

Whisper

- Créé par OpenAI, en open source : <https://github.com/openai/whisper>.
- Intelligence artificielle entraînée sur plusieurs tâches langagières.
- Basée sur Python.

Multitask training data (680k hours)

English transcription

- 🗣️ "Ask not what your country can do for ..."
- 📝 Ask not what your country can do for ...

Any-to-English speech translation

- 🗣️ "El rápido zorro marrón salta sobre ..."
- 📝 The quick brown fox jumps over ...

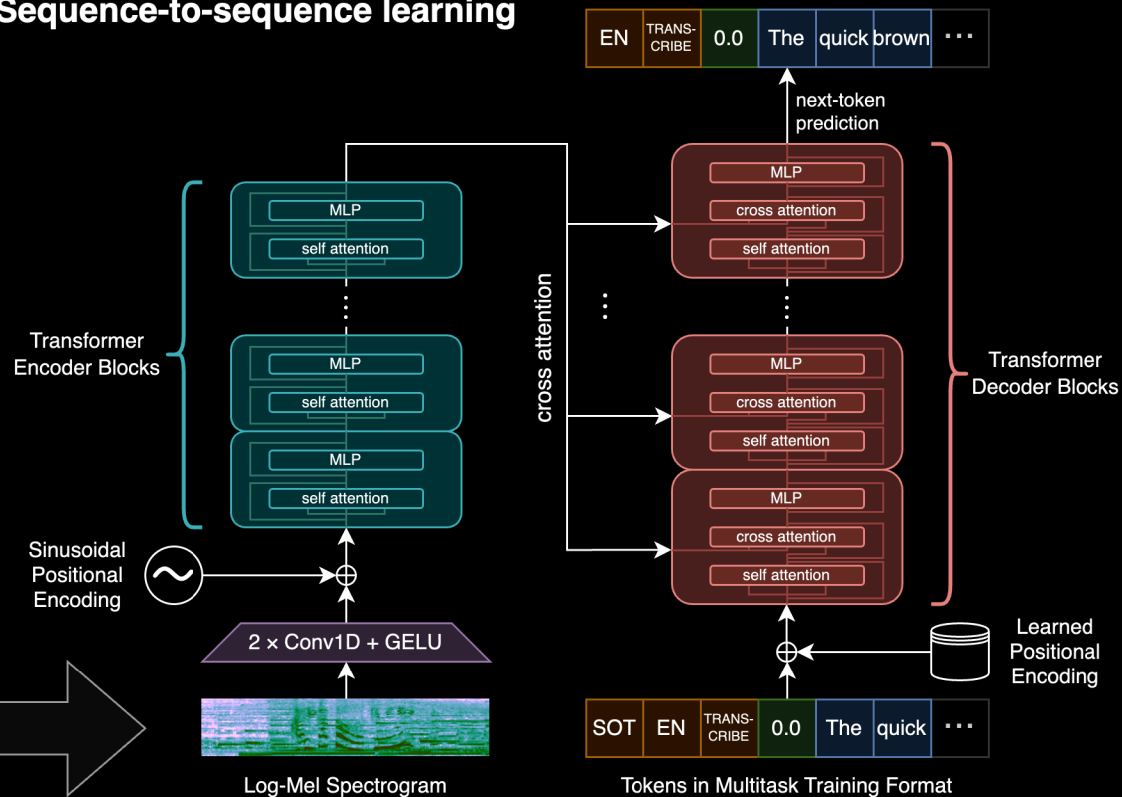
Non-English transcription

- 🗣️ "언덕 위에 올라 내려다보면 너무나 넓고 넓은 ..."
- 📝 언덕 위에 올라 내려다보면 너무나 넓고 넓은 ...

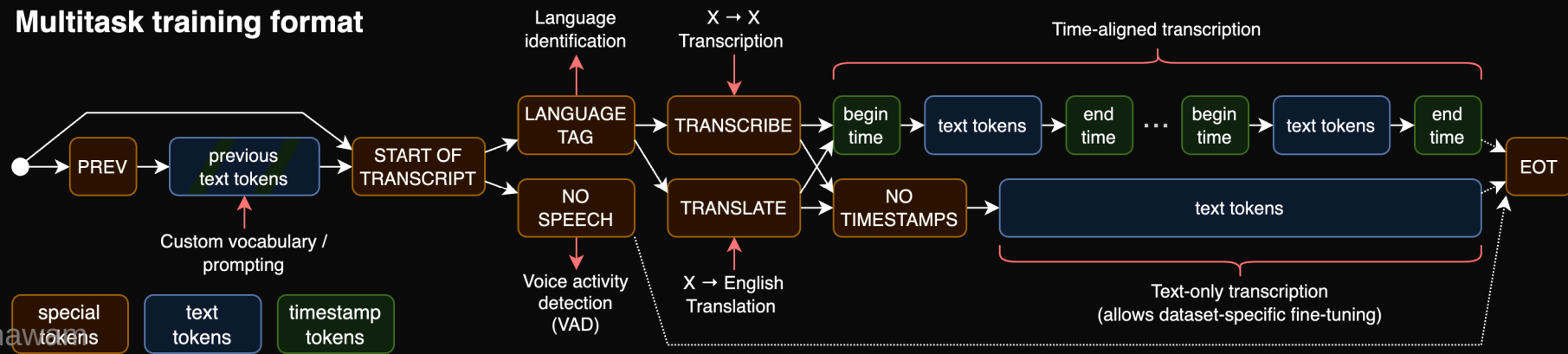
No speech

- 🎧 (background music playing)
- 📝 ∅

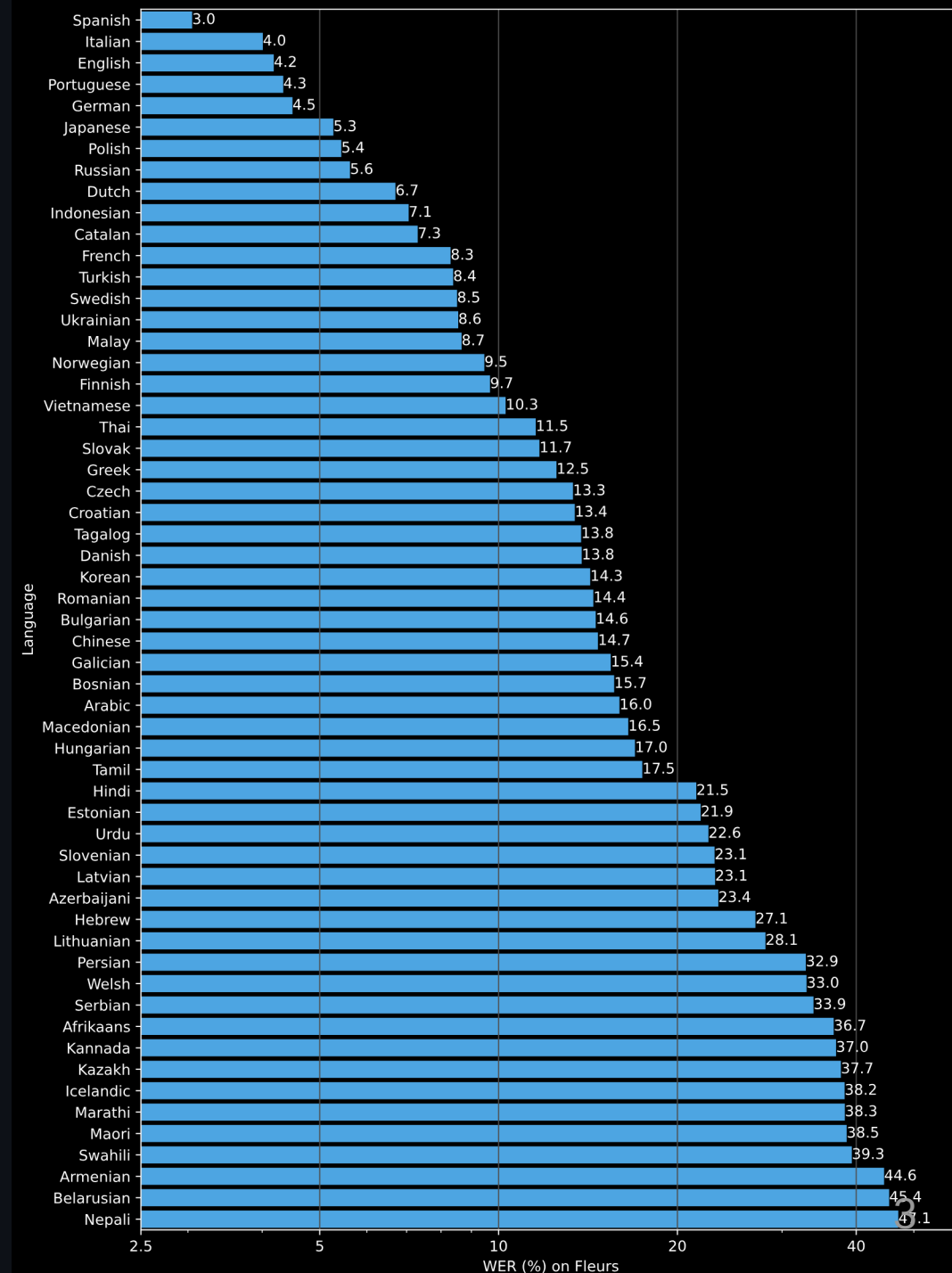
Sequence-to-sequence learning



Multitask training format



"Whisper's performance varies widely depending on the language. The figure shows a WER (Word Error Rate) breakdown by languages of the Fleurs dataset using the large-v2 model (The smaller the numbers, the better the performance)."



Google Colab

- Pour ne pas trop faire carburer l'ordinateur et plus rapide.
- Instructions détaillées ici : <https://bytexd.com/how-to-use-whisper-a-free-speech-to-text-ai-tool-by-openai>.
- Suivre : <https://colab.research.google.com/#create=true>.
- Autoriser GPU (Graphics Processing Unit) : Runtime > Change runtime (ou Exécution > Modifier le type d'exécution) et dans le menu déroulant Hardware Accelerator (ou Accélérateur matériel) choisir GPU puis enregistrer.

Utiliser whisper

- Installer whisper : `!pip install git+https://github.com/openai/whisper.git` . Et ffmpeg : `!sudo apt update && sudo apt install ffmpeg` . Puis `ctrl+enter` .
- Taper `!whisper fichier.format` . Ex. : `!whisper test.mp3` .
- Si espace dans le nom du fichier, le mettre entre guillemets. Ex. : `!whisper "mon entretien préféré.mp3"` .
- Pour la liste des commandes : `!whisper --help` .

Utiliser whisper

- Détection automatique de la langue d'origine mais possibilité d'indication au préalable. Ex. : `--language Spanish`. NB : toutes les commandes vont après le nom du fichier. Ex. : `!whisper test.mp3 --language Spanish`.
- Pour traduire, en anglais seulement : `--task translate`.
- Génère différents formats à la fin du processus, dont `.txt` et `.srt`.
- Difficulté dans la reconnaissance du point final et certains mots techniques et nom propres.

Modèles

- Arbitrage précision/temps de retranscription. `small` par défaut. Commande pour changer : `--model medium`.

Size	Parameters	English-only model	Multilingual model	Required VRAM	Relative speed
tiny	39 M	<code>tiny.en</code>	<code>tiny</code>	~1 GB	~32x
base	74 M	<code>base.en</code>	<code>base</code>	~1 GB	~16x
small	244 M	<code>small.en</code>	<code>small</code>	~2 GB	~6x
medium	769 M	<code>medium.en</code>	<code>medium</code>	~5 GB	~2x
large	1550 M	N/A	<code>large</code>	~10 GB	1x

Local

- Si souci de confidentialité, possibilité de le faire fonctionner en local.
- Besoin d'un terminal et d'un gestionnaire de paquets. Pour installer un gestionnaire :
 - MacOS : <https://brew.sh>
 - Windows : <https://scoop.sh> ou <https://chocolatey.org/>
 - Linux : déjà installé.

Local

- Installation de Python et de ses dépendances : `pip install git+https://github.com/openai/whisper.git` . Puis de ffmpeg:

```
# sur Ubuntu
sudo apt update && sudo apt install ffmpeg
# sur MacOS avec Homebrew (https://brew.sh/)
brew install ffmpeg
# sur Windows avec Scoop (https://scoop.sh/)
scoop install ffmpeg
```

- Commande sans point d'exclamation. Ex. : `whisper test.mp3 --language Spanish --model medium --task translate`