Review of: "Empowering Dysarthric Speech: Leveraging Advanced LLMs for Accurate Speech Correction and Multimodal Emotion Analysis"

Ammar Amjad[1]

1 National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Potential competing interests: No potential competing interests to declare.

Detailed Review of the Article

The paper presents an interesting approach using advanced LLMs (GPT-4, LLaMA) for dysarthric speech correction and emotion recognition. However, it lacks clarity in key areas, such as the validation of pseudo-labels and teacher-student model interaction. The methodology also lacks sufficient real-world validation, especially in noisy or diverse speech conditions.

While the concept is promising, the novelty is limited since similar techniques have been explored in speech recognition. The methodology lacks rigor, and the paper does not address potential real-world challenges, like non-native accents or noisy environments.

The examples provided are insufficient, relying heavily on synthetic datasets. More real-world data is needed to validate the method. The paper also struggles with clarity, particularly in explaining technical details and the mutual learning process. Simplifying these sections would improve comprehension.

The abstract is broad and doesn't clearly explain the contributions, while the introduction could further explore the limitations of existing methods. The paper's structure could be more concise, and redundant information should be trimmed for better focus.

Figures and captions need more detail, and better explanations of the diagrams are necessary for clarity. Additionally, while the keywords are relevant, more specific terms would enhance the paper's discoverability.

Overall, the paper requires significant revisions to improve clarity, methodology, and validation before being considered for publication.