



INTELLIGENCE ARTIFICIELLE ET SOCIÉTÉ

L'intelligence artificielle (IA) est en train de remodeler notre monde, offrant de nombreux avantages mais soulevant également des préoccupations critiques auxquelles il convient de répondre par une gouvernance inclusive et coordonnée à l'échelle mondiale. En octobre 2023, les dirigeants du G7 ont publié les principes directeurs internationaux du processus d'Hiroshima et le code de conduite qui les accompagne à l'intention des organisations publiques et privées qui développent et utilisent des systèmes d'IA avancés. Avec la présente déclaration commune, nous exprimons la position des Académies des sciences du G7 sur les préoccupations techniques et sociétales soulevées par l'IA, avec une liste de recommandations à l'intention des décideurs politiques.

1. Les puissants systèmes d'IA, susceptibles de poser des risques pour le public s'ils sont mal utilisés, devraient être protégés de manière appropriée contre les cyber-attaques et les attaques physiques. Pour les applications sensibles de l'IA, des cadres de protection des données (par exemple, le règlement général sur la protection des données de l'UE) proportionnels au niveau de risque et l'utilisation de technologies d'amélioration de la confidentialité sont essentiels pour protéger les données à caractère personnel. Lorsqu'ils interagissent avec des systèmes d'IA, les utilisateurs ont besoin de conseils clairs quant à la protection des données, sur la manière dont celles-ci seront utilisées, réutilisées et stockées, et sur la durée de cette utilisation.
2. Les systèmes d'IA devenant plus performants et présentant des risques plus importants pour le contrôle humain, des garanties de sécurité formelles seront nécessaires avant que les systèmes d'IA ne puissent être déployés. Il est essentiel d'adhérer à des normes strictes en matière de vérification des exigences, de validation et d'essai des systèmes, ainsi que de suivi post-déploiement, idéalement par des organismes indépendants qui vérifient également l'absence de biais cachés. Des lois devraient être promues pour renforcer le besoin de documentation, permettant aux utilisateurs de comprendre les capacités, les limites et les domaines d'application appropriés. La traçabilité des données de base sur lesquelles les modèles d'IA sont construits est essentielle. L'IA ayant un impact sur des décisions cruciales dans des domaines tels que la médecine, les sciences, le droit et la finance, la transparence et la responsabilité sont primordiales. Des voies claires de responsabilité devraient être créées au fur et à mesure que les systèmes d'IA

sont développés et déployés. Des caractéristiques telles que l'explicabilité, le recours algorithmique et les voies de recours permettent aux individus de contester les décisions de l'IA.

3. La fiabilité des systèmes d'IA est vitale. À mesure que les humains passent plus de temps à interagir avec les systèmes d'IA et que le contenu généré par l'IA constitue un pourcentage de plus en plus important de notre écosystème d'information, les risques pour la démocratie, la société et la compréhension humaine deviennent critiques. Les gouvernements devraient promouvoir des normes légales pour les contenus générés par l'IA. Les structures institutionnelles qui promeuvent l'exactitude et l'authenticité deviendront de plus en plus importantes. En outre, il sera essentiel d'améliorer les connaissances des populations tout au long de leur vie afin qu'elles comprennent comment identifier les contenus générés par l'IA et interagir avec eux. Les développeurs de technologies et les plateformes de médias en ligne devraient prendre en compte la valeur des systèmes de provenance du contenu numérique ainsi que l'inclusion d'experts du domaine dans les évaluations socio-techniques (par exemple, par des tests de piratage).
4. L'IA générative offre un vaste potentiel en matière de création de contenu et de développement de logiciels, mais soulève des questions quant à la protection des droits de propriété intellectuelle lorsque l'IA crée un contenu apparemment nouveau à partir d'œuvres existantes. Les industries créatives et la recherche scientifique comptent parmi les principaux secteurs dans lesquels l'IA générative a un impact significatif. Le cadre juridique existant en matière de droits d'auteur, d'utilisation équitable et de protection de la propriété intellectuelle doit être revu pour l'ère de l'IA. Pour résoudre ces problèmes, il faut une combinaison de législation claire, d'autorégulation de l'industrie et de contrôle indépendant.
5. Les systèmes d'IA à grande échelle nécessitent des ressources informatiques coûteuses, ce qui constitue une barrière à l'entrée. Il est essentiel pour l'avenir équitable de l'IA d'élargir l'accès mondial à des moyens de calcul performants et d'établir des centres de recherche intergouvernementaux. L'expansion des ressources informatiques de l'IA soulève des préoccupations environnementales en raison de l'augmentation de la consommation d'énergie. La conception durable des centres de données, les emplacements stratégiques pour l'efficacité énergétique et la coopération, ainsi que les nouvelles techniques d'IA économes en énergie sont essentiels pour aligner l'IA sur l'éthique environnementale.
6. L'accès à des bases de données libres de droits d'auteur est en principe un bien public à encourager, les grands modèles d'IA à code source ouvert présentent

un compromis complexe entre les avantages (par exemple, pour les chercheurs du secteur public et les petits développeurs en aval) et les risques d'utilisation abusive qui sont supportés par le public. Les décisions autorisant ou limitant la diffusion de systèmes d'IA puissants en libre accès doivent être soumises à un contrôle démocratique, et les règles de sécurité qui s'appliquent aux systèmes propriétaires doivent également s'appliquer à tous les cas de systèmes en libre accès.

7. L'essor des systèmes d'armes autonomes suscite des préoccupations éthiques et militaires urgentes. Leur autonomie pourrait abaisser le seuil de déclenchement des conflits, stimuler une course aux armements et risquer d'être utilisée à mauvais escient par des régimes voyous ou des terroristes. Alors que certains préconisent l'interdiction de ces dispositifs, d'autres réclament des réglementations strictes conformes au droit humanitaire international, garantissant un contrôle humain. Une action rapide de la société est essentielle pour relever ce défi important. La transparence et les discussions publiques sont essentielles pour relever ces défis, afin de défendre les valeurs humaines et les normes internationales, ainsi que la sécurité physique aux niveaux personnel, local, national et international.
8. L'alignement de l'IA sur les valeurs humaines et l'éthique est nécessaire car les systèmes d'IA vont gagner en influence. La collaboration intersectorielle est essentielle à la mise en œuvre d'une IA éthique, intégrant à la fois les aspects techniques et socioculturels tout en tenant compte de la diversité culturelle. Il est essentiel d'aborder les risques à court terme, qui peuvent avoir un impact négatif sur les groupes sociaux vulnérables, afin de maintenir la confiance et d'encourager l'adoption de technologies bénéfiques. Les coûts économiques et sociaux de l'alignement devraient être transparents et inclus dans le processus de prise de décision concernant le moment et la manière dont les systèmes d'IA sont utilisés.
9. Une gestion efficace est essentielle pour une répartition équitable des bénéfices de l'IA. Les systèmes d'IA de plus en plus performants auront un impact significatif sur les marchés du travail, et une approche conjointe avec les gouvernements, l'industrie, les éducateurs et la société civile est essentielle pour garantir une répartition équitable des bénéfices. Les gouvernements devraient envisager des incitations pour les systèmes d'IA socialement utiles qui répondent à des besoins non satisfaits plutôt que de simplement remplacer les humains dans les rôles existants. Les politiques devraient donner la priorité au développement de systèmes d'IA qui répondent aux défis les plus importants pour la planète, la société et l'économie.

10. La coopération entre les secteurs public et privé favorise le développement responsable de l'IA. Alors que les entités privées innovent rapidement, les institutions publiques offrent des cadres réglementaires et éthiques et encouragent la recherche indépendante à long terme au profit de l'humanité. L'échange de talents permet de fusionner les meilleures pratiques et les objectifs. Cette synergie associe l'innovation à la responsabilité, optimisant ainsi les avantages sociétaux de l'IA. Elle est d'autant plus efficace que la société, par l'intermédiaire des régulateurs et des processus démocratiques, fixe les lignes rouges de ce qui est acceptable ou non, et qu'il incombe aux développeurs de démontrer que leurs systèmes ne franchissent pas ces lignes rouges.

11. Le développement et le déploiement responsables et sûrs de l'IA ne sont pas seulement une question scientifique et technique : ils nécessitent une approche collaborative entre les différentes disciplines universitaires, y compris les sciences humaines et sociales, pour comprendre et façonner le système sociotechnique complexe au sein duquel l'IA fonctionne et pour concevoir des accords de gouvernance appropriés. En combinant les points de vue techniques et non techniques, nous pouvons relever les défis de l'IA tels que les préjugés et l'équité, en garantissant un déploiement inclusif, solide, transparent et équitable de l'IA.

12. L'éducation devrait clarifier les capacités de l'IA, démystifier les mythes et peser le pour et le contre. Les citoyens devraient avoir des connaissances techniques et être conscients des implications sociales et éthiques de l'IA. Le dialogue public, facilité par diverses parties prenantes, permet aux individus de façonner activement les politiques en matière d'IA et d'aligner l'IA sur l'intérêt public. Les risques et les opportunités liés à l'intégration de puissants systèmes d'IA dans le cadre éducatif doivent être soigneusement équilibrés.