Accessible Voice Control: Fine-Tuning OpenAI's Whisper ASR Model for Improved Speech Recognition in Individuals With Intellectual Disabilities

Cavalluzzi, Ian (School: Riverside High School)

This study investigates the application of OpenAI's Whisper Automatic Speech Recognition (ASR) model, fine-tuned to improve speech recognition accuracy for individuals with moderate to mild intellectual disabilities. 81.8% of children with mild intellectual disability (ID) have motor control issues that can prevent them from typing on a computer. Typically, voice control would be the replacement for typing, but 49.4% of people with a mild intellectual disability and 93.9% of people with a moderate intellectual disability have disordered speech, making standard voice-control tools ineffective. There are ASR models focused on disordered speech, but in large datasets audio is categorized by disorder type and not intellectual disability level. The resulting models are not tailored to specific groups of people with intellectual disabilities. The models are also too large to be ran locally on computers, leading to better Word Error Rate (WER) outcomes but less real-world applications. By fine-tuning Whisper-Small through parameter-efficient fine-tuning (PEFT), the research significantly reduced WER, enabling more intelligible transcriptions that can be ran locally on consumer-level hardware. Fine-tuning methods included the use of anonymized, high-quality audio recordings from participants, processed into a private dataset to train the model effectively while maintaining privacy. The fine-tuned model achieved a WER of 27.3%, outperforming Microsoft's built-in ASR system and the untrained Whisper-Small model, which had WERs of 50.6% and 46.4%, respectively. This enhanced transcription accuracy facilitates voice control for independent computer use. This also enables future research with a larger crowd-sourced dataset for improved WER.

Awards Won: