Task 2

Read up on any innovative technology using NLP (by companies such as Google or IBM for instance) and write a brief summary about the technology, what it achieves/does and an overview of how it works (250-500 words).

In 2022, research lab OpenAI (in partnership with Microsoft) released a speech recognition model called Whisper. They "trained and are open-sourcing a neural net called Whisper that approaches human level robustness and accuracy on English speech recognition."[1]

Whisper is a speech-to-text system that takes a spoken language audio input and produces a transcript of the content. It has many potential applications, for example: speaking is generally much quicker than writing or typing so ideas can be conveyed more efficiently in particular for dictation where documentation is key such as in the Legal and Medical professions, it can be used for translation to help people without a common language communicate or also in accessibility for the visually impaired or those who cannot type,

It works by taking audio inputs and converting them into a spectrogram. Then it makes predictions about which words most closely match the spectrographic representation and which words it expects to come next in the sentence producing the most probable output words. The process includes the possibility of outputting time-aligned transcriptions as well as a text-only version.

Thanks to the use of unsupervised learning techniques, OpenAI was able to build a training set of "680,000 hours of multilingual and multitask"[2] audio data which did not require any human labeling. This methodology carries the dual advantage of Whisper being better than other models at recognising "audio from many different environments, recording setups, speakers, and languages."[3] While Whisper didn't outperform systems which were purpose built to meet popular benchmarks such as LibriSpeech they do note that "when we measure Whisper's zero-shot performance across many diverse datasets we find it is much more robust and makes 50% fewer errors than those models."[4]

References

1. OpenAI (2022). "Introducing Whisper"

2. Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever (2022). "Robust Speech Recognition via Large-Scale Weak Supervision" p. 1

3. Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever (2022). "Robust Speech Recognition via Large-Scale Weak Supervision" p. 2

4. OpenAI (2022). "Introducing Whisper"