

4. Grenzen, Risiken und Verantwortung von KI

Künstliche Intelligenz kann uns beim Lernen, im Alltag und im Beruf unterstützen. Ihre Ergebnisse sind jedoch nicht immer richtig oder vollständig. Auch beim Datenschutz, beim Energieverbrauch und bei der fairen Behandlung von Menschen gibt es Herausforderungen.

In diesem Kapitel lernst du mögliche Risiken von KI kennen. Du erfährst, wie falsche Informationen entstehen können, worauf du beim Datenschutz achten solltest und wie du KI-generierte Inhalte kritisch beurteilst.

4.1 Wenn die KI eine falsche Information liefert: KI-Halluzinationen

KI-Systeme können Antworten erzeugen, die überzeugend klingen, aber falsche oder erfundene Informationen enthalten. Solche Fehler werden als KI-Halluzinationen bezeichnet.

Warum entstehen KI-Halluzinationen?

Sprachmodelle erzeugen Texte auf Grundlage erlernter Muster und berechnen dabei unter anderem, welche Wörter oder Wortfolgen wahrscheinlich passen. Sie überprüfen nicht automatisch, ob jede Aussage der Wahrheit entspricht. Deshalb können sie auch dann falsche Antworten erzeugen, wenn Informationen fehlen oder Zusammenhänge unklar sind.

Beispiele für KI-Halluzinationen sind:

- Eine KI erfindet Buchtitel oder wissenschaftliche Artikel und nennt dazu angebliche Autorinnen und Autoren.
- Eine KI gibt Internetadressen an, die nicht existieren, oder verweist auf Quellen, in denen die behaupteten Informationen nicht stehen.
- Eine KI nennt falsche Jahreszahlen oder beschreibt historische Ereignisse mit erfundenen Details.
- Eine KI schreibt einer bekannten Person ein Zitat zu, das diese nie gesagt hat.
- Eine KI erfindet Gesetze oder Gerichtsurteile und stellt sie als tatsächlich gültig dar.

Eine überzeugend formulierte Antwort ist nicht automatisch richtig. Überprüfe wichtige Aussagen einer KI anhand verlässlicher Quellen, zum Beispiel in Schulbüchern, auf offiziellen Webseiten oder in seriösen Fachpublikationen.

4.2 Datenqualität: Das GIGO-Prinzip

Damit ein Computersystem brauchbare Ergebnisse liefern kann, benötigt es geeignete Daten. In der Informatik beschreibt das GIGO-Prinzip einen wichtigen Zusammenhang.

Garbage In – Garbage Out (GIGO)

Übersetzt bedeutet das: „Müll hinein – Müll hinaus“.

Wenn ein Computersystem mit falschen, veralteten oder unvollständigen Daten arbeitet, können auch seine Ergebnisse fehlerhaft sein. Selbst eine leistungsfähige Software kann schlechte Eingangsdaten nicht immer ausgleichen.

Beispiel: Ein Computersystem erstellt eine Wettervorhersage. Ein Temperatursensor ist defekt und liefert falsche Messwerte. Werden diese Werte nicht erkannt und korrigiert, kann die Wettervorhersage ungenau werden.

Das Prinzip gilt auch für KI-Systeme: Die Qualität der Daten beeinflusst häufig die Qualität der Ergebnisse. Dabei sind sowohl die Daten wichtig, mit denen eine KI trainiert wird, als auch die Daten, die sie später bei ihrer Anwendung erhält.

4.3 Datenschutz und Datensicherheit

Bei der Nutzung von KI-Diensten solltest du darauf achten, welche Informationen du eingibst. Je nach Anbieter, Dienst und Einstellungen können Eingaben gespeichert, zur Verbesserung des Dienstes verwendet oder von dazu berechtigten Personen überprüft werden.

Das bedeutet nicht, dass deine Eingaben automatisch anderen Nutzern angezeigt werden. Trotzdem solltest du vertrauliche Informationen nur dann übermitteln, wenn dies erlaubt und notwendig ist und die Datenschutzbedingungen dafür geeignet sind.

Diese Informationen solltest du nicht unbedacht in ein KI-Tool eingeben:

- Deinen vollständigen Namen zusammen mit weiteren persönlichen Angaben, etwa deiner Adresse
- Passwörter, Zugangsdaten und Geheimcodes
- Private Fotos von dir, deinen Freunden oder deiner Familie
- Vertrauliche Verträge, Zeugnisse oder andere persönliche Dokumente

Wenn du beispielsweise ein Bewerbungsschreiben oder eine Entschuldigung für die Schule von einer KI überprüfen lassen möchtest, kannst du Namen und andere persönliche Angaben durch Platzhalter ersetzen. Dadurch verringerst du das Risiko, dass die betroffenen Personen erkannt werden.

Anonymisierung bedeutet, dass eine Person anhand der verbleibenden Informationen nicht mehr identifiziert werden kann. Bei der **Pseudonymisierung** werden Angaben zwar ersetzt, eine Zuordnung zur Person ist aber mit zusätzlichen Informationen weiterhin möglich.

- **Ungünstig:** „Ich, Max Mustermann, geboren am 12.03.2008 in Wien, bewerbe mich ...“
- **Besser:** „Ich, [NAME], bewerbe mich ...“

Prüfe außerdem, ob du ein Dokument überhaupt an einen externen Dienst übermitteln darfst. Das ist besonders bei vertraulichen Informationen aus der Schule oder dem Beruf wichtig.

4.4 Der ökologische Fußabdruck von KI

KI-Dienste benötigen Computer, Speicherplatz und Netzwerke. Große KI-Modelle werden in Rechenzentren trainiert und betrieben. Dort arbeiten zahlreiche leistungsfähige Server, die elektrische Energie benötigen und Wärme erzeugen.

Der Einsatz von KI kann die Umwelt auf verschiedene Weise belasten:

1. **Stromverbrauch:** Das Training großer KI-Modelle und die anschließende Verarbeitung von Anfragen benötigen elektrische Energie. Wie viel Strom verbraucht wird, hängt unter anderem von der Größe des Modells, der verwendeten Hardware und der Zahl der Anfragen ab.

2. **Wasserverbrauch und Kühlung:** Rechenzentren müssen ausreichend gekühlt werden. Je nach Kühlsystem wird dafür auch Wasser eingesetzt. Der tatsächliche Wasserverbrauch unterscheidet sich von Standort zu Standort.
3. **Rohstoffe und Herstellung:** Für Server, Prozessoren und andere elektronische Bauteile werden Rohstoffe benötigt. Ihre Gewinnung und Verarbeitung sowie die Herstellung und Entsorgung der Geräte können Umweltbelastungen verursachen.

Der ökologische Aufwand hängt somit von verschiedenen Faktoren ab. Auch die Energieversorgung, die Auslastung der Rechenzentren und die Lebensdauer der Geräte spielen eine Rolle.

Was bedeutet das für die Nutzung?

KI sollte dort eingesetzt werden, wo sie einen sinnvollen Nutzen bietet. Es ist hilfreich, unnötige Anfragen und aufwendige Berechnungen zu vermeiden. Gleichzeitig arbeiten Forschung und Unternehmen daran, KI-Systeme effizienter zu gestalten.

4.5 Ethische Herausforderungen und gefälschte Inhalte

Beim Einsatz von KI geht es nicht nur darum, ob ein System technisch funktioniert. Wichtig ist auch, ob es sicher, fair und verantwortungsvoll eingesetzt wird.

Damit KI vertrauenswürdig ist, sollte sie drei grundlegende Merkmale erfüllen:

- **Rechtmäßigkeit:** Die geltenden Gesetze müssen eingehalten werden.
- **Ethisches Verhalten:** Menschen sollen fair behandelt und ihre Rechte respektiert werden.
- **Robustheit:** Das System soll zuverlässig funktionieren und möglichst gut vor Fehlern und Angriffen geschützt sein.

Weitere wichtige Aspekte sind Datenschutz, Transparenz, menschliche Kontrolle und der verantwortungsvolle Umgang mit Umwelt und Ressourcen.

Algorithmic Bias (Voreingenommenheit und Verzerrung)

KI-Systeme können Verzerrungen aus den verwendeten Daten übernehmen. Solche Verzerrungen werden als Algorithmic Bias bezeichnet. Sie können entstehen, wenn bestimmte Personengruppen in den Daten unterrepräsentiert sind oder frühere Benachteiligungen in den Daten enthalten sind.

Beispiel: Ein Unternehmen verwendet eine KI, um Bewerbungen vorzusortieren. In den bisherigen Daten des Unternehmens wurden überwiegend Männer für Führungspositionen ausgewählt. Wenn das System diese Muster einfach übernimmt, kann es Bewerbungen von Frauen schlechter bewerten, obwohl diese genauso geeignet sind. Dadurch kann die KI bestehende Benachteiligungen verstärken.

Deepfakes und Desinformation

KI kann zur Erstellung oder Veränderung von Bildern, Videos, Stimmen und Texten verwendet werden. Solche Inhalte können sinnvoll sein, beispielsweise in Filmen oder bei kreativen Projekten. Sie können aber auch dazu dienen, Menschen zu täuschen.

- **Deepfakes:** Dabei handelt es sich um künstlich erzeugte oder veränderte Bild-, Ton- oder Videoinhalte, die reale Personen oder Ereignisse nachahmen. Beispielsweise kann ein Video den Eindruck erwecken, eine bekannte Person habe etwas gesagt, das sie tatsächlich nie gesagt hat.
- **Fake News und Desinformation:** Als Fake News werden häufig bewusst irreführende oder erfundene Nachrichten bezeichnet. Desinformation ist die gezielte Verbreitung falscher oder irreführender Informationen. KI kann dabei helfen, solche Inhalte zu erstellen oder in großer Zahl zu verbreiten.

Nicht jeder falsche Inhalt wurde mit KI erstellt, und nicht jeder KI-generierte Inhalt ist falsch. Entscheidend ist, ob die Informationen zutreffen und in welchem Zusammenhang sie verwendet werden.

Inhalte im Internet kritisch überprüfen

Wenn du ein Bild, Video oder einen Text im Internet siehst, solltest du die Informationen nicht ungeprüft übernehmen. Diese Fragen können dir dabei helfen:

1. Wer hat den Inhalt veröffentlicht? Ist die Quelle bekannt und vertrauenswürdig?
2. Berichten auch andere seriöse Nachrichtenmedien darüber?
3. Gibt es auffällige Fehler oder Unstimmigkeiten im Bild oder Video?
4. Könnte der Inhalt mit KI erzeugt oder verändert worden sein?
5. Ist der Inhalt als KI-generiert gekennzeichnet?

Achte darauf, dass keine dieser Fragen allein die Echtheit eines Inhalts beweist. Auch echte Bilder können ungewöhnlich aussehen, und gut gemachte KI-Fälschungen müssen keine sichtbaren Fehler enthalten. Umgekehrt bedeutet eine fehlende Kennzeichnung nicht, dass ein Inhalt echt ist.

Prüfe deshalb nach Möglichkeit die ursprüngliche Quelle, den Zusammenhang und weitere verlässliche Informationen.

Merke: KI kann Menschen beim Lernen, Arbeiten und im Alltag unterstützen. Ihre Ergebnisse sollten jedoch kritisch geprüft werden. Wichtig ist außerdem der Schutz persönlicher Daten und ein verantwortungsvoller Umgang mit den erzeugten Inhalten.

6 Teste dein Wissen! (Quiz)

Frage 1: Was versteht man unter einer „KI-Halluzination“?

- **A)** Das gezielte Anpassen eines Sprachmodells an ein spezielles Fachgebiet.
- **B)** Das Erzeugen von überzeugend klingenden, aber frei erfundenen oder falschen Antworten durch die KI.
- **C)** Die automatische Korrektur von Grammatik- und Rechtschreibfehlern im Eingabetext.
- **D)** Ein Computervirus, der den Monitor flackern lässt.

Frage 2: Was bedeutet die Regel „Garbage In – Garbage Out“ (GIGO)?

- **A)** Eine KI kann völlig ohne Daten arbeiten.
- **B)** Wenn eine KI mit schlechten oder falschen Daten gefüttert wird, sind auch die Ergebnisse schlecht.
- **C)** KI löscht automatisch überflüssige Dateien vom Computer.
- **D)** KI funktioniert nur, wenn man Bilder hochlädt.

Frage 3: Welche Informationen solltest du nicht unbedacht in ein öffentliches KI-Tool eingeben?

- **A)** Eine Mathe-Aufgabe aus dem Schulbuch.
- **B)** Eine Frage zum Wetter in Wien.
- **C)** Private und vertrauliche Daten (wie deine Adresse, Passwörter oder private Fotos).
- **D)** Den Titel eines bekannten Buches.

Frage 4: Was ist ein sogenannter „Algorithmic Bias“?

- **A)** Eine besonders schnelle Art der Datenübertragung.
- **B)** Eine Voreingenommenheit oder Verzerrung der KI-Ergebnisse aufgrund von unfairen Trainingsdaten.
- **C)** Eine Programmiersprache für KI-Entwickler.
- **D)** Ein neuer Typ von Computerviren.

Frage 5: Welche drei Grundprinzipien sind für eine vertrauenswürdige KI wichtig?

- **A)** Schnell – günstig – groß
- **B)** Rechtmäßigkeit – Ethisches Verhalten – Robustheit
- **C)** Modern – autonom – kreativ
- **D)** Intelligent – selbstständig – lernfähig

Lösungsschlüssel: 1B, 2B, 3C, 4B, 5B.