

So funktionieren KI-Chatbots

Sechs Mechanismen, die Schüler:innen kennen sollten, bevor sie mit KI arbeiten: wie sie funktionieren, was sie in der Praxis bedeuten und ein Beispiel für jeden. Dazu Beobachtungen aus dem Unterricht und ein Vorschlag für eine Stunde.

Die meisten Schüler:innen nutzen KI-Chatbots regelmäßig, mit einer Vorstellung im Kopf, die nicht stimmt: eine Suchmaschine, die alles weiß, oder ein Gegenüber, das nachdenkt. Gegen beides hilft weniger eine Erklärung als das Verständnis für sechs Mechanismen, aus denen sich fast alle typischen Fehler ableiten lassen. Fachbegriffe braucht es dafür keine; Wahrscheinlichkeit und Trainingstexte reichen.

1 Es lernt aus Texten

So funktioniert es: Ein Sprachmodell wird mit riesigen Textmengen trainiert und lernt daraus, welche Wörter und Aussagen typischerweise zusammen vorkommen. Häufiges prägt sich stark ein, Seltenes schwach, Fehlendes gar nicht.

Was das bedeutet: Bei viel beschriebenen Themen antwortet der Chatbot gut; bei seltenen Themen, kleinen Orten, unbekannten Personen und Details ergänzt er aus Ähnlichem, also erfindet er. Und er übernimmt die Schiefheiten der Texte, die überwiegend englisch und westlich sind.

BEISPIEL

Du Was ist ein typisches Frühstück?

KI Ein typisches Frühstück besteht aus Kaffee oder Tee, Brot oder Toast mit Butter, Marmelade oder Käse, oft Müsli mit Milch oder Joghurt, manchmal Eier ...

Typisch für wen? Der Chatbot beschreibt, was in seinen Texten am häufigsten vorkam.

IM UNTERRICHT

Das Frühstück ist ein harmloser Einstieg in das Thema Bias, der ohne Fachbegriff auskommt. Wer weitergehen will, lässt „einen typischen Chef“ und „eine typische Pflegekraft“ beschreiben und die Klasse zählen, welche Geschlechter, Namen und Herkunftsvorkommen. Die Diskussion, warum das so ist, führt direkt zu Mechanismus 1.

2 Es merkt sich Zusammenhänge, nicht Texte

So funktioniert es: Das Modell speichert nicht die Texte, sondern wie stark Wörter und Aussagen zusammenhängen, so wie man sich an ein Buch vom Vorjahr erinnert. Auch Zahlen hat es als Zusammenhänge gelernt, „7 mal 8“ und „56“, wie ein auswendig gelerntes Einmaleins, nicht als Rechenregel.

Was das bedeutet: Der Chatbot kann nichts nachschlagen, auch nicht in Texten, die er beim Training gelesen hat; wer ihn mit einem Text arbeiten lassen will, muss ihm den Text geben. Beim Rechnen liegt er bei größeren Zahlen nahe am Ergebnis und daneben. Viele Chatbots haben einen eingebauten Rechner; ob er verwendet wurde, ist meist an einem Hinweis in der Antwort erkennbar.

BEISPIEL

Du Rechne im Kopf, ohne Rechenwerkzeug: 4.817 mal 2.639.

KI $4.817 \times 2.639 = 12.711.863.$

Richtig wäre 12.712.063. Größenordnung und erste Ziffern stimmen, das Ergebnis nicht.

IM UNTERRICHT

Der Zusatz „ohne Rechenwerkzeug“ ist nötig, weil aktuelle Chatbots sonst oft ein Werkzeug einsetzen und richtig rechnen. Genau das ist der zweite Teil der Demo: dieselbe Rechnung ohne den Zusatz, und die Klasse sucht in der Antwort den Hinweis auf das Werkzeug. So wird Mechanismus 6 (Motor und Auto) vorbereitet, ohne dass er schon erklärt werden muss.

3 Es erzeugt Text Wort für Wort

So funktioniert es: Das Modell berechnet bei jedem Wort, welches als Nächstes am wahrscheinlichsten ist, und wählt mit etwas Zufall nicht immer das wahrscheinlichste, sonst würde es immer dieselben Sätze bauen.

Was das bedeutet: Es entsteht immer ein plausibler Text, ob er stimmt oder nicht; das Modell unterscheidet nicht zwischen „das weiß ich“ und „das passt hierher“. Und dieselbe Frage bekommt jedes Mal eine etwas andere Antwort. Deshalb ist ein Chatbot keine zitierfähige Quelle: Was er heute sagt, sagt er morgen anders.

BEISPIEL

Du Nenne die drei wichtigsten Ursachen des Ersten Weltkriegs.

- KI 1. Das Bündnissystem der europäischen Großmächte, 2. der Nationalismus auf dem Balkan, 3. das Wettrüsten.
- Du [Neuer Chat] Nenne die drei wichtigsten Ursachen des Ersten Weltkriegs.
- KI 1. Der Imperialismus und die Konkurrenz um Kolonien, 2. das Attentat von Sarajevo als Auslöser, 3. die starren Bündnisse.

Dieselbe Frage, zwei Listen. Beide vertretbar, keine ist „die“ Antwort.

IM UNTERRICHT

Diese Demo klärt nebenbei die Frage, warum man ChatGPT nicht wie ein Buch zitieren kann, die bei jeder Vorwissenschaftlichen oder Diplomarbeit auftaucht. Sie eignet sich für jedes Fach: dieselbe Frage in drei Chats parallel am Beamer, die Klasse markiert, was in allen drei vorkommt und was nur in einem. Das Gemeinsame ist meist das Belastbare.

4 Es ist für Sprache gebaut

So funktioniert es: Sprachmodelle sind für Sprache entwickelt, nicht zum Rechnen, Zählen oder Sortieren, und sie verarbeiten Sprache anders als wir: in Wortstücken statt in Buchstaben, in einem Zug von vorn nach hinten, ohne den Satz vorher zu planen.

Was das bedeutet: Alles, was Exaktheit verlangt, ist eine Schwachstelle: Wörter oder Zeichen zählen, Buchstaben in einem Wort zählen, exakte Längen einhalten, alphabetisch sortieren. Erklären, umformulieren, zusammenfassen, übersetzen und Ideen sammeln sind die Stärken.

BEISPIEL

- Du Schreib einen Satz über den Herbst mit genau zwölf Wörtern.
- KI Im Herbst färben sich die Blätter bunt und fallen langsam von den Bäumen herab.

Vierzehn Wörter. Der Satz ist gut, die Zahl stimmt nicht.

IM UNTERRICHT

Die Wortzahl-Demo ist die schnellste im ganzen Handout und wirkt gerade deshalb: Die Klasse zählt selbst nach. Wer die Buchstaben-Variante kennt (wie viele r in „Erdbeere“), sollte sie vorher testen; neuere Modelle lösen sie oft richtig, und eine gelungene Demo des Gegenteils verwirrt mehr, als sie erklärt.

5 Erstaunlich stark, erstaunlich schwach

Bei Menschen hängen Fähigkeiten zusammen; bei Sprachmodellen nicht. Die Grenze zwischen Können und Nichtkönnen verläuft nicht dort, wo man sie erwartet, und sie verschiebt sich mit jeder Version. Derselbe Chatbot erklärt die Relativitätstheorie in Reimen und verliert beim Tic-Tac-Toe, übersetzt fehlerfrei ins Japanische und kann nicht verlässlich sagen, wie oft der Buchstabe e in einem Absatz vorkommt, liest ein 30-seitiges PDF in Sekunden und weiß selbst nicht, welche seiner Aussagen darüber aus dem PDF stammen und welche er ergänzt hat.

Was das bedeutet: Weder „das kann er sicher“ noch „das kann er sicher nicht“ ist ein verlässliches Urteil. Jede Aufgabe muss neu ausprobiert und alles, was stimmen muss, geprüft werden.

IM UNTERRICHT

Dieser Punkt ist der wichtigste für die Bewertung von KI-Nutzung: Schüler:innen (und Lehrpersonen) überschätzen die Verlässlichkeit dort, wo die Antworten flüssig klingen, und unterschätzen sie dort, wo der Chatbot unerwartet gut ist. Eine Sammlung „Das hat überraschend funktioniert / das überraschend nicht“ an der Klassenpinnwand hält das Thema über das Schuljahr präsent.

6 Sprachmodell und Chatbot sind nicht dasselbe

So funktioniert es: Das Sprachmodell ist der Motor, der Text erzeugt, mit allen Eigenschaften der Punkte 1 bis 5. Der Chatbot ist das ganze Auto: Oberfläche plus Extras wie Websuche, Dateiupload, Rechner, Erinnerungen, Bilderzeugung. Die Extras fangen Schwächen des Motors ab: Die Websuche holt Aktuelles und Seltenes, der Rechner rechnet richtig, der Upload gibt dem Modell den Text, den es nicht gespeichert hat.

Was das bedeutet: Wie gut ein Chatbot antwortet, hängt davon ab, welche Extras er hat und ob sie eingeschaltet sind. Derselbe Motor verhält sich in zwei Apps verschieden, und dieselbe Frage bekommt mit und ohne Websuche eine andere Antwort.

BEISPIEL

Du [Websuche aus] Was weißt du über den steirischen Maler Leopold Wintersberger (1871–1934)?

- KI Leopold Wintersberger war ein österreichischer Landschaftsmaler, der vor allem für seine Darstellungen der südsteirischen Weinberge bekannt wurde ...
- Du [Websuche ein] Was weißt du über den steirischen Maler Leopold Wintersberger (1871–1934)?
- KI Zu einem steirischen Maler dieses Namens finde ich keine verlässlichen Informationen. Bei der Suche tauchen nur Namensvetter auf ...

Den Maler gibt es nicht. Der Motor erfindet ihn, das Extra Websuche verhindert es.

IM UNTERRICHT

Die erfundene Person funktioniert nur mit ausgeschalteter Websuche verlässlich; mit Websuche antworten aktuelle Chatbots meist korrekt, dass sie nichts finden. Genau dieser Unterschied ist die Demo. Prüfen Sie vor der Stunde, wo sich die Websuche in dem Chatbot Ihrer Schule ein- und ausschalten lässt; das ist bei jedem Anbieter woanders und ändert sich.

AUF EINEN SATZ GEBRACHT

Ein Chatbot hat aus sehr vielen Texten gelernt, wie Sprache und Wissen üblicherweise zusammenhängen, und erzeugt daraus Wort für Wort einen Text, der gut passt. Ob er auch stimmt, prüft er nicht.

Eine Unterrichtseinheit (50 Minuten)

- Einstieg (5 Min.)** „Was passiert, wenn ihr ChatGPT eine Frage stellt?“ Antworten unkommentiert an der Tafel sammeln.
- Drei Demos live (15 Min.)** Kopfrechnen ohne und mit Werkzeug, dieselbe Frage in zwei Chats, erfundene Person ohne und mit Websuche. Echter Chatbot am Beamer, keine Screenshots.
- Mechanismen (10 Min.)** Zu jeder Demo den Mechanismus dahinter benennen. Der Satz aus dem Kasten oben als Ergebnis, das die Klasse selbst formuliert.
- Experimente (15 Min.)** Die drei Aufgaben der Schüler:innen-Fassung in Partnerarbeit, jeweils mit Vermutung vor dem Ausprobieren.
- Abgleich (5 Min.)** Zurück zur Tafel: Welche Antworten vom Einstieg stimmen noch?

Für die Demos reicht ein Konto der Lehrperson; für die Partnerarbeit gelten die Altersgrenzen der Anbieter (meist ab 13 mit Zustimmung der Eltern, Claude erst ab 18). Mehr dazu und zum Datenschutz in Handout 7.

W O R A N S I E M E R K E N , D A S S E S A N G E K O M M E N I S T

Schüler:innen können in eigenen Worten sagen, warum der Chatbot bei seltenen Themen erfindet, warum er sich verrechnet, warum man ihn nicht zitieren kann und was die Websuche daran ändert. Das reicht für die weiteren Handouts.

Weiterlesen: freinhofer.substack.com (Kontextfenster · Tokens · Modellvarianten). Die Reihe „KI-Grundlagen“ bereitet auf PCRR vor: <https://pcrr.info/>