**Training report**

Disclaimer: I am using the base model rather than the large model due to the lack of compute resources on my local machine.

The base original model obtained a Word Error Rate (WER) of 0.14385, while the base fine-tuned version obtained a WER of 0.10821, which is considerably better. Of course, using the large model will yield much better results.

We utilized most of the previous configurations: processor, tokenizer, and architecture. We simply fine-tuned the model on the dataset over 1 epoch to avoid overfitting. However, we could have tested for:

- More epochs
- Various learning rates
- Warm up steps
- floating-point precision
- Various models

Additionally, to improve the model, we can fine-tune the model with more data:

- LibriSpeech
- Common Voice (current dataset)
- TED-LIUM
- GIgaSpeech
- Lirbi-Light
- Multilingual LibriSpeech
- Switchboard
- TIMIT
- LJ Speech
- FLERUS
- CHILDES
- Buckeye Corpus