

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

Fredrik Nylén¹

¹ Umea University, Sweden

Potential competing interests: No potential competing interests to declare.

The manuscript describes an effort to assess emotion in the speech of persons with dysarthria using large language models (LLMs) and automatic speech recognition (ASR) tools. While I applaud the authors' effort to contribute to our understanding of the effects of speech impairment on the ability to convey appropriate emotional content, the work in its present form has serious flaws:

1. The authors demonstrate a profound lack of understanding of what dysarthria is, that there are many forms of dysarthria, and the major domains of speech in which emotion in speech is conveyed.
2. The source material is constructed, and the study, therefore, has no ensured level of validity.
3. Human annotations were performed but not described in detail. Most importantly, the (inter-rater and intra-rater) reliability of the human annotation was not assessed.

Using a series of LLMs (and one ASR), the authors demonstrate that they can predict the prediction in synthesised samples, but no information other than the synthesised sentences was provided to the analysis toolchain. In the end, I fail to see how any level of validity and reliability (or indeed causality) has been ensured by the authors.