Dear authors,

Thank you for making the updates to the document based on the reviewers' comments. We appreciate your efforts. However, there are still a few points to address:

1. In the abstract, could you please clarify the study participants? If there are character constraints, I would suggest removing the explicit reference to Jupyter Notebook, as it may not contribute significantly to the understanding of the study's use.

2. *"However, RAT is not supposed to be limited to the evaluation of search systems. <u>When looking at the field of health, for instance, we found studies performing quality evaluations (e.g., Janssen et al., 2018) or content analyses (e.g., Rachul et al., 2020) of health-related search results.</u> Studies in the field of media and communication studies classify search results, e.g., regarding content type and <u>ideological bias</u> (Ballatore, 2015). We see great potential for the RAT here, as studies based on search results are often based on small datasets that were collected, evaluated, and analyzed manually, especially in studies based on search results from commercial search engines, because researchers usually do not have access to such data. In these cases where researchers do not have access to a search engine provider's data, <u>questions arise on the evidential significance of the conclusions acquired using small data sets. Unfortunately, most researchers do not have access to these datasets.</u> Having such access would help conduct studies that search engine providers do not carry out. These studies could cover information retrieval aspects, but also areas where the search engine providers have a vested interest in keeping the results hidden because they <u>may contradict their self-interests. The phrase "self-interest" refers to the major financial interests of search engine companies.</u>"*

   The following clarifications refer to the underlined sentences above:

   - When examining the field of health, for example, we found studies conducting quality evaluations (e.g., Janssen et al., 2018) or content analyses (e.g., Rachul et al., 2020) of   health- related search results, which are particularly relevant given the increasing importance of assessing online health information for accuracy and reliability in guiding

public health decisions. Overall, this passage lacks detail on how RAT could be implemented in practice. For example, it would be helpful to explain how RAT could be adapted for studies in fields like health and media, including any specific modifications or applications that would make it suitable for these areas of research.

    - The text mentions ideological bias in search results, but it does not directly tie this to the application of RAT. This point feels a bit disconnected. Could you please explain how RAT could help identify or mitigate these biases?

    - Could you please elaborate on why access to larger datasets is so difficult or problematic? Are there any legal or technical barriers?

    - Additionally, the concept of "self-interest" could be made more explicit and more directly tied to the potential consequences for research transparency. While the term "self-interest" is mentioned, the relationship between the financial motivations of search engine companies and the withholding of data could be explained more clearly. Could you also provide evidence showing how the financial interests of search engine providers might influence the ranking or visibility of search results? For example, this could include the prioritization of paid advertisements or content from partners.

    - Finally, there is no explicit discussion of how RAT could directly improve or be integrated into existing studies. For instance, you could provide a comparison with traditional methods to highlight how RAT offers potential improvements.

3. *" Even though software cannot remove this barrier, RAT makes the process of assessing results efficient by removing duplicates from the results, providing a user interface for study participants (…)".* Could you please provide more information regarding the study participants? While the groups of researchers and participants are discussed later in the text (in the 'User Journey in RAT' section), it would be helpful to clarify their roles and how they interact with RAT earlier in the document.

4. The following statement is not clear*: "We also resolved an issue where the scraper returned fewer results than specified—e.g., scraping only 24 results when the limit was set to 30."* Could you please provide more details on the issue? Specifically, how can scraping fewer results than expected represent a problem? For example, if fewer results are

scraped than the specified limit, how might this impact the quality or accuracy of the data?

5. Please avoid the use of contracted forms (for example, "does not" instead of "doesn't", 'it is' instead of 'it's')

Could you explain here what the focus of the software is? *"We have limited ourselves to studies using data from commercial search engines such as Google or Microsoft Bing and library search systems, as this is the focus of the software."*