

Künstliche Intelligenz

Hintergrundinformationen

Schlüsselbegriffe

Künstliche Intelligenz (KI) ist ein schwer zu definierendes Konzept, da es keine einheitliche, allgemein akzeptierte Beschreibung gibt, sondern eine **Vielzahl von Ansätzen**. Im Großen und Ganzen kann KI als die **Fähigkeit von Maschinen** definiert werden, die von Menschen ausgeführten **Aufgaben zu imitieren**, wie z. B. Denken, Lernen, Kreativität, Planung oder Problemlösung. Die **Europäische Union (EU)** hat Anfang 2024 das **erste KI-Gesetz weltweit** verabschiedet. Darin sind Grundregeln festgelegt, wie und wo Künstliche Intelligenz eingesetzt werden darf.

Generative KI-Modelle lernen Muster aus großen **Datensätzen** (Texte, Bilder oder Audiodateien) und **verwenden künstliche neuronale Netze**, die versuchen, ein menschliches Gehirn zu imitieren. So können sie Informationen verarbeiten und verstehen, Antworten generieren oder originelle Inhalte erstellen. Inhalte, die durch generative KI erzeugt werden, werden als künstlich definiert. Sie **müssen** laut KI-Gesetz **gekennzeichnet werden**.

Wie lernen KI-Modelle?

KI-Modelle werden mit riesigen Mengen an **Daten aus verschiedenen Quellen** trainiert, wie Texten, Bildern oder Nutzerverhalten im Internet. Diese Daten sind ein Spiegel der realen Welt, einschließlich bestehender sozialer und kultureller Vorurteile. Wenn die Trainingsdaten unausgewogen oder voreingenommen sind, kann dies dazu führen, dass die KI diese Vorurteile lernt und in ihren Ergebnissen widerspiegelt.

Ein Experiment der Washington Post hat beispielsweise gezeigt, dass Midjourney rassistische Vorurteile aufweist. Auf die Frage nach Bildern von Nutzern sozialer Dienstleistungen zeigte sie nicht-weiße Menschen, während auf die Frage nach Bildern einer „produktiven Person“ ausschließlich weiße Männer zu sehen waren.

Dies kann zu **verzerrten und letztlich unfairen Entscheidungen** führen, die bestimmte Gruppen benachteiligen. Daher ist es entscheidend, dass Entwickler*innen darauf achten, **diverse und ausgewogene Datensätze** zu verwenden, um die Risiken von Verzerrungen in KI-Systemen zu minimieren.



Künstliche Intelligenz

Wodurch kann KI beeinflusst sein?



Vorurteile

sind systematische Tendenzen in unseren Entscheidungen, Meinungen oder Denkprozessen, die zu falschen Schlussfolgerungen oder Handlungen führen können. Sie sind das Ergebnis automatischer, unbewusster mentaler Prozesse, die uns unbeabsichtigt beeinflussen können.



Vorgefasste Meinungen

sind ungerechtfertigte Meinungen oder Einstellungen, die wir gegenüber bestimmten Menschen oder Gruppen haben und die in der Regel auf Merkmalen wie Geschlecht, Alter oder ethnischer Zugehörigkeit beruhen. Vorurteile können beeinflussen, wie wir Menschen behandeln und wie wir Entscheidungen über sie treffen.



Stereotype

sind verallgemeinerte Überzeugungen über Mitglieder einer bestimmten Gruppe, die oft auf vereinfachten und übertriebenen Merkmalen dieser Gruppe beruhen. Diese Stereotypen können positiv, negativ oder neutral sein und beeinflussen, wie wir die Mitglieder dieser Gruppe wahrnehmen und mit ihnen umgehen.

Voreingenommenheit, Vorurteile und Stereotype sind eng miteinander verbunden, da sie alle die Art und Weise beeinflussen, wie wir Informationen und Menschen um uns herum wahrnehmen, interpretieren und auf sie reagieren.

Wie können wir überhaupt noch erkennen, ob etwas KI-generiert ist?

Künstliche Intelligenz wird immer besser. Gleichzeitig kommen immer mehr Apps und Werkzeuge auf den Markt, die uns dabei helfen wollen, KI-generierte Medien als künstlich zu identifizieren. Diese Hilfsmittel gibt es beispielsweise für Bilder oder Texte. Ihre Anfälligkeit für falsch-positive Ergebnisse ist allerdings noch recht hoch, weswegen wir uns nicht darauf verlassen können.

Mit Logik können wir aber in den meisten Fällen noch herausfinden, ob ein Inhalt echt ist oder von KI erzeugt wurde. KI-generierte Inhalte können auf logische Fehler oder Unstimmigkeiten hin überprüft und diese aufgedeckt werden.



Künstliche Intelligenz

KI-Checkliste

- **Schau dir die Umgebung auf Bildern an!**
Gibt es Details, die nicht zusammenpassen? Zum Beispiel ein angeblicher Waldbrand, aber es gibt keine Rauchentwicklung und die Bäume in der Nähe sind nicht betroffen? Eine berühmte Person soll verhaftet worden sein, aber die umstehenden Personen nehmen kaum Notiz davon? Das sind Indizien, dass die Aufnahme vermutlich nicht echt ist.
- **Stelle Dir grundsätzliche Logikfragen!**
Ist es wahrscheinlich, dass jemand in dieser Situation genau ein Foto gemacht hat? Wenn eine Aussage auf einer Tonspur oder ein Video aufsehenerregend oder skurril ist – ist die Situation überhaupt wahrscheinlich? Oder wenn es brennt: Sind aufgeregte Rufe zu hören oder Feuerwehrautos zu sehen, die den Brand löschen wollen?
- **Achte auf Details...**
...wie ungewöhnliche Muster in Bildern, fehlerhafte Details bei Gesichtern oder Händen oder Texte, die seltsam formuliert sind oder nicht zum Kontext passen!
- **Recherchiere weiter!**
Wenn Du keine Fehler in dem Text, dem Bild oder Video findest, dann schaue online, ob Du weitere Bilder oder Videos (aus einer anderen Perspektive) findest, die die gleiche Situation zeigen. In vielen Fällen müssten weitere Personen vor Ort gewesen sein, die ihre Aufnahmen veröffentlichen.

Weiterführende Links

- ▶ [Studie: Schule und KI - Ein praxisorientierter Leitfaden](#)
- ▶ [Washington Post - This is how AI image generators see the world](#)
- ▶ [KI-Campus - Künstliche Intelligenz in der Bildung](#)

Materialien: Lizenz CC BY 4.0 | Video: Lizenz CC BY-ND | ThinkTwice (dpa-Faktencheck)

