

"cdli/kenyan_english_nonstandard_speech_v0.9" Large

Saving to: /jupyter_kernel/trained_models/en_nonstandard_tune_whisper_large_4/best_model

[200/200 1:53:35, Epoch 0/1]

Step	Training Loss	Validation Loss	Wer	Cer	Lattescore	Input Tokens Seen
0	No log	1.291891	0.190313	0.110266	66.081871	0
25	1.084300	1.163285	0.174630	0.104459	69.590643	19200000
50	1.564100	1.032564	0.167183	0.094805	76.315789	38400000
75	0.683400	0.829263	0.169513	0.095383	74.561404	57600000
100	0.780200	0.811014	0.165127	0.093572	77.485380	76800000
125	0.986400	0.795741	0.167143	0.095396	76.608187	96000000
150	0.696200	0.788244	0.166380	0.094734	76.900585	115200000
175	1.052500	0.786887	0.166095	0.094580	76.900585	134400000
200	0.857800	0.786678	0.166095	0.094580	76.900585	153600000

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1903

Adjusted CER: 0.1103

LATTEScore: 66.08%

Un-adjusted WER: 0.1948

Un-adjusted CER: 0.1103

=====

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1746

Adjusted CER: 0.1045

LATTEScore: 69.59%

Un-adjusted WER: 0.1825

Un-adjusted CER: 0.1063

=====

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1672

Adjusted CER: 0.0948

LATTEScore: 76.32%

Un-adjusted WER: 0.1716

Un-adjusted CER: 0.0948

=====

```
/usr/local/lib/python3.11/site-packages/transformers/modeling_utils.py:3464: UserWarning:
Moving the following attributes in the config to the generation config: {'max_length':
448, 'begin_suppress_tokens': [220, 50257]}. You are seeing this warning because you've
set generation parameters in the model config, as opposed to in the generation config.
  warnings.warn(
```

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1695

Adjusted CER: 0.0954

```

LATTEScore: 74.56%
Un-adjusted WER: 0.1740
Un-adjusted CER: 0.0954
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.1651
Adjusted CER: 0.0936
LATTEScore: 77.49%
Un-adjusted WER: 0.1718
Un-adjusted CER: 0.0954
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.1671
Adjusted CER: 0.0954
LATTEScore: 76.61%
Un-adjusted WER: 0.1808
Un-adjusted CER: 0.1009
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.1664
Adjusted CER: 0.0947
LATTEScore: 76.90%
Un-adjusted WER: 0.1801
Un-adjusted CER: 0.1002
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.1661
Adjusted CER: 0.0946
LATTEScore: 76.90%
Un-adjusted WER: 0.1799
Un-adjusted CER: 0.1001
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.1661
Adjusted CER: 0.0946
LATTEScore: 76.90%
Un-adjusted WER: 0.1799
Un-adjusted CER: 0.1001
=====

```

There were missing keys in the checkpoint model loaded: ['proj_out.weight'].

```

TrainOutput(global_step=200, training_loss=0.9404865527153015, metrics={'train_runtime':
7639.8177, 'train_samples_per_second': 0.052, 'train_steps_per_second': 0.026,

```

```
'total_flos': 1.358999322624e+18, 'train_loss': 0.9404865527153015, 'epoch':  
0.12779552715654952, 'num_input_tokens_seen': 153600000})
```

Post training evaluation

On DEV set

```
# (should give the same result shown in training progress on dev set)  
trainer.evaluate(dev_dataset, language=LANGUAGE)
```

[342/342 46:44]

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1651

Adjusted CER: 0.0936

LATTEScore: 77.49%

Un-adjusted WER: 0.1718

Un-adjusted CER: 0.0954

=====

```
{'eval_loss': 0.8110142946243286,  
 'eval_wer': 0.1651269165690267,  
 'eval_cer': 0.09357179082818226,  
 'eval_lattescore': 77.48538011695906,  
 'eval_runtime': 812.3512,  
 'eval_samples_per_second': 0.421,  
 'eval_steps_per_second': 0.421,  
 'epoch': 0.12779552715654952,  
 'num_input_tokens_seen': 153600000}
```

On TEST set

```
: # run on dev-set
# (should give the same result shown in training progress on dev set)
trainer.evaluate(test_dataset, language=LANGUAGE)

Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.1075
Adjusted CER: 0.0571
LATTEScore: 80.85%
Un-adjusted WER: 0.1076
Un-adjusted CER: 0.0571
=====
: {'eval_loss': 0.7319276928901672,
  'eval_wer': 0.107495999781176,
  'eval_cer': 0.05714562031408691,
  'eval_lattescore': 80.85106382978722,
  'eval_runtime': 1742.0687,
  'eval_samples_per_second': 0.405,
  'eval_steps_per_second': 0.405,
  'epoch': 0.12779552715654952,
  'num_input_tokens_seen': 153600000}
```

speaker_metadata.tsv: 100%

9.27k/9.27k [00:00<00:00, 788kB/s]

Generating train split:

52/0 [00:00<00:00, 1165.09 examples/s]

Speaker metadata loaded: (52, 6)

Columns: ['speaker_id', 'gender', 'age', 'severity_speech_impairment',
'type_nonstandard_speech', 'etiology']

	speaker_id	gender	age	\
0	KES001	Female	30-40	
1	KES002	Female	30-40	
2	KES003	Male	25-30	
3	KES004	Male	25-30	
4	KES005	Male	18-24	

		severity_speech_impairment	\
0		Severe (frequent breakdowns)	
1		Severe (frequent breakdowns)	
2	Profound (communication very difficult or impo...		
3		Severe (frequent breakdowns)	
4	Moderate (requires effort to understand)		

type_nonstandard_speech

etiology

```
0          Dysarthria          Cerebral Palsy
1          Dysarthria          Cerebral Palsy
2 Stuttering (Disfluency Disorders)          Cerebral Palsy
3 Stuttering (Disfluency Disorders) Neurological disorder
4 Stuttering (Disfluency Disorders) Neurological disorder
```

Generating predictions...

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1651

Adjusted CER: 0.0936

LATTEScore: 77.49%

Un-adjusted WER: 0.1718

Un-adjusted CER: 0.0954

=====

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.1075

Adjusted CER: 0.0571

LATTEScore: 80.85%

Un-adjusted WER: 0.1076

Un-adjusted CER: 0.0571

=====

Dev WER: 0.311 | Word Accuracy: 68.9% | LATTEScore: 54.7%

Test WER: 0.276 | Word Accuracy: 72.4% | LATTEScore: 60.1%

Dev DataFrame shape: (342, 11)

Columns: ['speaker_id', 'reference', 'prediction', 'wer', 'word_accuracy',
'lattescore_meaning_preserved', 'gender', 'age', 'severity_speech_impairment',
'type_nonstandard_speech', 'etiology']


Test DataFrame shape: (705, 11)

Columns: ['speaker_id', 'reference', 'prediction', 'wer', 'word_accuracy',
'lattescore_meaning_preserved', 'gender', 'age', 'severity_speech_impairment',
'type_nonstandard_speech', 'etiology']

=== Model Deployment Analysis ===

LATTEScore: 60.1%

Deployment Threshold: 80.0%

 RECOMMENDATION: Model does not meet quality standards

Consider: More training data, hyperparameter tuning, or different architecture

=== FILES SAVED ===

dev_predictions.csv: 342 samples

test_predictions.csv: 705 samples

=== DATA PREVIEW ===

speaker_id

reference \

```

0    KES004  it seems like some some fish or some seafood i...
1    KES004  maybe it is kinda some milk or some some some ...
2    KES004  it s a cake of course but am not a good fan of...
3    KES004  evening walks maybe around the city or in the ...
4    KES004  maasai culture of course you will find them do...
5    KES004  i can see a giraffe there i think it it it is ...
6    KES004  i can say thats a a burger there on the left h...
7    KES004  fancy and large it is a a large fish there som...
8    KES004  the maasai the kids i think they were in a gat...
9    KES004  whenever i go to the local market i make sure ...

```

		prediction	wer \
0	It seems like some some face of some seafood ...	0.400000	
1	Maybe it is kind of some meal or some ice som...	0.258065	
2	It's a cake of course but I'm not a big fan o...	0.312500	
3	evening walks maybe around the city or in the ...	0.090909	
4	Maasai culture of course you will find them d...	0.294118	
5	I can see a giraffe there. I think it it it i...	0.321429	
6	I can see that there is a burger there on the...	0.206897	
7	fancy and large it was a large fish there som...	0.125000	
8	The Maasai, the kids, I think they were in a ...	0.461538	
9	Whenever I go to the local market I make sure...	0.170732	

	lattescore_meaning_preserved
0	0
1	1
2	0
3	1
4	1
5	0
6	1
7	1
8	0
9	1

=== DOWNLOAD LINKS ===

[dev_predictions.csv](#)

[test_predictions.csv](#)

=== NEXT STEPS ===

1. Analyze LATTEScore by speaker metadata (etiology, severity, gender)
2. Compare LATTEScore with WER to see if meaning preservation differs from word accuracy
3. Use LATTEScore for model deployment decisions
4. Calculate LATTEScore breakdown by speaker characteristics