**Plan prévisionnel**

## **Dataset retenu**

*Pour évaluer nos modèles speech to text, nous avons besoin d’un dataset contenant des extraits audio et les transcriptions correspondantes. J’ai choisi le Multilingual LibriSpeech (MLS), téléchargeable à l’adresse :* [*https://openslr.org/94*](https://openslr.org/94)

**Identifier:** SLR94

**Summary:** A large multilingual corpus derived from LibriVox audiobooks

**Category:** Speech

**License:** CC BY 4.0

*Ce dataset contient des jeux de training, validation et test, pour plusieurs langues. Nos deux modèles sont préentrainés, nous allons donc utiliser une partie des fichiers test, en français.*

## **Modèle envisagé**

**Whisper, an automatic speech recognition (ASR) system. The speech-to-text model is open-sourced. A révolutionné le domaine (voir refs)**

**L’extrait suivant provient d’un article publié sur kdnuggets le 2 juin 2023 :**

*« Whisper is a model based on neural networks developed by OpenAI to solve speech-to-text tasks. It belongs to the GPT-3 family and has become very popular for its ability to transcribe audio into text with very high accuracy.*

It doesn’t limit handling English, but its ability is extended to more than 50 languages. Furthermore, it can translate any language audio into English.

Like other OpenAI products, there is an API to get access to these speech recognition services, allowing developers and data scientists to integrate Whisper into their platforms and apps.

[…]

I found the outputs obtained after the transcription and the translation very impressive! This AI tool is surely helping a lot of people right now. The only limit is the fact that it’s only possible to translate to English text and not vice versa, but I am sure that OpenAI will provide it soon. »

*En effet, quelques mois plus tard Whisper est désormais disponible via l’API ou en local, capable de transcrire 99 langages, et de traduire non seulement vers l’anglais, mais aussi vers toutes les 98 autres langues.*

## **Références bibliographiques**

*- OpenAI’s Whisper API for Transcription and Translation, by* [***Eugenia Anello***](https://www.kdnuggets.com/author/eugenia-anello)***,***

***2 juin 2023***

***Présentation du modèle.***

***(Source :*** [***https://www.kdnuggets.com/2023/06/openai-whisper-api-transcription-translation.html***](https://www.kdnuggets.com/2023/06/openai-whisper-api-transcription-translation.html) ***)***

***- Houston we have a Divergence: A Subgroup Performance Analysis of ASR Models,***

***by Alkis Koudounas, Flavio Giobergia, 31 mars 2024***

***La Nasa utilise Whisper !***

***(source :*** [***https://arxiv.org/abs/2404.07226***](https://arxiv.org/abs/2404.07226) ***)***

***- WavLLM: Towards Robust and Adaptive Speech Large Language Model, by Shujie Hu, 31 mars 2024***

***L’université chinoise Hong Kong utilise le bloc encoder de whisper pour intégrer le speech to text à leur projet de LLM.***

***(source :*** [***https://arxiv.org/abs/2404.00656***](https://arxiv.org/abs/2404.00656) ***)***

***+ de nombreux articles récents / applications passionnantes, par exemple ce mois-ci : Cross-Domain Audio Deepfake Detection, by Yuang Li , le 7 avril 2024***

***ou encore : Transfer Learning from Whisper for Microscopic Intelligibility Prediction***

***, by Santiago Cuervo, le 2 avril.***

## **Explication de votre démarche de test du nouvel algorithme (votre preuve de concept)**

*Nous allons évaluer la performance des 2 modèles sur le même dataset de test, en mesurant le WER (le word error rate entre la prédiction du modèle et le « true » transcript) comme métrique spécifique.*