VoiceAuth: Differentiating AI-Generated Voices from Real Human Speech Using Machine Learning

Khan, Zohaib (School: Beaconhouse School System Valencia Town Campus)
Kashif, Umer (School: Beaconhouse School System Valencia Town Campus)

In recent years, advancement in artificial intelligence and speech synthesis technologies, have made it possible to clone human voices with remarkable accuracy. While this opens many beneficial applications (e.g virtual assistants), it also poses significant security threats. This project was motivated by a personal incident, where a relative's voice was cloned for ransom scam, as well as similar cases of people getting calls from their alleged 'relatives' who are then blackmailed into giving them money. The research aimed to develop and validate a reliable system for distinguishing between AI-generated and human voices. A labeled dataset of humans and AI-generated voice recordings was collected (using publicly available datasets). In addition, key acoustic features were extracted such as Mel-Frequency Cepstral Coefficients, spectral properties, and amplitude metrics. Moreover, a logistic regression classification model was then trained on these features to accurately predict the nature of a voice. The research controlled for consistency in audio format and duration, while also employing standard data preprocessing steps. Model evaluation showed the trained classifier successfully separated AI-generated from real human voices with a high degree of accuracy. Precision and recall metrics both indicated that the classifier was able to effectively reduce false positives and false negatives, which highlight that certain spectral and cepstral features appear consistently different between synthetic and genuine speech. The investigation demonstrates that targeted acoustic feature analysis can serve as a line of defense against emerging voice-cloning threats.

Awards Won: