

# Kognition & Künstliche Intelligenz

Denken, Fühlen und Verhalten im Zeitalter der AI

---

Seminar

7. Januar 2026

# Kognition & Künstliche Intelligenz

Denken, Fühlen und Verhalten im Zeitalter der AI

---

Seminar

7. Januar 2026

# Themenstellung

## 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 Modell Familien?
- Spielt die Größe des Modells eine Rolle?
- Welche Modellgröße erlaubt eine automatische Auswertung der Ergebnisse?

# 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

# 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.)

# Prompt – 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 – Wordassoziation

## 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:

{cue}

# 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: {w1}

Word 2: {w2}

# Experimentelles Setup / Verfahrensweise

1. Setup von **Ollama** mit verschiedenen Open-LLMs
2. Erste manuelle Tests
3. Erstellung der Pipelines für automatisierte Experimente
4. Datenerhebung mit LLMs
5. Statistische Analyse und Vergleich mit menschlichen Datensätzen

# Appendix

---

# Prompt – Systemanweisung

## Prompt

«SYS»

You MUST follow these rules:

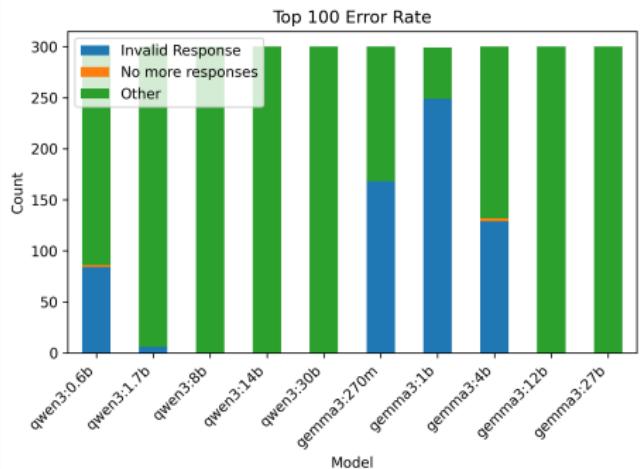
1. Do NOT output reasoning, chain-of-thought, thinking process, analysis, hidden thoughts, XML tags like <think>, or any extra formatting.
2. Output ONLY one single line with exactly four semicolon-separated fields.
3. Format: cue;A1;A2;A3
4. A1-A3 MUST be exactly one word each (no spaces).
5. If you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

«/SYS»

## Ergebnisse: Freie Assoziation

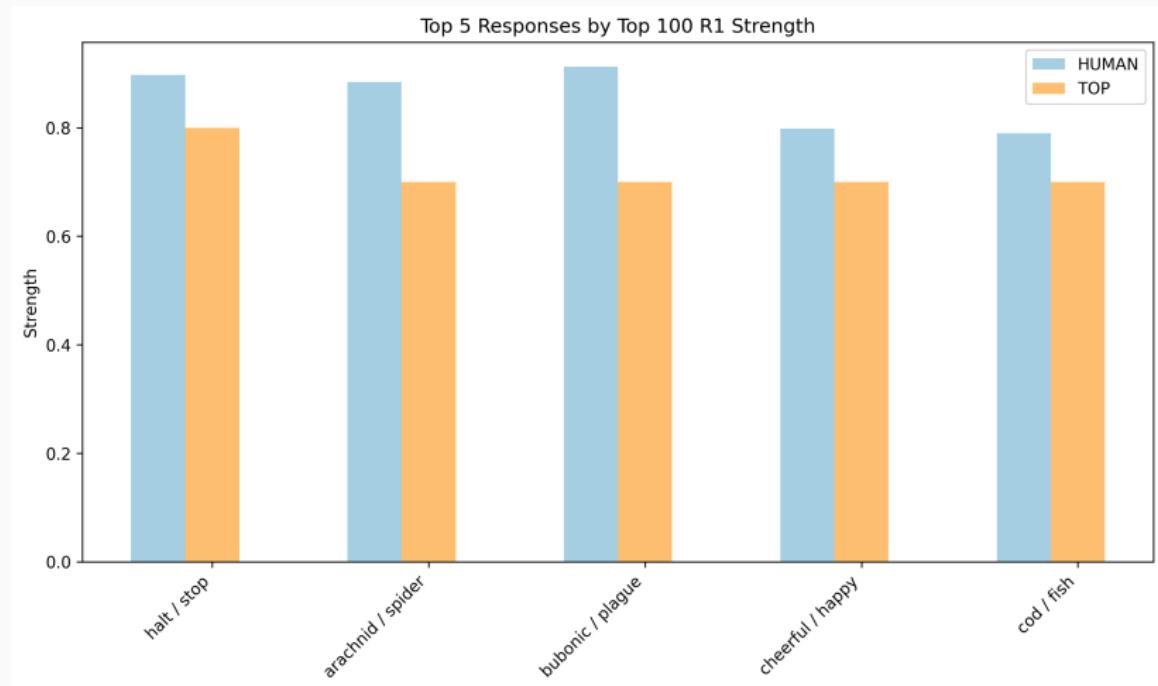
---

# Fehler-Rate

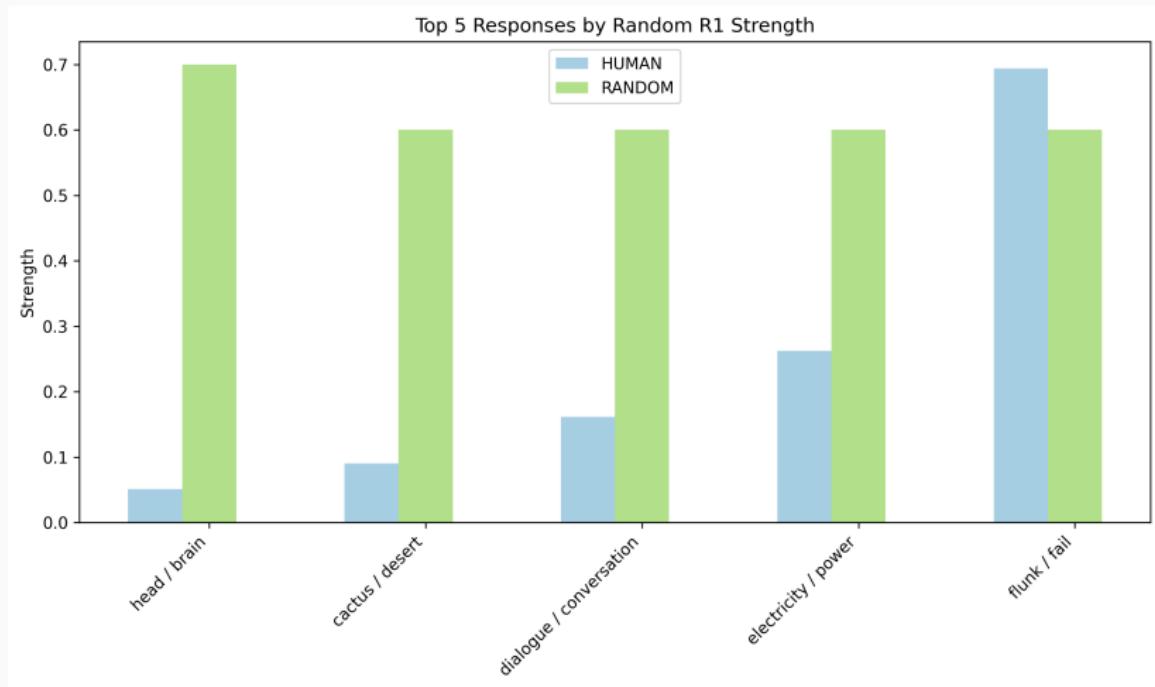


→ Mehr Parameter führt generell zu niedrigere Fehler-Rate

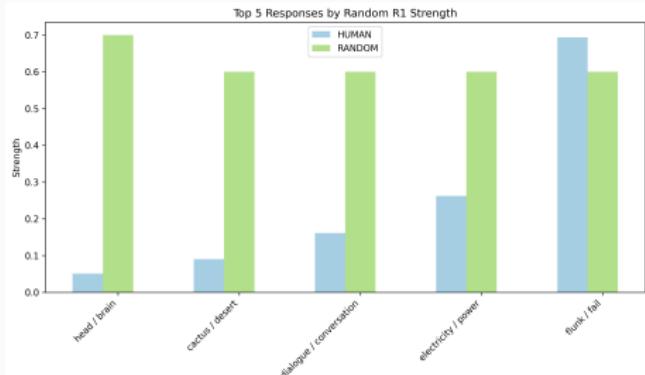
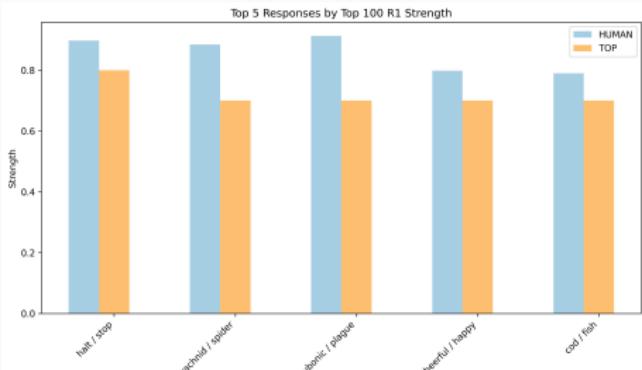
# Vergleich Top Antworten nach R1 (KI Top 100)



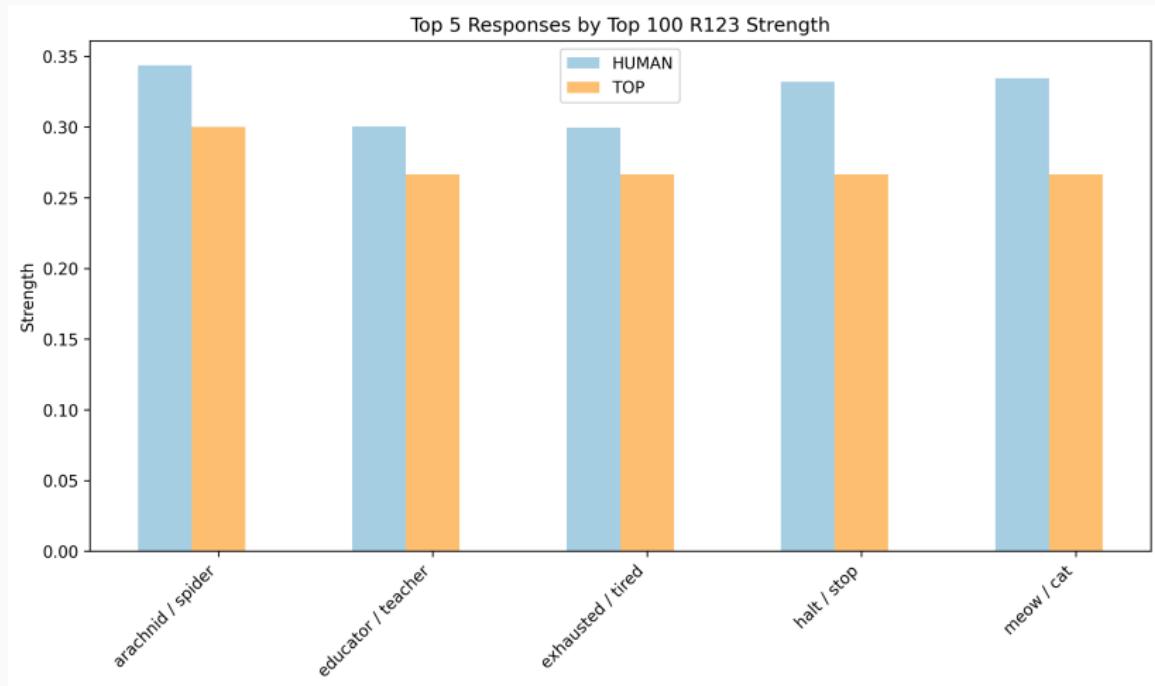
# Vergleich Top Antworten nach R1 (KI Random 100)



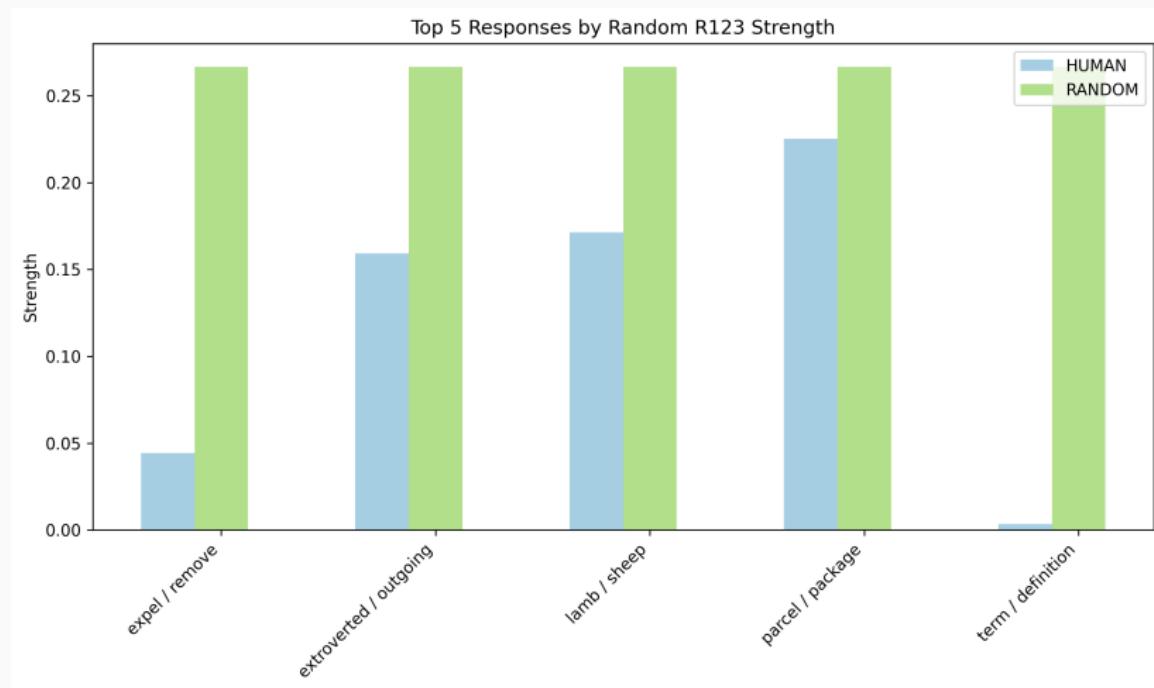
# Vergleich Top Antworten nach R1



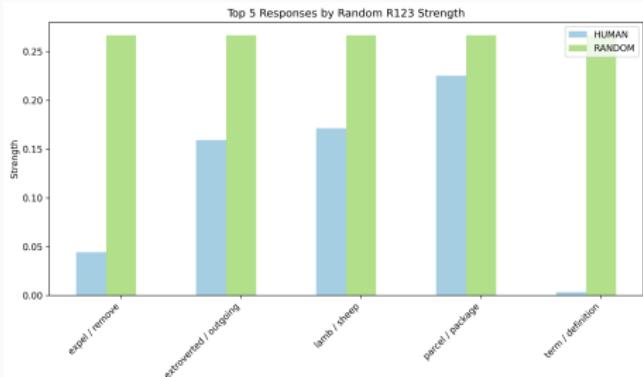
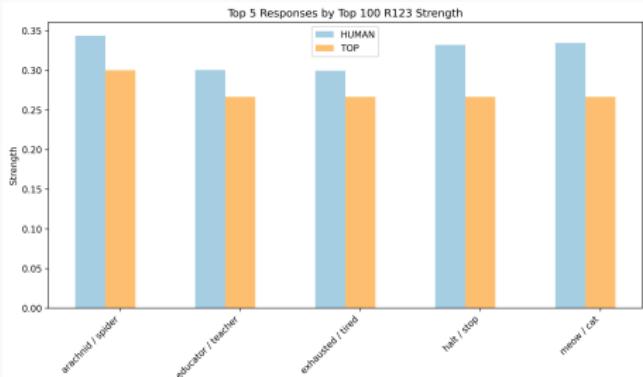
# Vergleich Top Antworten nach R123 (KI Top 100)



# Vergleich Top Antworten nach R123 (KI Random 100)



# Vergleich Top Antworten nach R123

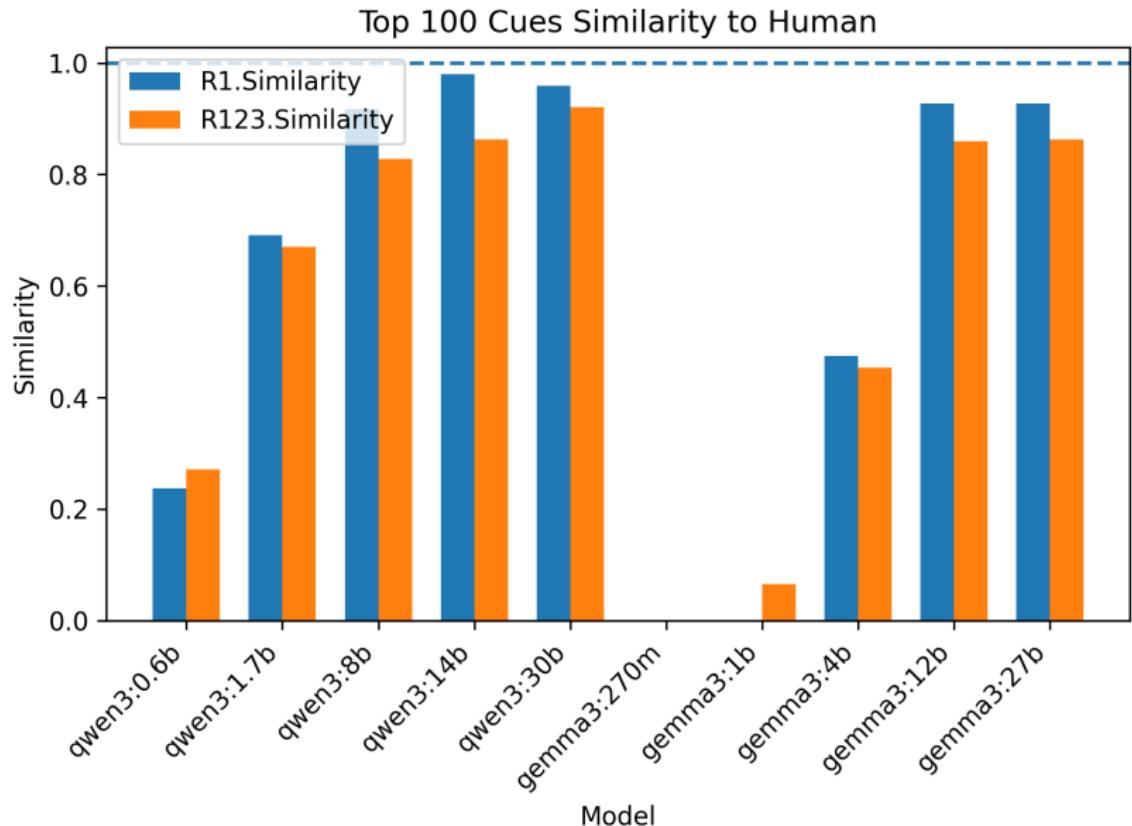


# Vergleich Top Antworten nach R1 und R123

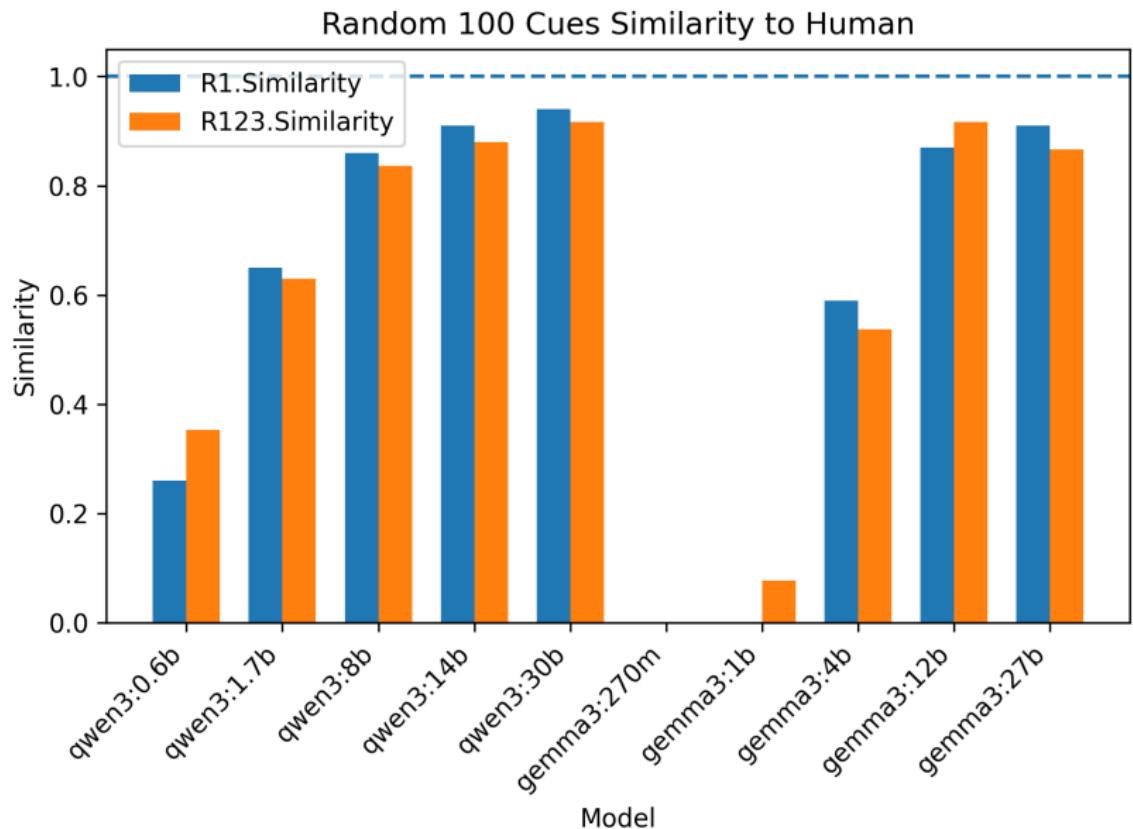
Beobachtungen:

- Modelle assoziert ähnlich zu Menschen gut, wenn sie mit Wörtern gepromted werden, bei denen der Mensch eine starke Assoziation hat
- Aber assoziert unterschiedlich zu Menschen bei Wörtern mit schwächerer/zufälliger menschlicher Assoziation

# Freie Assoziation Top 100



# Freie Assoziation Random 100



# Vergleich Freie Assoziation von Top 100 und Random 100

Beobachtungen:

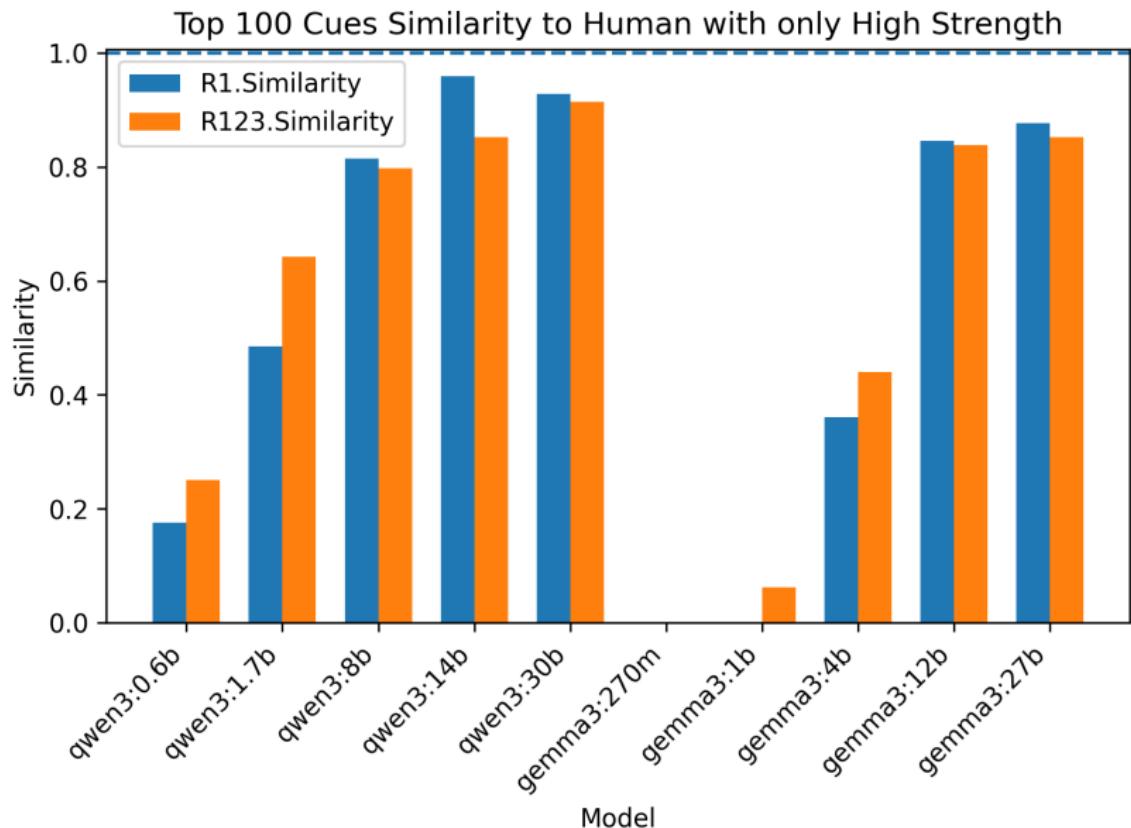
- Beide sehr ähnlich zu Menschen
- Bewertungs Algorithmus Fehlerbehaftet!

# Freie Assoziation Algorithmus

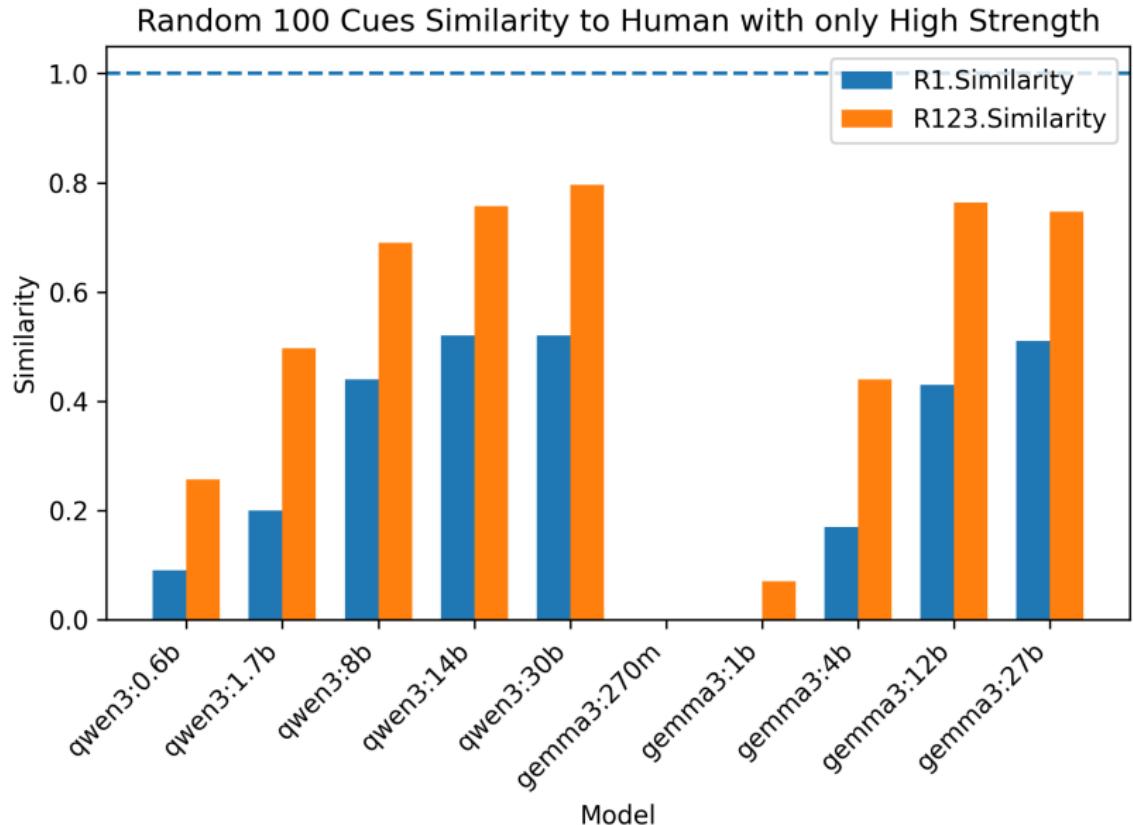
Kategorie	Wort	Wert1	Wert2	Anteil
arachnid	spider	84	95	0.884
arachnid	afraid	1	95	0.011
arachnid	disgusting	1	95	0.011
arachnid	fear	1	95	0.011
arachnid	Harry Potter	1	95	0.011
arachnid	insect	1	95	0.011
...				

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

# Anangepasste Freie Assoziation Top 100



# Angepasste Freie Assoziation Random 100



# Vergleich angepasste Freie Assoziation von Top 100 und Random 100

Beobachtungen:

- Top 100 immernoch sehr ähnlich zu Menschen
- Random 100 deutlich weniger ähnlich zu Menschen

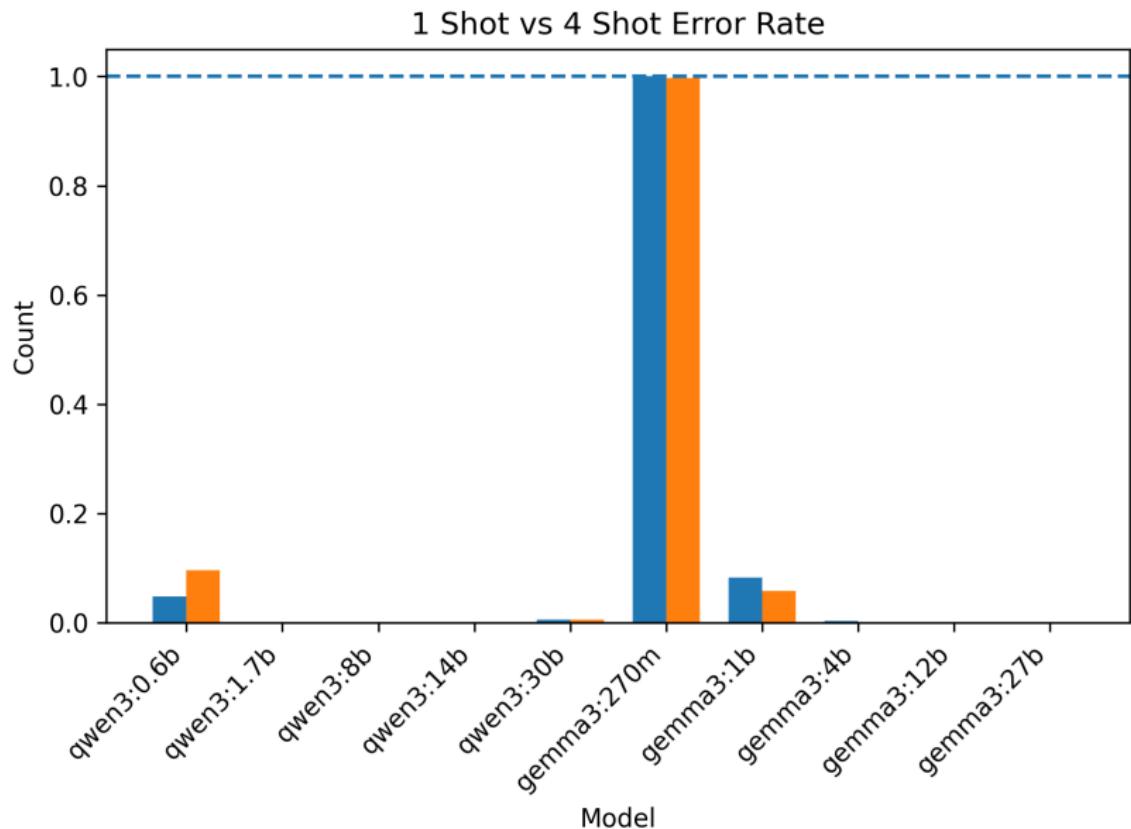
# Fazit Freie Assoziation

- Bei starken Assoziationen zeigen größere Modelle ähnliche Assoziationen wie Menschen
- Bei schwächeren/zufälligen Assoziationen weichen Modelle stärker von Menschen ab

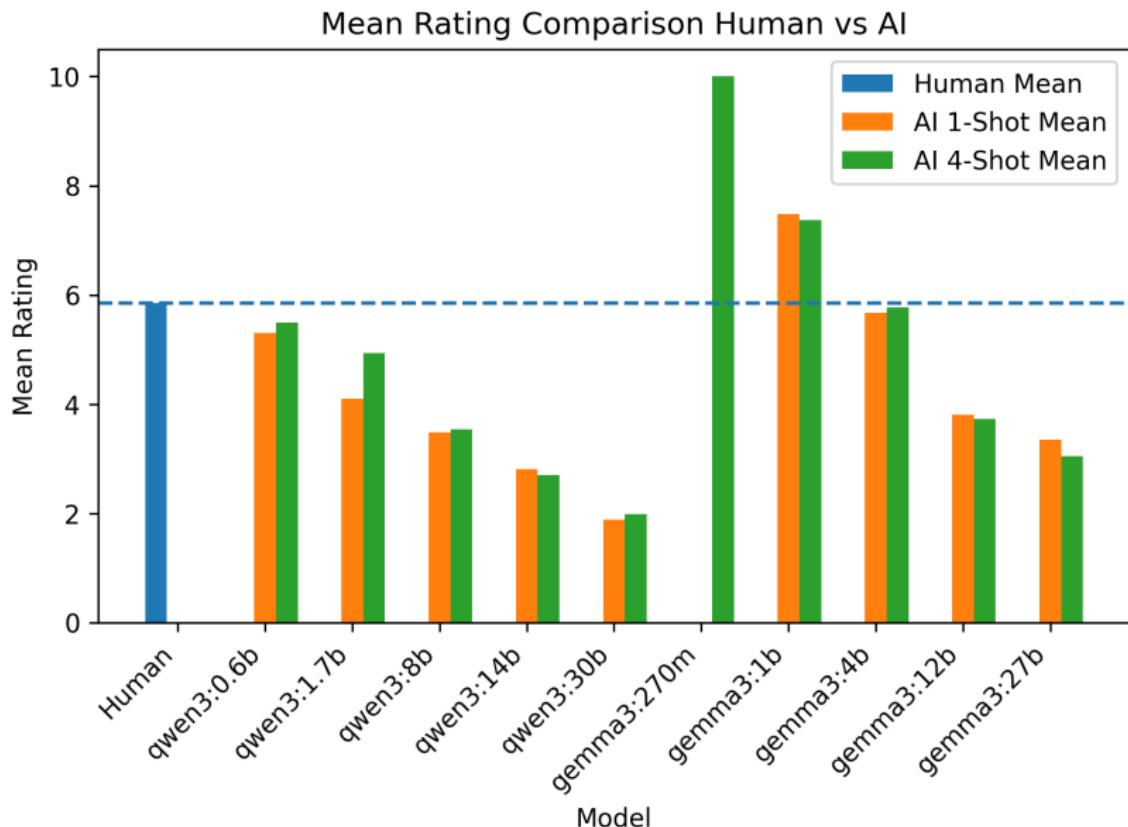
# Ergebnisse: Ähnlichkeitsbewertung

---

# 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 als nein:

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