



Ergänzung

Kapitel 2: KI durchschauen

Welche KI wo arbeitet, wie ein Chatbot rechnet und wer Fakes verbreitet.

- 1 KI jenseits von Chatbots
- 2 Wie ein Chatbot antwortet
- 3 Algorithmen denken, KI verbessern
- 4 Desinformation und Demokratie

Lektion 1 von 4

KI jenseits von Chatbots

Ziel: Ich kann erklären, wie eine erkennende oder empfehlende KI entscheidet und welche Fehler sie dabei macht.

Sortieren mit Wahrscheinlichkeit

Eine erkennende KI ordnet Dinge in Schubladen: Spam oder keine Werbung, Hund oder Katze, Betrug oder normaler Kauf. Fachleute sagen dazu Klassifikation.

Sie liefert dabei keinen sicheren Befund, sondern einen Wert, wie gut jede Schublade passt. Erst eine Schwelle, die Menschen festlegen, macht daraus eine Entscheidung.

Merksatz

Eine erkennende KI rechnet eine Wahrscheinlichkeit aus. Entschieden wird erst an einer Schwelle, die Menschen festlegen.

Zwei Arten, sich zu irren

Ein Fehllalarm: Etwas Harmloses wird als Treffer markiert, etwa eine echte Mail im Spamordner. Übersehen: Ein Treffer rutscht durch, etwa Werbung im Posteingang.

Senkt man die Schwelle, gibt es mehr Fehllalarme und weniger Übersehenes. Hebt man sie, ist es umgekehrt. Beides zugleich auf null bringt keine Schwelle.

Merksatz

Wer die Schwelle verschiebt, tauscht eine Fehlerart gegen die andere.

Was eine Bilderkennung wirklich sieht

Eine Bilderkennung sieht keine Dinge, sondern Muster aus Pixeln. Sie lernt, was in ihren Beispielen oft zusammen vorkam.

In einem Versuch von 2016 sollte ein Programm Huskys von Wölfen trennen. Die Forschenden hatten die Fotos absichtlich so gewählt, dass Wölfe im Schnee standen, und das Programm achtete prompt auf den Schnee statt auf das Tier.

Merksatz

Eine Bilderkennung lernt, was in ihren Beispielen zusammen vorkam, auch wenn es mit der Sache nichts zu tun hat.

Empfehlen: Leute wie du

Viele Empfehlungssysteme vergleichen dein Verhalten mit dem vieler anderer. Leute, die ähnliche Songs streamen oder ähnliche Dinge kaufen wie du, bestimmen mit, was dir als Nächstes vorgeschlagen wird.

Ohne Verhalten weiß so ein System nichts über dich. Deshalb fragen neue Apps am Anfang oft nach zwei oder drei Interessen.

Merksatz

Eine Empfehlung sagt, was Leuten mit ähnlichem Verhalten gefiel. Ob es zu dir passt, bleibt eine Vermutung.

Erst die Aufgabe, dann die Sorte

Welche Sorte KI passt, hängt davon ab, was herauskommen soll.

Manchmal ist gar keine KI nötig. Wenn sich eine Regel vollständig aufschreiben lässt, etwa „nach 22 Uhr stumm schalten“, ist ein festes Programm verlässlicher.

- In feste Gruppen einsortieren: erkennende KI
- Eine Zahl vorhersagen, etwa die Wartezeit: vorhersagende KI
- Aus vielen Möglichkeiten Passendes vorschlagen: Empfehlungssystem
- Neuen Text oder ein Bild herstellen: erzeugende KI

Merksatz

Erst die Aufgabe klären, dann die Sorte KI wählen. Reicht eine feste Regel, braucht es keine KI.

Lektion 2 von 4

Wie ein Chatbot antwortet

Ziel: Ich kann erklären, was zwischen meiner Frage und der Antwort eines Chatbots passiert, und meine Aufträge danach richten.

Wörter in Stücken

Ein Chatbot liest nicht Buchstabe für Buchstabe. Er zerlegt Text in Token: Häufige Wörter sind ein Stück, seltene werden in mehrere Teile zerschnitten, und gerechnet wird mit Nummern dieser Stücke.

Deshalb tut er sich schwer, wo es auf einzelne Buchstaben ankommt: Buchstaben zählen, ein Wort rückwärts schreiben, Silben reimen.

Merksatz

Ein Chatbot rechnet mit Textstücken. Einzelne Buchstaben sieht er darin oft gar nicht.

Das Kontextfenster

Alles, was ein Chatbot beim Antworten berücksichtigt, heißt Kontextfenster. Dazu gehören dein bisheriger Verlauf, eingefügte Texte und Anweisungen. Es ist groß, aber begrenzt.

In sehr langen Gesprächen fallen frühe Teile heraus oder werden schwächer beachtet. Eine Studie von 2023 fand, dass Angaben in der Mitte langer Texte schlechter genutzt werden als am Anfang oder am Ende.

Merksatz

Was nicht im Kontextfenster steht, gibt es für den Chatbot in diesem Gespräch nicht.

Von links nach rechts

Ein Chatbot schreibt seine Antwort Token für Token. Jedes neue Stück richtet sich nach dem, was schon dasteht, und während des Schreibens nimmt er nichts zurück.

Nennt er zuerst ein Ergebnis, baut er die Begründung danach um dieses Ergebnis herum. Deshalb lohnt es sich, erst die Schritte und dann das Ergebnis zu verlangen. Manche Chatbots rechnen vorher unsichtbar Zwischenschritte, aber auch dann kannst du nur prüfen, was sichtbar dasteht.

Merksatz

Verlang erst die Schritte, dann das Ergebnis. Sonst begründet der Chatbot oft nur, was er schon hingeschrieben hat.

Anweisungen, die du nicht siehst

Bevor du etwas schreibst, hat der Anbieter dem Chatbot schon Anweisungen mitgegeben: welchen Ton er anschlägt, was er ablehnt, ob er im Netz sucht. Man nennt das oft Systemanweisung.

Außerdem legt der Anbieter fest, wie viel Zufall bei der Auswahl der Token mitspielt. Diese Einstellung heißt oft Temperatur. Beides erklärt, warum zwei Chatbots dieselbe Frage verschieden beantworten.

Merksatz

Hinter jeder Antwort stehen deine Nachricht, der Verlauf und Anweisungen des Anbieters, die du nicht siehst.

Wissen aus dem Training, Treffer aus der Suche

Ohne Websuche antwortet ein Chatbot aus dem, was er aus seinen Trainingsdaten gelernt hat. Manche Chatbots suchen zusätzlich im Netz und fassen Treffer zusammen, dann stehen oft Links dabei.

Auch eine Zusammenfassung kann eine Quelle falsch wiedergeben oder einen Link nennen, der etwas anderes sagt.

Merksatz

Links in einer Antwort sind ein Anfang zum Nachprüfen, kein Beweis.

Lektion 3 von 4

Algorithmen denken, KI verbessern

Ziel: Ich kann einen Ablauf eindeutig in Schritte zerlegen und sagen, wie man prüft, ob eine KI allen nützt.

Was ein Algorithmus ist

Ein Algorithmus ist eine Folge eindeutiger Schritte, die zu einem Ergebnis führt. Wie ein Rezept, bei dem niemand nachfragen muss.

Drei Bausteine reichen für erstaunlich viel: Schritte der Reihe nach, Bedingungen nach dem Muster „wenn, dann“ und Wiederholungen nach dem Muster „so lange, bis“.

Merksatz

Ein Algorithmus ist so genau, dass ihn auch jemand ausführen kann, der nicht mitdenkt.

Zerlegen

Große Aufgaben werden lösbar, wenn man sie in Teilschritte zerlegt und bei jedem fragt, was hineingeht und was herauskommt.

So sind auch KI-Anwendungen gebaut. Eine App, die gesprochene Sätze übersetzt, besteht aus drei Teilen: Sprache erkennen, übersetzen, vorlesen. Läuft etwas schief, sucht man den Teil, an dem es passiert.

Merksatz

Wer zerlegt, findet den Teilschritt, an dem ein Fehler entsteht.

KI für das Gemeinwohl

KI kann Menschen helfen, die sonst zu kurz kommen: automatische Untertitel für Gehörlose, Bildbeschreibungen für Blinde, Hochwasserwarnungen für Gegenden ohne Messstationen, wie eine Studie von 2024 zeigte.

Ob sie hilft, hängt aber davon ab, für wen sie gebaut und an wem sie getestet wurde. Eine Studie von 2020 fand bei fünf Spracherkennungen fast doppelt so viele Fehler bei schwarzen wie bei weißen Sprechenden.

Merksatz

Eine KI dient dem Gemeinwohl nicht von selbst. Es kommt darauf an, wem sie nützt und wen sie übersieht.

So wird eine KI besser

Ein System mit gutem Durchschnitt kann für eine kleine Gruppe trotzdem schlecht funktionieren. Deshalb zählt man Fehler getrennt, statt nur einen Gesamtwert zu nennen. Fünf Schritte helfen, eine KI gezielt zu verbessern.

- Ziel klären: Wem soll sie helfen, und wer ist betroffen?
- Testfälle sammeln, auch ungewöhnliche und aus allen Gruppen
- Fehler je Gruppe zählen, nicht nur im Durchschnitt
- Rückmeldungen von Nutzenden einbauen
- Festlegen, wo ein Mensch entscheidet

Merksatz

Ein guter Durchschnitt kann eine Gruppe verdecken, für die das System schlecht funktioniert.

Lektion 4 von 4

Desinformation und Demokratie

Ziel: Ich kann erklären, wie KI Falschmeldungen verbreiten hilft, und eine Behauptung vor dem Teilen von außen prüfen.

Falsch verbreitet sich schneller

Eine große Auswertung von Meldungen auf Twitter aus den Jahren 2006 bis 2017 fand: Falschmeldungen wurden deutlich häufiger weitergeleitet als wahre. Sie wirkten neuer und überraschender.

Den Unterschied machten Menschen, nicht Programme. Wer weiterleitet, macht mit.

Merksatz

Was starke Gefühle weckt, wird schneller geteilt, ob es stimmt oder nicht.

Was KI daran ändert

Früher kostete eine gute Fälschung Zeit und Können. Mit KI entstehen Texte, Bilder und Stimmen in Minuten und in vielen Varianten, dazu Profile, die posten wie echte Menschen.

Vor der Parlamentswahl in der Slowakei im September 2023 verbreitete sich kurz vor dem Wahltag eine gefälschte Tonaufnahme, in der ein Kandidat angeblich über Wahlbetrug sprach.

Merksatz

KI macht Fälschungen nicht erst möglich, aber billig und massenhaft.

Die Lügendividende

Wenn alles gefälscht sein könnte, können Leute auch echte Aufnahmen als Fake abtun. Forschende haben dafür 2019 den Begriff Lügendividende geprägt.

Das schadet der Demokratie doppelt: Falsches wird geglaubt, und Echtes wird angezweifelt.

Merksatz

Fälschungen schaden auch dadurch, dass danach Echtes nicht mehr geglaubt wird.

Seitlich lesen

Eine Studie von 2019 zeigte: Profis in Faktenprüfung bleiben nicht auf der Seite, die sie prüfen. Sie öffnen neue Tabs und fragen: Wer steckt dahinter? Was sagen andere seriöse Quellen über die Behauptung?

Dazu gehören eine Suche nach dem Thema zusammen mit dem Wort Faktencheck, eine umgekehrte Bildersuche und ein Blick auf Datum und Ort.

Merksatz

Prüf eine Behauptung von außen: Was sagen andere über die Quelle und über die Sache?

Regeln für Plattformen und Fälschungen

Sehr große Plattformen müssen in der EU Risiken für Wahlen und öffentliche Debatte bewerten und eindämmen (Gesetz über digitale Dienste, Artikel 34 und 35). Beiträge kannst du dort melden.

Seit dem 2. August 2026 muss offengelegt werden, wenn Bilder, Töne oder Videos mit KI erzeugt oder verändert wurden und echt wirken (KI-Verordnung, Artikel 50). Wer fälschen will, hält sich daran allerdings nicht.

Merksatz

Eine fehlende Kennzeichnung beweist nicht, dass etwas echt ist.

Das nimmst du mit

- Ich kann erklären, wie eine erkennende oder empfehlende KI entscheidet und welche Fehler sie dabei macht.
- Ich kann erklären, was zwischen meiner Frage und der Antwort eines Chatbots passiert, und meine Aufträge danach richten.
- Ich kann einen Ablauf eindeutig in Schritte zerlegen und sagen, wie man prüft, ob eine KI allen nützt.
- Ich kann erklären, wie KI Falschmeldungen verbreiten hilft, und eine Behauptung vor dem Teilen von außen prüfen.

Die Foto-App auf deinem Handy schlägt vor, ein Bild im Ordner „Haustiere“ abzulegen. Was hat sie dafür berechnet?

- A** Gar nichts, sie legt alle Bilder mit Fell nach einer festen Regel ab, die ein Mensch geschrieben hat
- B** Einen sicheren Befund, weil sie das Tier mit einer vollständigen Liste aller Hunderassen verglichen hat
- C** Einen Wert, wie gut das Bild zum Ordner passt, und dieser Wert lag über einer festgelegten Schwelle

Auflösung

Richtig ist C: Einen Wert, wie gut das Bild zum Ordner passt, und dieser Wert lag über einer festgelegten Schwelle

Eine vollständige Liste aller Hunde gibt es nicht, und eine Fellregel würde jeden Teddybären einsortieren. Die App schätzt, wie gut das Bild passt, und schlägt ab einer Schwelle vor.