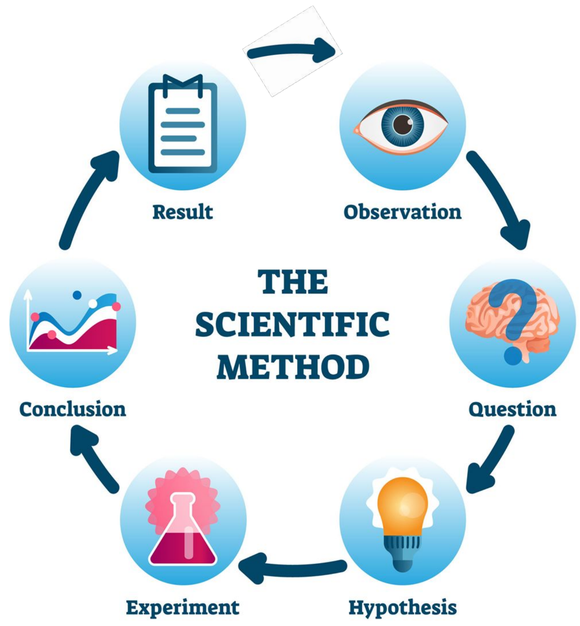
Using AI to validate assertions in science

Rotem Shimony, Avishai Henik, Moti Salti, Yoel Shilat

How can we use AI to improve validation in the scientific community?

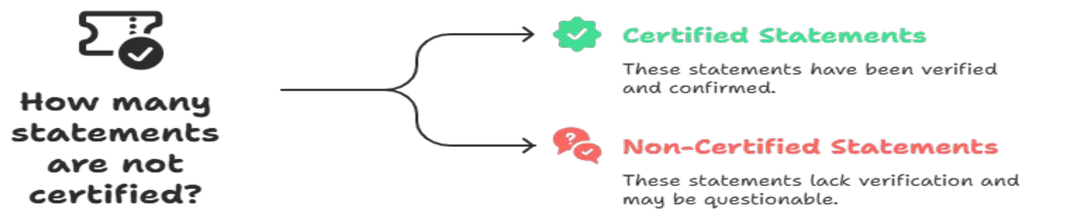
1. The Challenge

- Scientific knowledge relies on the collection of empirical results.
- In the scientific community there are appearances of statements not properly backed by evidence, these are called assertions.
- “An assertion is a statement acknowledging that something is true” (Frege,1918).
- These assertions present themselves as true, yet lack supporting evidence — contradicting the principles of the scientific method.
- By identifying such assertions, we can prevent future research from heading down unsupported or misleading paths.



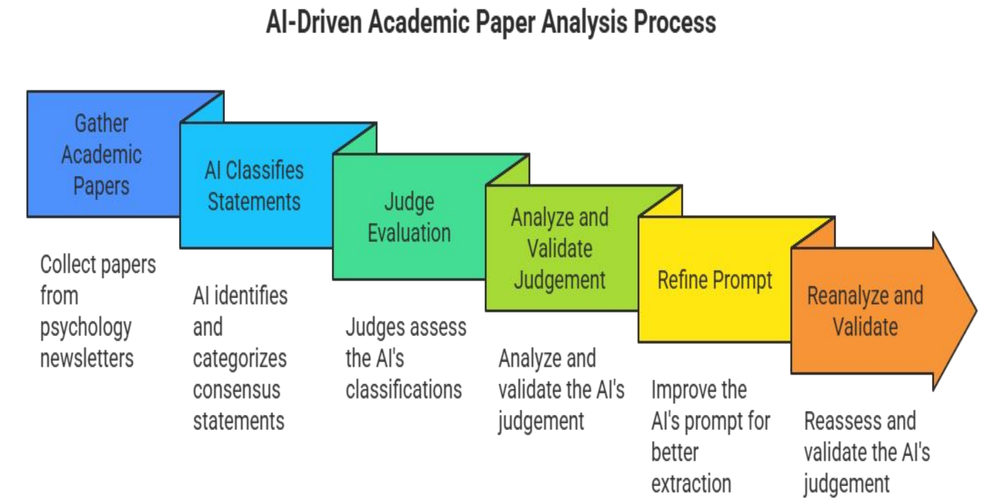
2. The research’s objective

- We explore how an AI large language model (Gemini 2.5) can classify scientific statement.
- Specifically we focused on identifying assertions on scientific consensus. Such as "the majority of studies show that...", "The dominant view is.." etc.



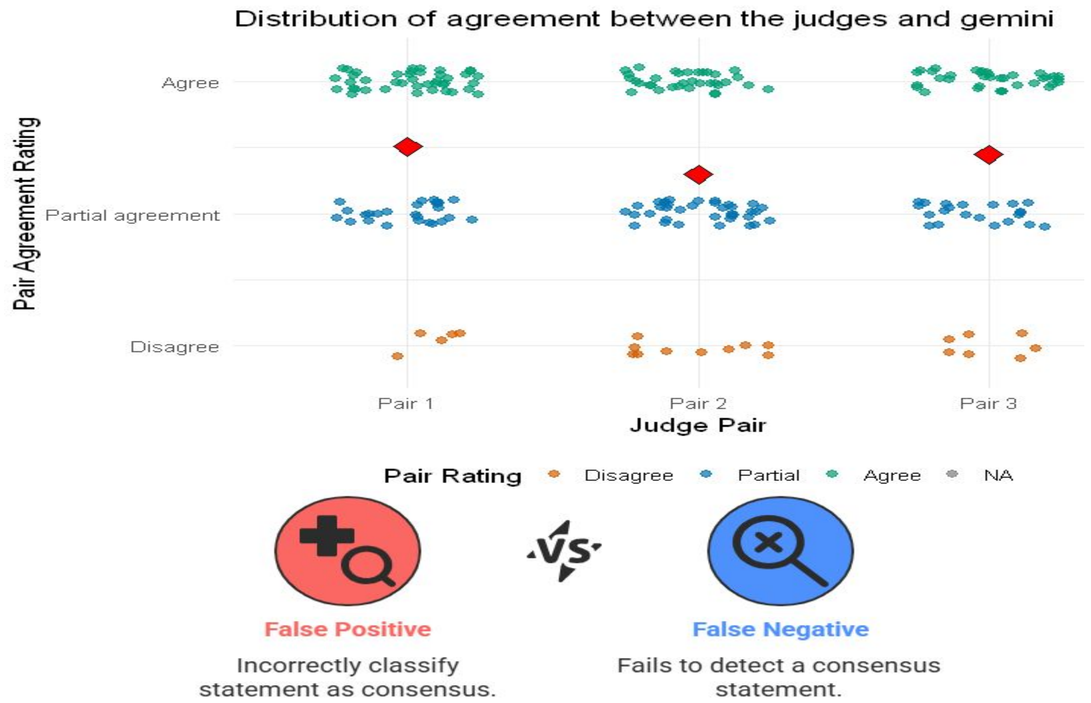
3. Method

- Scientific statements - from papers published since 2015 in the top 3 psychology journals (JCR ranking)



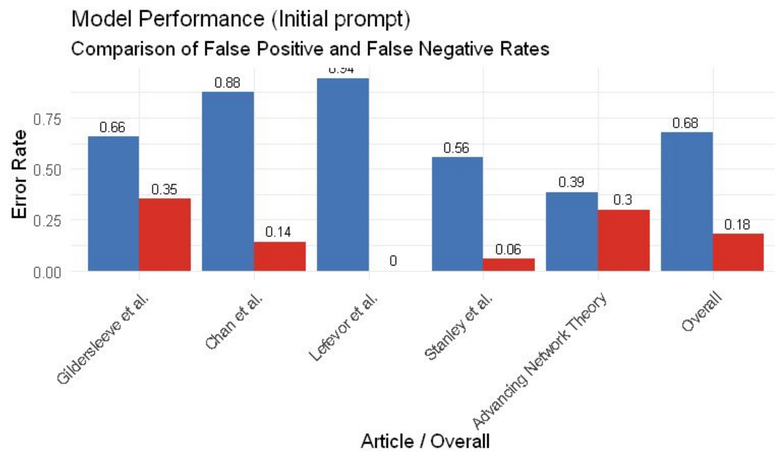
4. Results

- We found 1623 statements expressing consensus.
- 47.5% of them have been backed by citation.
- Human judge’s judgement of AI assertions inter-rate reliability - 0.603 (SE = 0.303, *P* = 0.002)
- Agreement between judges and gemini - 0.7 (SE = 0.309, *P* < 0.05)



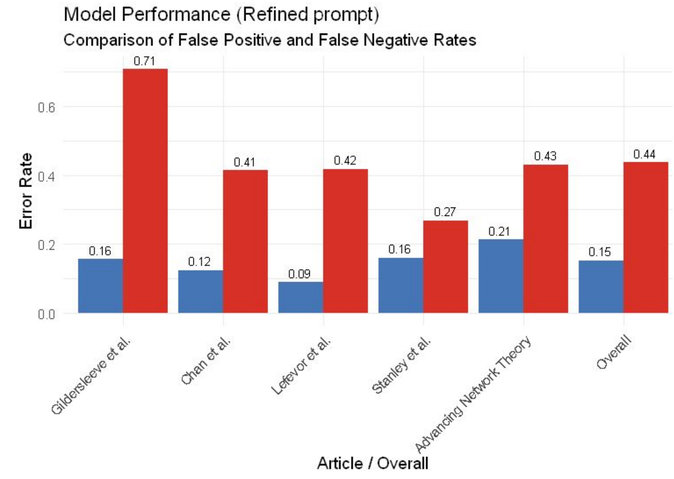
Initial prompt

Higher false negative and lower false positive



Refined prompt

Higher false positive and lower false negative - reduced misses



Overall reduction in missed statements

5. Discussion

- In the current study we created a database of assertions reflecting the scientific dynamic of consensus.
- Improved prompt (few-shot learning and chain-of-thought) and improved training set led to improved performance - Higher detection rates and fewer missed assertions.
- Study provided an initial proof of classification ability - model achieved 70% agreement with human annotators
- Overall, this study demonstrates the potential of using AI to classify unsupported scientific assertions.
- Less than 50% of scientific assertions were supported by evidence, this finding highlights the need for better validation practices.