



Conclusion

Elicit could be used as a tool for individual steps in the review process, but not as a stand-alone application.

Diskussion

There was a higher agreement between reviewers and AI than between R1 and R2 in the abstract and title screening. The AI had similar sensitivity and specificity rates like humans. The best results were achieved by all R1+R2 together. Time savings are significant – you can save at least 50% of time, but finding your way around the user interface and entering the right prompts took a lot of time at the beginning. In addition, it takes a person to make the right selection of studies beforehand. Furthermore, the results are only comparable to a limited extent, as the AI more or less combines the TAS and FTS steps. Furthermore, Elicit didn't find for all studies the full-text article.

Results II

- **TAS time needed**
- R1-R3: 32min
- AI: 17min (without preparation time, all prompts are already defined)

FTS of Trail 1:

- 2 full-texts R1-R2 agreed
- 3 full-texts had to be discussed
- consensus: N=5
- AI: 11
- time needed: R1-R2: 1,3h
- AI: 0min

FTS of T1:	Sensitivity	Spezifität
Reviewer1	0.80	0.88
Reviewer2	0.80	0.88
AI	1.0	0.25

Introduction

Systematic reviews are crucial for synthesizing existing evidence on a given topic. They help identify research gaps in the literature and guide future research. However, as the volume and complexity of published studies continue to rise, traditional approaches to systematic reviews have become increasingly time-consuming and less efficient. AI-based tools present an opportunity to streamline the review process and enhance efficiency.

Objectives

- Using AI-software vs. manual screening
- on studies about fear/threat conditioning + fMRI
- quality of outcomes, e.g. specificity/sensitivity
- potential time savings

Methoden

- Title- & Abstract-Screening (TAS)
 - Trial 1: n = 50 (T1)
 - Trial 2: n = 50 (T2)
- Full-Text-Screening of T1(FTS)
 - Trail 3: n = 13
- Reviewer 1, 2, (R1, R2, blinded) vs. AI
- Disagreement → Consens (R1-R2)
- Statistics:
 - specificity (true negative rate)
 - sensitivity (true positive rate)
 - cohens kappa (match R1, R2, AI)

Sensitivität = $TP / (TP + FN)$
Spezifität = $TN / (TN + FP)$, wobei $TN = 39 - FP$

$$\kappa = \frac{P_o - P_c}{1 - P_c}$$

Results I

- T1 – TAS**
 - 7 abstracts R1+R2 agreed
 - 6 abstracts had to be discussed
 - **Cohens Kappa: 52-78% match R1/R2-AI**
 - **59% R1-R1**
 - Consensus T1: N=13 → full text screening
- T2 – TAS**
 - 4 abstracts all R1 + R2 agreed
 - 7 abstracts had to be discussed
 - **Cohens Kappa: 60-69% match R1/R2-AI**
 - **55% R1-R1**
 - Consensus T1: N=13 → full text screening

Prompt:
Does the study use fMRI OR functional magnetic resonance imaging OR other functional neuroimaging methods while conducting fear/threat conditioning paradigms?

Trail 2: Person	True positive	False negative	False positive	True negative	Sensitivity	Spezifität
R1	10	1	5	34	0.91	0.87
R2	9	3	3	36	0.82	0.92
AI	8	2	3	36	0.73	0.92

Trail 1: Person	True positive	False negative	False positive	True negative	Sensitivity	Spezifität
R1	8	5	4	33	8 / (8+5) 0.62	33 / (33+4) 0.89
R2	13	0	0	37	1.00	1.00
AI	3	0	6	31	1.00	0.84

• Include 5 / 5

This study represents a primary research investigation using fMRI to examine fear conditioning in human participants with specific phobia, comparing them to healthy controls. The research comprehensively incorporates neuroimaging techniques while exploring fear acquisition, generalization, and extinction learning mechanisms in a psychiatric population. All screening criteria are definitively met through the study's rigorous design and clear methodological approach.

• Complete Components • Human Psychiatric Population • Neuroimaging And Conditioning • Study Type • Species