

Psychosemantische Modellierung

Lukas Göbl & Peer Schäfer

Fragestellung

Forschungsfrage

Zeigen LLMs ähnliche **semantische Verwandtschaften, Assoziationen und Bedeutungscluster** wie Menschen?

- Gibt es überlappende Bedeutungsräume?
- Wo treten systematische Abweichungen auf?
- Gibt es Unterschiede zwischen verschiedenen Modellfamilien?
- Spielt die Größe des Modells eine Rolle?
- Welche Modellgröße erlaubt eine automatische Auswertung der Ergebnisse?

Aufgabenformate

Freie Assoziation

Geben Sie zu folgendem Stichwort bis zu drei Assoziationen mit jeweils einem Wort an, wobei die erste Antwort die stärkste Assoziation ist.

Datensatz: SWOW-18EN (De Deyne et al.)

Ähnlichkeitsbewertung

Bewerten Sie das folgende Wortpaar nach Ähnlichkeit auf einer Skala von 0-10.

Datensatz: Wordsim353 (Finkelstein, Lev, et al.)

Modelle

Modellfamilie	Modell
Qwen3	qwen3:0.6b qwen3:1.7b qwen3:8b qwen3:14b qwen3:30b
Gemma3	gemma3:270m gemma3:1b gemma3:4b gemma3:12b gemma3:27b

Systemanweisung Wortassoziation

Prompt

«SYS»

You MUST follow these rules:

1. Do NOT output reasoning, chain-of-thought, thinking process, analysis, hidden thoughts, XML tags like <think>, or any extra formatting.
2. Output ONLY one single line with exactly four semicolon-separated fields.
3. Format: cue;A1;A2;A3
4. A1-A3 MUST be exactly one word each (no spaces).
5. If you cannot generate A2 or A3, use exactly: No more responses
6. Any extra text makes the output INVALID.

«/SYS»

Prompt Wortassoziation

Prompt

You will perform a word association task.

Task:

Given a cue word, produce up to three single-word associations:

A1 = strongest association

A2 = second association

A3 = third association

Output format (MANDATORY):

cue;A1;A2;A3

Cue:

arachnid

Antwort des Modells

arachnid;spider;insect;No more responses

Systemanweisung Ähnlichkeitsbewertung

Prompt

«SYS»

You MUST follow these rules:

1. Rate similarity on a scale from 0 to 10.
2. 0 = completely unrelated
3. 10 = identical in meaning
4. Only use integers (0-10).
5. Consider semantic similarity, not association or co-occurrence.
6. Do NOT explain your reasoning.
7. Output must be exactly one line:

[word1];[word2];[rating]

Example output:

car;automobile;10

«/SYS»

Prompt Ähnlichkeitsbewertung

Prompt

You will perform a word similarity rating task.

Task:

You will be given a pair of English words.

Your job is to judge how similar their meanings are.

Now rate the following word pair:

Word 1: sword

Word 2: knife

Antwort des Modells

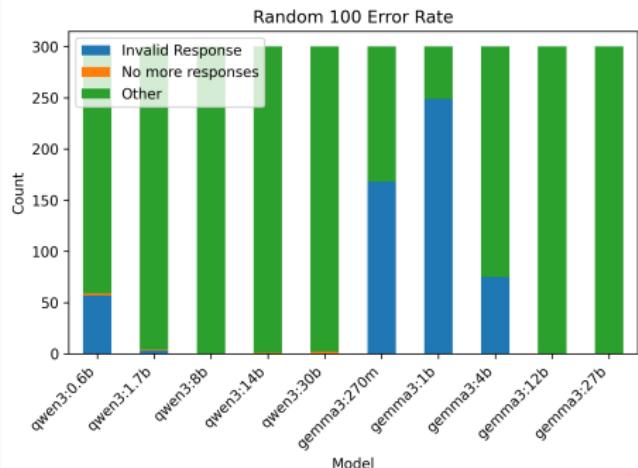
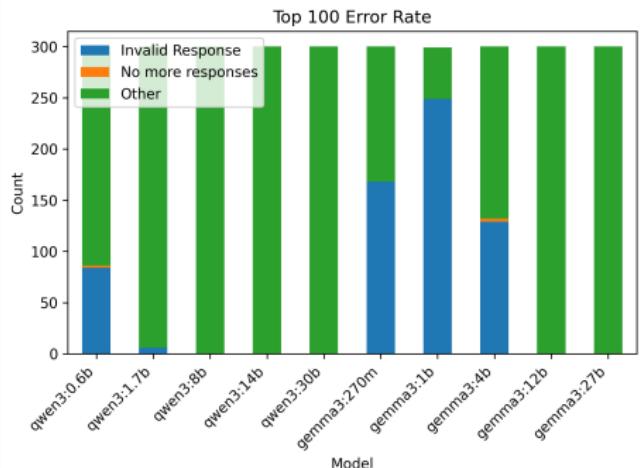
sword;knife;7

Experimentelles Setup / Vorgehensweise

1. Setup von **Ollama** mit verschiedenen Open-LLMs
2. Erste manuelle Tests
3. Erstellung der Pipelines für automatisierte Experimente
 - Freie Assoziation: Die Top-100 R1-Strength und 100 zufällige Stichwörter aus SWOW-18EN
 - Ähnlichkeitsbewertung: Vollständiger Wordsim353 Datensatz
4. Datenerhebung mit LLMs
5. Statistische Analyse und Vergleich mit menschlichen Datensätzen

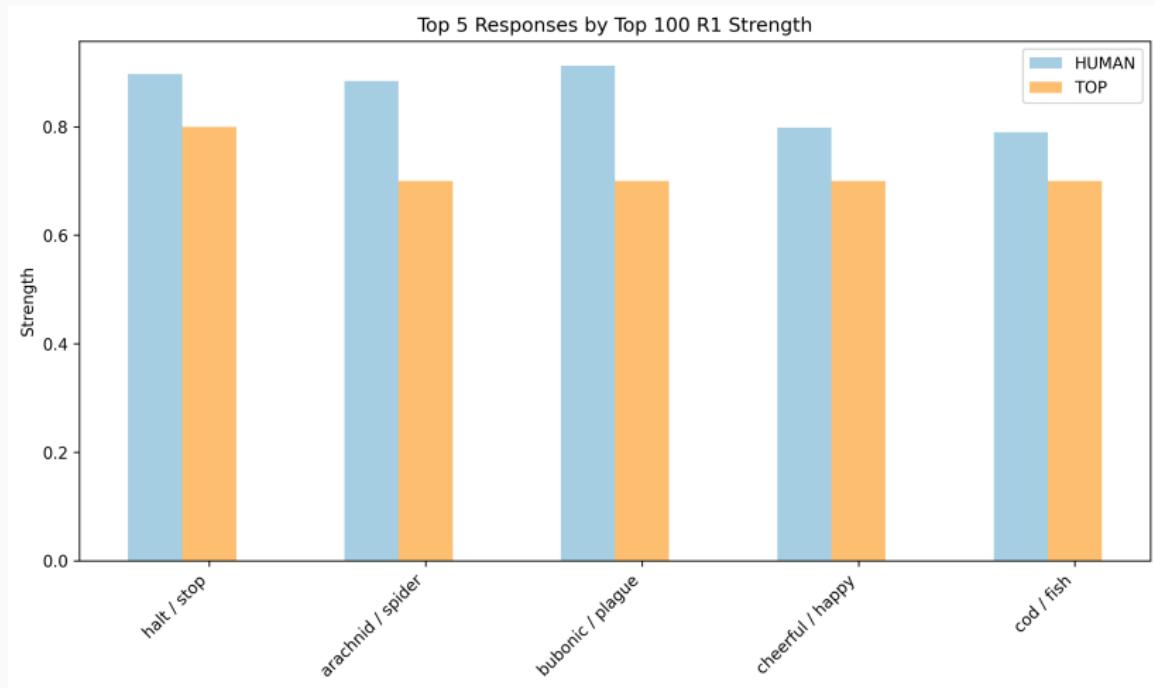
Ergebnisse: Freie Assoziation

Fehler-Rate

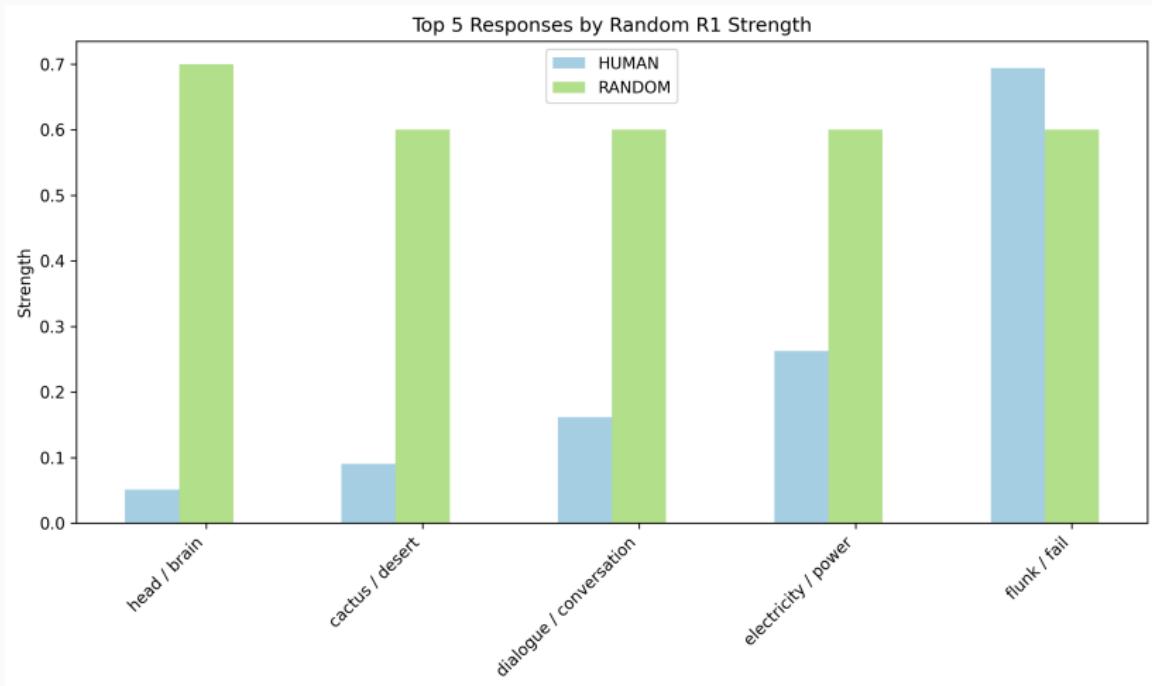


→ Mehr Parameter führen generell zu einer niedrigeren Fehlerrate

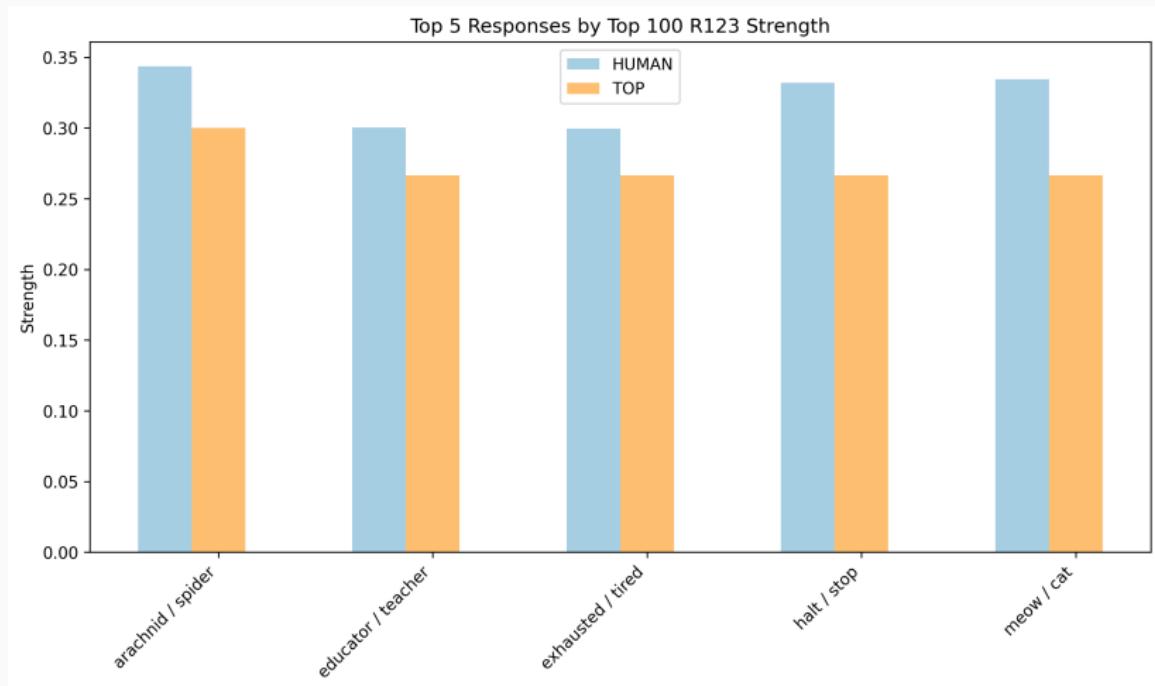
Antworten von 'Top-100' nach R1 sortiert



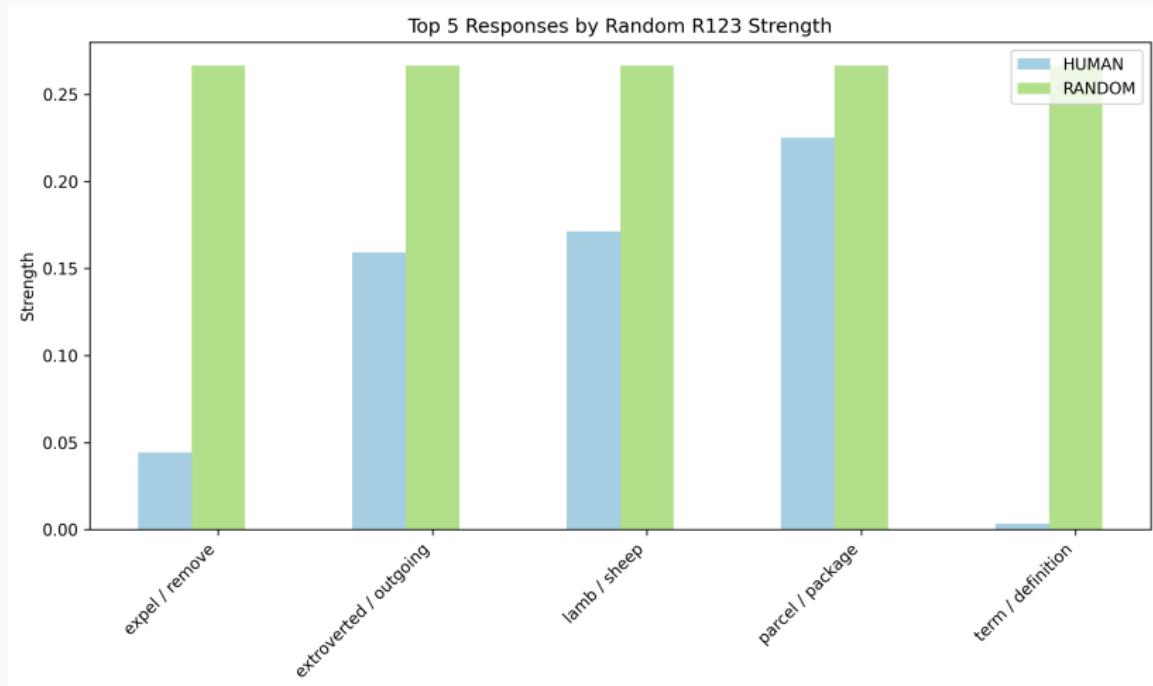
Antworten von 'Random-100' nach R1 sortiert



Antworten von 'Top-100' nach R123 sortiert



Antworten von 'Random-100' nach R123 sortiert

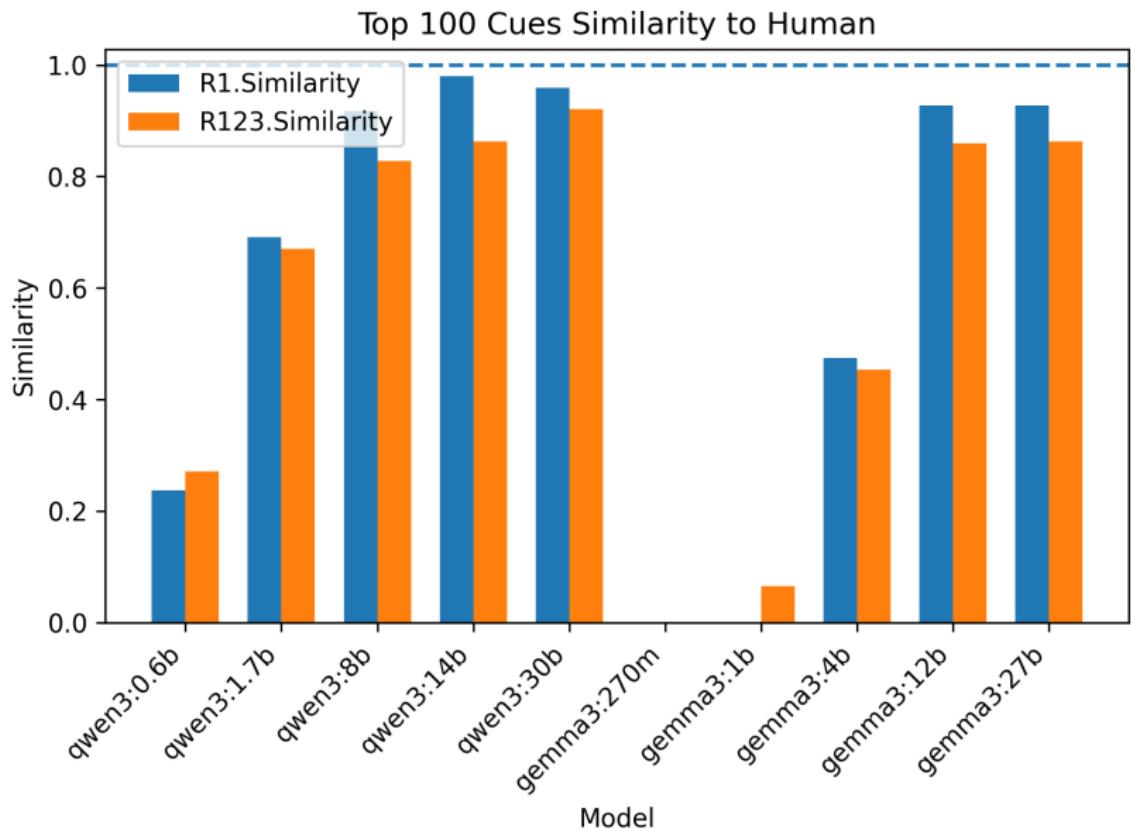


Vergleich Top-Antworten nach R1 und R123

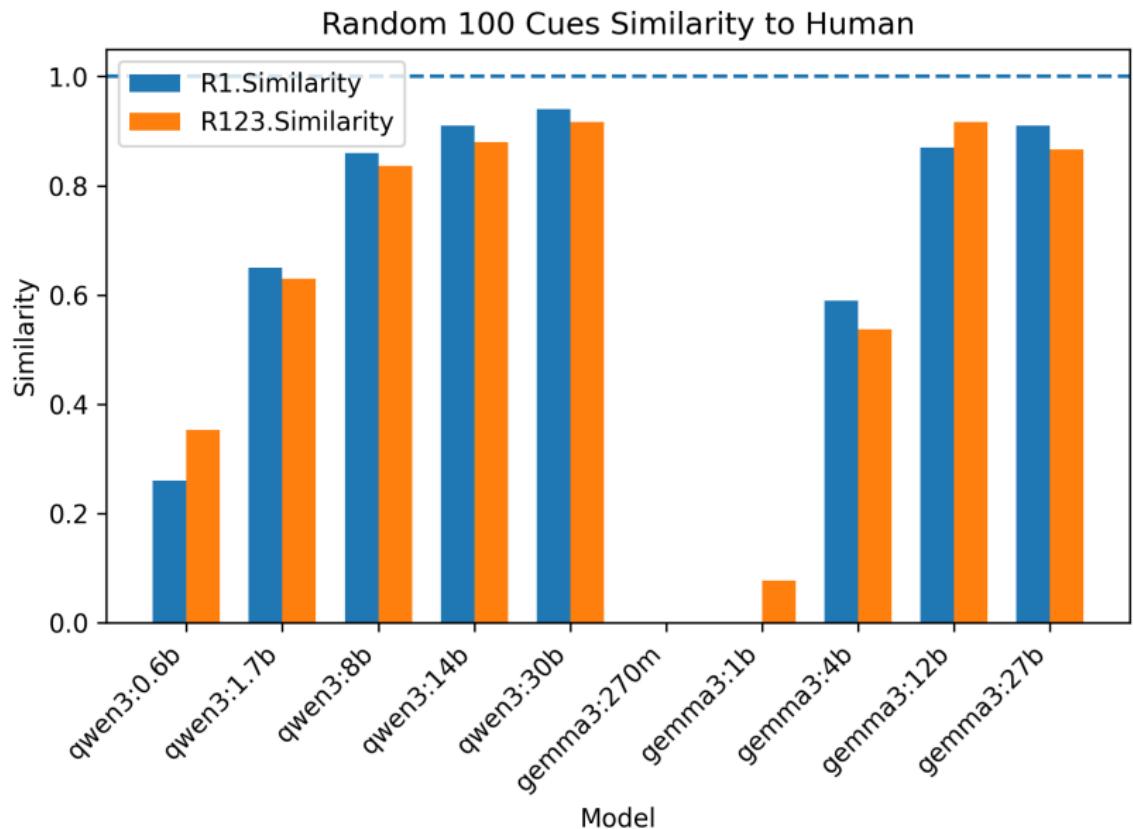
Beobachtungen:

- Modelle assoziieren ähnlich wie Menschen, wenn Menschen eine eindeutige starke Assoziation haben
- Aber assoziieren unterschiedlich zu Menschen bei Wörtern mit schwächerer/zufälliger menschlicher Assoziation

Ähnlichkeits-Analyse 'Top-100'



Ähnlichkeits-Analyse 'Random-100'

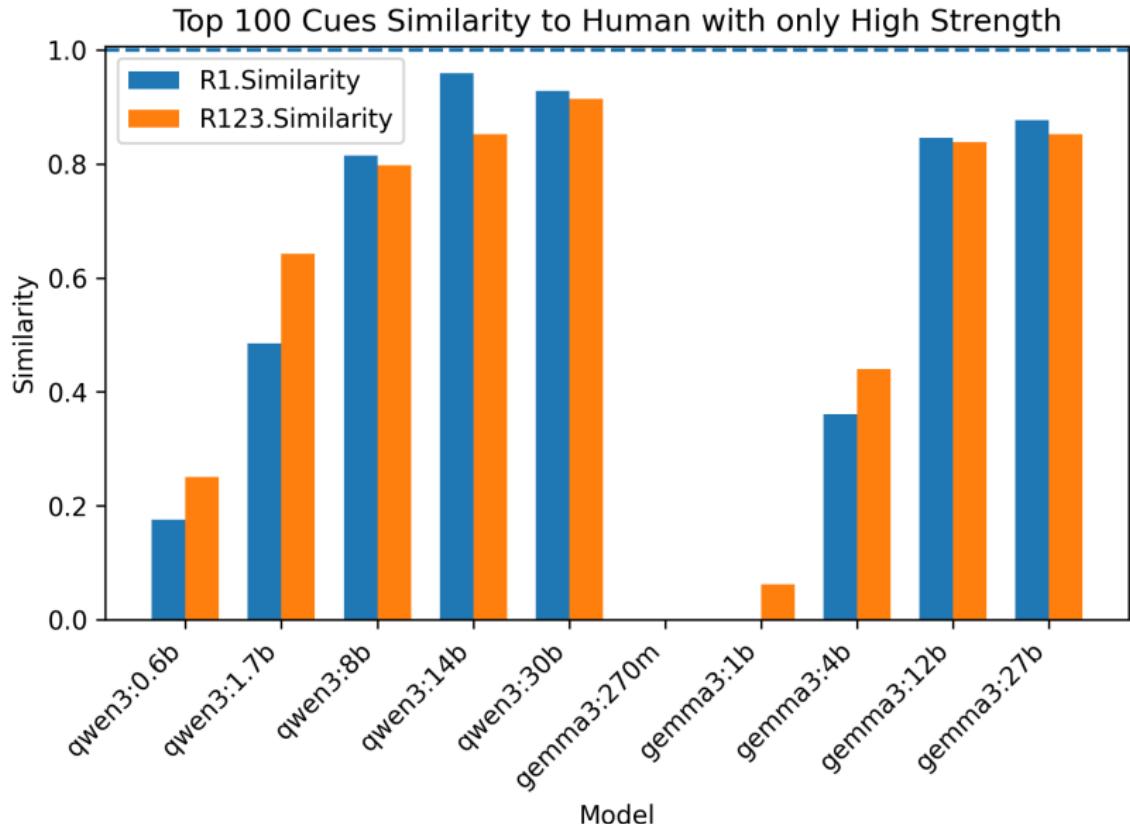


Freie Assoziation Algorithmus

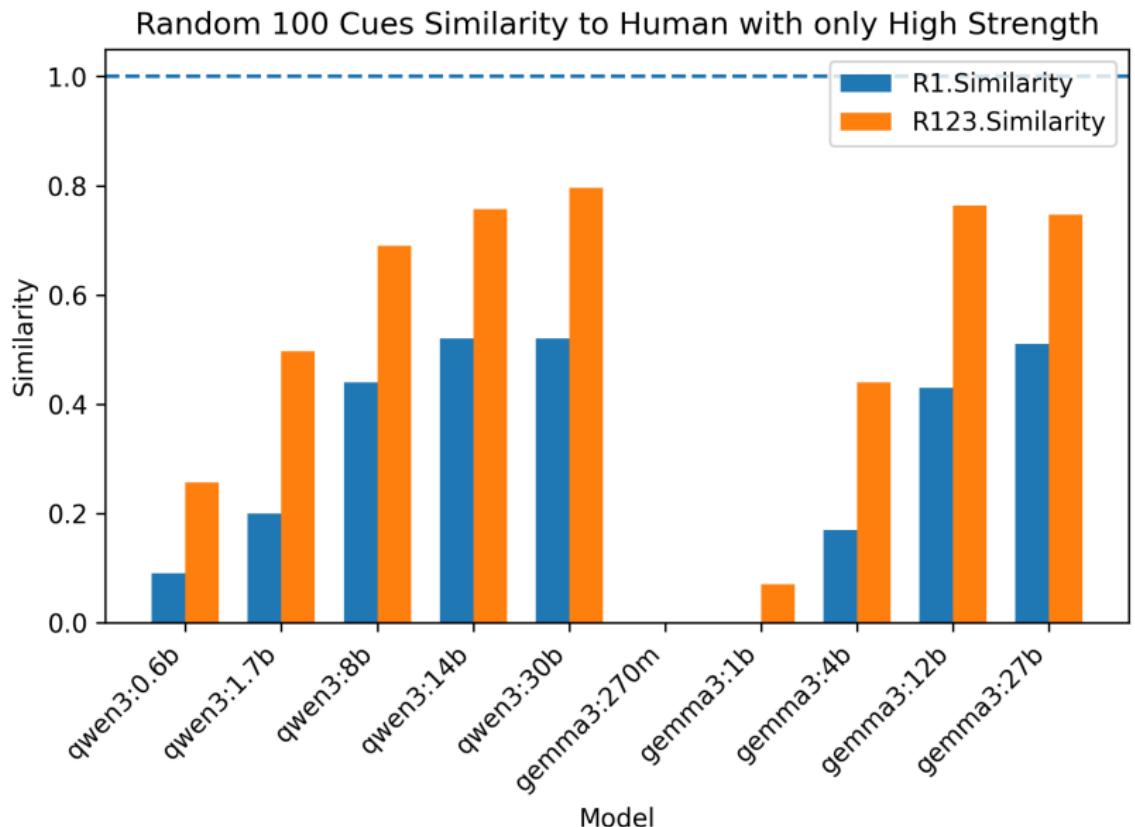
cue	response	R1	N	R1.Strength
arachnid	spider	84	95	0.884
arachnid	afraid	1	95	0.011
arachnid	disgusting	1	95	0.011
arachnid	fear	1	95	0.011
arachnid	Harry Potter	1	95	0.011
arachnid	insect	1	95	0.011
...				

→ Schwache Assoziationen werden genauso gewertet wie "spider", wodurch die Ähnlichkeit zu Menschen verfälscht wird

Anangepasste Ähnlichkeits-Analyse 'Top-100'



Anangepasste Ähnlichkeits-Analyse 'Random-100'



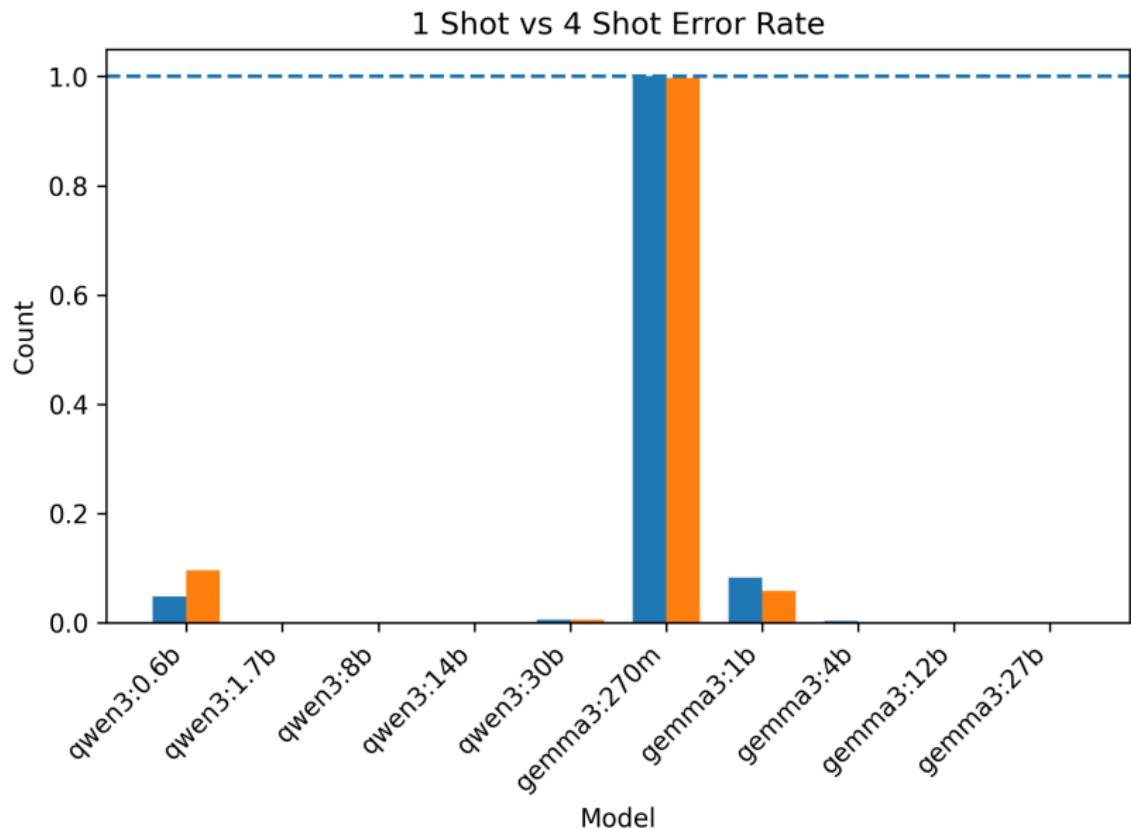
Fazit Freie Assoziation

Vermutung

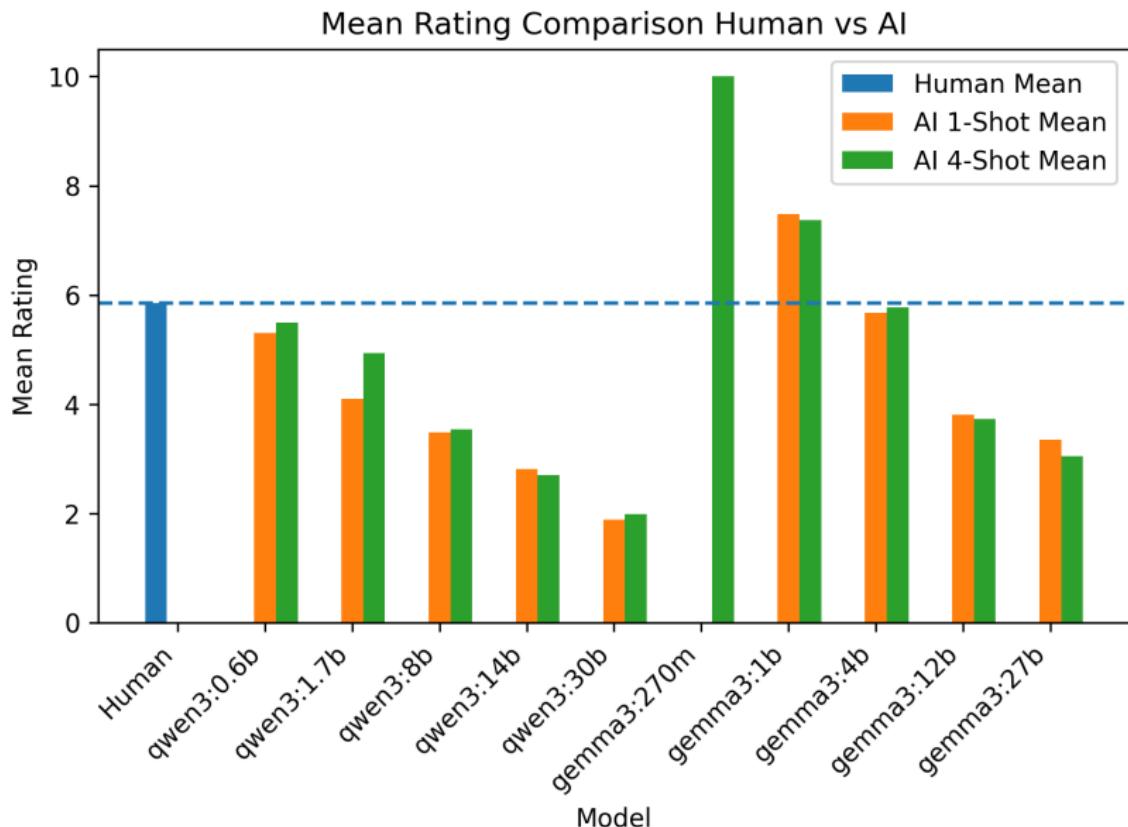
- Wenn Menschen eine eindeutige starke Assoziation haben, ist die Wahrscheinlichkeit höher, dass Modelle diese auch treffen
- Je stärker sich menschliche R1-Antworten streuen, desto geringer ist die Chance, dass das Modell eine menschtypische R1-Antwort trifft
- R123 bleibt in beiden Fällen ähnlich

Ergebnisse: Ähnlichkeitsbewertung

Fehler-Rate



Vergleich Durchschnitt



Fazit Ähnlichkeitsbewertung

Vermutung:

- Menschen komprimieren Bedeutungsräume stärker und vergeben tendenziell höhere Ähnlichkeit
- Größere Modelle können Bedeutungen repräsentieren, was den Unterschied zwischen Wörtern vergrößert

Fazit Ähnlichkeitsbewertung

Beispiel: "Schwert" und "Messer"

- Mensch: "Beides Waffen, also ähnlich" → Hohe Bewertung
- KI:
 - "Schwert ist eine Waffe, Messer ein Haushaltsgegenstand"
 - "Schwert ist größer und gefährlicher als Messer"
 - "Schwert ist eher mittelalterlich, Messer modern"
 - ...
- → Niedrigere Bewertung

Finales Fazit des Projekts

Forschungsfrage

Zeigen LLMs ähnliche **semantische Verwandtschaften, Assoziationen und Bedeutungscluster** wie Menschen?

Eher ja und nein:

- Freie Assoziation: Größere Modelle zeigen ähnliche Assoziationen wie Menschen
- Ähnlichkeitsbewertung: Größere Modelle zeigen niedrigere Bewertungen als Menschen

Quellen zu Datensätzen

- De Deyne, S., Navarro, D., Perfors, A., Brysbaert, M. & Storms, G. 2018. Measuring the associative structure of English: The “Small World of Words” norms for word association. Manuscript submitted for publication.
- Finkelstein, Lev, et al. "Placing search in context: The concept revisited." Proceedings of the 10th international conference on World Wide Web. ACM, 2001.