaine, valeur ou moyens d'action (manipulation, tromperie, incitation, condamnation)
  - Vie, sécurité, sûreté (armes létales autonomes ; voitures autonomes ; interaction avec des systèmes de défense chimique, biologique, radiologique ou nucléaire)
  - Intégrité physique et mentale, santé et sécurité (diagnostics, incitation, neurotechnologie)

- Droits humains et libertés civiles (autres), tels que le droit à un procès équitable (prédiction de la récidive) et les droits à la présomption d'innocence (police prédictive), à la liberté d'expression (incitation) et à la vie privée (reconnaissance biométrique)
- Chances à saisir dans la vie (éducation, emploi, stabilité financière)

• Groupes

- Discrimination et traitement inéquitable de sous-groupes, y compris sur la base du genre
- Isolement et marginalisation du groupe
- Fonctionnement d'un groupe
- Égalité et équité sociales (traitement inéquitable de groupes, y compris sur la base du genre)
- Enfants, personnes âgées et personnes handicapées

• Société

- Sécurité internationale et nationale (armes autonomes, désinformation)
- Démocratie (élections, confiance)
- Intégrité de l'information (mésinformation ou désinformation, hypertrucages, nouvelles personnalisées)
- État de droit (fonctionnement des institutions et du système judiciaire et confiance à leur égard)
- Sécurité (utilisations militaires et policières)
- Diversité culturelle et évolution des relations humaines (homogénéité, faux amis)
- Cohésion sociale (bulles de filtres, baisse de la confiance dans les nouvelles et les sources d'information)

• Économie

- Concentration du pouvoir
- Dépendance technologique
- Débouchés économiques inégaux
- Distribution et répartition des ressources
- Sous-utilisation ou surutilisation de l'IA, « solutionnisme technique »

• (Éco)systèmes

- Stabilité des systèmes financiers
- Risques pour les infrastructures critiques
- Pression sur l'environnement, le climat et les ressources naturelles

• Valeurs et normes

- Valeurs éthiques

- Valeurs morales
- Valeurs sociales
- Valeurs culturelles
- Normes juridiques

31. Pour l'instant, il n'y a pas de consensus sur la manière d'évaluer ou de traiter les risques susmentionnés. Néanmoins, il en va de même qu'avec le principe de précaution dans le droit de l'environnement : l'incertitude scientifique relative aux risques ne doit pas conduire à une paralysie de la gouvernance. Pour parvenir à un consensus et agir en conséquence, il faut instaurer une coopération et une coordination à l'échelle mondiale, notamment au moyen de mécanismes communs de surveillance des risques. Les organisations internationales ont des décennies d'expérience utile en matière de technologies à double usage (armes chimiques et biologiques, énergie nucléaire), fondée sur le droit des traités et d'autres cadres normatifs, qui pourrait être appliquée pour gérer les risques créés par l'IA.

32. L'Organe consultatif sait aussi qu'il faut être proactif. Des leçons importantes doivent être tirées des expériences faites récemment avec d'autres technologies à grande échelle et à fort impact, telles que les médias sociaux. Alors même que diverses sociétés étudient l'impact et les répercussions de l'IA, il apparaît clairement qu'une gouvernance mondiale efficace permettant de partager les préoccupations et de coordonner les solutions est nécessaire.

33. Il faut mettre au jour, classer et traiter les risques liés à l'IA, notamment en trouvant un consensus sur les risques inacceptables et sur la manière de les prévenir ou de les anticiper. La vigilance et les analyses prospectives sont de mise pour ce qui est des conséquences inattendues de l'IA, ces systèmes étant introduits dans des contextes de plus en plus divers, sans tests préalables. Des difficultés techniques, politiques et sociales doivent être résolues pour parvenir à une compréhension commune des risques.

B. Difficultés à résoudre

34. De nombreux systèmes d'IA sont opaques, soit en raison de leur complexité intrinsèque, soit en raison du secret commercial qui entoure leur fonctionnement interne. Les chercheurs et les organes de gouvernance ont du mal à accéder aux informations ou à interroger pleinement les jeux de données, les modèles et les systèmes protégés par des droits exclusifs. De plus, cette science n'en est qu'à ses balbutiements et on ne comprend pas encore bien comment fonctionnent les systèmes avancés. Ce manque de transparence est associé à un accès inégal, notamment aux ressources de calcul et autres, et à une compréhension insuffisante, et empêche donc de repérer l'origine des risques et d'établir la responsabilité de la prise en charge de ces risques (ou de la compensation en cas de dommages).

35. Malgré la portée mondiale de l'IA, la gouvernance demeure territoriale et fragmentée. En matière de réglementation, les approches nationales s'arrêtent généralement aux frontières physiques et peuvent entraîner des tensions ou des conflits si l'IA ne respecte pas ces frontières. Des efforts d'autoréglementation, de réglementation nationale et de gouvernance internationale seront nécessaires pour détecter, prévenir et atténuer les risques. L'obligation de rendre compte ne devrait pas faire défaut.

36. Il est important de tenir compte du stade de développement et des contextes propres aux États Membres afin de fournir une assistance adaptée, en gardant à l'esprit les contraintes qui pèsent sur eux en matière de participation et d'adhésion à la gouvernance de l'IA. Cette approche est préférable à celle qui consiste à leur dire à quel stade ils devraient se trouver et ce qu'ils devraient faire sur la base de situations qui ne s'appliquent pas à eux.

37. Outre les obstacles techniques et politiques, on constate aussi des difficultés dans un contexte social plus large. Les technologies numériques affectent le « logiciel » (ou mœurs sociales) des sociétés, remettant en cause la gouvernance dans tous les domaines. De plus, les coûts humains et environnementaux de l'IA (qu'il s'agisse des coûts matériels ou logiciels) doivent être pris en compte tout au long de son cycle de vie, dans la mesure où les vies humaines et l'environnement sont des éléments communs au début et à la fin de tous les processus intégrés à l'IA.

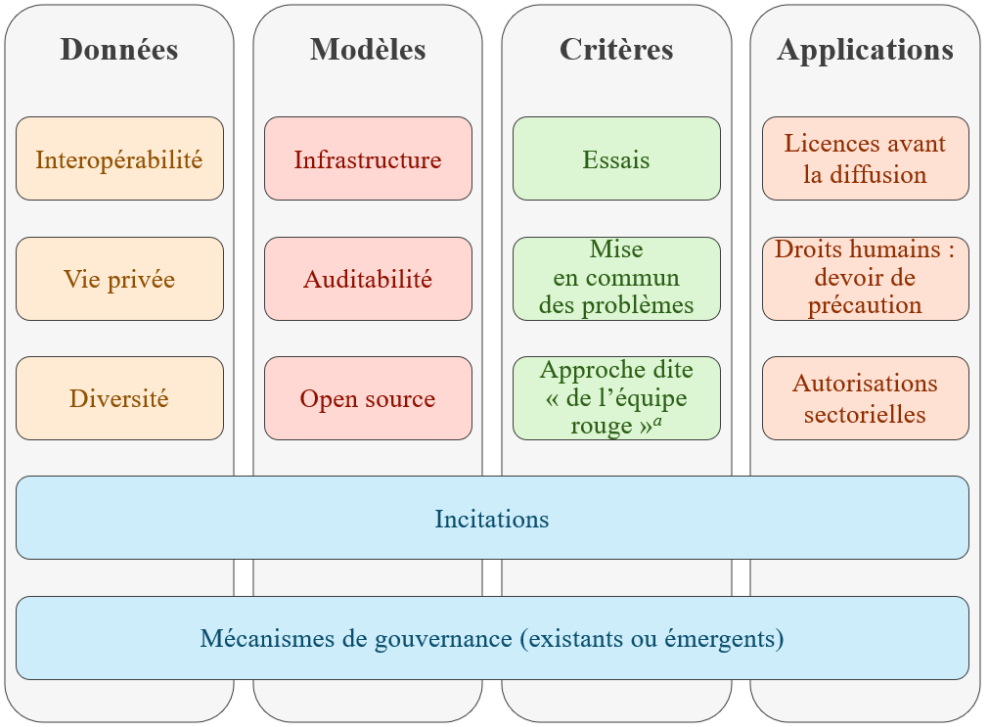
38. Aux inquiétudes liées à l'utilisation abusive s'ajoutent les préoccupations liées aux utilisations manquées, c'est-à-dire le fait de ne pas exploiter et partager les bienfaits de l'IA en raison d'une prudence excessive. L'utilisation de l'IA aux fins de l'amélioration de l'accès à l'éducation pourrait susciter des inquiétudes quant à la confidentialité des données des jeunes et au rôle des enseignants. Cependant, dans un monde où des centaines de millions d'élèves ne peuvent accéder à des ressources éducatives de qualité, ne pas utiliser la technologie pour combler les lacunes peut avoir des inconvénients. Les efforts déployés pour s'accorder sur ces arbitrages et en tenir compte bénéficieront de mécanismes de gouvernance internationale permettant le partage d'informations, la mise en commun de ressources et l'adoption de stratégies communes.

V. Une gouvernance internationale de l'intelligence artificielle

A. Situation en matière de gouvernance de l'intelligence artificielle

39. Les guides, cadres et principes portant sur la gouvernance de l'IA ne manquent pas aujourd'hui. Des documents ont été rédigés par le secteur privé et la société civile, ainsi que par des organismes nationaux, régionaux et multilatéraux, et ont eu des effets variables. Sur le plan technologique, les efforts de gouvernance ont été axés sur les données, les modèles et les critères de référence ou les évaluations. Les cas d'utilisation ont aussi fait l'objet d'une attention particulière, notamment lorsque des accords de gouvernance sectorielle ont été conclus, comme dans les domaines de la santé ou des technologies à double usage. Ces efforts peuvent être rattachés à des dispositifs de gouvernance particuliers, tels que la loi sur l'intelligence artificielle adoptée par l'Union européenne et le décret nº 14110 promulgué par les États-Unis d'Amérique sur le développement et l'utilisation sûrs, sécurisés et fiables de l'IA, et associés à des mesures d'incitation à la participation et au respect des règles. On trouvera dans la figure 1 ci-dessous un schéma simplifié passant en revue le paysage qui se dessine en matière de gouvernance de l'IA, que l'Organe consultatif approfondira au cours de la prochaine phase de ses travaux.

Figure 1
Interopérabilité entre les différents efforts de gouvernance de l'intelligence artificielle : schéma triple simplifié



[a] L'approche dite « de l'équipe rouge » (Red Teaming) consiste à remettre rigoureusement en question les plans, les politiques, les systèmes et les hypothèses en adoptant une démarche contradictoire.

40. Les efforts existants en matière de gouvernance de l'IA ont débouché sur des similitudes de langage, telles que l'importance de l'équité, de la responsabilité et de la transparence. Cependant, il n'y a pas d'harmonisation d'ensemble en ce qui concerne la mise en œuvre, qu'il s'agisse de l'interopérabilité entre les juridictions ou des incitations à la conformité dans les juridictions. Certains sont favorables à l'adoption de règles contraignantes, tandis que d'autres préfèrent des incitations non contraignantes. Des arbitrages, portant notamment sur la manière d'équilibrer l'accès et la sécurité et sur la question de savoir si l'accent doit être mis sur les dommages actuels ou sur ceux qui pourraient se produire, font l'objet de débats. Des modèles différents peuvent aussi nécessiter des cadres de gouvernance différents. L'absence de normes et de références communes entre les cadres nationaux et multinationaux de gestion des risques ainsi que les multiples définitions de l'IA utilisées dans ces cadres ont rendu l'environnement de gouvernance de l'IA plus complexe. Dans le même temps, il est nécessaire de disposer d'un espace dans lequel différentes approches réglementaires peuvent coexister et qui refléterait ainsi la diversité sociale et culturelle du monde.

41. Dans le même temps, les progrès techniques de l'IA et son utilisation continuent de s'accélérer, creusant le fossé en termes de compréhension et de capacité entre les entreprises technologiques qui la développent, les entreprises et autres organisations qui l'utilisent dans divers secteurs et espaces sociétaux, et les personnes qui souhaitent réguler son développement, son déploiement et son usage.

42. Il en résulte que, dans de nombreuses juridictions, la gouvernance de l'IA peut se résumer à une autosurveillance de la part des développeurs, des personnes chargées du déploiement et des utilisateurs de ces systèmes. Même en supposant que ces organisations et ces personnes agissent de bonne foi, une telle situation n'encourage pas une vision à long terme du risque ou la participation de diverses parties prenantes, en particulier celles du monde du Sud. Cette situation doit changer.

B. Vers des principes et des fonctions de gouvernance internationale de l'intelligence artificielle

43. L'Organe consultatif est chargé de présenter des choix pour la gouvernance internationale de l'IA. Il a examiné, entre autres, les fonctions exercées par les institutions de gouvernance existantes qui ont une dimension technologique, dont le Groupe d'action financière, le Conseil de stabilité financière, l'AIEA, la Société pour l'attribution des noms de domaine et des numéros sur Internet, l'OACI, l'OIT, l'OMI, le Groupe d'experts intergouvernemental sur l'évolution du climat, l'UIT, la Société de télécommunications interbancaires mondiales et le Bureau des affaires spatiales de l'Organisation des Nations Unies. Elles constituent une source d'inspiration et servent d'exemples de gouvernance et de coordination mondiales.

44. La variété des parties prenantes et des applications potentielles de l'IA et son utilisation dans des contextes très variés font qu'il ne serait pas approprié de reproduire un modèle de gouvernance existant. On peut néanmoins tirer des enseignements des exemples d'entités qui ont cherché à : a) établir un consensus scientifique sur les risques, les répercussions et la politique (Groupe d'experts intergouvernemental sur l'évolution du climat) ; b) établir des normes mondiales, les modifier et les adapter (OACI, OMI, UIT) ; c) assurer le renforcement des capacités, l'assurance mutuelle et la surveillance (AIEA, OACI) ; d) mettre en réseau et en commun les ressources de la recherche (CERN) ; e) mobiliser des parties prenantes diverses (OIT, Société pour l'attribution des noms de domaine et des numéros sur Internet) ; f) faciliter les flux commerciaux et traiter les risques systémiques (Société de télécommunications interbancaires mondiales, Groupe d'action financière, Conseil de stabilité financière).

45. Plutôt que de proposer un modèle unique de gouvernance de l'IA à ce stade, les recommandations préliminaires formulées dans le présent rapport d'étape portent principalement sur les principes qui devraient guider la création de nouvelles institutions mondiales de gouvernance de l'IA et sur les fonctions générales que ces institutions devraient remplir. Les sous-fonctions énumérées dans le tableau ci-dessous découlent d'un examen de la recherche menée sur la gouvernance de l'IA et d'une analyse des lacunes de neuf initiatives actuelles en matière de gouvernance de l'IA, à savoir les mesures conservatoires prises par la Chine pour la gestion des services d'IA, le projet de convention sur l'IA du Conseil de l'Europe, la loi sur l'intelligence artificielle adoptée par l'Union européenne, le processus d'Hiroshima du Groupe des Sept, le Partenariat mondial sur l'intelligence artificielle, les principes sur l'IA de l'Organisation de coopération et de développement économiques, le Partnership on AI, le Foundation Model Forum, le Sommet sur la sécurité de l'IA du Royaume-Uni et le décret n° 14110 promulgué par les États-Unis.

C. Recommandations préliminaires

1. Principes directeurs

a) Principe directeur nº 1 : l'intelligence artificielle doit être gouvernée de manière inclusive, par toutes et par tous et dans l'intérêt commun

46. Malgré le potentiel de l'IA, une grande partie de la population mondiale ne peut encore y accéder et l'utiliser de manière à améliorer significativement le quotidien. Il est essentiel d'exploiter pleinement ce potentiel et de permettre une large participation au développement, au déploiement et à l'utilisation de l'IA pour trouver des solutions durables aux problèmes mondiaux. Tous les citoyens, y compris ceux du monde du Sud, devraient pouvoir créer leurs propres débouchés, les exploiter et atteindre la prospérité grâce à l'IA. Tous les pays, quelle que soit leur taille, doivent pouvoir participer à la gouvernance de l'IA.

47. Il faudra prendre des mesures positives et correctives, notamment en matière d'accès et de renforcement des capacités, pour remédier à l'exclusion historique et structurelle de certaines populations (telles que les femmes et les acteurs de la diversité de genre) du développement, du déploiement, de l'utilisation et de la gouvernance des technologies, et pour transformer les fractures numériques en opportunités numériques associant toutes les parties.

b) Principe directeur nº 2 : l'intelligence artificielle doit être gouvernée dans l'intérêt général

48. Le développement des systèmes d'IA est largement concentré entre les mains des entreprises du secteur des technologies. D'autres acteurs participeront au perfectionnement, au déploiement et à l'utilisation de l'IA, y compris, entre autres, les développeurs initiaux (qu'il s'agisse d'entreprises, de petits laboratoires, d'autres organisations ou même de pays), les personnes chargées du déploiement et les utilisateurs, qui seront aussi bien des particuliers que des entreprises, des organisations et des gouvernements, et qui apporteront une grande variété d'incitations à leurs approches.

49. Comme l'a montré l'expérience des médias sociaux, les produits et services d'IA peuvent être rapidement déployés à grande échelle, au-delà des frontières et à l'intention de toutes les catégories d'utilisateurs. Cette réalité ainsi que des considérations plus larges liées aux débouchés et aux risques expliquent pourquoi l'IA doit être gouvernée dans l'intérêt général. Le principe « ne pas nuire » doit certes être respecté, mais cela ne suffit pas. Un cadre plus large est nécessaire pour responsabiliser les entreprises et autres organisations qui construisent, déploient et contrôlent l'IA, ainsi que celles qui l'utilisent dans de multiples secteurs de l'économie et de la société tout au long de son cycle de vie. On ne peut pas se fier uniquement à l'autorégulation ; il faut instaurer des normes contraignantes appliquées de manière cohérente par les États Membres pour garantir que c'est l'intérêt public qui prévaudra, et non les intérêts privés.

50. L'IA sera utilisée dans de multiples secteurs par des personnes et des organisations, qui auront chacune des cas d'utilisation, des difficultés et des risques différents. Dans les activités de gouvernance, il faut garder à l'esprit les objectifs de politique publique relatifs à la diversité, à l'équité, à l'inclusion, à la durabilité, au bien-être sociétal et individuel, aux marchés compétitifs et aux écosystèmes de l'innovation saine. Les incidences des utilisations manquées sur le développement économique et social devront aussi être prises en compte. Dans ce contexte, la gouvernance devrait élargir la représentation des diverses parties prenantes et offrir une plus grande clarté pour ce qui est de délimiter les responsabilités entre les acteurs

du secteur public et du secteur privé. Gouverner dans l'intérêt public nécessite aussi des investissements dans la technologie publique, l'infrastructure et les capacités des agents publics.

c) Principe directeur nº 3 : la gouvernance de l'intelligence artificielle doit être articulée avec la gouvernance des données et la promotion des biens communs en matière de données

51. Les données sont essentielles pour de nombreux systèmes d'intelligence artificielle. La gouvernance et la gestion des données dans l'intérêt public ne peuvent être dissociées des autres composantes de la gouvernance de l'IA (voir figure 1 ci-dessus). Les cadres réglementaires et les dispositions technicojuridiques qui protègent la confidentialité et la sécurité des données à caractère personnel, dans le respect des lois applicables, tout en facilitant activement l'utilisation de ces données, constitueront un complément essentiel aux dispositions de gouvernance de l'IA respectant le droit local ou régional. Il convient également d'encourager le développement de biens communs en lien avec les données publiques, en accordant une attention particulière aux données publiques (destinées à être utilisées par de multiples parties prenantes) qui sont essentielles pour aider à surmonter les obstacles sociétaux tels que les changements climatiques, la santé publique, le développement économique, le renforcement des capacités et la résolution des crises.

d) Principe directeur nº 4 : la gouvernance de l'intelligence artificielle doit être universelle, en réseau et ancrée dans une collaboration multipartite adaptative

52. Tout effort de gouvernance de l'IA devrait avoir pour priorité d'assurer l'adhésion universelle des différents États Membres et parties prenantes. Il devrait aussi garantir une participation inclusive, notamment par l'abaissement des barrières à l'entrée pour les populations précédemment exclues dans le monde du Sud (voir principe directeur nº 1, paragraphes 46 et 47, ci-dessus). L'adoption d'une telle approche est essentielle pour que les nouvelles réglementations en matière d'IA soient harmonisées de manière à éviter les lacunes en matière de responsabilité.

53. Une gouvernance efficace doit s'appuyer sur les institutions existantes, qui devront revoir leurs fonctions actuelles en tenant compte des incidences de l'IA. Cependant, le financement à lui seul ne suffira pas. De nouvelles fonctions de coordination horizontale et de supervision sont aussi nécessaires et devraient être confiées à une nouvelle structure organisationnelle. Les institutions nouvelles et existantes pourraient former des nœuds dans un réseau de structures de gouvernance. Une dynamique claire s'est installée dans différents États pour obtenir ce résultat et le secteur privé est de plus en plus conscient de la nécessité d'un cadre de gouvernance bien coordonné et interopérable. Les préoccupations exprimées par la société civile au sujet des effets de l'IA sur les droits humains vont dans le même sens.

54. Le dispositif de gouvernance de l'IA peut s'inspirer des meilleures pratiques et de l'expertise recueillie dans le monde entier. Il doit aussi s'appuyer sur une compréhension des différentes idéologies culturelles qui sous-tendent le développement, le déploiement et l'utilisation de l'IA. Il faudrait mettre en place des structures innovantes dans le dispositif de gouvernance afin de mobiliser le secteur privé, le monde universitaire et la société civile aux côtés des gouvernements. On peut s'inspirer des efforts déployés par le passé pour faire participer le secteur privé à la recherche de biens collectifs, tels que la structure tripartite de l'OIT et le Pacte mondial des Nations Unies.

e) Principe directeur nº 5 : la gouvernance de l'intelligence artificielle devrait prendre ses racines dans la Charte des Nations Unies, le droit international des droits de l'homme et d'autres engagements convenus au niveau international, tels que les objectifs de développement durable

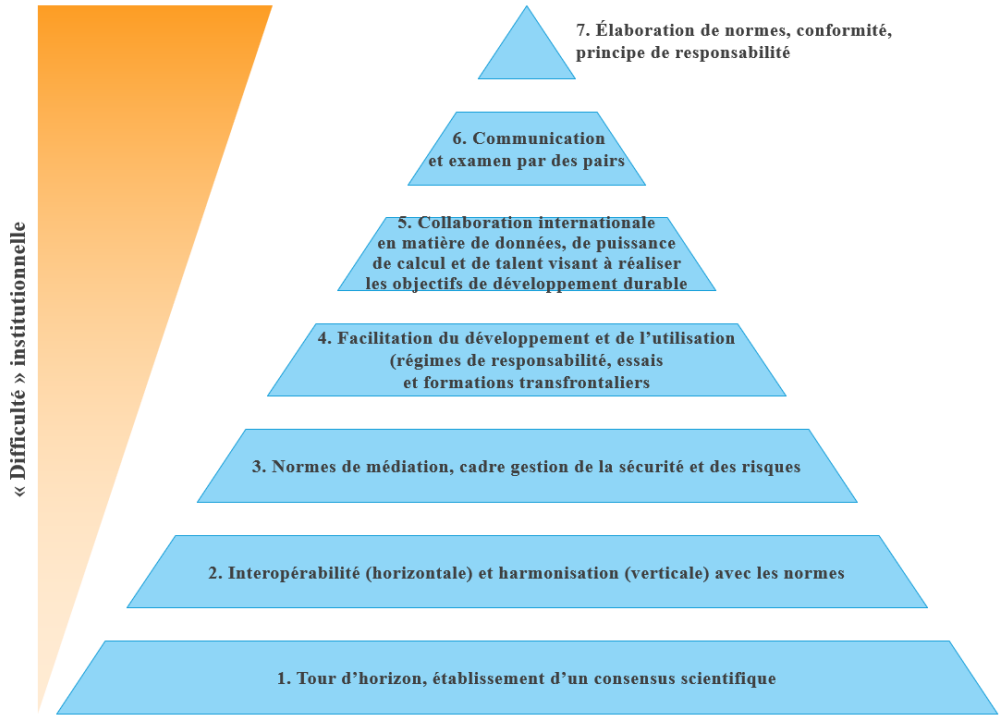
55. L'Organisation des Nations Unies a un rôle normatif et institutionnel unique à jouer ; par la mise en cohérence de la gouvernance de l'IA avec les valeurs fondamentales de l'Organisation, notamment la Charte et l'engagement en faveur de la paix et de la sécurité, des droits humains et du développement durable, qui constituent un socle et une boussole solides. L'Organisation est en mesure d'examiner les effets de l'IA sur diverses conditions économiques, sociales, sanitaires, sécuritaires et culturelles au niveau mondial, qui reposent toutes sur la nécessité de maintenir le respect universel et l'application des droits humains et de l'État de droit. Plusieurs entités des Nations Unies ont déjà mené des travaux importants sur les répercussions de l'IA dans des domaines tels que l'éducation et le contrôle des armements.

56. Le pacte numérique mondial et le Plan d'action de coopération numérique sont des exemples de délibérations multipartites sur un dispositif de gouvernance mondiale applicable aux technologies, y compris l'IA. La forte mobilisation des États Membres de l'Organisation des Nations Unies, les moyens d'action donnés aux entités des Nations Unies et la participation de diverses parties prenantes seront essentiels pour que la réponse mondiale en matière de gouvernance de l'IA dispose des moyens et des ressources nécessaires.

2. Fonctions institutionnelles

57. Pour être en mesure de gouverner correctement l'IA dans l'intérêt de l'humanité, un régime international de gouvernance de l'IA devrait au moins remplir les fonctions présentées dans la figure 2 (ci-dessous). Ces fonctions peuvent être exercées par des institutions à titre individuel ou par un réseau d'institutions.

Figure 2
Fonctions de gouvernance de l'intelligence artificielle, réparties en fonction de la « difficulté » institutionnelle[a]



[a] La « difficulté » fait référence aux pouvoirs présents à chaque niveau institutionnel et à la difficulté de parvenir à un consensus quant à la création de la fonction de gouvernance concernée. Le partage d'informations, par exemple, serait relativement peu controversé, alors que l'adoption d'un traité serait plus difficile.

58. On trouvera à la figure 2 un résumé des fonctions institutionnelles recommandées par l'Organe consultatif pour la gouvernance internationale de l'IA. Au niveau mondial, les organisations internationales, les gouvernements et le secteur privé seraient les premiers exécutants de ces fonctions. La société civile, y compris le monde universitaire et les scientifiques indépendants, jouerait un rôle clé en communiquant des données et des preuves aux fins de l'élaboration des politiques et de l'évaluation des incidences, et en demandant aux principaux acteurs de rendre des comptes pendant la mise en œuvre. Chaque ensemble de fonctions aurait différents lieux de responsabilité à différents niveaux de gouvernance, à savoir le secteur privé, les gouvernements et les organisations internationales. Au cours de la prochaine phase de ses travaux, l'Organe consultatif approfondira le concept de responsabilités partagées et différenciées pour les multiples parties prenantes à différents niveaux de la structure de gouvernance.

a) Fonction institutionnelle n° 1 : évaluer régulièrement les orientations et les implications futures de l'intelligence artificielle

59. Il n'existe actuellement aucune fonction institutionnelle faisant autorité pour des évaluations indépendantes, inclusives et multidisciplinaires de la trajectoire et des implications futures de l'IA. Un consensus sur l'orientation et le rythme d'évolution de ces technologies, ainsi que sur les risques et les opportunités qui y sont associés, pourrait constituer une ressource sur laquelle les décideurs pourraient s'appuyer lors

de l'élaboration de programmes nationaux d'IA visant à encourager l'innovation et à gérer les risques.

60. Comme pour le Groupe d'experts intergouvernemental sur l'évolution du climat, une fonction spécialisée de connaissance et de recherche en matière d'IA nécessiterait un processus indépendant, dirigé par des experts, qui fournirait tous les six mois (par exemple) des informations scientifiques fondées sur des données probantes, afin d'informer les dirigeants de la trajectoire future du développement, du déploiement et de l'utilisation de l'IA (voir les sous-fonctions n^{os} 1 à 3 du tableau ci-dessous). À cette fin, des accords devraient être conclus avec les entreprises pour garantir l'accès à l'information à des fins de recherche et d'analyse prospective. Cette fonction spécialisée aiderait le public à mieux comprendre l'IA et favoriserait un consensus dans la communauté internationale sur la rapidité et les incidences de son évolution. Elle produirait des estimations des risques régulières, les diffuserait et établirait des normes pour mesurer les effets de l'IA sur l'environnement et dans d'autres domaines. L'Organe consultatif est en quelque sorte le point de départ d'un tel processus dirigé par des experts, qui devrait être doté de ressources suffisantes et institutionnalisé.

61. L'ampleur des externalités négatives de l'IA n'est pas encore tout à fait claire. Le rôle de l'IA dans la désintermédiation des aspects de la vie qui sont au cœur du développement humain pourrait changer fondamentalement la façon dont les personnes et les groupes fonctionnent. À mesure que les capacités de l'IA progressent, il est possible de procéder à des ajustements structurels profonds de la manière dont les personnes vivent, travaillent et interagissent. Un observatoire analytique mondial pourrait coordonner les activités de recherche sur les répercussions sociales critiques de l'IA, y compris ses effets sur le travail, l'éducation, la santé publique, la paix et la sécurité, et la stabilité géopolitique. En s'appuyant sur l'expertise et le partage des connaissances à travers le monde, une telle fonction pourrait faciliter l'émergence de meilleures pratiques et de réponses communes.

b) Fonction institutionnelle n^{o} 2 : renforcer l'interopérabilité des efforts de gouvernance et leur ancrage dans les normes internationales grâce à un cadre mondial de gouvernance de l'intelligence artificielle

62. Les dispositifs de gouvernance de l'IA devraient être interopérables entre les juridictions et fondés sur des normes internationales, telles que la Déclaration universelle des droits de l'homme (voir le principe directeur n^{o} 4, paragraphes 52 à 54, ci-dessus). Ils devraient tirer parti de l'expertise des entités et des forums existants du système des Nations Unies, tels que l'UIT et l'UNESCO, afin de renforcer l'interopérabilité des mesures réglementaires entre les juridictions. Les efforts de gouvernance de l'IA pourraient également être coordonnés par un organe chargé d'harmoniser les politiques, de développer une compréhension commune, de mettre en lumière les meilleures pratiques, de soutenir l'exécution et d'encourager l'échange de connaissances entre pairs (voir les sous-fonctions n^{os} 7 à 10 du tableau ci-dessous). Un cadre mondial de gouvernance pourrait soutenir l'élaboration des politiques et guider la mise en œuvre afin d'éviter les fractures en matière d'IA et les lacunes de gouvernance entre les secteurs public et privé, les régions et les pays. Il pourrait aussi préciser les principes et les normes qui régissent le fonctionnement de différentes organisations. Dans ce cadre, une attention particulière devrait être accordée au renforcement des capacités dans les secteurs privé et public, à la diffusion des connaissances et à la sensibilisation à l'échelle mondiale. Les meilleures pratiques, telles que les études d'impact sur les droits humains réalisées par les développeurs de systèmes d'IA des secteurs privé et public, pourraient être mises en commun via un tel cadre, qui pourrait reposer sur un accord international.

c) Fonction institutionnelle nº 3 : élaborer et harmoniser des normes et des cadres de gestion de la sécurité et des risques

63. Plusieurs initiatives importantes visant à élaborer des normes techniques et normatives et des cadres de sécurité et de gestion des risques pour l'IA sont en cours, mais on constate un manque d'harmonisation au niveau mondial (voir la sous-fonction nº 11 dans le tableau ci-dessous). L'Organisation des Nations Unies réunissant tous les pays du monde, elle peut jouer un rôle essentiel en rapprochant les États, en élaborant des normes sociotechniques communes et en veillant à l'interopérabilité juridique et technique.

64. Les nouveaux instituts de sécurité de l'IA pourraient, par exemple, être mis en réseau afin de réduire le risque de cadres concurrents, la fragmentation des pratiques de normalisation entre les juridictions et l'avènement d'un patchwork mondial présentant trop de lacunes. Il convient toutefois de veiller à ne pas accorder trop d'importance à l'interopérabilité technique sans agir simultanément sur d'autres fonctions et normes. Les normes sociotechniques sont mieux connues, mais il est nécessaire de poursuivre la recherche, de faire participer activement la société civile et de mettre en place une coopération transdisciplinaire afin d'élaborer de telles normes.

65. De plus, de nouvelles normes et de nouveaux indicateurs mondiaux permettant de mesurer et de suivre les retombées environnementales de l'IA, ainsi que sa consommation d'énergie et de ressources naturelles (électricité et eau), pourraient être définis afin d'orienter son développement et de contribuer à la réalisation des objectifs de développement durable liés à l'environnement.

d) Fonction institutionnelle nº 4 : faciliter le développement, le déploiement et l'utilisation de l'intelligence artificielle dans l'intérêt de l'économie et de la société grâce à une coopération internationale multipartite

66. Outre les normes visant à prévenir les dommages et les abus, les développeurs et les utilisateurs, en particulier dans le monde du Sud, ont besoin d'outils essentiels, tels que des normes pour la catégorisation et le test des données, la protection des données et les protocoles d'échange, qui permettent aux start-ups de tester et de déployer leurs produits au-delà des frontières, ainsi que des mécanismes de responsabilité juridique, de résolution des litiges, de développement des entreprises et d'autres mécanismes d'appui. Les dispositions juridiques, financières et techniques existantes doivent évoluer afin d'anticiper les systèmes d'IA complexes et adaptatifs à venir. À cette fin, il faudra tenir compte des enseignements tirés de forums tels que le Groupe d'action financière, la Société de télécommunications interbancaires mondiales et d'autres mécanismes équivalents. De plus, dans de nombreux pays et régions il est urgent de développer les capacités du secteur public, pour faciliter l'utilisation responsable et bénéfique de l'IA et pour être en mesure de participer de manière significative aux cadres de coopération internationaux multipartites visant à développer des outils pour l'IA (voir les sous-fonctions nºs 4, 5 et 11 dans le tableau ci-dessous).

e) Fonction institutionnelle nº 5 : promouvoir la collaboration internationale en matière de développement des talents, d'accès aux infrastructures de calcul, de constitution de divers jeux de données de haute qualité, de partage responsable de modèles libres de droits et de biens collectifs utilisant l'intelligence artificielle aux fins de la réalisation des objectifs de développement durable

67. Un ou plusieurs nouveaux mécanismes sont nécessaires pour faciliter l'accès aux données, à la puissance de calcul et aux talents afin de développer, de déployer et d'utiliser des systèmes d'IA aux fins de la réalisation des objectifs de développement durable grâce à des chaînes de valeur locales améliorées, grâce à quoi les chercheurs universitaires indépendants, les entrepreneurs sociaux et la société civile pourront accéder à l'infrastructure et aux jeux de données nécessaires à la construction de leurs propres modèles et à la conduite de recherches et d'évaluations. Un tel mécanisme pourrait nécessiter des ressources et des efforts coordonnés qui permettraient de créer des jeux de données communs et des données communes à utiliser dans l'intérêt général, d'assurer le partage responsable de modèles et de ressources de calcul libres de droits et d'intensifier l'éducation et la formation.

68. La mise en commun de connaissances spécialisées et de ressources analogues à celles du CERN, du Laboratoire européen de biologie moléculaire ou de l'ITER ainsi que les fonctions de diffusion des technologies de l'AIEA pourraient donner un coup de fouet indispensable aux efforts déployés pour atteindre les objectifs de développement durable (voir la sous-fonction nº 6 dans le tableau ci-dessous). Ces fonctions pourraient aussi être complétées par la création d'incitations à partager et à mettre à disposition des outils de recherche et de développement, ciblant les acteurs du secteur privé. Les experts du monde du Sud sont souvent invisibles dans les conférences mondiales sur l'IA : cela doit changer.

69. L'ouverture de l'accès aux données et à la puissance de calcul devrait aussi s'accompagner d'un renforcement des capacités, en particulier dans le monde du Sud. Pour faciliter la création, l'adoption et la mise au point de modèles propres au contexte local, il serait important de suivre les utilisations positives de l'IA et d'encourager et d'évaluer les biens collectifs utilisant l'IA. La mobilisation du secteur privé serait cruciale pour mettre l'IA au service des objectifs de développement durable. Comme dans les engagements pris par les entreprises dans le cadre du Pacte mondial des Nations Unies, cette approche pourrait englober des engagements publics pris par les entreprises du secteur des technologies et autres de développer, de déployer et d'utiliser l'IA pour le bien de toutes et tous. Dans le contexte plus large du pacte numérique mondial, il pourrait aussi s'agir de communiquer des informations sur la manière dont l'IA contribue à la réalisation des objectifs de développement durable.

f) Fonction institutionnelle nº 6 : surveiller les risques, signaler les incidents et coordonner les interventions d'urgence

70. Le fait que les outils d'IA soient, par nature, sans frontières et puissent proliférer à l'échelle mondiale sur simple pression d'une touche crée de nouveaux défis en matière de sécurité internationale et de stabilité mondiale. Les modèles d'IA pourraient réduire les obstacles à l'accès aux armes de destruction massive. Les cyberoutils utilisant l'IA augmentent le risque d'attaques ciblant des infrastructures critiques, et l'IA à double usage peut être utilisée pour alimenter des armes létales autonomes, ce qui pourrait compromettre le respect du droit international humanitaire et d'autres normes. Des robots ayant des caractéristiques de plus en plus humaines peuvent rapidement diffuser des informations nuisibles d'une manière qui peut causer des dommages importants aux marchés et aux institutions publiques. Il n'est pas exclu que des IA malveillantes puissent échapper aux contrôles et faire peser des risques

encore plus grands. Compte tenu de ces difficultés, des capacités doivent être créées au niveau mondial pour surveiller les vulnérabilités systémiques et les perturbations de la stabilité internationale, en rendre compte et y répondre rapidement (voir les sous-fonctions nos 13 et 14 du tableau ci-dessous).

71. Par exemple, un modèle « technoprudentiel »[3], semblable au cadre macroprudentiel utilisé pour accroître la résilience des systèmes bancaires centraux et rassembler les cadres élaborés au niveau national, pourrait contribuer à protéger de la même manière la stabilité mondiale contre les risques liés à l'IA. Ce modèle doit être fondé sur les principes des droits humains.

72. Les pratiques adoptées par l'AIEA en matière de réassurance mutuelle pour ce qui est de la sûreté et de la sécurité nucléaires, et par l'OMS en matière de surveillance des maladies, pourraient servir d'inspiration pour les dispositifs de communication de l'information.

g) Fonction institutionnelle no 7 : conformité et responsabilité fondées sur des normes

73. La nécessité de normes juridiquement contraignantes et d'une exécution au niveau mondial ne peut être exclue. Un effort régional en faveur d'un traité sur l'IA est déjà en cours, et la question des technologies émergentes dans le domaine des systèmes d'armes létaux autonomes est examinée dans le cadre d'un traité sur les armes classiques. Des normes non contraignantes pourraient aussi jouer un rôle important, seules ou en association avec des normes contraignantes. L'Organisation des Nations Unies ne peut et ne doit pas chercher à être le seul arbitre de la gouvernance de l'IA. Toutefois, dans certains domaines tels que la sécurité internationale, elle dispose d'une légitimité unique en matière d'élaboration de normes (voir la sous-fonction no 12 dans le tableau ci-dessous). L'Organisation peut aussi veiller à ce qu'il n'y ait pas de lacunes en matière de responsabilité, par exemple en encourageant les États à rendre compte en suivant l'exemple des rapports soumis sur les progrès accomplis dans la réalisation des objectifs de développement durable ou dans le cadre de l'examen périodique universel, qui facilite le suivi, l'évaluation et la communication d'informations pour ce qui est des pratiques en matière de droits humains (voir la sous-fonction no 15 dans le tableau ci-dessous). Ces rapports devraient être établis en temps utile et de manière précise. Par ailleurs, on pourrait s'inspirer de mécanismes de règlement des différends existants, tels que ceux de l'Organisation mondiale du commerce, et utiliser des forums mondiaux pour faciliter le règlement des différends.

74. Dans le même temps, la légitimité de toute institution de gouvernance mondiale dépend de la responsabilité effective de cette institution. Les objectifs et processus liés à la gouvernance internationale doivent être transparents afin d'obtenir la confiance des citoyens concernés, notamment en prévenant les conflits d'intérêts.

[3] On trouvera des informations complémentaires à l'adresse www.gzeromedia.com/ai/what-is-a-techno-prudential-approach-to-ai-governance (non disponible en français).

Sous-fonctions pour la gouvernance internationale de l'intelligence artificielle et délais de réalisation possibles

Sous-fonction	*Description*	*Catégorie*	*Délai possible pour l'institutionnalisation de la sous-fonction proposée*
1. Évaluation scientifique	Passer en revue les politiques internationales, régionales et nationales en matière d'IA et élaborer un rapport public, au moins tous les six mois.	Recherche et analyse	6 à 12 mois
2. Tour d'horizon prospectif	Élaborer un rapport d'analyse prospective recensant les risques qui dépassent les frontières et pourraient affecter toutes les juridictions.	Recherche et analyse	6 à 12 mois
3. Classement des risques	Évaluer les modèles d'IA existants et futurs sur une échelle de risques : risques insoutenables, risques élevés, risques moyens et risques faibles ou nuls.	Recherche et analyse	6 à 12 mois
4. Accès aux bienfaits	Accès équitable à la technologie et aux bienfaits de l'IA, qui accélèrent la réalisation des objectifs de développement durable.	Catalyseurs	12 à 24 mois
5. Renforcement des capacités	Programmes et ressources facilitant le développement des technologies et activités liées à l'IA, ainsi que les capacités de gouvernance et de promotion des États.	Catalyseurs	12 à 24 mois
6. Recherche et développement conjoints	Mettre en place la capacité d'entreprendre une recherche et un développement collaboratifs de l'IA au profit de celles et ceux qui n'ont pas accès aux outils ou à l'expertise en la matière.	Catalyseurs	12 à 24 mois
7. Participation sans exclusive	Veiller à la participation de tous les groupes de parties prenantes et de tous les pays et régions à la gouvernance collective, à la gestion des risques et à la concrétisation des opportunités ; s'efforcer de mettre en place une gouvernance innovante.	Gouvernance	6 à 12 mois
8. Réunion ; apprentissage international	Réunir régulièrement les parties prenantes pour examiner les politiques en matière d'IA dans les différentes juridictions ; rechercher un consensus sur une terminologie et des définitions communes ; échanger des connaissances entre pairs.	Gouvernance	6 à 12 mois
9. Coordination internationale	Déconflictualiser les travaux et créer des synergies entre les organismes internationaux existants qui poursuivent leurs travaux sur l'IA.	Gouvernance	6 à 12 mois

Sous-fonction	*Description*	*Catégorie*	*Délai possible pour l'institutionnalisation de la sous-fonction proposée*
10. Harmonisation des politiques, mise en cohérence des normes	Mettre en évidence les meilleures pratiques en matière de normes et de règles, notamment pour ce qui concerne l'atténuation des risques et la croissance économique ; mettre en cohérence, exploiter et inclure les normes, méthodes et cadres juridiques contraignants et non contraignants élaborés aux niveaux régional, national et sectoriel pour soutenir l'interopérabilité.	Gouvernance	12 à 24 mois
11. Établissement de normes	Trouver un consensus mondial sur les normes d'utilisation de l'IA entre les groupes de parties prenantes en travaillant avec les organismes normatifs nationaux. Mises à jour régulières.	Gouvernance	12 à 24 mois
12. Élaboration de normes	Convoquer les parties prenantes pour évaluer la nécessité de cadres, traités ou autres régimes contraignants et non contraignants pour l'IA et mener des négociations portant sur ces cadres, traités ou autres régimes.	Gouvernance	24 à 36 mois
13. Application	Élaborer des mécanismes de réassurance mutuelle, des mécanismes de partage de l'information qui respectent les informations relatives à la sécurité commerciale et nationale, des mécanismes de règlement des différends et des systèmes ou régimes de responsabilité.	Gouvernance	Plus de 36 mois
14. Stabilisation et intervention	Développer et entretenir collectivement une capacité d'intervention en cas d'urgence, des « interrupteurs » et d'autres mesures de stabilisation.	Gouvernance	Plus de 36 mois
15. Suivi et vérification	Mettre en place des dispositifs de contrôle et de vérification, le cas échéant, afin de garantir que la conception, le déploiement et l'utilisation des systèmes d'IA respectent le droit international applicable.	Gouvernance	Plus de 36 mois

VI. Conclusions

75. L'ampleur des répercussions qu'a l'IA sur la vie des personnes – par exemple, sur la façon dont elles travaillent et nouent des relations sociales, sur la façon dont elles sont éduquées et gouvernées, sur la façon dont elles interagissent quotidiennement les unes avec les autres – soulève des questions plus fondamentales que la façon dont elle devrait être gouvernée. Ces questions, telles que celle de déterminer ce que signifie être humain(e) dans un monde entièrement numérique et en réseau, dépassent largement le cadre du mandat de l'Organe consultatif, mais sont

liées aux décisions prises aujourd'hui. En effet, la gouvernance n'est pas une fin mais un moyen : un ensemble de mécanismes destinés à exercer un contrôle ou une direction sur un élément qui a le potentiel d'être bénéfique ou néfaste.

76. L'Organe consultatif souhaite évaluer de façon exhaustive les effets de l'IA sur la vie des personnes et mettre au jour de façon ciblée la contribution que peut apporter l'Organisation des Nations Unies. Il espère ainsi montrer qu'il est conscient des bienfaits réels de l'IA, tout en étant lucide sur les risques qu'elle comporte.

77. Les risques créés par l'inaction sont évidents. L'Organe consultatif estime qu'une gouvernance mondiale de l'IA est essentielle pour tirer parti des débouchés importants et gérer les risques que cette technologie présente pour chaque État, groupe et personne aujourd'hui, ainsi que pour les générations à venir.

78. Pour être efficace, la gouvernance internationale de l'IA doit être guidée par des principes et assurée par des fonctions claires. Ces fonctions globales doivent apporter une valeur ajoutée, combler les lacunes recensées et faciliter une action interopérable aux niveaux régional, national, sectoriel et communautaire. Elles doivent être menées de concert dans les institutions internationales, les dispositifs nationaux et régionaux et le secteur privé. Dans ses recommandations préliminaires, l'Organe consultatif expose ce qu'il considère comme les principes et fonctions essentiels de tout dispositif de gouvernance de l'IA à l'échelle mondiale.

79. L'Organe consultatif a adopté une approche selon laquelle la forme suit la fonction et il ne propose pas, à ce stade, de modèle unique de gouvernance de l'IA. À terme, cependant, cette gouvernance doit apporter des bienfaits et des garanties tangibles aux personnes et aux sociétés. Un cadre de gouvernance mondiale efficace doit combler le fossé entre les principes et les répercussions concrètes. Au cours de la prochaine phase de ses travaux, l'Organe consultatif étudiera les possibilités de formes institutionnelles de gouvernance mondiale de l'IA, en s'appuyant sur les points de vue de diverses parties prenantes dans le monde.

VII. Les prochaines étapes

80. Plutôt que de proposer un modèle unique de gouvernance de l'IA à ce stade, les recommandations préliminaires précédemment exposées dans le présent document sont axées sur les principes et les fonctions auxquels tout régime de ce type doit aspirer.

81. Au cours des prochains mois, l'Organe consultatif prendra l'avis (individuellement et en groupe) de diverses parties prenantes partout dans le monde. Des événements seront organisés dans le but de débattre des questions abordées dans le présent rapport d'étape, et de dialoguer avec les gouvernements, le secteur privé, la société civile et les communautés des secteurs technique et de la recherche. L'Organe consultatif poursuivra également des travaux de recherche, notamment sur les méthodes d'évaluation des risques et l'interopérabilité de la gouvernance. Des études de cas seront élaborées pour faciliter la réflexion sur les questions mises au jour dans le présent rapport, dans des contextes particuliers. L'Organe consultatif a aussi l'intention d'approfondir plusieurs questions, qui concernent entre autres l'open source, l'IA et le secteur financier, la définition de normes, la propriété intellectuelle, les droits humains et l'avenir du travail, en s'appuyant sur les travaux et les institutions existants.

82. L'Organe consultatif encourage un dialogue constructif avec toute personne intéressée par l'IA. On trouvera des informations sur les possibilités de participer aux travaux qu'il mène actuellement à l'adresse www.un.org/en/ai-advisory-body (non disponible en français).

83. L'Organe consultatif attend avec intérêt de dialoguer avec diverses parties prenantes au cours de ses travaux visant à répondre plus complètement aux questions recensées dans le présent rapport d'étape (voir encadré 4), à l'appui des efforts actuellement déployés par l'Organisation des Nations Unies en matière de coopération numérique, de progrès social et d'amélioration des conditions de vie dans une liberté plus grande.

Encadré 4
Questions clés à approfondir lors de la prochaine phase des travaux

Débouchés et catalyseurs

- Le développement de l'IA peut-il être rendu plus inclusif en facilitant les écosystèmes de construction de modèles, par exemple grâce à la protection des données et aux cadres d'échange, avec un accès partagé à la puissance de calcul ?
- Des normes communes en matière de catégorisation et de test des données encourageraient-elles les start-ups du secteur de l'IA à tester et à déployer leurs activités dans un plus grand nombre de pays et de régions ?
- Quels mécanismes permettraient de promouvoir un accès équitable à la puissance de calcul et le partage de jeux de données dans le respect de la vie privée entre les parties prenantes et les États Membres ?
- Comment développer et diffuser les talents en matière d'IA ? Les entités du système des Nations Unies ou d'autres institutions peuvent-elles faciliter les échanges d'étudiants, les programmes de doctorat conjoints et le développement de talents dans plusieurs domaines (par exemple, la santé et l'IA, l'agriculture et l'IA) ?
- Comment la collaboration internationale peut-elle exploiter les talents, les données et la puissance de calcul liés à l'IA aux fins de la recherche scientifique et de la réalisation des objectifs de développement durable ?
- Comment inciter les gouvernements et le secteur privé à investir dans d'autres infrastructures de base qui stimulent le développement de l'IA dans le monde ?

Risques et difficultés

- Quel est le meilleur moyen de parvenir à un consensus sur la détection, la classification et le traitement des risques liés à l'IA ?
- Comment les évaluations des risques et des problèmes devraient-elles être liées à des cas d'utilisation plus précis de l'IA, tels que les systèmes d'armes létaux autonomes ?
- Quel devrait être le seuil ou le déclencheur qui constituerait le franchissement d'une ligne rouge (par analogie, peut-être, avec l'interdiction du clonage humain dans la recherche biomédicale) ? Comment contrôler le franchissement de cette ligne rouge et la faire respecter ?

Gouvernance internationale

- Les principes énumérés ci-dessus reflètent-ils correctement les aspirations qui devraient être celles d'un régime de gouvernance mondiale de l'IA ?
- Les fonctions décrites ci-dessus reflètent-elles correctement ce que la gouvernance mondiale de l'IA peut et doit faire ?
- Quelles sont les dispositions structurelles qui permettraient le mieux à une nouvelle institution ou à un nouvel ensemble d'institutions de défendre ces principes et de remplir ces fonctions ?
- Il existe dans le système des Nations Unies une série de modèles permettant de faire participer les acteurs du secteur aux travaux sectoriels (OACI, UIT, OMS, etc.).
- Quel type de mécanisme pourrait le mieux soutenir la participation du secteur à la gouvernance internationale de l'IA ? Parmi les instruments normatifs, politiques et d'information qui existent aujourd'hui, quels sont ceux qui pourraient favoriser la cohérence de la gouvernance technologique entre les gouvernements, le secteur privé et la société civile ?
- Quels types de mécanismes de financement et de renforcement des capacités seraient nécessaires pour que des accords internationaux efficaces puissent remplir les fonctions décrites ci-dessus ?

Annexe I

À propos de l’Organe consultatif de haut niveau sur l’intelligence artificielle

Initialement proposé en 2020 dans le cadre du Plan d’action de coopération numérique (A/74/821), l’Organe consultatif de haut niveau multipartite sur l’intelligence artificielle a été créé en octobre 2023 pour analyser la gouvernance internationale de l’IA et formuler des recommandations.

Ses membres sont nommés à titre individuel et ne représentent pas leur organisation. Le rapport d’étape représente un consensus majoritaire ; aucun membre n’est censé approuver chacun des points contenus dans le document. En publiant ce rapport, les membres de l’Organe consultatif affirment qu’ils approuvent de manière générale, mais non unilatérale, les conclusions et recommandations formulées. Les termes utilisés dans le rapport n’impliquent pas l’approbation institutionnelle des différentes organisations dont sont issus les membres de l’Organe.

Annexe II

Membres de l'Organe consultatif de haut niveau sur l'intelligence artificielle

Carmen Artigas (Coprésidente)
James Manyika (Coprésident)
Anna Abramova
Omar Sultan Al Olama
Latifa al-Abdulkarim
Estela Aranha
Ran Balicer
Paolo Benanti
Abeba Birhane
Ian Bremmer
Anna Christmann
Natasha Crampton
Nighat Dad
Vilas Dhar
Virginia Dignum
Arisa Ema
Mohamed Farahat
Amandeep Singh Gill
Wendy Hall
Rahaf Harfoush
Hiroaki Kitano
Haksoo Ko
Andreas Krause
María Vanina Martínez Posse
Seydina Moussa Ndiaye
Mira Murati
Petri Myllymäki
Alondra Nelson
Nazneen Rajani
Craig Ramlal
He Ruimin
Emma Ruttkamp-Bloem
Marietje Schaake
Sharad Sharma
Jaan Tallinn
Philip Thigo
Jimena Sofía Viveros Álvarez
Yi Zeng
Zhang Linghan

Annexe III

Mandat de l'Organe consultatif de haut niveau sur l'intelligence artificielle

L'Organe consultatif de haut niveau sur l'intelligence artificielle, créé par le Secrétaire général, analysera la gouvernance internationale de l'intelligence artificielle et formulera des recommandations à ce sujet. Les premiers rapports de l'Organe consultatif apporteront des contributions d'experts de haut niveau et des contributions indépendantes aux débats actuellement menés aux niveaux national, régional et multilatéral.

L'Organe consultatif sera composé de 38 membres issus de gouvernements, du secteur privé, de la société civile et du milieu universitaire, ainsi que d'un membre de droit venu du Secrétariat. Sa composition sera équilibrée en termes de genre, d'âge, de représentation géographique et de domaine d'expertise lié aux risques et aux applications de l'intelligence artificielle. Les membres de l'Organe consultatif siègent en leur nom personnel.

L'Organe consultatif dialoguera largement avec les gouvernements, le secteur privé, les universités, la société civile et les organisations internationales, et les consultera. Il sera agile et innovant dans ses échanges avec les processus et les plateformes existants, ainsi que dans l'exploitation des contributions de diverses parties prenantes. Il peut créer des groupes de travail ou des groupes sur des sujets particuliers.

Les membres de l'Organe consultatif sont sélectionnés par le Secrétaire général sur la base des candidatures proposées par les États Membres et d'un appel public à candidatures. L'Organe consultatif sera composé de deux coprésidents et d'un Comité exécutif. Tous les groupes de parties prenantes seront représentés au Comité exécutif.

L'Organe consultatif est créé pour une période initiale d'un an, le Secrétaire général ayant la possibilité de décider d'une prolongation. Il tiendra des réunions en personne et en ligne.

L'Organe consultatif élaborera un rapport d'étape, d'ici le 31 décembre 2023, pour examen par le Secrétaire général et les États Membres de l'Organisation des Nations Unies. Le rapport d'étape présentera une analyse des solutions possibles en matière de gouvernance internationale de l'IA.

L'Organe consultatif utilisera les commentaires reçus sur le rapport pour soumettre un rapport final, d'ici le 31 août 2024, qui pourrait donner des recommandations détaillées sur les fonctions, la forme et le calendrier d'une nouvelle institution internationale chargée de la gouvernance de l'intelligence artificielle.

L'Organe consultatif doit éviter tout double emploi avec les forums et processus existants dans le cadre desquels les questions relatives à l'intelligence artificielle sont examinées. Il s'efforcera plutôt de tirer parti des plateformes et des partenaires existants, y compris les entités du système des Nations Unies, qui travaillent dans des domaines connexes. Il respectera pleinement les structures actuelles des Nations Unies ainsi que les prérogatives nationales, régionales et sectorielles en matière de gouvernance de l'intelligence artificielle.

Les travaux de l'Organe consultatif seront appuyés par un petit secrétariat basé au Bureau de l'Envoyé du Secrétaire général pour les technologies et financés par des ressources extrabudgétaires fournies par des donateurs.

Annexe IV

Groupes de travail et thèmes transversaux

Les travaux actuellement menés par l'Organe consultatif s'articulent autour de cinq groupes de travail et de dix thèmes transversaux. Les applications sectorielles et les thèmes supplémentaires seront examinés en détail lors de la prochaine phase.

Groupes de travail

Débouchés et catalyseurs

Risques et difficultés

Interopérabilité

Harmonisation avec les normes et les valeurs

Institutions internationales

Questions intersectorielles

Culture

Équité

Déontologie

L'avenir du travail

Capacités des gouvernements

Questions de genre

Droits humains, démocratie et état de droit

Open source

Répercussions sociétales

Durabilité

Applications sectorielles et thèmes supplémentaires à approfondir

Agriculture

Éducation

Environnement

Finances

Santé

Propriété intellectuelle

Sécurité nationale

Établissement de normes