



Princípios de IA

Objetivos para criar uma
IA benéfica

Responsabilidade: nossos princípios

Estamos otimistas em relação ao potencial da IA, mas reconhecemos que tecnologias avançadas podem trazer grandes desafios que precisam ser abordados de maneira clara, ponderada e afirmativa. Estes princípios de IA descrevem nosso compromisso em desenvolver tecnologias com responsabilidade e trabalhar para estabelecer áreas de aplicação específicas que não vamos explorar.

Objetivos das aplicações de IA

1. Ser socialmente benéfica.

O alcance expandido das novas tecnologias afeta cada vez mais a sociedade como um todo. Os avanços em IA causarão impactos transformadores em uma ampla gama de áreas, incluindo saúde, segurança, energia, transporte, manufatura e entretenimento. Ao considerarmos o potencial de desenvolvimento e uso das tecnologias de IA, levaremos em conta vários fatores sociais e econômicos e prosseguiremos quando acreditarmos que os prováveis benefícios gerais excedem de modo significativo os riscos e as desvantagens previsíveis.

A IA também aumenta nossa capacidade de entender o significado de conteúdos em grande escala. Faremos o possível para disponibilizar prontamente informações precisas e de alta qualidade usando a IA, respeitando sempre as normas culturais, sociais e legais nos países onde atuamos. Além disso, vamos avaliar com cuidado quando é adequado disponibilizar nossas tecnologias de modo não comercial.

2. Evitar criar ou reforçar vieses injustos.

Algoritmos e conjuntos de dados de IA podem refletir, reforçar ou reduzir vieses injustos. Reconhecemos que a distinção entre vieses justos e injustos nem sempre é simples e varia de acordo com as culturas e sociedades. Vamos evitar impactos injustos sobre as pessoas, particularmente aqueles relacionados a características sensíveis como raça, etnia, gênero, nacionalidade, renda, orientação sexual, deficiências e crença política ou religiosa.

3. Ser criada e testada visando a segurança.

Vamos continuar a desenvolver e aplicar práticas de segurança sólidas para evitar resultados indesejados que criem riscos de danos. Vamos projetar nossos sistemas de IA para que sejam devidamente cautelosos e desenvolvidos de acordo com as práticas recomendadas pelas pesquisas sobre segurança de IA. Quando for necessário, vamos testar tecnologias de IA em ambientes restritos e monitorar a operação delas após a implantação.

4. Ser responsável perante as pessoas.

Vamos projetar sistemas de IA que ofereçam oportunidades adequadas para comentários, explicações relevantes e contestação. Nossas tecnologias de IA vão estar sujeitas à orientação e ao controle humanos apropriados.

5. Incorporar princípios de design de privacidade.

Incorporaremos nossos princípios de privacidade ao desenvolvimento e uso das tecnologias de IA. Vamos proporcionar opções de notificação e consentimento, além de incentivar arquiteturas com proteções de privacidade e oferecer transparência e controle sobre o uso de dados.

6. Manter altos padrões de excelência científica.

A inovação tecnológica tem como base o método científico e um compromisso com investigação aberta, rigor intelectual, integridade e colaboração. As ferramentas de IA têm o potencial de abrir novos horizontes para a pesquisa científica e ampliar os conhecimentos em domínios importantes como biologia, química, medicina e ciências ambientais. Pretendemos alcançar altos padrões de excelência científica enquanto trabalhamos para avançar no desenvolvimento da IA.

7. Ser disponibilizada para usos que estejam de acordo com esses princípios.

Muitas tecnologias têm vários usos. Trabalharemos para limitar aplicações que sejam potencialmente prejudiciais ou abusivas. À medida que desenvolvermos e implementarmos tecnologias de IA, vamos avaliar usos prováveis com base nos seguintes fatores:

- Finalidade e uso principais: a finalidade e o provável uso principais de uma tecnologia e aplicação, incluindo o quanto a solução está relacionada ou pode ser adaptada a um uso prejudicial
- Natureza e caráter único: se estamos disponibilizando tecnologia única ou disponível de forma mais geral
- Escala: se o uso dessa tecnologia vai ter um impacto significativo
- Natureza do envolvimento do Google: seja oferecendo ferramentas de uso geral, integrando ferramentas para clientes ou desenvolvendo soluções personalizadas

Aplicações de IA que não vamos adotar

Além dos objetivos acima, não vamos projetar nem implantar IA nas seguintes áreas de aplicação:

1. Tecnologias que causam ou podem causar danos gerais. Quando houver um risco significativo de dano, vamos dar continuidade somente quando acreditarmos que os benefícios superam substancialmente os riscos e vamos incorporar as restrições de segurança adequadas.
2. Armas ou outras tecnologias que sejam criadas ou implementadas principalmente para causar ou facilitar diretamente ferimentos às pessoas.
3. Tecnologias que coletam ou usam informações para vigilância, violando as normas aceitas internacionalmente.
4. Tecnologias que têm um objetivo contrário aos princípios amplamente aceitos das legislações internacionais e dos direitos humanos.

À medida que adquirirmos mais experiência nessa área, essa lista poderá aumentar.