Google

Principios para la IA

Objetivos de desarrollar una IA beneficiosa

Responsabilidad: Nuestros principios

Aunque vemos con optimismo el potencial de la IA, somos conscientes de que las tecnologías avanzadas pueden plantear desafíos importantes que se deben abordar de forma clara, meditada y positiva. Estos principios para la IA describen nuestro compromiso de desarrollar tecnología de manera responsable y sirven para establecer áreas concretas de aplicación que no abordaremos.

Objetivos de las aplicaciones de la IA

1. Beneficiar a la sociedad.

El gran alcance de las nuevas tecnologías afecta cada vez más a toda la sociedad. Los avances en IA tendrán un impacto transformador en multitud de sectores, como la sanidad, la seguridad, la energía, el transporte, la fabricación y el entretenimiento. Cuando nos planteemos el desarrollo y los usos potenciales de las tecnologías de IA, tendremos en cuenta una amplia variedad de factores sociales y económicos, y seguiremos adelante cuando consideremos que los beneficios generales probables superan considerablemente a los riesgos e inconvenientes predecibles.

La IA también mejora nuestra capacidad de comprender el significado del contenido a gran escala. Haremos todo lo posible para poner a disposición de las personas información precisa y de alta calidad mediante IA, respetando las normas culturales, sociales y jurídicas de los países en los que operamos. Además, seguiremos evaluando de forma meditada en qué momento ofrecer nuestras tecnologías de manera no comercial.

2. Evitar crear o reforzar sesgos injustos.

Los algoritmos y conjuntos de datos de la IA pueden reflejar, reforzar o reducir sesgos injustos. Admitimos que no siempre es fácil distinguir la justicia en esos sesgos y que varían dependiendo de las culturas y las sociedades. Buscaremos la manera de evitar impactos injustos sobre las personas, en particular los relacionados con características sensibles, como raza, etnia, sexo, nacionalidad, ingresos, orientación sexual, capacidades y creencias políticas o religiosas.

3. Diseñarse y probarse pensando en la seguridad.

Seguiremos desarrollando y aplicando prácticas sólidas de seguridad y protección para evitar resultados no deseados que conlleven riesgos de causar daños. Diseñaremos nuestros sistemas de IA para que mantengan una precaución adecuada y buscaremos la manera de desarrollarlos de acuerdo con las prácticas recomendadas de las investigaciones sobre seguridad de la IA. En los casos pertinentes, probaremos las tecnologías de IA en entornos limitados y monitorizaremos su funcionamiento una vez implementadas.

4. Responder ante los usuarios.
Diseñaremos sistemas de IA que permitan a los usuarios dar feedback, obtener explicaciones pertinentes y reclamar si es necesario. Nuestras tecnologías de IA estarán dirigidas y controladas de forma competente por humanos.

5. Incorporar principios de diseño de la privacidad.
Incluiremos nuestros principios de privacidad en el desarrollo y el uso de nuestras tecnologías de IA. Ofreceremos la oportunidad de notificar y dar consentimiento, fomentaremos las arquitecturas que protejan la privacidad y proporcionaremos la transparencia y el control adecuados sobre el uso de los datos.

6. Cumplir estándares rigurosos de excelencia científica.
La innovación tecnológica se basa en el método científico y en el compromiso de plantear preguntas y mantener una actitud de rigor intelectual, integridad y colaboración. Las herramientas de IA tienen el potencial de abrir las puertas a nuevos campos de investigación y conocimiento científico de disciplinas fundamentales como la biología, la química, la medicina y las ciencias medioambientales. Aspiramos a alcanzar unos estándares altos de excelencia científica a medida que trabajamos para seguir desarrollando la IA.

7. Destinarse a usos acordes con estos principios.
Muchas tecnologías tienen varios usos. Trabajaremos para limitar las aplicaciones potencialmente dañinas o abusivas. Conforme desarrollemos e implementemos tecnologías de IA, evaluaremos los usos probables teniendo en cuenta los siguientes factores:

- El propósito y uso principales: el propósito principal y el uso probable de una tecnología y su aplicación, incluido el grado en que la solución esté relacionada o se pueda adaptar a un uso dañino
- La naturaleza y la singularidad: si presentamos una tecnología que sea única o que esté disponible de forma más general
- La escala: si el uso de esa tecnología tendrá o no un impacto significativo
- La naturaleza de la implicación de Google: si proporcionamos herramientas de uso general, si integramos herramientas para clientes o si desarrollamos soluciones personalizadas

Aplicaciones de la IA que no vamos a abordar

Además de los objetivos anteriores, no diseñaremos ni implementaremos IA en los siguientes ámbitos:

1. Tecnologías que causen o puedan causar daño en general. Cuando exista un riesgo importante de causar algún daño, seguiremos adelante solo cuando consideremos que los beneficios compensan considerablemente los riesgos, e incluiremos restricciones adecuadas para asegurar la seguridad.

2. Armamento u otras tecnologías cuya finalidad o aplicación principal sea ocasionar daños o herir a personas.

3. Tecnologías que recopilen o usen información para vigilar incumpliendo las normas aceptadas internacionalmente.

4. Tecnologías cuya finalidad contravenga los principios generalmente aceptados del derecho internacional y los derechos humanos.

Esta lista puede cambiar a medida que adquiramos experiencia con estas tecnologías.