Qeios

Peer Review

Review of: "PANDA – Paired Anti-hate Narratives Dataset from Asia: Using an LLM-as-a-Judge to Create the First Chinese Counterspeech Dataset"

Samsul Arifin[1]

1. Binus University, Indonesia

- **Title**: The title is clear and informative, accurately reflecting the research focus. However, it is quite long and could be streamlined for better readability, such as omitting redundant phrases or focusing on the core novelty of the study.
- **Abstract**: The abstract effectively summarizes the research but lacks a concise statement of key findings and contributions. Consider emphasizing the dataset's uniqueness, its impact, and potential applications in a clearer, more structured manner.
- **Introduction**: The introduction provides strong motivation for the study but could better articulate the research gap. While it highlights the lack of Chinese counterspeech datasets, it should explicitly contrast existing approaches and clearly define the novelty of this work.
- **Related Works**: The literature review is well-researched but could benefit from a more structured comparison of existing datasets and approaches. A table summarizing dataset characteristics and their limitations compared to the proposed work would enhance clarity.
- **Methodology**: The methodology is detailed and technically sound. However, the description of simulated annealing and ranking mechanisms is somewhat dense. Including a flowchart or diagram would improve readability and help readers grasp the process more efficiently.
- **Results**: The results section presents valuable insights but lacks statistical significance tests or comparative evaluations against baseline methods. Adding these would strengthen the validity of the findings. Additionally, a discussion on the potential biases of the LLM-as-a-Judge approach should be expanded.

doi.org/10.32388/1I1MD2

- **References**: The references are relevant and up-to-date, though a few key works on counterspeech evaluation metrics and multilingual NLP models could be included to provide additional context. Formatting should be checked for consistency.
- **Readability**: The paper is well-written but includes some complex, technical explanations that could be broken down for better accessibility. Some long sentences and dense paragraphs could be restructured for easier comprehension.
- **Open Problems**: The paper highlights challenges in dataset labeling and model biases but could further discuss broader implications, such as ethical considerations and the applicability of this approach to other non-Western languages. Suggesting concrete future research directions would strengthen this section.

Declarations

Potential competing interests: No potential competing interests to declare.