

# Whisper : retranscription et traduction

Programme python



Whisper est un modèle de reconnaissance de la parole très performant, car il a été formé sur un grand jeu de données audio diversifié. Il est capable de reconnaître plusieurs langues, ce qui le rend très utile pour les applications qui nécessitent une reconnaissance de la parole multilingue. De plus, Whisper est également un modèle multitâche, il peut donc être utilisé pour effectuer non seulement la reconnaissance de la parole, mais aussi la traduction de la parole et l'identification de la langue. Cela en fait un outil très polyvalent pour les applications de traitement de la parole.

En général, Whisper est un modèle avancé qui permet de transcrire et traduire plusieurs langues et identifier la langue d'une parole en entrée, cela peut être utilisé dans de nombreux domaines tels que la traduction automatique, la reconnaissance de commandes vocales, les assistants virtuels, les assistants de traduction pour les personnes sourdes et malentendantes, etc.

—

Pour l'installer sous Fedora Linux, voici les commandes :

```
pip install -U openai-whisper
sudo dnf install ffmpeg
```

```
> pip install -U openai-whisper
Defaulting to user installation because normal site-packages is not writeable
Collecting openai-whisper
  Downloading openai-whisper-20230117.tar.gz (1.2 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.2/1.2 MB 709.6 kB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting numpy
  Downloading numpy-1.24.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.3 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 17.3/17.3 MB 706.9 kB/s eta 0:00:00
Collecting torch
  Downloading torch-1.13.1-cp311-cp311-manylinux1_x86_64.whl (887.4 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 887.4/887.4 MB 222.4 kB/s eta 0:00:00
Collecting tqdm
  Downloading tqdm-4.64.1-py2.py3-none-any.whl (78 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 78.5/78.5 kB 404.9 kB/s eta 0:00:00
Collecting more-itertools
  Downloading more_itertools-9.0.0-py3-none-any.whl (52 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 52.8/52.8 kB 471.9 kB/s eta 0:00:00
Collecting transformers>=4.19.0
  Downloading transformers-4.25.1-py3-none-any.whl (5.8 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.8/5.8 MB 294.6 kB/s eta 0:00:00
Collecting ffmpeg-python==0.2.0
  Downloading ffmpeg_python-0.2.0-py3-none-any.whl (25 kB)
Collecting future
  Downloading future-0.18.3.tar.gz (840 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 840.9/840.9 kB 693.2 kB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting filelock
  Downloading filelock-3.9.0-py3-none-any.whl (9.7 kB)
Collecting huggingface-hub<1.0,>=0.10.0
  Downloading huggingface_hub-0.11.1-py3-none-any.whl (182 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 182.4/182.4 kB 492.9 kB/s eta 0:00:00
Requirement already satisfied: packaging>=20.0 in /usr/lib/python3.11/site-packages (from transformers>=4.19.0->openai-whisper) (21.3)
Requirement already satisfied: pyyaml>=5.1 in /usr/lib64/python3.11/site-packages (from transformers>=4.19.0->openai-whisper) (6.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/lib64/python3.11/site-packages (from transformers>=4.19.0->openai-whisper) (2022.10.31)
Requirement already satisfied: requests in /usr/lib/python3.11/site-packages (from transformers>=4.19.0->openai-whisper) (2.28.1)
Collecting tokenizers!=0.11.3,<0.14,>=0.11.1
  Downloading tokenizers-0.13.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (7.6 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.6/7.6 MB 647.3 kB/s eta 0:00:00
```

Pour utiliser la fonction de transcription auditive en texte, utilisant un modèle *medium* :

```
whisper audio.flac audio.mp3 audio.wav --model medium
```

Il existe cinq tailles de modèles, dont quatre avec des versions anglaises uniquement, offrant des compromis entre vitesse et précision. Ci-dessous se trouvent les noms des modèles disponibles et leurs exigences de mémoire approximatives et leur vitesse relative.

Taille	Paramètres	Modèle en anglais seulement	Modèle multi-langues	VRAM nécessaire	Vitesse relative
tiny	39 M	tiny.en	tiny	~1 GB	~32x
base	74 M	base.en	base	~1 GB	~16x
small	244 M	small.en	small	~2 GB	~6x
medium	769 M	medium.en	medium	~5 GB	~2x
large	1550 M	N/A	large	~10 GB	1x

Par défaut, le modèle utilisé est *small*, en langue anglaise. Pour transcrire un fichier audio contenant autre chose que de l'anglais, il faut spécifier la langue avec l'option `--language` :

```
whisper french.wav --language French
```

Le programme Whisper générera 3 fichiers :

- Fichier de sous-titre SRT
- Fichier de sous-titre VTT

