



Principios de la IA

Objetivos para crear
IA beneficiosa

Responsabilidad: Nuestros principios

Si bien somos optimistas respecto de las posibilidades de la IA, reconocemos que las tecnologías avanzadas pueden generar desafíos importantes que deben abordarse con un enfoque claro, cuidadoso y positivo. Estos Principios de la IA describen nuestro compromiso de desarrollar tecnología de forma responsable y trabajar para definir áreas de aplicación específicas que no exploraremos.

Objetivos de los usos de la IA

1. Ser beneficiosa a nivel social.

El alcance ampliado de las nuevas tecnologías llega cada vez más a la sociedad en su totalidad. Los avances en IA tendrán impactos transformadores en una amplia variedad de campos, entre los que se incluyen salud, seguridad, energía, transporte, industria manufacturera y entretenimiento. A medida que consideremos el potencial desarrollo y los usos de las tecnologías de IA, tomaremos en cuenta una amplia variedad de factores sociales y económicos, y continuaremos únicamente cuando creamos que los beneficios superan considerablemente los riesgos y las desventajas previsibles.

La IA también mejora nuestra capacidad para comprender el significado del contenido a escala. Nos esforzaremos por lograr que haya información precisa y de alta calidad disponible para su uso a través de IA y, a su vez, seguir respetando las normas culturales, sociales y legales de los países donde operamos. Seguiremos evaluando con dedicación cuándo poner nuestras tecnologías a disposición con objetivos no comerciales.

2. Evitar crear o reforzar un sesgo parcial.

Los conjuntos de datos y los algoritmos de IA pueden reflejar, reforzar o reducir los sesgos parciales. Reconocemos que la distinción entre sesgo parcial o imparcial no siempre es simple, y varía entre las culturas y sociedades. Intentaremos evitar los impactos injustos sobre las personas, en particular aquellos aspectos relacionados con características sensibles, como raza, etnia, género, nacionalidad, ingresos, orientación sexual, habilidad y creencia política o religiosa.

3. Estar creada y probada para ofrecer seguridad.

Seguiremos desarrollando y aplicando sólidas prácticas de seguridad y protección para evitar resultados no intencionados que puedan producir daños. Diseñaremos nuestros sistemas de IA para que sean debidamente prudentes y buscaremos desarrollarlos en conformidad con las prácticas recomendadas para las investigaciones sobre seguridad de la IA. Cuando corresponda, probaremos las tecnologías de IA en entornos restringidos y supervisaremos su funcionamiento después de la implementación.

4. Ser responsable frente a las personas.

Diseñaremos sistemas de IA que brinden oportunidades adecuadas de comentarios, explicaciones relevantes y apelación. Nuestras tecnologías de IA estarán dirigidas y controladas de manera adecuada por personas.

5. Incorporar principios de diseño de privacidad.

Incorporaremos nuestros principios de privacidad al desarrollo y uso de nuestras tecnologías de IA. Daremos oportunidades de aviso y consentimiento, incentivaremos arquitecturas con protecciones de privacidad y proporcionaremos la transparencia y el control adecuados sobre el uso de los datos.

6. Mantener estándares altos de excelencia científica.

La innovación tecnológica está fundada en el método científico y un compromiso de investigación abierta, rigor intelectual, integridad y colaboración. Las herramientas de IA tienen el potencial de develar nuevos ámbitos de investigación científica y conocimientos en campos críticos, como la biología, la química, la medicina y las ciencias ambientales. A medida que trabajamos para progresar en el desarrollo de la IA, aspiramos a altos estándares de excelencia científica.

7. Estar disponible para usos que concuerden con estos principios.

Las diferentes tecnologías pueden tener una amplia variedad de usos. Trabajaremos para limitar las aplicaciones potencialmente dañinas o abusivas. A medida que desarrollemos tecnologías de IA y las implementemos, evaluaremos sus usos probables teniendo en cuenta los siguientes factores:

- Propósito y uso principales: El propósito principal y el uso probable de una tecnología y su aplicación, incluso hasta qué nivel la solución está relacionada con un uso dañino o se puede adaptar para este tipo de uso
- Naturaleza y exclusividad: Si estamos creando tecnología disponible de manera exclusiva o general
- Escala: Si el uso de esta tecnología tendrá un impacto significativo
- Naturaleza de la participación de Google: Si proporcionamos herramientas de uso general, integramos herramientas para los clientes o desarrollamos soluciones personalizadas

Usos que no le daremos a la IA

Además de los objetivos que se indicaron anteriormente, no diseñaremos ni implementaremos la IA en las siguientes áreas de aplicación:

1. Tecnologías que causen o puedan causar daño general. Cuando haya un riesgo concreto de daño, continuaremos únicamente cuando creamos que los beneficios superan considerablemente los riesgos y, además, incorporaremos las restricciones de seguridad adecuadas.
2. Armas o cualquier otra tecnología cuyo propósito o implementación principal sea causar lesiones a las personas o permitir esto de manera directa.
3. Tecnologías que recopilen o usen información para vigilancia que infrinjan las normas aceptadas a nivel internacional.
4. Tecnologías cuyo propósito desobedezca los principios ampliamente aceptados de leyes internacionales y derechos humanos.

A medida que adquiramos mayor experiencia en este ámbito, es posible que esta lista se extienda.