

Was sind Halluzinationen bei KI?

Was bedeutet 'Halluzination' bei Kunstlicher Intelligenz?

Eine 'Halluzination' bei KI bedeutet, dass die KI falsche oder erfundene Informationen liefert - auch wenn sie ueberzeugend oder wahr klingen. Die KI macht das nicht absichtlich, sondern weil sie Texte auf Basis von Wahrscheinlichkeiten erstellt - nicht auf echtem Wissen.

Beispiele fuer KI-Halluzinationen

- Die KI behauptet, ein Buch existiere, das nie geschrieben wurde.
- Sie erfindet ein Zitat, das niemand gesagt hat.
- Sie gibt falsche Jahreszahlen an oder mischt Ereignisse durcheinander.

Beispiel 1 - Geschichte

Frage: Wann wurde Deutschland gegrundet?

Antwort der KI: 1914.

Richtig waere: 1871. Die KI hat sich hier geirrt - das ist eine Halluzination.

Beispiel 2 - Literatur

Frage: Hat Goethe ein Buch ueber Vampire geschrieben?

Antwort der KI: Ja, sein Werk 'Der Blutmond' handelt von Vampiren.

Fakt: Dieses Buch existiert nicht - es wurde erfunden.

Warum ist das wichtig?

Man sollte KI-Antworten nie blind glauben. Auch wenn sie ueberzeugend klingen, koennen sie Fehler enthalten. Deshalb ist es wichtig, Aussagen der KI immer zu ueberpruefen - mit Buechern, serioesen Quellen oder durch eigenes Nachdenken.

Merksatz

Die KI weiss nicht, was wahr ist - sie raet, was wahrscheinlich klingt.