HUMAN VS AI: COMPARISON OF SCIENTIFIC PAPER DRAFTING CAPABILITIES

TsankoYotsov[1], Martin Petkov[2], PetarBonev[3]
1.Ruse University "Angel Kanchev", Department of Health care, UMHAT Medica Ruse
2.Independent writer and marketer, author of the "Metaverse AI" book trilogy
3.Ruse University "Angel Kanchev", Department of Medical and Clinical Diagnostic Activities, UMHAT Medica Ruse
tsankoyotsov@gmail.com

INTRODUCTION

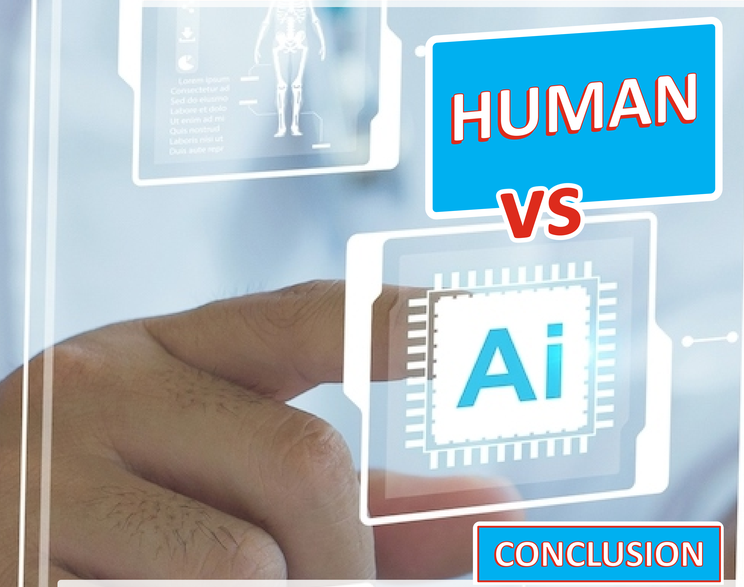
Artificial Intelligence (AI) is widely used in research and has recently been applied to writing scientific papers.

While AI can generate drafts rapidly, it requires well-structured, detailed prompts to produce high-quality content. This study evaluates the capabilities of two free to use AI platforms, **ChatGPT-4o** and **Claude 3.5 Sonnet**, in generating scientific papers compared to human-written articles.

MATERIALS AND METHODS

Our objective was to generate the entire article from a single prompt, encompassing uploaded data (a coded dataset related to the study), a codebook explaining the data structure, and a questionnaire detailing the research methodology. This approach aimed to simulate a realistic scenario where AI is tasked with producing a comprehensive scientific output based on provided information. We designed a detailed prompt that provided all necessary guidance and contextual information for both platforms.

AI outputs were compared for **quality, references, statistical analysis,** and **completeness.**



RESULTS

Both **ChatGPT** and **Claude** successfully generated initial drafts. **ChatGPT's** output was slightly more thorough and included visualizations of the provided data. However, the initial output lacked references. **Claude**, while unable to visualize the data, included five references in its initial draft. Notably, these references were not fabricated, but drawn from its training data.

Subsequent refinement with additional prompts was necessary for both platforms.

ChatGPT's web access facilitated real-time literature searches and integration of relevant studies. **Claude**, lacking this capability, produced a longer and more comprehensive article after additional prompting. However, its final reference list contained some fabricated articles and discrepancies between the referenced articles and the provided DOIs. They both provided short reference lists, incomparable to the one provided by human both in quality and quantity.

Both AI tools used **descriptive statistics** and **Chi-square tests,** but lacked **detailed equations**. **Claude** also applied the **McNemar test**, aligning with human statistical choices.

CONCLUSION

Both platforms successfully generated a first draft from a single, comprehensive prompt. However, the quality of these initial drafts was not comparable to human-generated articles and **would likely not have passed peer review**. Subsequent refinement with follow-up prompts significantly improved both the length and quality of the generated text. LLMs can serve as helpful aids in the drafting and refinement stages of scientific writing. However, the tested free versions of ChatGPT and Claude, in their current state**, lack the capabilities to produce articles that meet the rigor and completeness standards of human-authored scientific publications.**