

"cdli/kenyan\_swahili\_nonstandard\_speech\_v0.9" Large

Saving to: /jupyter\_kernel/trained\_models/sw\_nonstandard\_tune\_whisper\_large\_4/best\_model

[200/200 1:52:29, Epoch 0/1]

Step	Training Loss	Validation Loss	Wer	Cer	Lattescore	Input Tokens Seen
0	No log	1.849143	0.722794	0.304201	1.838235	0
25	2.398700	1.506467	0.630979	0.221502	4.044118	19200000
50	1.587400	1.418386	0.621938	0.229396	5.147059	38400000
75	1.783200	1.336483	0.602169	0.221205	4.779412	57600000
100	1.034600	1.303896	0.585901	0.214827	6.250000	76800000
125	1.350300	1.281153	0.582280	0.221534	7.352941	96000000
150	1.097800	1.272910	0.583561	0.221364	7.352941	115200000
175	1.264900	1.270432	0.581967	0.221552	7.720588	134400000
200	1.182300	1.269624	0.582217	0.221454	7.352941	153600000

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.7228

Adjusted CER: 0.3042

LATTEScore: 1.84%

Un-adjusted WER: 0.7964

Un-adjusted CER: 0.3585

=====

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.6310

Adjusted CER: 0.2215

LATTEScore: 4.04%

Un-adjusted WER: 0.6455

Un-adjusted CER: 0.2227

=====

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.6219

Adjusted CER: 0.2294

LATTEScore: 5.15%

Un-adjusted WER: 0.7031

Un-adjusted CER: 0.2655

=====

```
/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: 0.6022

Adjusted CER: 0.2212

```
LATTEScore: 4.78%
Un-adjusted WER: 0.6323
Un-adjusted CER: 0.2412
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.5859
Adjusted CER: 0.2148
LATTEScore: 6.25%
Un-adjusted WER: 0.6644
Un-adjusted CER: 0.2618
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.5823
Adjusted CER: 0.2215
LATTEScore: 7.35%
Un-adjusted WER: 0.7025
Un-adjusted CER: 0.2856
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.5836
Adjusted CER: 0.2214
LATTEScore: 7.35%
Un-adjusted WER: 0.7041
Un-adjusted CER: 0.2854
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.5820
Adjusted CER: 0.2216
LATTEScore: 7.72%
Un-adjusted WER: 0.7031
Un-adjusted CER: 0.2856
=====
Sentence transformers not available, using WER-based LATTEScore approximation
=== Metrics ===
Adjusted WER: 0.5822
Adjusted CER: 0.2215
LATTEScore: 7.35%
Un-adjusted WER: 0.7033
Un-adjusted CER: 0.2855
=====
```

There were missing keys in the checkpoint model loaded: ['proj\_out.weight'].

```
TrainOutput(global_step=200, training_loss=1.4087182784080505, metrics={'train_runtime':
7612.5076, 'train_samples_per_second': 0.053, 'train_steps_per_second': 0.026,
```

```
'total_flos': 1.358999322624e+18, 'train_loss': 1.4087182784080505, 'epoch':  
0.1400560224089636, 'num_input_tokens_seen': 153600000})
```

Post Training evaluation

## On DEV set

```
: # (should give the same result shown in trainig progress on dev set)  
trainer.evaluate(dev_dataset, language=LANGUAGE)  
  
Sentence transformers not available, using WER-based LATTEScore approximation  
=== Metrics ===  
Adjusted WER: 0.5859  
Adjusted CER: 0.2148  
LATTEScore: 6.25%  
Un-adjusted WER: 0.6644  
Un-adjusted CER: 0.2618  
=====
```

```
: {'eval_loss': 1.3038960695266724,  
  'eval_wer': 0.5859010886650711,  
  'eval_cer': 0.21482694968073462,  
  'eval_lattescore': 6.25,  
  'eval_runtime': 804.9829,  
  'eval_samples_per_second': 0.338,  
  'eval_steps_per_second': 0.338,  
  'epoch': 0.1400560224089636,  
  'num_input_tokens_seen': 153600000}
```

## On TEST set

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

Sentence transformers not available, using WER-based LATTEscore approximation

=== Metrics ===

Adjusted WER: 0.4158

Adjusted CER: 0.1528

LATTEscore: 22.02%

Un-adjusted WER: 0.4594

Un-adjusted CER: 0.1701

=====

```
{'eval_loss': 1.0432450771331787,
 'eval_wer': 0.41582398430788187,
 'eval_cer': 0.1527811293851456,
 'eval_lattescore': 22.021660649819495,
 'eval_runtime': 1831.4445,
 'eval_samples_per_second': 0.302,
 'eval_steps_per_second': 0.302,
 'epoch': 0.1400560224089636,
 'num_input_tokens_seen': 153600000}
```

speaker\_metadata.tsv: 100%

9.27k/9.27k [00:00<00:00, 836kB/s]

Generating train split:

52/0 [00:00<00:00, 1238.47 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: 0.5859

Adjusted CER: 0.2148

LATTEScore: 6.25%

Un-adjusted WER: 0.6644

Un-adjusted CER: 0.2618

=====

Sentence transformers not available, using WER-based LATTEScore approximation

=== Metrics ===

Adjusted WER: 0.4158

Adjusted CER: 0.1528

LATTEScore: 22.02%

Un-adjusted WER: 0.4594

Un-adjusted CER: 0.1701

=====

Dev WER: 0.737 | Word Accuracy: 26.3% | LATTEScore: 1.8%

Test WER: 0.569 | Word Accuracy: 43.1% | LATTEScore: 11.6%

Dev DataFrame shape: (272, 11)

Columns: ['speaker\_id', 'reference', 'prediction', 'wer', 'word\_accuracy',  
'lattescore\_meaning\_preserved', 'gender', 'age', 'severity\_speech\_impairment',  
'type\_nonstandard\_speech', 'etiology']

Test DataFrame shape: (554, 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: 11.6%

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: 272 samples

test\_predictions.csv: 554 samples

=== DATA PREVIEW ===

	speaker_id	reference \
0	KES004	kila disemba familia yetu hukutana kijijini kw...

```

1 KES004 katika shule ya msingi ya langas ni ni ni nili...
2 KES004 nilipofanya safari ya kitalii katika mbuga ya ...
3 KES004 kila jumapili mama huamka mapena na kutengenez...
4 KES004 nilipokua chuo kikuu nairobi tulikua na na uta...
5 KES004 katika kampaini za usafi mtaani kibera sisi ka...
6 KES004 nilipokuwa soko la gikomba nilikutana na mama ...
7 KES004 siku moja nilisaidia shangazi yangu kuvuna mah...
8 KES004 nilihudhuria harusi ya ndugu yangu katika kaun...
9 KES004 katika safari yangu ya ya kwenda kakamega tuli...

```

```

                                prediction      wer \
0 Kila didisema familia yetu ukutana kijijini k... 0.500000
1 Katika shule ya msingi ya lelidanga nilijifun... 0.405405
2 Nilipofanya safari ya chitalika katika mbuga ... 0.409091
3 Kila jumapili mama wamuka mapema na kutengenz... 0.565217
4 Nilipokuwa chukitu huu Nairobi tulikona na ut... 0.555556
5 Katika kampeni za usafi mtani kimbera sisi ka... 0.375000
6 Nilipokuwa sokula jikombani nilikuwa kitanya ... 0.633333
7 Siku mmoja ni nisaaidia shangazi yangu kufuna ... 0.593750
8 Niliudhuria arusi ya ndugu yangu katika count... 0.531250
9 Kiatika safari ya konda kamega tulisemama kwe... 0.437500

```

```

lattescore_meaning_preserved
0
1
2
3
4
5
6
7
8
9

```

=== 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