Council of Presidents of the United Nations General Assembly
Conseil des Présidents de l'Assemblée Générale des Nations Unies

Gouvernance de la Transition vers l'Intelligence Artificielle Générale (AGI)

Considérations Urgentes pour le Rapport de l'Assemblée Générale des Nations Unies destiné au Conseil des Présidents de l'Assemblée Générale des Nations Unies (UNCPGA)

Résumé

Les systèmes d'IA progressent rapidement vers l'intelligence artificielle générale (AGI), caractérisée par des systèmes capables d'égaler ou de surpasser l'intelligence humaine dans diverses tâches cognitives. Grâce aux investissements financiers les plus importants de l'histoire, qui stimulent des efforts de R&D sans précédent, les leaders et experts du secteur prévoient que l'AGI pourrait voir le jour au cours de cette décennie[1],apportant des avantages extraordinaires à l'humanité. Parmi ces avantages, l'AGI pourrait accélérer les découvertes scientifiques liées à la santé publique, transformer de nombreux secteurs et augmenter la productivité, et contribuer à la réalisation des objectifs de développement durable.

Néanmoins, l'AGI pourrait également créer des risques uniques et potentiellement catastrophiques. Contrairement à l'IA traditionnelle, l'AGI pourrait exécuter de manière autonome des actions nuisibles échappant au contrôle humain, entraînant des impacts irréversibles, des menaces provenant de systèmes d'armes avancés et des vulnérabilités dans les infrastructures critiques. Nous devons veiller à atténuer ces risques si nous voulons profiter des avantages extraordinaires de l'AGI.

Pour relever efficacement ces défis mondiaux, une action internationale immédiate et coordonnée soutenue par les Nations Unies est essentielle. Ces actions devraient être initiées par une Assemblée générale spéciale des Nations Unies consacrée spécifiquement à l'AGI afin de discuter des avantages et des risques de l'AGI et de la création éventuelle d'un observatoire mondial de l'AGI, d'un système de certification pour une AGI sûre et fiable, d'une convention des Nations Unies sur l'AGI et d'une agence internationale de l'AGI. Sans une gestion mondiale proactive, la concurrence entre les nations et les entreprises accélérera le développement risqué de l'AGI, compromettra les protocoles de sécurité et exacerbera les tensions géopolitiques.

[1] METR: Measuring AI Ability to Compete Long Tasks https://arxiv.org/abs/2503.14499

Une action internationale coordonnée peut empêcher ces conséquences, en favorisant le développement et l'utilisation sécurisés de l'AGI, la répartition équitable des avantages et la stabilité mondiale.

Introduction

Les progrès en matière d'IA ont été rapides ces dernières années et ces derniers mois[2] et pourraient s'accélérer encore davantage, en partie parce que les entreprises spécialisées dans l'IA investissent des sommes considérables dans la création d'agents IA plus performants et plus autonomes, et en raison de l'utilisation croissante des modèles d'IA les plus puissants pour faire avancer la recherche sur l'IA elle-même[3].

Il est largement prévu que ces améliorations des capacités de l'IA conduiront à l'« intelligence artificielle générale » (AGI) : des systèmes d'IA qui égalent ou dépassent les performances humaines dans la plupart des tâches cognitives.

Bien qu'il y ait des désaccords sur la date à laquelle l'AGI est attendue, tous les experts de ce panel estiment qu'elle pourrait bien être développée au cours de cette décennie. Les entreprises spécialisées dans l'IA engagent des centaines de milliards de dollars pour parvenir très rapidement à l'AGI, ce qui en fait de loin le plus grand effort de R&D de l'histoire de l'humanité. Le secteur privé a la responsabilité de développer des technologies beaucoup plus sûres, et il devrait être incité à le faire ; mais la course à la compétitivité pour être le premier à atteindre l'AGI le pousse à concentrer tous ses efforts sur les capacités plutôt que sur la sécurité, afin de « gagner la course ».

Les risques actuels liés à l'IA proviennent principalement d'une mauvaise utilisation de la technologie par l'homme. Cependant, l'AGI présente également un risque fondamentalement différent, car ses menaces potentielles vont au-delà de l'utilisation abusive par l'homme. L'AGI pourrait générer et exécuter de manière autonome des plans aux conséquences catastrophiques, dépassant la capacité humaine à reconnaître, analyser et répondre aux menaces émergentes et aux perturbations sans précédent[4]. Combiné à la tendance à l'autoconservation récemment observée[5] chez les IA avancées, cela pourrait conduire à des situations où l'AGI deviendrait incontrôlable.

[2] See the International AI Safety Report, Bengio et al 2025.
[3] https://www.forethought.org/research/will-ai-r-and-d-automation-cause-a-software-intelligence-explosion.pdf
[4] "Claude 3.7 (often) Knows When it is in Alignment Evaluations" https://www.apolloresearch.ai/blog/claude-sonnet-37-often-knows-when-its-in-alignment-evaluations
[5] See Meinke et al 2024, Frontier Models are Capable of In-context Scheming, https://arxiv.org/abs/2412.04984.

Cela devrait être une préoccupation mondiale commune. Les risques liés à l'AGI ne se limitent pas à des industries ou à des sociétés spécifiques, mais ont des implications mondiales, quel que soit leur lieu d'origine. Garantir une intégration sûre et harmonieuse de l'AGI nécessite non seulement des efforts nationaux ou privés, mais aussi une gouvernance internationale proactive, menée par les Nations unies. Les Nations Unies sont particulièrement bien placées pour faciliter un accord scientifique sur les risques et les stratégies d'atténuation, établir un consensus politique autour d'une approche commune de l'atténuation des risques, coordonner les politiques, promouvoir des normes ou des garde-fous, répondre aux urgences et éventuellement mener ou coordonner des recherches conjointes en matière de sûreté ou de sécurité. Sans gouvernance mondiale, le potentiel transformateur de l'AGI pour relever les défis mondiaux pourrait être sous-utilisé ou mal orienté. De plus, la coordination mondiale sera essentielle pour gérer les menaces catastrophiques mondiales que l'AGI est susceptible de poser. Il est difficile d'imaginer que cette coordination puisse être réalisée à l'échelle mondiale sans le leadership actif des Nations Unies.

I. Urgence d'une action de l'Assemblée générale des Nations unies sur la gouvernance de l'AGI et conséquences probables si aucune mesure n'est prise

Dans un contexte géopolitique complexe et en l'absence de normes internationales cohérentes et contraignantes, la course effrénée au développement de l'AGI sans mesures de sécurité adéquates augmente le risque d'accidents ou d'utilisation abusive, de militarisation et de défaillances existentielles[6] .

Les nations et les entreprises privilégient la rapidité au détriment de la sécurité, sapant ainsi les cadres réglementaires nationaux et reléguant les protocoles de sécurité au second plan derrière les avantages économiques ou militaires.

Étant donné que de nombreuses formes d'AGI provenant des gouvernements et des entreprises pourraient voir le jour avant la fin de cette décennie, et que la mise en place de systèmes de gouvernance nationaux et internationaux prendra des années, il est urgent d'entamer les procédures nécessaires pour éviter les conséquences suivantes :

[6] OpenAI response to US Office of Science and Technology Policy's AI Action Plan
[OpenAI Response] OSTP/NSF RFI: Notice Request for Information on the Development of an Artificial Intelligence (AI) Action Plan - Google Docs

1. **Conséquences irréversibles** - Une fois l'AGI atteinte, son impact pourrait être irréversible. De nombreuses formes d'IA de pointe affichent déjà des comportements trompeurs et d'autoprotection, et la tendance vers des IA plus autonomes, interactives et capables de s'améliorer elles-mêmes, intégrées aux infrastructures, pourrait rendre les impacts et la trajectoire de l'AGI incontrôlables. Si cela se produit, il pourrait être impossible de revenir à un état de surveillance humaine fiable. Une gouvernance proactive est essentielle pour garantir que l'AGI ne franchisse pas nos lignes rouges[7], conduisant à des systèmes incontrôlables sans moyen clair de revenir au contrôle humain.

2. **Armes de destruction massive** - L'AGI pourrait permettre à certains États et acteurs non étatiques malveillants de fabriquer des armes chimiques, biologiques, radiologiques et nucléaires. De plus, de grands essaims d'armes autonomes létales contrôlés par l'AGI pourraient constituer eux-mêmes une nouvelle catégorie d'armes de destruction massive.

3. **Vulnérabilités des infrastructures critiques** - Les systèmes nationaux critiques (par exemple, les réseaux énergétiques, les systèmes financiers, les réseaux de transport, les infrastructures de communication et les systèmes de santé) pourraient être soumis à de puissantes cyberattaques lancées par ou avec l'aide de l'AGI. Sans dissuasion nationale et coordination internationale, des acteurs non étatiques malveillants, des terroristes aux organisations criminelles transnationales, pourraient mener des attaques à grande échelle.

4. **Concentration du pouvoir, inégalités mondiales et instabilité** - Le développement et l'utilisation incontrôlés de l'AGI pourraient exacerber les disparités de richesse et de pouvoir à une échelle sans précédent. Si l'AGI reste entre les mains de quelques nations, entreprises ou groupes d'élite, cela pourrait renforcer la domination économique et créer des monopoles mondiaux sur l'intelligence, l'innovation et la production industrielle. Cela pourrait entraîner un chômage massif, une perte de pouvoir généralisée affectant les fondements juridiques, une perte de confidentialité et un effondrement de la confiance dans les institutions, les connaissances scientifiques et la gouvernance.

Cela pourrait miner les institutions démocratiques par la persuasion, la manipulation et la propagande générée par l'IA, et accroître l'instabilité géopolitique d'une manière qui augmente les vulnérabilités systémiques. Un manque de coordination pourrait entraîner des conflits autour des ressources, des capacités ou du contrôle de l'AGI, pouvant dégénérer en guerre.

[7] International Dialogues on AI Safety (2024): https://idais.ai/dialogue/idais-beijing/

L'AGI mettra à rude épreuve les cadres juridiques existants : de nombreuses questions nouvelles et complexes liées à la propriété intellectuelle, à la responsabilité, aux droits de l'homme et à la souveraineté pourraient submerger les systèmes juridiques nationaux et internationaux.

5. **Risques existentiels** - L'AGI pourrait être utilisée à mauvais escient pour causer des dommages massifs ou développée d'une manière qui ne correspond pas aux valeurs humaines ; elle pourrait même agir de manière autonome, hors du contrôle humain, en développant ses propres objectifs en fonction de ses objectifs d'autoconservation déjà observés dans les IA de pointe actuelles. L'AGI pourrait également rechercher le pouvoir comme moyen de s'assurer qu'elle peut atteindre les objectifs qu'elle se fixe, indépendamment de l'intervention humaine. Les gouvernements nationaux, les principaux experts et les entreprises qui développent l'AGI ont tous déclaré que ces tendances pourraient conduire à des scénarios dans lesquels les systèmes AGI chercheraient à dominer les humains. Il ne s'agit pas d'hypothèses de science-fiction farfelues sur un avenir lointain : de nombreux experts de premier plan considèrent que ces risques pourraient tous se concrétiser au cours de cette décennie, et leurs prémices se manifestent déjà. De plus, les principaux développeurs d'IA n'ont jusqu'à présent aucune proposition viable pour prévenir ces risques avec un haut degré de confiance.

6. **Perte d'avantages extraordinaires pour l'humanité tout entière** - Une IA générale correctement gérée promet des améliorations dans tous les domaines, pour tous les peuples, de la médecine personnalisée, au traitement du cancer et à la régénération cellulaire, en passant par les systèmes d'apprentissage individualisés, l'éradication de la pauvreté, la lutte contre le changement climatique et l'accélération des découvertes scientifiques, avec des avantages inimaginables. Garantir un avenir aussi magnifique pour tous nécessite une gouvernance mondiale, qui commence par une meilleure prise de conscience mondiale des risques et des avantages.

Les Nations unies jouent un rôle essentiel dans cette mission.

II. Objectif de la gouvernance des Nations Unies dans le cadre de la transition vers l'AGI

Étant donné que l'AGI pourrait bien être développée au cours de cette décennie, il est à la fois scientifiquement et éthiquement impératif que nous mettions en place des structures de gouvernance solides afin de nous préparer à la fois aux avantages extraordinaires et aux risques extraordinaires qu'elle pourrait entraîner. L'objectif de la gouvernance des Nations Unies dans la transition vers l'AGI est de garantir que le développement et l'utilisation de l'AGI soient conformes aux valeurs humaines, à la sécurité et au développement mondiaux.

Cela implique : 1) de faire progresser la recherche sur l'alignement et le contrôle de l'IA afin d'identifier des méthodes techniques permettant de diriger et/ou de contrôler des systèmes d'IA de plus en plus performants ; 2) de fournir des orientations pour le développement de l'AGI, en établissant des cadres visant à garantir que l'AGI soit développée de manière responsable, avec des mesures de sécurité robustes, de la transparence et en accord avec les valeurs humaines ; 3) Développer des cadres de gouvernance pour le déploiement et l'utilisation de l'AGI, afin de prévenir les abus, de garantir un accès équitable et de maximiser ses avantages pour l'humanité tout en minimisant les risques ; 4) Favoriser des visions d'avenir bénéfiques pour l'AGI, avec de nouveaux cadres pour le développement social, environnemental et économique ; et 5) Fournir une plateforme neutre et inclusive pour la coopération internationale, en établissant des normes mondiales, mettre en place un cadre juridique international et créer des incitations à la conformité ; favorisant ainsi la confiance entre les nations afin de garantir l'accès mondial aux avantages de l'AGI.

III. Session de l'Assemblée générale des Nations unies sur les considérations clés relatives à l'AGI

L'un des plus grands défis de la gouvernance de l'AGI est l'incertitude qui entoure son développement technologique futur. Il est donc difficile de prédire avec précision les avantages et les risques potentiels. Par conséquent, un cadre de réponse large et complet doit être mis en place pour anticiper et atténuer les menaces envisageables tout en renforçant les avantages potentiels. Les Nations unies peuvent assurer la coordination internationale indispensable au développement et à l'utilisation de l'AGI. Il est particulièrement important que toutes les nations soient représentées dans ce processus et que celui-ci réduise les divisions géopolitiques ; à l'heure actuelle, seule l'ONU semble bien placée pour jouer ce rôle.

Les points suivants devraient être examinés lors d'une session de l'Assemblée générale des Nations unies consacrée spécifiquement à l'AGI :

A. Observatoire mondial de l'AGI

Un observatoire mondial de l'AGI est nécessaire pour suivre les progrès de la recherche et du développement liés à l'AGI et fournir des alertes précoces sur la sécurité de l'IA aux États membres. Cet observatoire devrait tirer parti de l'expertise d'autres initiatives des Nations Unies, telles que le Panel scientifique international indépendant sur l'IA créé par le Global Digital et la méthodologie d'évaluation de l'état de préparation de l'UNESCO.

B. Système international de bonnes pratiques et de certification pour une AGI sûre et fiable

Étant donné que l'AGI pourrait bien être développée au cours de cette décennie, il est à la fois scientifiquement et éthiquement impératif que nous mettions en place des structures de gouvernance solides afin de nous préparer à la fois aux avantages extraordinaires et aux risques extraordinaires qu'elle pourrait entraîner.

C. Convention-cadre des Nations unies sur l'AGI

Une convention-cadre sur l'AGI[8]est nécessaire pour établir des objectifs communs et des protocoles flexibles afin de gérer les risques liés à l'AGI et de garantir une répartition équitable des avantages à l'échelle mondiale. Elle devrait définir des niveaux de risque clairs nécessitant une action internationale proportionnée, allant de la mise en place de normes et de régimes d'autorisation à la création d'installations de recherche communes pour l'AGI à haut risque, en passant par des lignes rouges ou des seuils de déclenchement[9] pour le développement de l'AGI. Une convention fournirait la base institutionnelle adaptable essentielle à une gouvernance mondiale légitime, inclusive et efficace de l'AGI, minimisant les risques mondiaux et maximisant la prospérité mondiale grâce à l'AGI.

D. Étude de faisabilité sur une agence des Nations Unies dédiée à l'AGI

Compte tenu de l'ampleur des mesures nécessaires pour se préparer à l'AGI et de l'urgence de la question, des mesures doivent être prises pour étudier la faisabilité d'une agence des Nations Unies dédiée à l'AGI, idéalement dans le cadre d'un processus accéléré. Une structure similaire à l'AIEA a été suggérée, sachant que la gouvernance de l'AGI est bien plus complexe que celle de l'énergie nucléaire et qu'elle nécessite donc des considérations particulières dans le cadre d'une telle étude de faisabilité.

---

[8] Cass-Beggs, Duncan, Stephen Clare, Dawn Dimowo, and Zaheed Kara. 2024. "Framework Convention on Global AI Challenges." Center for International Governance Innovation. https://www.cigionline.org/publications/framework-convention-on-global-ai-challenges/.

[9] Russell, Stuart, Edson Prestes, Mohan Kankanhalli, Jibu Elias, Constanza Gómez Mont, Vilas Dhar, Adrian Weller, Pascale Fung, and Karim Beguir, "AI red lines: The opportunities and challenges of setting limits." World Economic Forum, 11 March 2025. https://www.weforum.org/stories/2025/03/ai-red-lines-uses-behaviours/
Karnofsky, Holden. 2024. "A Sketch of Potential Tripwire Capabilities for AI." Carnegie Endowment for International Peace. December 10, 2024. A Sketch of Potential Tripwire Capabilities for AI | Carnegie Endowment for International Peace

IV. Ces recommandations contribuent à la mise en œuvre du Pacte pour l'avenir des Nations unies et d'autres initiatives des Nations unies

De multiples initiatives des Nations unies appellent au développement d'une IA sûre, sécurisée et fiable. Parmi celles-ci, les résolutions de l'Assemblée générale des Nations unies sur l'IA – A/78/L.49, A/78/L.86 et A/C.1/79/L.43 – ainsi que le Pacte des Nations unies pour l'avenir, le Pacte numérique mondial et la Recommandation de l'UNESCO sur l'éthique de l'IA appellent à une coopération internationale afin de développer une IA bénéfique pour toute l'humanité, tout en gérant de manière proactive les risques mondiaux.

Ces initiatives ont attiré l'attention du monde entier sur les formes actuelles d'IA. Le présent rapport s'appuie sur ces initiatives des Nations unies en abordant spécifiquement le développement de l'AGI dans un avenir proche. Les engagements pris dans le cadre du Pacte pour l'avenir sont développés de plusieurs manières dans ce rapport. Une session de l'Assemblée générale des Nations unies consacrée à l'AGI répond à l'engagement du Pacte pour l'avenir en faveur d'un dialogue mondial sur la gouvernance de l'IA. Les recommandations de ce rapport concernant une convention-cadre des Nations unies sur l'AGI et une étude de faisabilité pour la création d'une agence des Nations unies sur l'AGI. L'Observatoire que nous avons proposé soutiendrait les travaux du futur Panel scientifique international indépendant sur l'IA, l'un des principaux résultats du Pacte numérique mondial. Enfin, le Système international de bonnes pratiques et de certification pour une AGI sûre et fiable contribuerait à la confiance et à la transparence, comme le demandent les résolutions de l'Assemblée générale des Nations unies, l'UNESCO et le Pacte pour l'avenir.

V. Conclusion

Il est urgent de sensibiliser les dirigeants nationaux et internationaux aux avantages et aux risques de l'AGI future, qui se distingue des formes actuelles d'IA. L'Assemblée générale des Nations unies est l'instance appropriée pour lancer un tel débat mondial.

Une coordination internationale du développement et de l'utilisation de l'AGI sera nécessaire pour tirer parti des avantages extraordinaires de l'AGI tout en préservant les droits de l'homme et la sécurité. Le groupe d'experts sur l'AGI recommande vivement à l'Assemblée générale des Nations unies d'agir de toute urgence pour traiter ces questions lors d'une session de l'Assemblée générale consacrée spécifiquement à un cadre de gouvernance mondiale pour l'AGI. Sans une telle action, les risques liés au développement et à l'utilisation incontrôlés de l'AGI, allant d'une augmentation spectaculaire des inégalités mondiales à des menaces existentielles, sont immenses.

Cette approche menée par les Nations unies, qui comprend un observatoire mondial, une certification internationale, une convention des Nations unies sur l'AGI et une agence dédiée à l'AGI, augmente la probabilité que l'AGI soit développée et utilisée de manière à bénéficier à l'ensemble de l'humanité tout en minimisant les risques. Ce cadre doit être inclusif, transparent et applicable afin de favoriser la confiance et la coopération entre les nations.

Annexe

Termes de référence : Panel de haut niveau sur l'intelligence artificielle générale (IAG) pour le Conseil des présidents de l'Assemblée générale des Nations unies (UNCPGA)

Mandat

Groupe de haut niveau sur l'intelligence artificielle générale (AGI) pour le Conseil des présidents de l'Assemblée générale des Nations unies (UNCPGA).

Contexte

La Déclaration de Séoul 2024 de l'UNCPGA appelle à la création d'un groupe d'experts en intelligence artificielle générale (AGI) chargé de fournir un cadre et des lignes directrices à l'Assemblée générale des Nations Unies pour l'aider à traiter les questions urgentes liées à la transition vers l'intelligence artificielle générale.

Ce travail doit s'appuyer sur les efforts considérables déjà déployés en matière de valeurs et de principes de l'IA par l'UNESCO, l'OCDE, le G20, le G7, le Partenariat mondial sur l'IA et la Déclaration de Bletchley, ainsi que sur les recommandations du Groupe consultatif de haut niveau du Secrétaire général des Nations unies sur l'IA, le Pacte mondial des Nations unies pour le numérique, le Réseau international des instituts de sécurité de l'IA, la Convention-cadre du Conseil européen sur l'IA et les deux sur l'IA, le Pacte mondial numérique des Nations unies, le Réseau international des instituts de sécurité de l'IA, la Convention-cadre du Conseil européen sur l'IA et les deux résolutions de l'Assemblée générale des Nations unies sur l'IA. Ceux-ci se sont davantage concentrés sur des formes plus restreintes d'IA. Il existe actuellement un manque d'attention similaire envers l'AGI.

L'IA est bien connue dans le monde d'aujourd'hui et souvent utilisée, mais l'AGI ne l'est pas et n'existe pas encore. De nombreux experts en AGI estiment qu'elle pourrait être mise au point d'ici un à cinq ans et qu'elle pourrait finalement évoluer vers une superintelligence artificielle échappant à notre contrôle.

Il n'existe pas de définition universellement acceptée de l'AGI, mais la plupart des experts s'accordent à dire qu'il s'agirait d'une IA polyvalente capable d'apprendre, de modifier son code et d'agir de manière autonome pour résoudre de nombreux problèmes nouveaux à l'aide de solutions novatrices similaires ou supérieures aux capacités humaines.

L'IA actuelle ne dispose pas de ces capacités, mais la trajectoire des progrès techniques va clairement dans ce sens. Le Pacte mondial numérique des Nations unies appelle à un dialogue mondial sur la gouvernance de l'IA au sein des Nations unies. Les experts du secteur privé en matière d'AGI ont souligné l'urgence d'un débat mondial afin de mieux comprendre les opportunités et les risques liés à l'AGI. Une session extraordinaire de l'Assemblée générale des Nations unies sur l'AGI est probablement le moyen le plus rapide, le plus rentable et le plus court pour stimuler un tel débat.

Objectif

En réponse à la Déclaration de Séoul 2024 de l'UNCPGA, produire un rapport initial à l'intention du président de l'UNCPGA et de ses membres pour la réunion de l'UNCPGA qui se tiendra à Bratislava du 8 au 10 avril 2025.

Le rapport doit identifier les risques, les menaces et les opportunités liés à l'AGI. Il doit mettre l'accent sur la sensibilisation à la mobilisation de l'Assemblée générale des Nations unies afin d'aborder la gouvernance de l'AGI de manière plus systématique. Il doit se concentrer sur l'AGI qui n'a pas encore été réalisée, plutôt que sur les formes actuelles de systèmes d'IA plus restreints. Il doit souligner l'urgence de traiter les questions relatives à l'AGI dès que possible, compte tenu des développements rapides de l'AGI, qui peuvent présenter des risques sérieux pour l'humanité ainsi que des avantages extraordinaires pour celle-ci.

Le rapport devrait également inclure à la fois les accords multilatéraux et les actions du secteur privé visant à relever ces défis sans précédent. Il devrait répondre aux appels lancés par les leaders du secteur privé dans le domaine de l'AGI en faveur d'une coordination internationale et d'une action multilatérale pour relever ce qui pourrait être le défi de gestion le plus difficile auquel l'humanité n'ait jamais été confrontée.

Procédures

● Convoquer un groupe d'experts internationaux de haut niveau (5 à 8 membres) sur l'IA générale afin d'examiner les menaces potentielles de l'IA générale pour l'humanité, les opportunités qu'elle pourrait offrir à l'humanité et les questions politiques connexes.

● Le groupe d'experts sur l'AGI se réunira virtuellement à intervalles réguliers à partir de janvier 2025 et achèvera son rapport initial pour la prochaine réunion de l'UNCPGA à Bratislava au printemps 2025.

● Sur la base des commentaires formulés sur le rapport initial lors de la réunion de l'UNCPGA à Bratislava, le groupe d'experts finalisera le rapport et le soumettra au secrétaire général de l'UNCPGA. S'il est accepté par le président de l'UNCPGA, il sera transmis au président de l'Assemblée générale des Nations unies, provisoirement avant le 1er mai 2025.

Membres du groupe d'experts indépendants de haut niveau sur l'AGI pour le Conseil des présidents de l'Assemblée générale des Nations unies

Jerome Glenn (EE.UU.), Président

Président Membre votant de l'IEEE Organizational Governance of AI ; auteur du document de l'Union européenne Horizon 2025-27 sur l'AGI : enjeux et opportunités ; PDG du Millennium Project et auteur de son document International Governance Issues of the Transition from Artificial Narrow à l'AGI, Requirements for Global Governance of AGI (Exigences pour la gouvernance mondiale de l'AGI) et Work/Technology 2050: Scenarios and Actions (Travail/Technologie 2050 : scénarios et actions). Auteur de Future Mind: Artificial Intelligence (Esprit futur : intelligence artificielle) (1989).

Renan Araujo (Brésil)

Directeur de recherche à l'Institute for AI Policy and Strategy (Institut pour la politique et la stratégie en matière d'IA), spécialisé dans la gestion des risques liés au développement de l'AGI. Il dirige actuellement les travaux de l'IAPS sur la gouvernance internationale de l'AGI. Il est membre de l'Oxford China Policy Lab, avocat, cofondateur de la Condor Initiative (qui met en relation des étudiants brésiliens avec des opportunités de classe mondiale pour façonner la recherche et la politique en matière d'IA) et a travaillé sur des programmes de gouvernance de l'IA chez Rethink Priorities et à l'Institute for Law and AI.

Yoshua Bengio (Canada)

Professeur d'informatique à l'Université de Montréal ; président du groupe consultatif sur la sécurité et la sûreté de l'IA pour le gouvernement canadien ; président du rapport international sur la sécurité de l'IA mandaté par 30 pays ainsi que l'ONU, l'OCDE et l'UE. Directeur scientifique de Mila, l'Institut québécois d'IA ; membre du Conseil consultatif scientifique du Secrétaire général des Nations unies pour les avancées scientifiques et technologiques ; lauréat du prix Turing et actuellement l'informaticien le plus cité au monde.

Joon Ho Kwak (République de Corée)

Conseiller technique de l'Institut coréen pour la sécurité de l'IA ; a joué un rôle de premier plan dans l'élaboration des lignes directrices de l'OCDE pour le développement d'une IA fiable ; participant au processus du G7 à Hiroshima, aux préparatifs du sommet de Paris sur l'IA, au groupe de travail Corée-États-Unis sur l'IA, et membre de la délégation coréenne auprès du Réseau international des instituts pour la sécurité de l'IA.

Lan Xue (Chine)

Président du Comité national d'experts sur la gouvernance de l'IA ; doyen de l'Institut pour la gouvernance internationale de l'IA International Governance de l'université Tsinghua ; membre du groupe consultatif de la direction STI de l'OCDE ; conseiller pour le China AI Safety Institute ; coprésident du Leadership Council du Réseau des solutions pour le développement durable des Nations unies (UNSDSN) ; lauréat du prix Fudan Distinguished Contribution Award for Management Science et du Distinguished Contribution Award de l'Association chinoise des sciences et des politiques scientifiques et technologiques.

Stuart Russell (Royaume-Uni et États-Unis)

Professeur émérite d'informatique et directeur du Center for Human-Compatible AI, Université de Californie, Berkeley ; auteur de Artificial Intelligence: A Modern Approach, le manuel de référence sur l'IA utilisé dans 1 500 universités à travers 135 pays et cité plus de 74 000 fois ; coprésident du groupe d'experts de l'OCDE sur l'avenir de l'IA et du Global AI Council du World Economic Forum.

Jaan Tallinn (Estonie)

Membre de l'organe consultatif des Nations unies sur l'IA ; a siégé au groupe d'experts de haut niveau de la CE sur l'IA ; cofondateur du Center for the Study of Existential Risk et du Future of Life Institute de l'université de Cambridge (deux institutions leaders dans le domaine de l'AGI) ; membre du conseil d'administration du Center for AI Safety ; investisseur estonien dans la sécurité de l'AGI ; ingénieur fondateur de Skype et de FastTrack/Kazaa ; et directeur investisseur fondateur de DeepMind Google.

Mariana Todorova (Bulgarie)

Représentante bulgare au sein du Groupe intergouvernemental de l'UNESCO sur les cadres éthiques de l'IA ; porte-parole de premier plan sur l'AGI dans les médias bulgares ; auteure et conférencière de renommée internationale sur les dimensions éthiques et technologiques de l'IA et de l'AGI ; ancienne membre du Parlement et conseillère du président de la République de Bulgarie.

José Jaime Villalobos (Costa Rica)

Responsable de la gouvernance multilatérale au Future of Life Institute ; chercheur associé principal au Centre for International Governance Innovation ; chercheur affilié à l'Oxford Martin AI Governance Initiative ; chercheur affilié à l'Institute for Law & AI ; titulaire d'un doctorat en droit international ; et coauteur d'ouvrages et d'articles de référence sur la gouvernance internationale de l'IA.