September 19, 2024

Att. Douglas Steinley

Editor, *Psychological Methods*

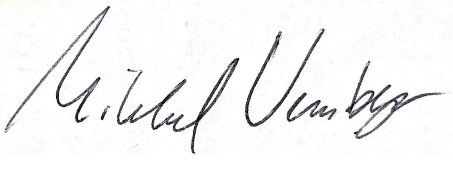
Dear Dr. Steinley

For your consideration, we are submitting the manuscript “GPT API Models Can Function as Highly Reliable Second Screeners of Titles and Abstracts in Systematic Reviews: A Proof of Concept and Common Guidelines.” The manuscript contains 34 double-spaced pages with 10.843 words, not including references, tables, figures, and text boxes. The article includes three figures, five tables, and three text boxes. R code to reproduce all parts of the article can be found at <https://osf.io/apdfw/> and https://github.com/MikkelVembye/screen\_benchmarks. Along with the paper, we have developed the AIscreenR R package that supports the major aims of the article. This open-source software can be assessed at <https://mikkelvembye.github.io/AIscreenR/>. A preprint of this article is available at <https://osf.io/preprints/osf/yrhzm>.

My co-authors and I believe that the article will be relevant for the readers of *Psychological Methods* for the following reasons. First of all, we introduce a novel method for screening titles and abstracts in systematic reviews with (OpenAI’s) GPT API models that among other things can release substantial costs in most systematic reviews. We validate this screening method by showing that GPT API models can yield screening performances on par with humans and in some cases even better than human screeners when compared to typical human screener performances in high-quality systematic reviews in psychology and the social sciences. To support future applications of this screening approach, we develop a novel statistical benchmark scheme and guideline for when it is appropriate to use GPT API models of title and abstract screening in high-quality reviews. Moreover, we present the workflow for how such screenings can be standardized, and present the R package AIscreenR to support this purpose. Altogether, we hope that these contributions can ease some of the challenges encountered when doing independent human duplicate screening of titles and abstracts in systematic reviews. For example, reducing rigid and costly human labor. Ultimately we hope to make this method acceptable within high-quality systematic reviews, such as the ones published in *Psychological Bulletin*.

Thank you for considering this work.

Sincerely,



Mikkel Helding Vembye

Researcher

VIVE - The Danish Center for Social Science Research   
Phone: (+45) 3131 9209

Mail: [mihv@vive.dk](mailto:mihv@vive.dk)