



KI- Grundsätze

Ziele für die Entwicklung
nutzbringender künstlicher
Intelligenz

Verantwortung: Unsere Grundsätze

Wir sind vom Potenzial künstlicher Intelligenz (KI) überzeugt. Gleichzeitig ist uns bewusst, dass fortschrittliche Technologien auch Herausforderungen mit sich bringen können, die eine entschlossene, wohlüberlegte und aktive Vorgehensweise erfordern. In diesen KI-Grundsätzen erklären wir, wie wir uns für eine verantwortungsvolle Nutzung der Technologie einsetzen und in welchen Bereichen wir ihre Anwendung nicht weiterverfolgen werden.

Zielsetzungen für die Entwicklung und Anwendung künstlicher Intelligenz

1. Nutzen für die Gesellschaft

Die gestiegene Verbreitung neuer Technologien betrifft immer mehr unsere Gesellschaft als Ganzes.

Die Fortschritte im Bereich künstliche Intelligenz werden sich zukünftig auf viele verschiedene Bereiche auswirken, egal ob Gesundheit, Sicherheit, Energie, Verkehr, Industrie oder Unterhaltung. Wenn wir über die Entwicklung und möglichen Anwendungsfelder von künstlicher Intelligenz nachdenken, werden wir dabei verschiedenste soziale und ökonomische Faktoren berücksichtigen. Nur wenn wir überzeugt sind, dass die insgesamt zu erwartenden Vorteile die absehbaren Risiken und Nachteile deutlich überwiegen, werden wir die entsprechende Idee weiterverfolgen.

Durch die Nutzung künstlicher Intelligenz werden wir immer besser darin, große Mengen an Informationen in kurzer Zeit zu erfassen. Wir werden alles daran setzen, mithilfe von KI qualitativ hochwertige und verlässliche Informationen für alle leicht zugänglich zu machen, und uns dabei an die kulturellen, sozialen und rechtlichen Normen der Länder halten, in denen wir tätig sind. Außerdem werden wir weiterhin sorgfältig erwägen, wann wir unsere Technologien auch auf nicht kommerzieller Basis verfügbar machen können.

2. Vermeidung von Bias

KI-Algorithmen und -Datensätze können unfaire Tendenzen oder Voreingenommenheiten (Bias) widerspiegeln, verstärken oder abschwächen. Wir wissen, dass es nicht immer leicht ist, fair von unfair zu unterscheiden und das Verständnis beider Begriffe kann je nach Kultur und Gesellschaft variieren. Wir werden uns dafür einsetzen, ungerechte Auswirkungen auf Menschen zu vermeiden. Dies gilt insbesondere im Zusammenhang mit so sensiblen Merkmalen wie ethnischer Herkunft und Zugehörigkeit, Geschlecht, Nationalität, Einkommen, sexueller Orientierung, individueller Leistungsfähigkeit, politischer Ansichten oder Religion.

3. Integrierte und geprüfte Sicherheit

Wir werden auch weiterhin umfangreiche Sicherheitsmechanismen entwickeln und anwenden, um unerwünschte, potenziell schädliche Auswirkungen zu vermeiden. Wir werden bei der Entwicklung unserer KI-Systeme alle gebotene Vorsicht walten lassen und dafür sorgen, dass sie den Best Practices auf dem Gebiet der KI-Sicherheitsforschung entsprechen. Wo es angemessen ist, werden wir KI-Technologien in geschützten Umgebungen testen und sie auch nach ihrer Einführung überwachen.

4. Menschen in der Verantwortung

Wir werden KI-Systeme entwickeln, die Nutzerinnen und Nutzern ausreichende Möglichkeiten für Feedback, relevante Erklärungen und Einspruch bieten. Unsere KI-Technologien werden in ausreichendem Umfang von Menschen kontrolliert und gesteuert werden.

5. Berücksichtigung von Datenschutzgrundsätzen

Wir werden bei der Entwicklung und Anwendung unserer KI-Technologien unsere Datenschutzgrundsätze berücksichtigen. Wir werden Nutzerinnen und Nutzern Gelegenheit zur Kenntnisnahme und Einwilligung geben, Architekturen mit geeigneten Datenschutzmaßnahmen fördern und angemessene Transparenz und Kontrolle über die Verwendung von Daten bieten.

6. Beachtung hoher wissenschaftlicher Standards

Technologische Innovation gründet auf wissenschaftlichen Methoden und einem Bekenntnis zu ergebnisoffener Forschung, intellektueller Gründlichkeit, Integrität und Zusammenarbeit. KI-Tools verfügen über das Potenzial, in entscheidenden Gebieten wie Biologie, Chemie, Medizin und Umweltwissenschaften die bisherigen Grenzen wissenschaftlicher Forschung und Erkenntnis zu erweitern. Bei unseren Bemühungen, die Entwicklung von KI-Technologien voranzutreiben, werden wir uns an hohen wissenschaftlichen Standards orientieren.

7. Einsatz im Einklang mit bestimmten Grundsätzen

Viele Technologien werden in unterschiedlichsten Bereichen angewendet. Wir werden uns dafür einsetzen, potenziell gefährliche oder missbräuchliche Verwendungsmöglichkeiten zu begrenzen. Bei der Entwicklung und Umsetzung von KI-Technologien werden wir deren wahrscheinliche Einsatzbereiche unter Berücksichtigung der folgenden Faktoren einschätzen:

- **Hauptsächlicher Verwendungszweck und Einsatzbereich:** der hauptsächliche Verwendungszweck und wahrscheinliche Einsatzbereich einer Technologie und Anwendung – inklusive der Frage, wie wahrscheinlich eine schädliche Verwendung ist bzw. wie leicht eine solche durch Modifikationen herbeigeführt werden kann
- **Art und Einzigartigkeit:** die Frage, ob wir eine Technologie zur Verfügung stellen, die einzigartig oder breiter zugänglich ist
- **Ausmaß:** die Frage, ob die Verwendung dieser Technologie erhebliche Auswirkungen haben wird
- **Art der Beteiligung von Google:** die Frage, ob wir universell einsetzbare Tools zur Verfügung stellen, Tools für Kunden integrieren oder individuelle Lösungen entwickeln

Anwendungsbereiche, in denen wir künstliche Intelligenz nicht einsetzen werden

Zusätzlich zu den obigen Grundsätzen werden wir künstliche Intelligenz in den folgenden Bereichen weder entwickeln noch einsetzen:

1. Technologien, die insgesamt oder mit hoher Wahrscheinlichkeit schädlich sind: Wenn ein beträchtliches Risiko besteht, dass eine Technologie Schaden verursachen kann, werden wir mit der Entwicklung nur fortfahren, sofern wir überzeugt sind, dass der Nutzen die Risiken erheblich überwiegt, und zudem angemessene Sicherheitsvorkehrungen treffen.
2. Waffen oder andere Technologien, die in erster Linie dazu bestimmt sind oder eingesetzt werden, andere Menschen zu verletzen oder direkt dazu beitragen
3. Technologien, die Informationen zu Überwachungszwecken sammeln oder verwenden und dabei international anerkannte Normen verletzen
4. Technologien, deren Zweck im Widerspruch zu allgemein akzeptierten Grundsätzen internationalen Rechts und der Menschenrechte steht

Mit zunehmender Erfahrung in diesem Bereich kann sich diese Liste weiterentwickeln.