

**Arbeitsblatt**

KI-Ergebnisse nach Kriterien bewerten

Ergänzung · Kapitel 7: Ergebnisse messen und beschreiben · Lektion 1 von 3

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

Name: _____ **Klasse:** _____ **Datum:** _____

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.

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

Aufgabe 1 · Kreuze die richtige Antwort an.

Eine KI soll 40 Bewertungen des Cafés, in dem du jobbst, zu den drei häufigsten Beschwerden zusammenfassen. Was legst du fest, bevor du ihr Ergebnis liest?

- ☐ **A** Welche Formulierung am Ende überzeugend genug klingt, um sie an die Chefin weiterzuleiten
- ☐ **B** Wie viele Wörter die Zusammenfassung höchstens haben darf, damit sie auf eine Folie passt
- ☐ **C** Welche Beschwerden du beim Überfliegen selbst gesehen hast und dass jede eine Anzahl hat

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

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



Aufgabe 2 · Kreuze alle richtigen Antworten an. Es sind mehrere.

Du testest eine Vorlage, die aus Stellenanzeigen die verlangten Anforderungen als Liste herauszieht. Welche Testfälle prüfen mehr als den normalen Fall?

- ☐ **A** Eine Anzeige, in der gar keine Anforderungen stehen, nur Aufgaben und Vorteile der Stelle
- ☐ **B** Eine Anzeige, in der „wünschenswert“ und „Pflicht“ bunt gemischt nebeneinanderstehen
- ☐ **C** Eine Anzeige mit nur zwei Zeilen Text, ohne Überschriften und ohne jede Aufzählung
- ☐ **D** Eine zweite übliche Anzeige für eine Bürostelle, ganz ähnlich aufgebaut wie die erste
- ☐ **E** Eine übliche Anzeige für eine Stelle im Einzelhandel mit sechs klar benannten Anforderungen

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.

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

Aufgabe 3 · Kreuze die richtige Antwort an.

Eine KI hat 150 Bildbeschreibungen für den Onlineshop in deinem Nebenjob geschrieben. Du hast Zeit, 15 davon zu lesen. Welche nimmst du?

- ☐ **A** Die ersten 15, weil sich ein Fehler der Vorlage gleich am Anfang zeigen müsste
- ☐ **B** 15 zufällig über alle 150 verteilte, auch solche zu Produkten mit wenig Angaben
- ☐ **C** Die 15 längsten, weil in langen Beschreibungen mehr Platz für Fehler ist



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?

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

Aufgabe 4 · Welche Fassung ist besser? Kreuze an und begründe kurz.

Du hast für deine Lerngruppe eine Vorlage gebaut, die Vorlesungsfolien in Lernkarten verwandelt. Welche Frage an die Gruppe bringt dir mehr?

Fassung 1

Bei welchen Karten musstet ihr in den Folien nachsehen, und welche habt ihr geändert oder gelöscht?

Fassung 2

Wie findet ihr die Lernkarten aus meiner Vorlage insgesamt, taugen die was für die Klausur im Februar?

☐ Fassung 1 ist besser

☐ Fassung 2 ist besser

Begründung:

Aufgabe 5 · Bring die Schritte in die richtige Reihenfolge: Schreib 1, 2, 3 ... in die Kästchen.

Du willst eine KI-Vorlage für Antworten auf Kundenmails in deinem Nebenjob bewerten. In welcher Reihenfolge gehst du vor?

- ☐ **A** Mit der vorher gesetzten Grenze entscheiden: übernehmen, überarbeiten oder verwerfen
- ☐ **B** Zwei Kolleginnen die Antworten lesen lassen und fragen, was sie ändern würden
- ☐ **C** Kriterien festlegen und zu jedem notieren, was eine gute Antwort enthält
- ☐ **D** Testfälle sammeln: normale Mails, Randfälle und mindestens ein Negativfall
- ☐ **E** Die Vorlage mit allen Testfällen laufen lassen und nach den Kriterien zählen



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

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

Aufgabe 6 · Kreuze die Zeilen an, die nicht passen.

Ein Bewertungsplan für eine Vorlage, die Protokolle von Teamsitzungen zusammenfasst. Welche Zeilen taugen nicht?

- | | |
|--------------------------|--|
| ☐ | 1 Aus 30 Protokollen des letzten Jahres werden 8 zufällig gezogen. |
| ☐ | 2 Ob die Zusammenfassung reicht, entscheide ich nach dem Lesen aus dem Bauch. |
| ☐ | 3 Ein Testprotokoll enthält keinen einzigen Beschluss, nur eine lange Diskussion. |
| ☐ | 4 Zwei Leute aus dem Team sagen, welcher Beschluss ihnen gefehlt hat. |
| ☐ | 5 Getestet wird mit dem Protokoll, bei dem die Vorlage schon einmal gut war. |
| ☐ | 6 Kriterium: Jeder Beschluss aus dem Protokoll steht in der Zusammenfassung. |

Warum passt eine der Zeilen nicht? Schreib es zu einer Zeile auf.



Aufgabe 7 · Schreib deinen eigenen Auftrag an eine KI.

Schreib einen Auftrag, mit dem eine KI dir einen Bewertungsplan für eine eigene Vorlage entwirft. Sag, wofür die Vorlage da ist, verlang eine Tabelle mit Kriterien und je einem Bewertungsanker und dazu Testfälle: normale Fälle, Randfälle und ein Negativfall.

Das muss drin sein:

- Ein Auftrag mit Verb
- Kontext, also worum es geht
- Ein Format
