



Ergänzung

Kapitel 7: Ergebnisse messen und beschreiben

KI-Ergebnisse nach Kriterien bewerten, Daten auswählen und ein Modell ehrlich beschreiben.

- 1 KI-Ergebnisse nach Kriterien bewerten
- 2 Daten kuratieren
- 3 Ein Modell beschreiben

Lektion 1 von 3

KI-Ergebnisse nach Kriterien bewerten

Ziel: Ich kann KI-Ergebnisse an vorher festgelegten Kriterien und Testfällen bewerten und begründet entscheiden, ob ich sie übernehme.

Erst der Maßstab, dann das Ergebnis

Wer ein KI-Ergebnis erst liest und danach überlegt, was gut wäre, bewertet nach Eindruck. Ein flüssiger Text wirkt dann besser, als er ist.

Leg deshalb vorher fest, woran du ein gutes Ergebnis erkennst, und schreib zu jedem Kriterium auf, was du erwartest. Drei bis fünf Kriterien reichen oft.

Merksatz

Was ein gutes Ergebnis ist, entscheidest du vor dem ersten Lesen, nicht danach.

Drei Sorten Testfälle

Ein einzelnes gelungenes Ergebnis sagt wenig. Prüf eine Vorlage an Fällen, bei denen du die richtige Antwort schon kennst, und zwar an drei Sorten.

- Der normale Fall: so, wie die Aufgabe gewöhnlich aussieht
- Der Randfall: sehr kurz, sehr lang, unvollständig oder ungewöhnlich aufgebaut
- Der Negativfall: ein Fall, bei dem eine falsche Antwort naheliegt, etwa eine Frage, deren Antwort nicht im Text steht

Merksatz

Ein Negativfall unter den Testfällen zeigt, ob eine Vorlage auch dann richtig reagiert, wenn die Antwort fehlt.

Stichprobe statt Einzelfall

Läuft eine Vorlage oft, etwa für 200 Produktbeschreibungen, kannst du nicht jedes Ergebnis lesen. Ein einziges zu lesen reicht aber auch nicht.

Zieh eine Stichprobe: zum Beispiel 20 Ergebnisse, zufällig über alle verteilt und nicht die ersten 20. Zähl, wie viele davon deine Kriterien erfüllen, und notier jeden Fehler mit seiner Art.

Merksatz

Eine zufällige Stichprobe sagt mehr über viele Ergebnisse als das eine, das du gerade vor dir hast.

Nutzende fragen

Deine Kriterien sind deine Sicht. Wer mit dem Ergebnis arbeitet, merkt oft etwas anderes: dass eine Zusammenfassung die wichtige Frist weglässt oder der Ton nicht zur Kundschaft passt.

Frag zwei oder drei Nutzende konkret statt „Passt das?“. Was hat dir gefehlt? Wo musstest du nachsehen? Was hast du geändert?

Merksatz

Rückmeldung von Nutzenden zeigt, was deine eigenen Kriterien übersehen haben.

Übernehmen, überarbeiten oder verwerfen

Am Ende steht eine Entscheidung mit Grund. Leg vorher fest, ab wann welche der drei Möglichkeiten gilt. Sonst verschiebt sich die Grenze, sobald dir ein Ergebnis gefällt.

- Übernehmen: alle Pflichtkriterien erfüllt, in der Stichprobe nur kleine Fehler
- Überarbeiten: die Fehler haben eine gemeinsame Ursache, die du beheben kannst
- Verwerfen: Fehler dort, wo ein Fehler teuer wird, oder keine erkennbare Ursache

Merksatz

Die Grenze fürs Übernehmen steht fest, bevor du die Zahlen siehst.

Lektion 2 von 3

Daten kuratieren

Ziel: Ich kann die Daten für ein KI-System auswählen, ihre Herkunft, ihren Stand und ihre Lücken aufschreiben und Personen darin schützen.

Was ein System zu sehen bekommt

Ein KI-System arbeitet mit dem, was du ihm gibst. Auf drei Wegen kommen Daten hinein, und bei allen dreien entscheidest du über die Auswahl.

- Beispiele im Prompt, an denen sich Form und Ton der Antwort ausrichten
- Eine Wissenssammlung, also Dokumente, die das System bei jeder Frage durchsucht. Der Fachbegriff dafür ist Retrieval-Augmented Generation
- Angehängte Dokumente, die nur für diesen einen Auftrag gelten

Merksatz

Die Auswahl der Daten bestimmt mit, wie ein System antwortet, auch wenn der Auftrag gleich bleibt.

Vier Fragen an jede Quelle

Bevor ein Dokument in die Auswahl kommt, stellst du ihm vier Fragen. Wo du keine Antwort findest, schreibst du das auf.

- Herkunft: Wer hat es geschrieben, und woher stammt es?
- Aktualität: Von wann ist es, und gilt es heute noch?
- Vollständigkeit: Fehlt ein Teil, ein Zeitraum oder eine Gruppe?
- Verzerrung: Kommt eine Sicht oder eine Gruppe zu oft oder zu selten vor?

Merksatz

Eine Quelle, deren Herkunft und Stand niemand kennt, gehört nicht in die Auswahl.

Personen bleiben draußen

Namen, Mailadressen, Kundennummern oder Noten sind personenbezogene Daten nach der DSGVO. In eine Sammlung für ein KI-System gehören sie nur mit einer klaren Erlaubnis, und darüber entscheidest nicht du allein.

Oft reicht der Inhalt ohne die Person. Pseudonymisieren heißt: Du ersetzt Namen und Nummern durch Platzhalter wie „Kundin A“ und bewahrst die Zuordnung getrennt auf. Solche Daten gelten weiter als personenbezogen, aber im Text steht niemand mehr.

Merksatz

Was ein System für seine Aufgabe nicht braucht, bekommt es nicht, auch keine Namen.

Lücken und Stand aufschreiben

Keine Sammlung ist vollständig. Schreib auf, was fehlt, etwa „nur Anfragen aus dem Onlineshop, keine aus dem Laden“ oder „nur Texte bis März“. Wer das System nutzt, weiß dann, wo es danebenliegen kann.

Zwei Angaben gehören außerdem zu jeder Sammlung: ein Datum als Versionsstand und eine kurze Notiz, was seit der letzten Fassung dazukam oder herausflog.

Merksatz

Eine Sammlung ohne Versionsstand und ohne Liste ihrer Lücken lässt sich nicht nachprüfen.

Lektion 3 von 3

Ein Modell beschreiben

Ziel: Ich kann eine Modellkarte lesen, eine Anbieterangabe kritisch prüfen und für eine eigene Vorlage eine ehrliche Karte schreiben.

Ein Steckbrief mit Grenzen

Eine Modellkarte, englisch Model Card, ist ein kurzer Steckbrief zu einem Modell: wofür es gebaut ist, woran es getestet wurde und wo es versagt. Vorgeschlagen haben das Margaret Mitchell und andere 2019.

Die Karte richtet sich an alle, die das Modell einsetzen wollen, auch an Leute ohne Technikwissen.

Merksatz

Eine Modellkarte sagt nicht nur, was ein Modell kann, sondern auch, wofür es nicht gedacht ist.

Sieben Teile einer Modellkarte

Die Teile heißen je nach Vorlage etwas anders. Gemeint ist fast immer dasselbe.

- Zweck: welche Aufgabe das Modell lösen soll
- Vorgesehene Nutzung: für wen und in welcher Lage
- Nicht vorgesehene Nutzung: wofür es ausdrücklich nicht taugt
- Daten: womit es gebaut und womit es getestet wurde
- Bewertung: welche Zahlen, an welchen Fällen gemessen
- Grenzen: wo es schwächer ist, etwa bei einer Sprache oder Gruppe
- Stand: Datum und Version der Karte

Merksatz

Fehlt auf einer Karte die nicht vorgesehene Nutzung, fehlt ihre wichtigste Warnung.

Anbieterangaben lesen

Anbieter beschreiben ihre Modelle oft selbst. Die Zahlen darin sind gemessen, aber an Aufgaben, die der Anbieter ausgewählt hat.

Drei Fragen helfen beim Lesen: Woran wurde gemessen? Passt das zu deiner Aufgabe? Was steht unter Grenzen, und gibt es den Abschnitt überhaupt?

Merksatz

Eine Zahl in einer Anbieterangabe gilt für die Aufgabe, an der sie gemessen wurde, nicht automatisch für deine.

Eine Karte für die eigene Vorlage

Eine Modellkarte lohnt sich auch im Kleinen. Wer eine Vorlage an andere weitergibt, schreibt dieselben Teile in Kurzform dazu.

Zehn Zeilen reichen oft. Sie ersparen die Frage, ob man die Vorlage auch für etwas anderes nehmen kann, und die Zahlen darauf stammen aus deinem eigenen Testlauf.

Merksatz

Eine weitergegebene Vorlage braucht eine Karte, auf der auch steht, wofür sie nicht taugt.

Das nimmst du mit

- Ich kann KI-Ergebnisse an vorher festgelegten Kriterien und Testfällen bewerten und begründet entscheiden, ob ich sie übernehme.
- Ich kann die Daten für ein KI-System auswählen, ihre Herkunft, ihren Stand und ihre Lücken aufschreiben und Personen darin schützen.
- Ich kann eine Modellkarte lesen, eine Anbieterangabe kritisch prüfen und für eine eigene Vorlage eine ehrliche Karte schreiben.

Deine Vorlage für Antworten auf Zimmeranfragen eurer WG antwortet heute deutlich anders als letzte Woche. Am Prompt hast du nichts geändert. Was prüfst du zuerst?

- A** Ob du heute ein anderes Gerät benutzt, weil das Gerät bestimmt, welche Antwort kommt
- B** Ob die Antwort von heute besser klingt, denn dann ist die Änderung kein Problem
- C** Ob sich die mitgegebenen Unterlagen geändert haben, etwa die Hausordnung

Auflösung

Richtig ist C: Ob sich die mitgegebenen Unterlagen geändert haben, etwa die Hausordnung

Antworten schwanken auch ohne jede Änderung ein wenig. Ein deutlicher Unterschied hat aber oft eine Ursache in den Daten, und die findest du nur, wenn die Sammlung einen Versionsstand hat.