**NER and STT models for Uzbek Language**

# RoBERTa-NER for Uzbek Language

This section covers the fine-tuning process, datasets, preprocessing steps, evaluation metrics, and results for the RoBERTa-based Named Entity Recognition (NER) model for the Uzbek language.

**Link for colab:** [ner\_roberta\_uz.ipynb](https://colab.research.google.com/drive/1kqHf417RMo5He1AdshL-V4GyKuYyhFiQ?usp=sharing)

## Fine-Tuning Process for RoBERTa-NER

The RoBERTa-NER model for the Uzbek language is fine-tuned from the pre-trained `FacebookAI/xlm-roberta-large` model. The fine-tuning is designed to make the model more specific to the task of Named Entity Recognition (NER) for the Uzbek language.

*Key Aspects of the Fine-Tuning Process for RoBERTa:*  
1. **Model Selection**: `**FacebookAI/xlm-roberta-large**` was chosen for its robust multilingual capabilities.  
2. **Dataset**: The model was fine-tuned using the `**risqaliyevds/uzbek\_ner**` dataset, specifically for NER.  
3. **Training Configuration**: Fine-tuning was performed for **6** epochs, with batch sizes of **32** for training and **16** for evaluation. The learning rate was set to **2e-5**, using a **cosine** scheduler for the learning rate and enabling **FP16** for faster processing.  
4. **Preprocessing**: Text data was changed to **BIO** labeling format, tokenized and padded to ensure uniformity.

## Datasets and Preprocessing for RoBERTa

The fine-tuning was performed on the `**risqaliyevds/uzbek\_ner**` dataset, which contains labeled data specifically for Uzbek NER tasks. This dataset includes a variety of named entities such as **persons, locations, dates, events and organizations** in the Uzbek language.

*Preprocessing Steps for RoBERTa:*

1. **Labeling**: Each token was assigned an NER label, following the **BIO** format.
2. **Tokenization**: Tokenization was performed using the **XLM-RoBERTa tokenizer**.
3. **Padding**: Text sequences were padded to ensure consistent input lengths.

## Evaluation Metrics and Results for RoBERTa

The evaluation of the RoBERTa-NER model was performed using the following overall metrics:  
**- Precision: 0.6217  
- Recall: 0.6262  
- F1 Score: 0.6239** - **Loss: 0.1413**

### Training hyperparameters

The following hyperparameters were used during training:

* learning\_rate: 2e-05
* train\_batch\_size: 32
* eval\_batch\_size: 32
* seed: 42
* gradient\_accumulation\_steps: 2
* total\_train\_batch\_size: 64
* optimizer: Use adamw\_torch with betas=(0.9,0.999) and epsilon=1e-08 and optimizer\_args=No additional optimizer arguments
* lr\_scheduler\_type: cosine
* lr\_scheduler\_warmup\_ratio: 0.1
* num\_epochs: 3
* mixed\_precision\_training: Native AMP

### 

### 

### 

### 

### Training results

| Training Loss | Epoch | Step | Validation Loss | Precision | Recall | F1 | Accuracy |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0.1671 | 0.5758 | 150 | 0.1632 | 0.5260 | 0.6425 | 0.5785 | 0.9402 |
| 0.1453 | 1.1497 | 300 | 0.1481 | 0.5935 | 0.6191 | 0.6061 | 0.9467 |
| 0.134 | 1.7255 | 450 | 0.1449 | 0.5936 | 0.6216 | 0.6073 | 0.9480 |
| 0.1273 | 2.2994 | 600 | 0.1413 | 0.6217 | 0.6262 | 0.6239 | 0.9493 |
| 0.1258 | 2.8752 | 750 | 0.1421 | 0.6071 | 0.6482 | 0.6270 | 0.9486 |

**XLM-ROBERTA-LARGE** vs **BERT-MULTILINGUAL-CASED** performance

You can see overfitting after 2-epoch from the loss graph for both models. And **RoBERTa** model’s overall F1 score is better than the **BERT** model’s performance.

# Whisper Small for Uzbek Language

This section describes the fine-tuning of the Whisper Small model on the Common Voice 17.0 dataset for Uzbek speech-to-text tasks.

**Link for colab**: [fine\_tune\_whisper.ipynb](https://colab.research.google.com/drive/1ElpM2JNW5m6XByMgESDgnE_zQXZtk39O?usp=sharing)

## Fine-Tuning Process for Whisper Small

The Whisper Small model was fine-tuned from the pre-trained `**openai/whisper-small**` model. Fine-tuning was performed using the **Common Voice 17.0** dataset, tailored for Uzbek speech recognition tasks.

*Key Aspects of the Fine-Tuning Process for Whisper Small:*1. Model Selection: `**openai/whisper-small**` was chosen as the base model due to its capability to handle diverse speech recognition tasks.  
2. Dataset: The model was fine-tuned using the **Common Voice 17.0** dataset’s 40K row data about **67 hours**, which includes labeled Uzbek speech data.  
3. Training Configuration: Fine-tuning was carried out with a batch size of **16** for training and **16** for evaluation, and a learning rate of **2e-5**.  
4. Preprocessing: The audio data was preprocessed and converted into **spectrograms** for input to the model.

## Datasets and Preprocessing for Whisper

The training dataset for Whisper was the Common Voice 17.0 dataset, which provides a large collection of spoken language data, including Uzbek. Preprocessing involved converting audio files into spectrograms to prepare the data for model training.

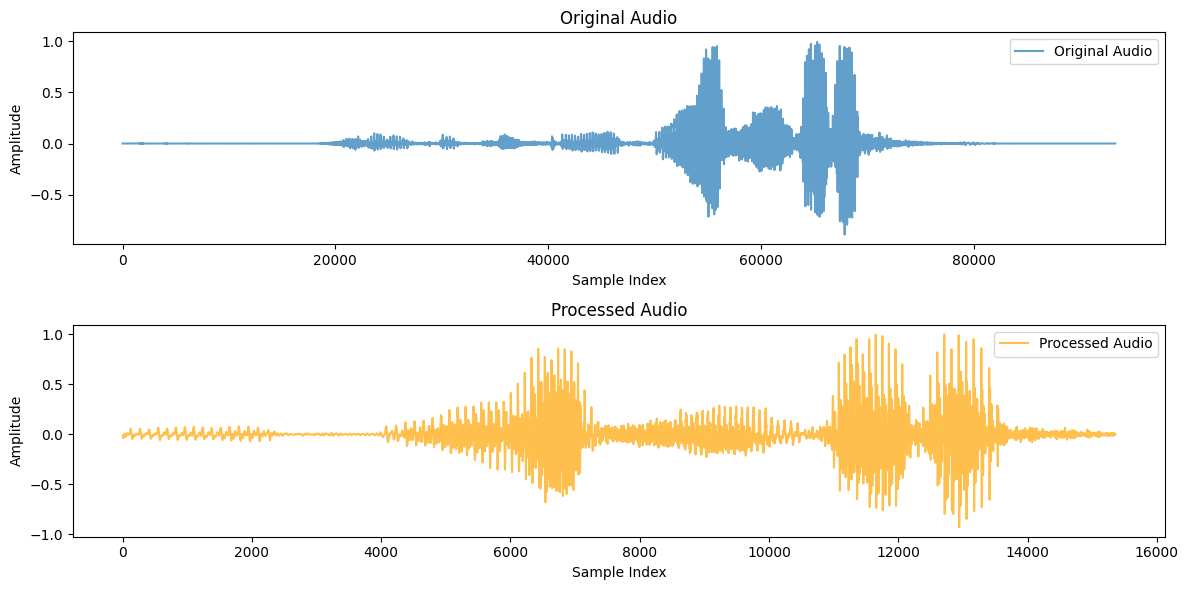
*Preprocessing Steps for Whisper:*1. **Audio Processing**: The audio files were converted into spectrograms.  
2. **Noise Reduction**: Techniques such as **noise reduction** were used to improve the robustness of the model.  
3. **Normalization**: Spectrograms were **normalized** to standardize the input features.

4. **Silence Trimming**: Silence audio detected and trimmed.

## Evaluation Metrics and Results for Whisper

The Whisper Small model was evaluated based on the following metrics:  
- **Loss**: **0.37**  
- **Word Error Rate** (WER): **35.86**%

The results indicate the model's performance on the Uzbek language speech-to-text task, with a WER of **35.8660**%, which is considered reasonable for speech recognition.

Before and after preprocessing:  


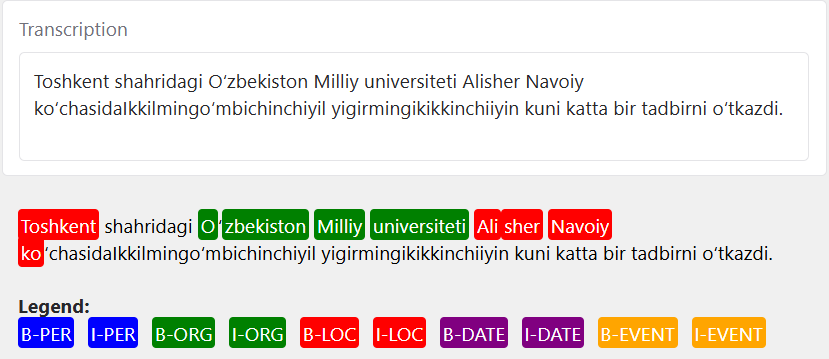
### Training hyperparameters

The following hyperparameters were used during training:

* learning\_rate: 2e-05
* train\_batch\_size: 16
* eval\_batch\_size: 16
* seed: 42
* optimizer: Use adamw\_torch with betas=(0.9,0.999) and epsilon=1e-08 and optimizer\_args=No additional optimizer arguments
* lr\_scheduler\_type: linear
* lr\_scheduler\_warmup\_steps: 1500
* training\_steps: 5500
* mixed\_precision\_training: Native AMP

### Training results

| **Training Loss** | **Epoch** | **Step** | **Validation Loss** | **Wer** |
| --- | --- | --- | --- | --- |
| 0.913 | 0.2 | 500 | 0.8213 | 62.5843 |
| 0.5734 | 0.6 | 1500 | 0.5458 | 48.0513 |
| 0.3092 | 1.4 | 3500 | 0.4184 | 40.1141 |
| 0.289 | 1.8 | 4500 | 0.3811 | 36.7950 |
| 0.1547 | 2.2 | 5500 | 0.3776 | 35.8660 |

**Examples:**

# 

# Conclusion

Both the RoBERTa-NER and Whisper Small models for the Uzbek language provide valuable tools for natural language processing and speech recognition tasks. Further improvements can be made by increasing the dataset diversity and optimizing the training processes.