Review of: "Data Science using openAI: testing their new capabilities focused on data science"

Adriano Bressane[1]

1 Universidade Estadual Paulista

Potential competing interests: No potential competing interests to declare.

Dear author(s) and editor,

Regarding the submission "*Data Science using openAI: testing their new capabilities focused on data science*" please find my comments below.

The study discusses the potential of openAI's coder interpreter API in performing statistical analysis on datasets, allowing non-experts to obtain evidence-based answers in various fields. This can be considered an interesting topic for the research field; however, it is necessary to clarify the following points:

1. The focus of the paper is on discussing the possibilities and potential benefits of the API. However, it is essential to also explore its limitations or address any potential challenges that may arise in its usage.
2. The abstract provides a comprehensive overview of the paper's objectives. However, it lacks precision in articulating the novelty of the OpenAI Codex Interpreter API in addressing these challenges. Please consider clearly defining the specific challenges faced by researchers in statistical analysis, and explicitly state the objectives of utilizing the OpenAI Codex Interpreter API to overcome these challenges.
3. The introduction provides a comprehensive overview of the evolution of chatGPT and its impact on coding practices, emphasizing the shift from basic coding to abstraction. Please consider articulating the primary purpose of the paper, which appears to be exploring the transformative impact of chatGPT on coding practices and its application in data science. Moreover, emphasize the interdisciplinary potential of the OpenAI API by illustrating how it can replace the need for a multidisciplinary team in certain data science tasks.
4. The Methods section outlines the testing procedures and details the use of the OpenAI API capabilities. Please consider elaborating on the technical aspects of utilizing the OpenAI API, especially the integration of the Assistant, Files, and Code Interpreter capabilities. Specify how these capabilities were employed in conjunction. In addition, offer more specific information on how the assistants were configured for different datasets.
5. Regarding the Results and Discussion, elaborate on how the Code Interpreter can help data scientists avoid internal biases and ideologies when interpreting datasets. Discuss the potential impact on the use of proprietary software in data science, considering the rise of open-source platforms and the role of the open-source community. Acknowledge the limitations of the Code Interpreter, such as the need for users to know what to ask and the stochastic behavior observed at times. Discuss any challenges faced during the analysis and provide insights into potential improvements or future developments.

6. The conclusion provides a concise summary of the paper's exploration of OpenAI's Codex Interpreter tool. Overall, the section effectively synthesizes the key findings but could benefit from minor refinements for clarity and specificity.

From the foregoing, I recommend reconsidering the paper after clarifying these issues.

Sincerely.