Unfake: A Solution for Detecting Audio Deepfakes Using Artificial Intelligence

Godinho Filho, Marcos (School: Colegio Tecnico de Campinas)
Carvalho Figueira, Eric (School: Colegio Tecnico de Campinas)

The growing evolution of artificial intelligence tools has culminated in the development of potentially harmful technologies, such as audio deepfakes. These artificially created media can clone an individual's identity, using it to damage their reputation, spread fake news, and promote fraud and security breaches. Solutions are therefore needed to detect this type of content. However, most existing technologies are focused on the English language and noise-free audio, hindering detection in media that does not follow these standards. The goal of this project was therefore to develop an artificial intelligence capable of identifying whether an audio contains human or synthetic speech. To this end, a pioneering dataset of Portuguese audio deepfakes was initially developed, based on a set of recorded speeches and deepfakes created from them using a Text-to-Speech model. The audio files were then converted into spectrograms, visual representations containing information about their time, frequency and amplitude. Furthermore, a CNN model to classify the audio spectrograms as real or fake was structured, trained and validated using Python libraries such as Tensorflow. A website and an API have been developed using Flask and Next.js frameworks in order to provide access to the classifier, allowing users to send audio and detect whether the content is a deepfake. Results showed this method can be applied to identify audio deepfakes, while more data would make it precise and robust. Finally, this solution may contribute to a safer and more reliable digital environment, as well as encouraging future research in the area using our dataset.

Awards Won: