



Principes concernant l'IA

Nos objectifs pour le développement
d'une IA bénéfique

Responsabilité : nos principes

Malgré notre optimisme quant au potentiel de l'IA, nous reconnaissons que les technologies avancées peuvent soulever des questions importantes auxquelles il faut répondre de manière claire, réfléchie et positive. Ces principes concernant l'IA décrivent notre engagement à développer la technologie de manière responsable et à définir les domaines d'application spécifiques que nous excluons de nos recherches.

Nos objectifs concernant les applications de l'IA

1. Être bénéfique à la société.

La portée accrue des nouvelles technologies touche de plus en plus la société dans son ensemble. Les avancées de l'IA transformeront un grand nombre de domaines, tels que la santé, la sécurité, l'énergie, les transports, l'industrie et les loisirs. Nous prendrons en compte un grand nombre de facteurs sociaux et économiques pour évaluer les développements potentiels des technologies d'IA et leurs utilisations. Nous consacrerons nos travaux aux domaines pour lesquels les avantages probables sont dans l'ensemble plus importants que les risques et les inconvénients prévisibles.

L'IA accroît également notre capacité à appréhender la signification de contenus à grande échelle. Nous ferons notre possible pour que des informations exactes et de qualité soient facilement accessibles à l'aide de l'IA, tout en respectant toujours les normes culturelles, sociales et légales des pays dans lesquels nous exerçons notre activité. En outre, nous continuerons à évaluer avec la plus grande rigueur le moment opportun pour rendre nos technologies disponibles pour un usage non commercial.

2. Éviter de créer ou de renforcer des injustices.

Les algorithmes et les ensembles de données de l'IA peuvent refléter, renforcer ou réduire les partis pris injustes. Nous reconnaissons que l'appréciation du caractère juste ou injuste d'une opinion n'est pas simple et qu'elle diffère selon les cultures et les sociétés. Nous nous engageons à éviter les impacts injustes sur les individus, en particulier ceux qui concernent des caractéristiques sensibles, telles que l'origine ethnique, l'appartenance ethnique, le genre, la nationalité, les revenus, l'orientation sexuelle, les aptitudes, ainsi que les convictions politiques ou les croyances religieuses.

3. Être conçue et testée pour garantir la sécurité.

Nous continuerons de développer et d'appliquer des pratiques rigoureuses en matière de protection et de sécurité afin d'éviter toute conséquence involontaire susceptible de nuire à autrui. Nous nous engageons à concevoir nos systèmes d'IA avec toute la prudence nécessaire et à les développer en respectant de bonnes pratiques de recherche sur la sécurité de l'IA. Le cas échéant, nous testerons les technologies d'IA dans des environnements restreints et nous surveillerons leur fonctionnement après déploiement.

4. Être humainement responsable.

Nous nous engageons à concevoir des systèmes d'IA qui laissent la possibilité d'adresser des commentaires, d'obtenir des explications pertinentes et de faire appel. Nos technologies d'IA feront l'objet d'une supervision et d'un contrôle humains appropriés.

5. Intégrer des principes de confidentialité dès la conception.

Nous nous engageons à intégrer nos principes de confidentialité au développement et à l'utilisation de nos technologies d'IA. Nous y intégrerons la possibilité d'être averti et de donner son consentement, et nous privilégierons les architectures protégeant la confidentialité. Nous offrirons également la transparence et le contrôle nécessaires quant à l'utilisation des données.

6. Respecter des normes élevées d'excellence scientifique.

L'innovation technologique se fonde sur la méthode scientifique et sur les principes de libre investigation, de rigueur intellectuelle, d'intégrité et de collaboration. Les outils d'IA sont susceptibles d'ouvrir de nouvelles portes en matière de recherche et de connaissance scientifiques dans des domaines essentiels, tels que la biologie, la chimie, la médecine et les sciences de l'environnement. Nous nous attachons à respecter des normes élevées d'excellence scientifique dans nos travaux de développement de l'IA.

7. Être disponible pour les utilisations qui respectent ces principes.

De nombreuses technologies permettent plusieurs usages. Nous nous emploierons à limiter les applications potentiellement dangereuses ou abusives. Lors du développement et du déploiement de technologies d'IA, nous évaluerons les utilisations probables en tenant compte des facteurs suivants :

- Finalité et utilisation principales : la finalité principale et l'utilisation probable d'une technologie et d'une application, y compris la facilité avec laquelle il est possible d'associer ou d'adapter la solution à une utilisation dangereuse
- Nature et caractère unique : détermine si la technologie que nous proposons est unique ou plus largement disponible
- Portée : détermine si l'utilisation de la technologie aura un impact significatif
- Nature de l'implication de Google : détermine si nous proposons des outils généraux, si nous intégrons des outils pour les clients ou si nous développons des solutions personnalisées

Applications d'IA que nous excluons de nos recherches

Outre les objectifs définis ci-dessus, nous nous engageons à ne pas concevoir ni déployer de technologies d'IA dans les domaines d'application suivants :

1. Les technologies qui présentent ou sont susceptibles de présenter un danger en général. En cas de risque substantiel, nous ne poursuivons nos recherches que si nous estimons que les avantages sont nettement plus importants que les risques. Dans ce cas, nous intégrerons des restrictions de sécurité adaptées.
2. Les armes ou d'autres technologies dont l'objectif principal ou la mise en œuvre causent des blessures aux individus ou les facilitent directement.
3. Les technologies qui recueillent ou utilisent des informations pour la surveillance, en violation des normes en vigueur à l'échelle internationale.
4. Les technologies dont l'objectif enfreint les principes largement acceptés du droit international et des droits humains.

Cette liste est susceptible d'évoluer à mesure que notre expérience dans ce domaine s'approfondit.