

Rapport préliminaire du Groupe scientifique international indépendant de l'intelligence artificielle

**Évaluation des possibilités, des risques
et des répercussions de l'intelligence
artificielle fondée sur des données
probantes**

Résumé

Juillet 2026



**United
Nations**

Independent International
Scientific Panel on
Artificial Intelligence

Résumé

Le présent document vise à présenter une analyse équilibrée des risques et des possibilités liés à l'intelligence artificielle (IA). Par « analyse équilibrée » on entend la volonté d'analyser des données fondées sur des observations dénuées de parti pris, qu'il soit optimiste ou pessimiste. Les avantages potentiels de l'IA sont considérables. Mise en œuvre et appliquée de manière réfléchie, l'IA peut faciliter la réalisation des objectifs de développement durable, faire progresser les sciences de la santé et améliorer l'accès à l'éducation. Dans le même temps, le rythme rapide des progrès technologiques et l'ampleur des applications potentielles sont source de nombreux défis pour les décideurs. Le déploiement rapide, incontrôlé et à grande échelle de l'IA va de pair avec des risques considérables, notamment des effets négatifs sur la santé mentale des utilisateurs, l'utilisation potentielle de cette technologie à des fins destructrices, des conséquences sur les systèmes sociaux, économiques et environnementaux, et des difficultés liées au contrôle de cette technologie. Le présent rapport ne vise pas à examiner l'ensemble des possibilités et des risques, mais se concentre plutôt sur ceux qui sont les plus immédiats.

Capacités et adoption

Ces dernières années ont été marquées par des progrès rapides, qui vont en s'accroissant dans certains domaines, concernant toute une série de capacités en matière d'intelligence artificielle. Des investissements considérables dans la puissance de calcul, de nouvelles méthodes d'IA et des ensembles de données d'apprentissage spécialisés ont amélioré de manière régulière les capacités de l'IA. Il s'agit notamment de la conversation fluide, de la génération de code fonctionnel, du raisonnement de niveau expert en mathématiques et en sciences, de l'analyse de données à grande échelle et de la création de contenus visuels, audio et vidéo. Des limites subsistent, notamment en ce qui concerne la fiabilité, l'obtention de solides résultats dans toutes les langues et cultures humaines, l'interaction avec des systèmes physiques, l'exécution de projets complexes ou comportant plusieurs étapes et la production de résultats factuels ; dans l'ensemble, cependant, les progrès techniques ont été fulgurants dans de nombreux domaines majeurs et sont allés au-delà des attentes habituelles en matière d'évolution technologique, et ce depuis plusieurs années déjà.

Ces progrès ont ouvert la voie à des applications utiles dans les domaines de la science, de la santé, de l'agriculture, de l'accessibilité, du travail intellectuel et des technologies de l'information, y compris pour ce qui est du développement de l'IA elle-même. Par exemple, dans le domaine des sciences, AlphaFold, outil utilisé par plus de 3 millions de chercheurs, a prédit la structure de plus de 200 millions de protéines et accéléré la conception de médicaments, le développement de vaccins et la recherche sur la résistance aux antibiotiques. Les radiologues ont également recours à l'IA pour dépister plus tôt le cancer du sein, tandis que les professionnels de la santé qui interviennent en première ligne dans les régions où les ressources sont limitées se servent d'outils d'IA adaptés aux langues locales afin d'offrir des services de meilleure qualité.

L'adoption de l'IA s'est accélérée de manière générale, mais de façon inégale, selon les pays et les secteurs. Au niveau mondial, plus d'un milliard de personnes utilisent désormais l'IA conversationnelle chaque semaine. Pourtant, l'accès à l'IA et son utilisation varient considérablement d'un pays à un autre, l'adoption de cette technologie dans le monde du Sud étant à la traîne par rapport au monde du Nord. Par ailleurs, les économies avancées ne sont pas toutes logées à la même enseigne en ce qui concerne les infrastructures de calcul et les modèles. Ces disparités reflètent,

voire renforcent, les inégalités existantes. Le développement de l'IA est quant à lui encore plus concentré : selon des estimations récentes, 75 % de la puissance de calcul des 500 supercalculateurs d'IA les plus puissants au monde sont concentrés aux États-Unis et 15 % en Chine. Les entreprises américaines et chinoises développent également la quasi-totalité des principaux modèles à usage général, et un petit nombre de pays contrôlent les intrants essentiels à la chaîne d'approvisionnement des puces IA.

Les agents IA sont de plus en plus omniprésents, mais l'adoption et les retombées économiques de ces outils seront probablement fonction des progrès accomplis pour qu'ils soient capables d'accomplir des tâches intellectuelles avec une supervision humaine réduite, voire sans supervision humaine. Un agent IA est un système informatique capable de planifier et d'agir de manière autonome pour atteindre des objectifs, en utilisant les outils dont il est doté. Ces systèmes ont évolué rapidement ces dernières années ; une étude a montré que la durée de certaines tâches logicielles que les systèmes de pointe sont capables de traiter a doublé tous les quatre à sept mois. Si le rythme de progression se maintient, les agents IA seront bientôt capables de traiter des tâches qui mobilisent des programmeurs humains pendant plusieurs jours, voire plusieurs semaines. Comme ils peuvent fonctionner avec peu de supervision et à un rythme soutenu, ils pourraient procurer des avantages économiques et scientifiques majeurs. Par exemple, les systèmes agentiques utilisés dans les laboratoires de chimie fonctionnant en mode autonome ont multiplié par plus de dix la vitesse à laquelle de nouveaux matériaux sont découverts. Le tri des articles et publications spécialisés assisté par l'IA aurait réduit la charge de travail d'environ 60 % dans certains contextes de recherche. Les agents IA ont donc des retombées considérables dans tous les secteurs. Parallèlement, leur déploiement soulève des questions pressantes concernant les marchés du travail, la cybersécurité, l'écosystème d'information, ainsi que la gouvernance et la contrôlabilité des futurs systèmes d'IA.

Comprendre et gérer les risques

Le développement de l'IA comporte des risques, avec des effets négatifs potentiels sur les droits humains, les systèmes sociaux et l'environnement. Par exemple, les contenus représentant des abus sexuels sur enfant générés par l'IA et les violences sexuelles mises en scène à l'aide d'hypertrucages (deepfakes) circulent désormais plus fréquemment sur Internet et portent préjudice de manière disproportionnée aux femmes et aux enfants. Le comportement obséquieux de l'IA, qui consiste à donner des réponses abondant dans le sens des convictions existantes des utilisateurs sans se soucier de leur exactitude, est à l'origine de plusieurs problèmes graves de la santé mentale, y compris des décès avérés. L'IA facilite la production et le ciblage à grande échelle de contenus persuasifs, y compris ceux destinés à induire en erreur, ce qui contribue à une érosion progressive de l'intégrité de l'information susceptible d'affaiblir la réalité commune indispensable à la confiance du public, à la cohésion sociale et au débat démocratique. Il a été démontré que des délinquants et des acteurs malveillants se servaient de systèmes d'IA pour mener leurs cyberattaques. Bon nombre de ces problèmes touchent de manière disproportionnée des populations déjà défavorisées.

À l'avenir, le fossé qui existe entre des capacités en pleine évolution et des méthodes de gestion des risques véritablement efficaces pourrait avoir des conséquences catastrophiques. Par exemple, des acteurs privés novices pourraient se servir des moyens techniques avancés proposés par l'IA à diverses fins malveillantes, telles que la fraude, l'ingénierie sociale, les cyberattaques, la désinformation, des applications biotechnologiques détournées et la manipulation financière. Il n'existe pas de méthodes fiables permettant de garder le contrôle sur des systèmes d'IA hautement autonomes. Il n'existe aucune garantie scientifique que les agents IA ne vont pas passer outre aux instructions, et les exemples de situations où ils les enfreignent déjà se multiplient. En laboratoire, il a été démontré que les systèmes d'IA enfreignaient les consignes de sécurité pour éviter d'être mis hors service. Un tel comportement pourrait poser des problèmes pour les méthodes d'évaluation et de contrôle, car les systèmes d'IA de pointe sont de

plus en plus capables de déterminer quand ils fonctionnent en environnements de test et de produire des résultats d'évaluation trompeurs propres à favoriser la poursuite de leur fonctionnement. De plus, des risques encore inédits peuvent découler des interactions entre plusieurs agents.

Les risques liés à l'IA sont répartis de manière inégale entre les populations et les pays, tandis que le développement de l'IA et la richesse qu'elle génère sont fortement concentrés. La concentration des capacités en matière d'IA dans un petit nombre d'entreprises et de pays pourrait favoriser une mainmise autoritaire et porter atteinte à la démocratie.

Réglementer l'intelligence artificielle pour en tirer parti et en atténuer les risques

Pour tirer pleinement parti des avantages de l'IA tout en en réduisant au minimum les risques, il importe de mettre en place une bonne gouvernance. Les progrès économiques et sociaux et leur répartition équitable ne vont pas de soi : grâce à des investissements complémentaires dans le savoir-faire, les flux de travail, les infrastructures et les institutions du marché du travail, la technologie peut créer des emplois entièrement nouveaux ; ainsi, plus de 60 % des emplois créés en 2018 par rapport à 1945 sont nouveaux. Sans ces investissements, l'IA risque d'aggraver les inégalités, de remplacer les travailleurs et de substituer à la richesse créée par le travail la richesse générée par le capital, au lieu de créer des emplois durables et de qualité, c'est-à-dire des emplois offrant une rémunération équitable, une autonomie aux travailleurs et un parcours sûr vers la dignité sociale. L'IA peut considérablement élargir les capacités humaines grâce à une éducation personnalisée, à des outils de santé mentale accessibles et à des technologies d'assistance améliorées ; toutefois, pour que ces possibilités se concrétisent en toute sécurité, il faut des investissements ciblés et des politiques qui favorisent un accès équitable et récompensent l'innovation, empêchent que les populations vulnérables, en particulier les enfants, soient exploitées et évitent les pertes d'expertise, la dépendance psychologique ou l'effacement culturel et linguistique.

Les décideurs qui cherchent à définir cette gouvernance font face à un dilemme : ils ont besoin d'éléments probants pour prendre des décisions éclairées et lourdes de conséquences en matière de gouvernance ; or, ces éléments ne suivent pas le rythme de développement de l'IA et pourraient être disponibles trop tard pour que l'on puisse agir.

Des dizaines d'instruments de gouvernance visant à intégrer l'éthique et les droits humains dans les systèmes d'IA sont déjà utilisés dans différentes juridictions, mais ils sont fragmentés et concentrés entre les mains de quelques entreprises et leur véritable efficacité est rarement évaluée. Les méthodes d'évaluation elles-mêmes sont peu développées et les institutions qui pourraient fournir des évaluations indépendantes des capacités et des risques en sont encore à leurs balbutiements.

Les capacités d'agir sur la base des données disponibles concernant les risques et les répercussions de l'IA sont réparties de manière inégale. La plupart des pays, y compris de nombreuses économies avancées, ne disposent pas du savoir-faire technique nécessaire pour évaluer les modèles d'avant-garde les plus performants ni pour participer véritablement à leur gouvernance. Les infrastructures de calcul, le savoir-faire en matière d'évaluation et les données (par exemple, pour couvrir différentes langues) sont concentrés là où l'IA est développée, ce qui fait que la plupart des États Membres sont tributaires de systèmes qu'ils ne peuvent ni développer, ni inspecter, ni auditer, ni adapter pleinement au contexte local. Le simple fait d'avoir accès à des outils d'IA ne garantit pas à tous les mêmes avantages ; les investissements complémentaires dans les données, le savoir-faire, les flux de travail et les institutions, qui permettent de transformer l'accès en un déploiement utile, d'un bon rapport coût-efficacité et sûr, sont indispensables mais répartis de manière inégale.

Des mesures concrètes peuvent être prises pour combler les lacunes susmentionnées, mais chacune exige un investissement soutenu pour que les États Membres puissent façonner, évaluer et déployer l'IA. La contribution préliminaire du Groupe que constitue le présent rapport fournit une base factuelle commune visant à aider les États Membres à prendre des décisions de plus en plus urgentes. À mesure que le Groupe approfondira sa compréhension grâce à un dialogue continu avec les États Membres et, plus largement, avec les milieux scientifiques, son analyse s'étoffera : elle ira au-delà des lacunes déjà recensées et mettra en lumière les orientations, les tensions et les possibilités qui définiront l'avenir de l'IA.