



Assessing ChatGPT's utility in systematic reviews: COVID-19 seroprevalence case study



Hankun Chen, Rachel Howard, Francisca Mutapi

1) ABSTRACT

Introduction: Advancements in artificial intelligence (AI) will undoubtedly have implications for scientific research. With the recent launch of the novel AI platform ChatGPT, the present study examined the feasibility of outsourcing the data extraction phase of systematic reviews to AI technologies.

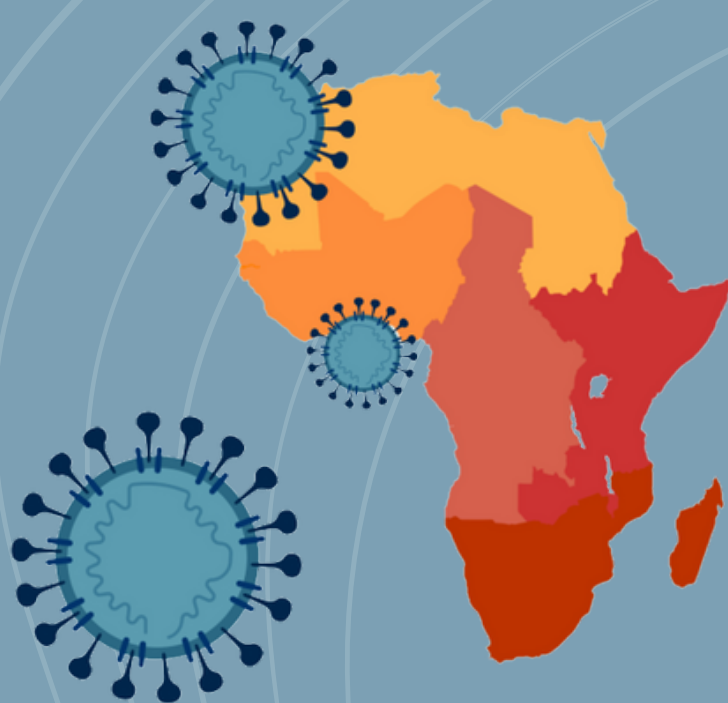
Methodology: Three versions of a "Task list" were formulated for ChatGPT to memorize, with the aim of establishing an assembly line approach.

Results: Extracting data based on providing only the paper title in the query was unsuccessful. Even following "Task list" modifications data discrepancies persisted.

Conclusion: Two main reasons for the failure of extraction were proposed: 1) the limited database of ChatGPT; 2) queries using only the paper title provide insufficient information to retrieve the relevant publication

2) INTRODUCTION

- To assess the novel applications of AI technologies to contribute to scientific research, our study investigated the utility of ChatGPT to conduct a systematic review characterizing the seroprevalence of SARS-CoV-2 on the African continent.
- ChatGPT is a novel AI platform launched by OpenAI that has been controversially incorporated into scientific research.
- The ChatGPT platform is pre-trained with text resources from online databases and programmed to respond to text prompts with humanised intuition.
- As the manual extraction of data from publications is a time-consuming process, we evaluated the proficiency of ChatGPT to accurately obtain data from online databases to streamline the process of conducting a systematic review



3) METHODS AND RESULTS

1 Methodology: A task list designated "Rule 1" was devised to instruct ChatGPT to extract data.

Results: Although ChatGPT could memorize the task list, incorporating only the paper title into the query search generated false data.

2 Methodology adaptation 1: The false data may be attributable to ChatGPT gathering information from unrelated papers. "Rule 2" to ensure that the AI presents the abstract and methodology corresponding to the provided title.

Results: ChatGPT's results still displayed considerable bias and inconsistencies. Moreover, ChatGPT altered the content and sequence of assigned tasks, occasionally omitting certain requirements.

3 Methodology adaptation 2: "Rule 3" was formulated to stabilise the task list, ensuring no data loss.

Results: The data still failed to match the original text. Numeric data was the primary source of mismatching.

4 Methodology adaptation 3: "Rule 3 v0.2" was created to refine the extraction scope to only text data to enhance precision.

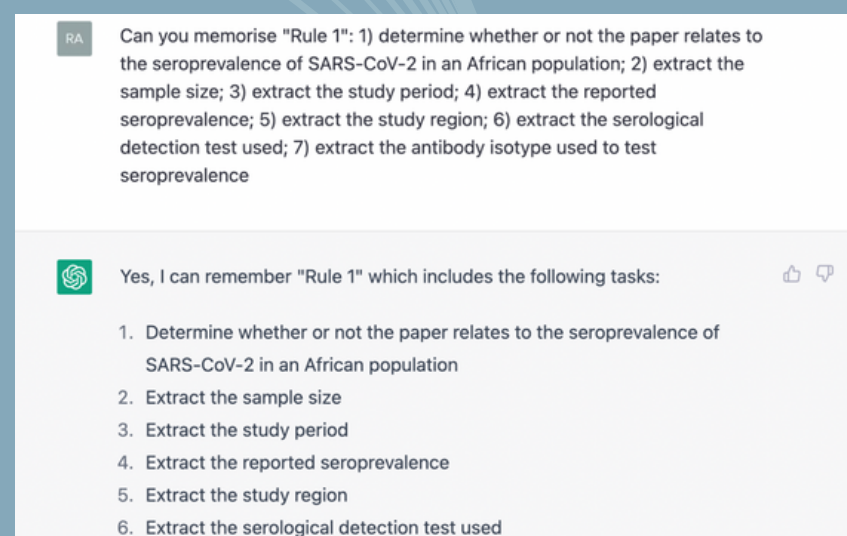
Results: Unfortunately, the implementation of "Rule 3 v0.2" yielded no improvements.

5 Methodology adaptation 4: Further investigation revealed that ChatGPT's accessible database is confined to information predating September 2021.

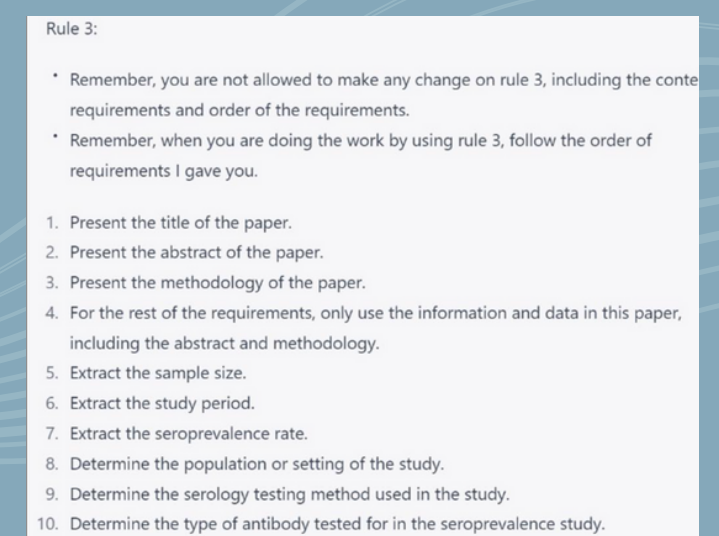
Results: Searching papers published prior to this date did not improve extraction success.

6 Methodology adaptation 5: A "WebChatGPT" Chrome plugin is available that allows ChatGPT to access current internet and database resources.

Results: However, the plug-in is primarily intended for general internet searches and permits the addition of only five papers to ChatGPT, rendering it inadequate for resolving the issue.



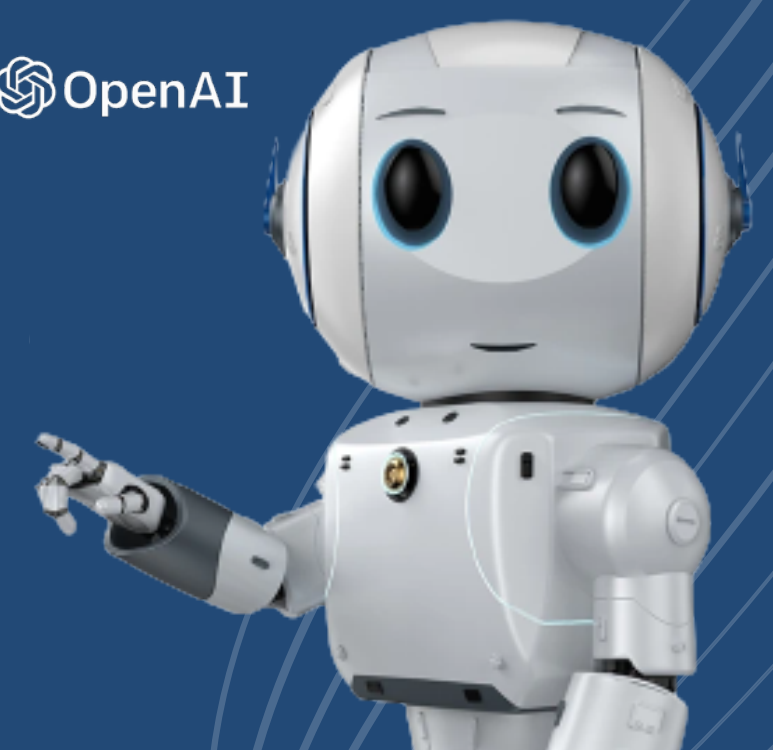
"Rule 1" text prompts



"Rule 3" text prompts

4) CONCLUSION AND DISCUSSION

- Our findings suggest data extraction should not be outsourced to ChatGPT, as the platform exhibited a propensity to present misleading data that was contextually appropriate but factually inaccurate.
- Two main problems were identified that undermined the utility of ChatGPT: 1) the limited database of ChatGPT; 2) queries using only the paper title provide insufficient information to retrieve the relevant publication.



Comparison of ChatGPT to the traditional data extraction software Data thief

- Data Extraction:** Data Thief extracts graphic data, whereas ChatGPT only analyses text and cannot process graphic data.
- Accuracy:** Data Thief's accuracy depends on image quality. ChatGPT's accuracy relies on information detail and its presence in the platform's pre-September 2021 training set.
- Active development:** Data Thief lacks updates and may face compatibility issues, whereas ChatGPT's development remains active.
- Calibration:** Data Thief requires a manual calibration process, which takes time before data extraction can occur.