

Capstone 1: Data Analysis of speech-to-text accuracy

March 20, 2025

0.0.1 Part 1: Data Analysis and Visualization

Task 1: Dataset Selection & Exploration Step 1: Import Datasets

```
[1]: import os

#Navigate. to the audio clip directory
audio_dir = '/content/drive/MyDrive/Colab_Notebooks/M2M Tech/Mozilla Common_
↳Voice/clips'

#List the audio file extensions
audio_extensions = [file.split('.')[1] for file in os.listdir(audio_dir)]

#Print the unique audio file extensions
print(set(audio_extensions))
```

{'mp3'}

Step 2: Inspect Transcription Text Files

The transcription text files are in TSV(Tab Separated Values) format. Lets load one of the files.

```
[2]: import pandas as pd
import numpy as np

#Load the validation.tsv file
validated_df = pd.read_csv('/content/drive/MyDrive/Colab_Notebooks/M2M Tech/
↳Mozilla Common Voice/validated.tsv', sep='\t')

copy_df = validated_df.copy()

#Display the first few rows of the dataframe
validated_df
```

```
[2]:
```

	client_id \
0	031903093b6fa1aeb0a243843eb9ed57baf6e99d1f8f92...
1	058fe5b1170aa09ef3f1092b179384639bc46ac53c1675...
2	08190396a5c298331813531d1a832b56d8ffe44aaedcb7...
3	14698ee63cabe08b43f0faa93304202d1e6ffeaa2cdf86...
4	28d8f8a88afad9eb9e5b36ee84bd4c5ba84137310da15f...

```

..
245 bf35656a89be5e1cff674565dbf2957250385cfde9d9bc...
246 bf35656a89be5e1cff674565dbf2957250385cfde9d9bc...
247 bf35656a89be5e1cff674565dbf2957250385cfde9d9bc...
248 bf35656a89be5e1cff674565dbf2957250385cfde9d9bc...
249 bf35656a89be5e1cff674565dbf2957250385cfde9d9bc...

```

```

                                path \
0   common_voice_en_41383256.mp3
1   common_voice_en_41823983.mp3
2   common_voice_en_41881685.mp3
3   common_voice_en_41799514.mp3
4   common_voice_en_41552032.mp3
..
245 common_voice_en_41764721.mp3
246 common_voice_en_41764867.mp3
247 common_voice_en_41764882.mp3
248 common_voice_en_41765127.mp3
249 common_voice_en_41765553.mp3

```

```

                                sentence_id \
0   f19a785911b1a3b1338e3eb5cc785e58b8381d21ec7c33...
1   f50360e1be367d8155b3c8340f0b3d38d1e6701df79dc5...
2   f4f3a5714cc36a9abbabf78a33feb4a9c368005f1f4bf5...
3   f4d04f6e48777c3ad180c629858a19fdfa4cb875d2bb22...
4   f262ed293fa5fe0986d1e7a80b5bbae11205f8089a1857...
..
245 0662495f0d7b39560a9762a9401bb66a7c8fd4f833e836...
246 f43ad81b60ed9f6d17c15b8607b0e4ed17e96e79479d59...
247 f43bd27fb270fea8b932b42fe50fddaa8403689f16db22...
248 f44951542856f68b0a69f3580f4a0a99708e59f44b2769...
249 f43db65048a3f6b3b448f29fa9394f137aaf2c03186680...

```

```

                                sentence sentence_domain \
0   The outer rim has undergone some erosion due t...      NaN
1   For purposes of this definition, the intent ma...      NaN
2   Bennett was educated at Lawnswood High School,...      NaN
3   These rules became known as Admiral-Lord Mount...      NaN
4   The grouping traditionally called apes is brac...      NaN
..
245 You can compare it with the marketing strategi...      NaN
246     Most other cities have similar problems.          NaN
247 He headed many lobbies to the provincial and n...      NaN
248 Behind this is Tai Mo Shan, the tallest hill i...      NaN
249 He ran two more races the next year, suffering...      NaN

```

```

up_votes  down_votes      age      gender \

```

0	2	0	NaN	NaN
1	3	0	fifties	female_feminine
2	2	0	fourties	female_feminine
3	2	0	twenties	NaN
4	2	0	NaN	NaN
..
245	2	0	sixties	female_feminine
246	2	0	sixties	female_feminine
247	2	0	sixties	female_feminine
248	2	0	sixties	female_feminine
249	2	0	sixties	female_feminine

			accents	variant	locale	segment
0			NaN	NaN	en	NaN
1	United States English,Washington State			NaN	en	NaN
2	United States English,southern draw			NaN	en	NaN
3	Japan English			NaN	en	NaN
4	Canadian English,English Hungarian			NaN	en	NaN
..		
245		United States English		NaN	en	NaN
246		United States English		NaN	en	NaN
247		United States English		NaN	en	NaN
248		United States English		NaN	en	NaN
249		United States English		NaN	en	NaN

[250 rows x 13 columns]

```
[3]: validated_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 250 entries, 0 to 249
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   client_id             250 non-null    object
1   path                  250 non-null    object
2   sentence_id           250 non-null    object
3   sentence               250 non-null    object
4   sentence_domain        0 non-null      float64
5   up_votes               250 non-null    int64
6   down_votes             250 non-null    int64
7   age                   230 non-null    object
8   gender                 122 non-null    object
9   accents                240 non-null    object
10  variant                0 non-null      float64
11  locale                 250 non-null    object
12  segment                0 non-null      float64
```

```
dtypes: float64(3), int64(2), object(8)
memory usage: 25.5+ KB
```

0.0.2 Task 2: Data Cleaning

Step 1: Check for na's

```
[4]: # Check for missing values
print(validated_df.isnull().sum())

# Drop rows with missing values
validated_df = validated_df.fillna('Unknown')
print('\n')
# Check the updated DataFrame
print(validated_df.head())
```

```
client_id      0
path           0
sentence_id    0
sentence       0
sentence_domain 250
up_votes       0
down_votes     0
age            20
gender         128
accents        10
variant        250
locale         0
segment        250
dtype: int64
```

```
client_id \
0  031903093b6fa1aeb0a243843eb9ed57baf6e99d1f8f92...
1  058fe5b1170aa09ef3f1092b179384639bc46ac53c1675...
2  08190396a5c298331813531d1a832b56d8ffe44aaedcb7...
3  14698ee63cabe08b43f0faa93304202d1e6ffea2cdf86...
4  28d8f8a88afad9eb9e5b36ee84bd4c5ba84137310da15f...
```

```
path \
0  common_voice_en_41383256.mp3
1  common_voice_en_41823983.mp3
2  common_voice_en_41881685.mp3
3  common_voice_en_41799514.mp3
4  common_voice_en_41552032.mp3
```

```
sentence_id \
0  f19a785911b1a3b1338e3eb5cc785e58b8381d21ec7c33...
```

```

1 f50360e1be367d8155b3c8340f0b3d38d1e6701df79dc5...
2 f4f3a5714cc36a9abbabf78a33feb4a9c368005f1f4bf5...
3 f4d04f6e48777c3ad180c629858a19fdfa4cb875d2bb22...
4 f262ed293fa5fe0986d1e7a80b5bbae11205f8089a1857...

```

	sentence	sentence_domain	\
0	The outer rim has undergone some erosion due t...	Unknown	
1	For purposes of this definition, the intent ma...	Unknown	
2	Bennett was educated at Lawnswood High School,...	Unknown	
3	These rules became known as Admiral-Lord Mount...	Unknown	
4	The grouping traditionally called apes is brac...	Unknown	

	up_votes	down_votes	age	gender	\
0	2	0	Unknown	Unknown	
1	3	0	fifties	female_feminine	
2	2	0	fourties	female_feminine	
3	2	0	twenties	Unknown	
4	2	0	Unknown	Unknown	

	accents	variant	locale	segment
0	Unknown	Unknown	en	Unknown
1	United States English,Washington State	Unknown	en	Unknown
2	United States English,southern draw	Unknown	en	Unknown
3	Japan English	Unknown	en	Unknown
4	Canadian English,English Hungarian	Unknown	en	Unknown

```
[5]: print(validated_df.columns)
```

```

Index(['client_id', 'path', 'sentence_id', 'sentence', 'sentence_domain',
      'up_votes', 'down_votes', 'age', 'gender', 'accents', 'variant',
      'locale', 'segment'],
      dtype='object')

```

Task 2: Clean up age column and change datatype

```

[6]: #Get unique values from the age column
unique_ages = validated_df['age'].unique()

#Print the unique ages
print(unique_ages)

```

```

['Unknown' 'fifties' 'fourties' 'twenties' 'sixties' 'thirties'
 'seventies' 'teens']

```

```

[7]: # Define the age mapping
age_mapping = {'Unknown':np.nan, 'teens':15, 'twenties':25, 'thirties':35,
               'fourties':45, 'fifties':55, 'sixties':65, 'seventies':75}

```

```
#Replace string values with numeric ages
validated_df['age'] = validated_df['age'].replace(age_mapping).astype('float64')

print(validated_df['age'])
print('\n')
print(validated_df['age'].describe())
```

```
0      NaN
1      55.0
2      45.0
3      25.0
4      NaN
...
245     65.0
246     65.0
247     65.0
248     65.0
249     65.0
Name: age, Length: 250, dtype: float64
```

```
count      230.00000
mean       42.00000
std        18.81071
min        15.00000
25%        25.00000
50%        35.00000
75%        65.00000
max        75.00000
Name: age, dtype: float64
```

```
<ipython-input-7-91966202e9d2>:5: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
validated_df['age'] =
validated_df['age'].replace(age_mapping).astype('float64')
```

Task 3: Clean up gender and accents column and change datatype

```
[8]: # Get unique values for gender
gender_uniques = validated_df['gender'].unique()
print("Unique values for gender:")
print(gender_uniques)
print('\n')

# Get unique values for accent
accent_uniques = validated_df['accents'].unique()
```

```
print("\nUnique values for accents:")
print(accent_uniques)
```

Unique values for gender:

```
['Unknown' 'female_feminine' 'male_masculine' 'non-binary' 'transgender']
```

Unique values for accents:

```
['Unknown' 'United States English,Washington State'
 'United States English,southern draw' 'Japan English'
 'Canadian English,English Hungarian' 'United States English'
 'England English,Kent' 'England English'
 'India and South Asia (India, Pakistan, Sri Lanka),South India,Kannadiga'
 'chicago italian' 'United States English,Chicago/Midwestern'
 'Southern African (South Africa, Zimbabwe, Namibia)'
 'India and South Asia (India, Pakistan, Sri Lanka)' 'Filipino'
 'French speaking english' 'Korean' 'Israeli English' 'Canadian English'
 'British with slightly Brazilian accent,British and slightly Brazilian '
 'Californian Accent' 'United States English,German,Russian'
 'United States English,learnt from Spanish' 'geordie ,England Northern'
 'United States English,Malaysian English' 'Japanese English'
 'Scottish English']
```

```
[9]: gender_mapping = {
    'Unknown': 'Unknown',
    'female_feminine': 'Female',
    'male_masculine': 'Male',
    'non-binary': 'Non-binary',
    'transgender': 'Transgender'
}

accent_mapping = {
    'Unknown': 'Unknown',
    'United States English,Washington State': 'US English',
    'United States English,southern draw': 'US English (Southern)',
    'United States English,Chicago/Midwestern': 'US English (Midwestern)',
    'United States English,German,Russian': 'US English (European influence)',
    'United States English,learnt from Spanish': 'US English (Latin American_
↳influence)',
    'United States English,Malaysian English': 'US English (Asian influence)',
    'Canadian English': 'Canadian English',
    'Canadian English,English Hungarian': 'Canadian English (Eastern European_
↳influence)',
    'England English': 'UK English',
    'England English,Kent': 'UK English (Southern)',
```

```

    'British with slightly Brazilian accent,British and slightly Brazilian ':␣
↪'UK English (Latin American influence)',
    'Scottish English': 'UK English (Scottish)',
    'geordie ,England Northern': 'UK English (Northern)',
    'India and South Asia (India, Pakistan, Sri Lanka)': 'South Asian English',
    'India and South Asia (India, Pakistan, Sri Lanka),South India,Kannadiga':␣
↪'South Asian English (Indian influence)',
    'Japan English': 'Japanese English',
    'Japanese English': 'Japanese English',
    'Korean': 'Korean English',
    'Israeli English': 'Middle Eastern English',
    'Filipino': 'Southeast Asian English',
    'French speaking english': 'European English (French influence)',
    'Southern African (South Africa, Zimbabwe, Namibia)': 'African English',
    'chicago italian': 'US English (Italian influence)',
    'Californian Accent': 'US English (West Coast)'
}

```

```

[10]: validated_df['gender'] = validated_df['gender'].replace(gender_mapping)
validated_df['accents'] = validated_df['accents'].replace(accent_mapping)

# Drop columns that we dont need and columns with no values
validated_df = validated_df.drop(['sentence_domain', 'up_votes', 'down_votes',␣
↪'variant', 'segment'], axis=1)

```

```

[11]: validated_df.head()

```

```

[11]:
                                     client_id \
0  031903093b6fa1aeb0a243843eb9ed57baf6e99d1f8f92...
1  058fe5b1170aa09ef3f1092b179384639bc46ac53c1675...
2  08190396a5c298331813531d1a832b56d8ffe44aaedcb7...
3  14698ee63cabe08b43f0faa93304202d1e6fffaa2cdf86...
4  28d8f8a88afad9eb9e5b36ee84bd4c5ba84137310da15f...

                                     path \
0  common_voice_en_41383256.mp3
1  common_voice_en_41823983.mp3
2  common_voice_en_41881685.mp3
3  common_voice_en_41799514.mp3
4  common_voice_en_41552032.mp3

                                     sentence_id \
0  f19a785911b1a3b1338e3eb5cc785e58b8381d21ec7c33...
1  f50360e1be367d8155b3c8340f0b3d38d1e6701df79dc5...
2  f4f3a5714cc36a9abbabf78a33feb4a9c368005f1f4bf5...
3  f4d04f6e48777c3ad180c629858a19fdfa4cb875d2bb22...
4  f262ed293fa5fe0986d1e7a80b5bbae11205f8089a1857...

```


	sentence	age	gender	\
0	The outer rim has undergone some erosion due t...	NaN	Unknown	
1	For purposes of this definition, the intent ma...	55.0	Female	
2	Bennett was educated at Lawnswood High School,...	45.0	Female	
3	These rules became known as Admiral-Lord Mount...	25.0	Unknown	
4	The grouping traditionally called apes is brac...	NaN	Unknown	

	accents	locale
0	Unknown	en
1	US English	en
2	US English (Southern)	en
3	Japanese English	en
4	Canadian English (Eastern European influence)	en

0.0.3 Task 3: Generate Basic Statistics

```
[12]: # Calculate word frequency
word_freq = validated_df['sentence'].apply(lambda x:len(x.split()))
print("word frequency:")
print(word_freq.describe())

print('\n')
# Calculate sentence length
sentence_length = validated_df['sentence'].apply(lambda x:len(x))
print("sentence length:")
print(sentence_length.describe())
print('\n')

#calculate punctuation use
punctuation_use = validated_df['sentence'].apply(lambda x: sum(not c.isalnum()
    ↪for c in x))
print("Punctuation Use:")
print(punctuation_use.describe())
```

```
word frequency:
count    250.000000
mean       9.964000
std        2.816101
min         3.000000
25%         8.000000
50%        10.000000
75%        12.000000
max        14.000000
Name: sentence, dtype: float64
```

```
sentence length:
count      250.000000
mean       61.100000
std        19.364657
min        10.000000
25%        48.000000
50%        61.000000
75%        74.750000
max        102.000000
Name: sentence, dtype: float64
```

```
Punctuation Use:
count      250.000000
mean       10.668000
std         3.113285
min         3.000000
25%         8.000000
50%        11.000000
75%        13.000000
max        17.000000
Name: sentence, dtype: float64
```

key insights:

Word frequency: 1. Most sentences have around 8-12 words (25th-75th percentile).

2. The average sentence has approximately 10 words (mean).
3. There's a relatively small range of word frequencies, with most sentences having between 3 and 14 words.

Sentence length:

1. Most sentences are around 48-75 characters long (25th-75th percentile). The average sentence is approximately 61 characters long (mean).
2. There's a wider range of sentence lengths, with some sentences being as short as 10 characters or as long as 102 characters.

Punctuation use:

1. Most sentences have around 8-13 punctuation marks (25th-75th percentile). The average sentence has approximately 11 punctuation marks (mean).
2. There's a relatively small range of punctuation use, with most sentences having between 3 and 17 punctuation marks.

0.0.4 Task 4: Visualizations

```
[13]: import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots

# Create a figure with three subplots
fig = make_subplots(rows=3, cols=1, subplot_titles=['Age Distribution', 'Gender_
↪Distribution', 'Accent Distribution'], vertical_spacing=0.1)

# Plot the age distribution
age_df = validated_df['age'].value_counts().reset_index()
age_df.columns = ['Age', 'Frequency']
fig.add_trace(go.Bar(x=age_df['Age'], y=age_df['Frequency']), row=1, col=1)

# Plot the gender distribution
gender_df = validated_df['gender'].value_counts().reset_index()
gender_df.columns = ['Gender', 'Frequency']
fig.add_trace(go.Bar(x=gender_df['Gender'], y=gender_df['Frequency']), row=2,
↪col=1)

# Plot the accent distribution
accent_df = validated_df['accents'].value_counts().reset_index()
accent_df.columns = ['Accent', 'Frequency']
fig.add_trace(go.Bar(x=accent_df['Accent'], y=accent_df['Frequency']), row=3,
↪col=1)

# Update the layout to use a log scale for the y-axis
fig.update_yaxes(type="log", row=3, col=1)

# Update the layout
fig.update_layout(height=1000, margin=dict(l=50, r=50, t=100, b=50))

# Show the plot
fig.show()
```

0.0.5 Summary

The analysis reveals imbalances in age, gender, and accent distributions. The dataset is dominated by users in their 20s and 60s, with underrepresentation in the 30s and 50s.

There's a bias towards female speakers, and while the dataset includes non-binary and transgender speakers, their numbers are small.

Accent distribution is also skewed, with “Unknown” accents dominating and underrepresentation of regional variants. These imbalances may impact speech model performance and robustness.

0.0.6 Part 2: AI vs Human Transcription Accuracy

Step 1: Transcribe Audio with Whisper AI

```
[14]: !pip install -U openai-whisper jiwer
import whisper
```

Collecting openai-whisper

Downloading openai-whisper-20240930.tar.gz (800 kB)
0.0/800.5

kB ? eta -:--:--

798.7/800.5

kB 25.7 MB/s eta 0:00:01

800.5/800.5 kB

17.4 MB/s eta 0:00:00

Installing build dependencies ... done

Getting requirements to build wheel ... done

Preparing metadata (pyproject.toml) ... done

Collecting jiwer

Downloading jiwer-3.1.0-py3-none-any.whl.metadata (2.6 kB)

Requirement already satisfied: numba in /usr/local/lib/python3.11/dist-packages
(from openai-whisper) (0.60.0)

Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages
(from openai-whisper) (2.0.2)

Requirement already satisfied: torch in /usr/local/lib/python3.11/dist-packages
(from openai-whisper) (2.6.0+cu124)

Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages
(from openai-whisper) (4.67.1)

Requirement already satisfied: more-itertools in /usr/local/lib/python3.11/dist-packages
(from openai-whisper) (10.6.0)

Collecting tiktoken (from openai-whisper)

Downloading tiktoken-0.9.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.7 kB)

Requirement already satisfied: triton>=2.0.0 in /usr/local/lib/python3.11/dist-packages
(from openai-whisper) (3.2.0)

Requirement already satisfied: click>=8.1.8 in /usr/local/lib/python3.11/dist-packages
(from jiwer) (8.1.8)

Collecting rapidfuzz>=3.9.7 (from jiwer)

Downloading rapidfuzz-3.12.2-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)

Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in
/usr/local/lib/python3.11/dist-packages (from numba->openai-whisper) (0.43.0)

Requirement already satisfied: regex>=2022.1.18 in
/usr/local/lib/python3.11/dist-packages (from tiktoken->openai-whisper)
(2024.11.6)

Requirement already satisfied: requests>=2.26.0 in
/usr/local/lib/python3.11/dist-packages (from tiktoken->openai-whisper) (2.32.3)

Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (3.17.0)

Requirement already satisfied: typing-extensions>=4.10.0 in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (4.12.2)

Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (3.4.2)

Requirement already satisfied: Jinja2 in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (3.1.6)

Requirement already satisfied: fsspec in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (2024.10.0)

Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch->openai-whisper)
 Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)

Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch->openai-whisper)
 Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)

Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch->openai-whisper)
 Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)

Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch->openai-whisper)
 Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)

Collecting nvidia-cublas-cu12==12.4.5.8 (from torch->openai-whisper)
 Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)

Collecting nvidia-cufft-cu12==11.2.1.3 (from torch->openai-whisper)
 Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)

Collecting nvidia-curand-cu12==10.3.5.147 (from torch->openai-whisper)
 Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)

Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch->openai-whisper)
 Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)

Collecting nvidia-cusparselt-cu12==12.3.1.170 (from torch->openai-whisper)
 Downloading nvidia_cusparselt_cu12-12.3.1.170-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)

Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (0.6.2)

Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (2.21.5)

Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (12.4.127)

Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch->openai-whisper)
 Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)

Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.11/dist-packages (from torch->openai-whisper) (1.13.1)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch->openai-whisper) (1.3.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken->openai-whisper) (3.4.1)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken->openai-whisper) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken->openai-whisper) (2.3.0)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken->openai-whisper) (2025.1.31)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2->torch->openai-whisper) (3.0.2)

Downloading jiwer-3.1.0-py3-none-any.whl (22 kB)

Downloading rapidfuzz-3.12.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.1 MB)

3.1/3.1 MB

73.5 MB/s eta 0:00:00

Downloading tiktoken-0.9.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2 MB)

1.2/1.2 MB

43.6 MB/s eta 0:00:00

Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl (363.4 MB)

363.4/363.4 MB

5.0 MB/s eta 0:00:00

Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (13.8 MB)

13.8/13.8 MB

73.9 MB/s eta 0:00:00

Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (24.6 MB)

24.6/24.6 MB

89.4 MB/s eta 0:00:00

Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)

883.7/883.7 kB

57.8 MB/s eta 0:00:00

Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl (664.8 MB)

664.8/664.8 MB

2.1 MB/s eta 0:00:00

Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl
(211.5 MB)

211.5/211.5 MB

5.4 MB/s eta 0:00:00

Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl (56.3 MB)

56.3/56.3 MB

12.9 MB/s eta 0:00:00

Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-manylinux2014_x86_64.whl (127.9 MB)

127.9/127.9 MB

7.4 MB/s eta 0:00:00

Downloading nvidia_cusparses_cu12-12.3.1.170-py3-none-manylinux2014_x86_64.whl (207.5 MB)

207.5/207.5 MB

5.8 MB/s eta 0:00:00

Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (21.1 MB)

21.1/21.1 MB

94.1 MB/s eta 0:00:00

Building wheels for collected packages: openai-whisper

Building wheel for openai-whisper (pyproject.toml) ... done

Created wheel for openai-whisper: filename=openai_whisper-20240930-py3-none-any.whl size=803406

sha256=64fba4796c4edfe1d8c77a5ad4abb4408e8471361d539c70fd5a44a6b78616a5

Stored in directory: /root/.cache/pip/wheels/2f/f2/ce/6eb23db4091d026238ce76703bd66da60b969d70bcc81d5d3a

Successfully built openai-whisper

Installing collected packages: rapidfuzz, nvidia-nvjitlink-cu12, nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-cu12, tiktoken, nvidia-cusparses-cu12, nvidia-cudnn-cu12, jiwer, nvidia-cusolver-cu12, openai-whisper

Attempting uninstall: nvidia-nvjitlink-cu12

Found existing installation: nvidia-nvjitlink-cu12 12.5.82

Uninstalling nvidia-nvjitlink-cu12-12.5.82:

Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82

Attempting uninstall: nvidia-curand-cu12

Found existing installation: nvidia-curand-cu12 10.3.6.82

Uninstalling nvidia-curand-cu12-10.3.6.82:

Successfully uninstalled nvidia-curand-cu12-10.3.6.82

Attempting uninstall: nvidia-cufft-cu12

Found existing installation: nvidia-cufft-cu12 11.2.3.61

Uninstalling nvidia-cufft-cu12-11.2.3.61:

Successfully uninstalled nvidia-cufft-cu12-11.2.3.61

Attempting uninstall: nvidia-cuda-runtime-cu12

Found existing installation: nvidia-cuda-runtime-cu12 12.5.82

Uninstalling nvidia-cuda-runtime-cu12-12.5.82:

Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82

```

Attempting uninstall: nvidia-cuda-nvrtc-cu12
Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
  Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
Attempting uninstall: nvidia-cuda-cupti-cu12
Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
  Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
Attempting uninstall: nvidia-cublas-cu12
Found existing installation: nvidia-cublas-cu12 12.5.3.2
Uninstalling nvidia-cublas-cu12-12.5.3.2:
  Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
Attempting uninstall: nvidia-cusparse-cu12
Found existing installation: nvidia-cusparse-cu12 12.5.1.3
Uninstalling nvidia-cusparse-cu12-12.5.1.3:
  Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
Attempting uninstall: nvidia-cudnn-cu12
Found existing installation: nvidia-cudnn-cu12 9.3.0.75
Uninstalling nvidia-cudnn-cu12-9.3.0.75:
  Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
Attempting uninstall: nvidia-cusolver-cu12
Found existing installation: nvidia-cusolver-cu12 11.6.3.83
Uninstalling nvidia-cusolver-cu12-11.6.3.83:
  Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Successfully installed jiwer-3.1.0 nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-
cupti-cu12-12.4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-
cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70 nvidia-cufft-cu12-11.2.1.3 nvidia-
curand-cu12-10.3.5.147 nvidia-cusolver-cu12-11.6.1.9 nvidia-cusparse-
cu12-12.3.1.170 nvidia-nvjitlink-cu12-12.4.127 openai-whisper-20240930
rapidfuzz-3.12.2 tiktoken-0.9.0

```

Step2: Import Libraries and load transcriptions

```
[15]: # Load Whisper model (medium for balance of speed/ accuracy)
```

```
model = whisper.load_model("medium")
```

```
100%|          | 1.42G/1.42G [00:30<00:00, 49.5MiB/s]
```

```
[16]: # Transcribe first 50 samples (adjust based on Colab resources)
```

```

subset = validated_df.head(50)
transcripts = []

for idx, row in subset.iterrows():
    audio_path = os.path.join(audio_dir, row['path'])
    result = model.transcribe(audio_path)
    transcripts.append(result['text'])

```



```
subset["ai_transcript"] = transcripts
subset.to_csv("ai_transcripts.csv", index=False)
```

<ipython-input-16-9e611312a6ab>:11: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
[17]: print(subset['ai_transcript'])
```

```
0    The out of rim has undergone some erosion due...
1    For the purposes of this definition, the inte...
2    Bennett was educated at Lonswood High School,...
3        These s became quas maelodmount Tabin's pad
4    The grouping traditionally called apes is bra...
5    The track, Coca Leaf, is made up of vocal sam...
6        He also taught at the same university.
7    Most Naruto video games have been released on...
8    Bold denotes players currently active in inte...
9    At Southside, Greer was an all-state selectio...
10        Riley and Buffy pursue in his car.
11    The combined company was later renamed Kimbal...
12    It also features puzzles and platforming elem...
13        It has three platforms.
14    After graduation, she became the administrato...
15        It airs in North America via Dish Network.
16    Instead of signing artists, he makes one-time...
17    After this, the town was rebuilt and simply r...
18        That was enough for a conflict.
19    In Korea, the carnations express admiration, ...
20        Naval Academy in Annapolis.
21    It also likes flower, nectar, fruit and some ...
22    After graduation he served with the Royal Pru...
23    He worked as a builder and was later employed...
24    It has also been proposed as a possible expla...
25        He later attended the University of Maryland.
26        Cuttings are slow to strike roots.
27    The undergraduate program in Theological Stud...
28    Balanjiga can be reached through public utili...
29    The old comedy was awkward and poor in its ve...
30        Considered a shining star of the core worlds.
31    A win earns a two-year exemption for most eve...
```

```

32     These may also be forgeries to conceal identi...
33         Moore lives in Toronto.
34         In Laos, represents a double word.
35     Many of the illustrations he created for the ...
36     You just didn't bother saying this sucks anym...
37     Its headquarters are in the town of Dematru, ...
38
39
40     He experienced weightlessness and saw the cur...
41         Holmes vs. Cooney was refereed by Mills Lane.
42         It is a freeway for its entire length.
43     This group became legendary in the dance world.
44     The mountains separate Luzon's central plain ...
45     Fundamental justice was thus seen as having b...
46     A mosque is situated inside Sitio Bolangan on...
47     The river bearing this name was renamed Pagan...
48     A similar example is Charlotte Perkins Gilman...
49     The establishment was commanded by an officer...
Name: ai_transcript, dtype: object

```

```
[18]: # Import Word Error(WER) and Character Error (CER)
```

```

!pip install -U jiwer
from jiwer import wer, cer

```

Requirement already satisfied: jiwer in /usr/local/lib/python3.11/dist-packages (3.1.0)

Requirement already satisfied: click>=8.1.8 in /usr/local/lib/python3.11/dist-packages (from jiwer) (8.1.8)

Requirement already satisfied: rapidfuzz>=3.9.7 in /usr/local/lib/python3.11/dist-packages (from jiwer) (3.12.2)

```
[26]: #Load the dataset with AI-generated transcripts
```

```

df = pd.read_csv('/content/drive/MyDrive/Colab_Notebooks/M2M Tech/Mozilla_
↳Common Voice/ai_transcripts.csv')

```

```
# Ensure columns exist
```

```

df = df.dropna(subset=["sentence", "ai_transcript"])
df = df.drop(['sentence_domain', 'up_votes', 'down_votes', 'variant',
↳'segment'], axis=1)
df.head(10)

```

```

[26]: client_id \
0  031903093b6fa1aeb0a243843eb9ed57baf6e99d1f8f92...
1  058fe5b1170aa09ef3f1092b179384639bc46ac53c1675...
2  08190396a5c298331813531d1a832b56d8ffe44aaedcb7...
3  14698ee63cabe08b43f0faa93304202d1e6ffeea2cdf86...
4  28d8f8a88afad9eb9e5b36ee84bd4c5ba84137310da15f...

```

5 29f0a09db1abdd5ad550cc754834822748ab17043e7b51...
6 302b51de89ba0e815cce38bba05ee858e556375d19c745...
7 31a787d9d9c7203d20fbd2b9590c92dda91ef69e2b36f7...
8 40ef8f7ec5e58f2ba7b9c934cf57ae928bae5b13662b8f...
9 4af8321a13399a2662bef236a736ad5329cb2edfe01eee...

path \
0 common_voice_en_41383256.mp3
1 common_voice_en_41823983.mp3
2 common_voice_en_41881685.mp3
3 common_voice_en_41799514.mp3
4 common_voice_en_41552032.mp3
5 common_voice_en_41827319.mp3
6 common_voice_en_41526838.mp3
7 common_voice_en_41435787.mp3
8 common_voice_en_41633128.mp3
9 common_voice_en_41586424.mp3

sentence_id \
0 f19a785911b1a3b1338e3eb5cc785e58b8381d21ec7c33...
1 f50360e1be367d8155b3c8340f0b3d38d1e6701df79dc5...
2 f4f3a5714cc36a9abbabf78a33feb4a9c368005f1f4bf5...
3 f4d04f6e48777c3ad180c629858a19fdfa4cb875d2bb22...
4 f262ed293fa5fe0986d1e7a80b5bbae11205f8089a1857...
5 f506be20f56b00999c323469c384114e27bb5fa0372d0e...
6 f21bd922ee542930c8c3914edad64557f31004a22a80a1...
7 f1b86bcf63efc1b0117b5587f8803741a5a6917ab44421...
8 f2d2020fe97305b41d4c626ee2963f0984e255e5599105...
9 f2872bf6e9d312de78bdcc664bc348936d9039966c4104...

	sentence	age	gender
0	The outer rim has undergone some erosion due t...	NaN	Unknown
1	For purposes of this definition, the intent ma...	55.0	Female
2	Bennett was educated at Lawnswood High School,...	45.0	Female
3	These rules became known as Admiral-Lord Mount...	25.0	Unknown
4	The grouping traditionally called apes is brac...	NaN	Unknown
5	The track "Coca Leaf" is made up of vocal samp...	NaN	Unknown
6	He also taught at the same university.	65.0	Female
7	Most "Naruto" video games have been released o...	NaN	Unknown
8	Bold denotes players currently active in inter...	35.0	Female
9	At South Side, Greer was an all state selectio...	55.0	Female

	accents	locale
0	Unknown	en
1	US English	en
2	US English (Southern)	en
3	Japanese English	en

4	Canadian English (Eastern European influence)	en
5	Unknown	en
6	United States English	en
7	Unknown	en
8	United States English	en
9	United States English	en

```

                                ai_transcript
0  The out of rim has undergone some erosion due...
1  For the purposes of this definition, the inte...
2  Bennett was educated at Lonswood High School,...
3  These little spochondras are mirror- Dais Mou...
4  The grouping traditionally called apes is bra...
5  The track, Coca Leaf, is made up of vocal sam...
6      He also taught at the same university.
7  Most Naruto video games have been released on...
8  Bold denotes players currently active in inte...
9  At Southside, Greer was an all-state selectio...

```

Step 3: Compute WER and CER for Each sample

```

[27]: # Ensure subset is a new copy of the DataFrame
subset = subset.copy()

# Compute WER and CER
subset.loc[:, "wer"] = subset.apply(lambda row: wer(row["sentence"],
    ↪row["ai_transcript"]), axis=1)
subset.loc[:, "cer"] = subset.apply(lambda row: cer(row["sentence"],
    ↪row["ai_transcript"]), axis=1)

# Save updated results
subset.to_csv("ai_transcripts_with_metrics.csv", index=False)

# Display average WER and CER
average_wer = subset["wer"].mean()
average_cer = subset["cer"].mean()

print(f"Average WER: {average_wer:.2%}")
print(f"Average CER: {average_cer:.2%}")

```

Average WER: 13.73%

Average CER: 7.08%

Step 4: Visualize the Error Rates

```

[28]: import seaborn as sns
import matplotlib.pyplot as plt

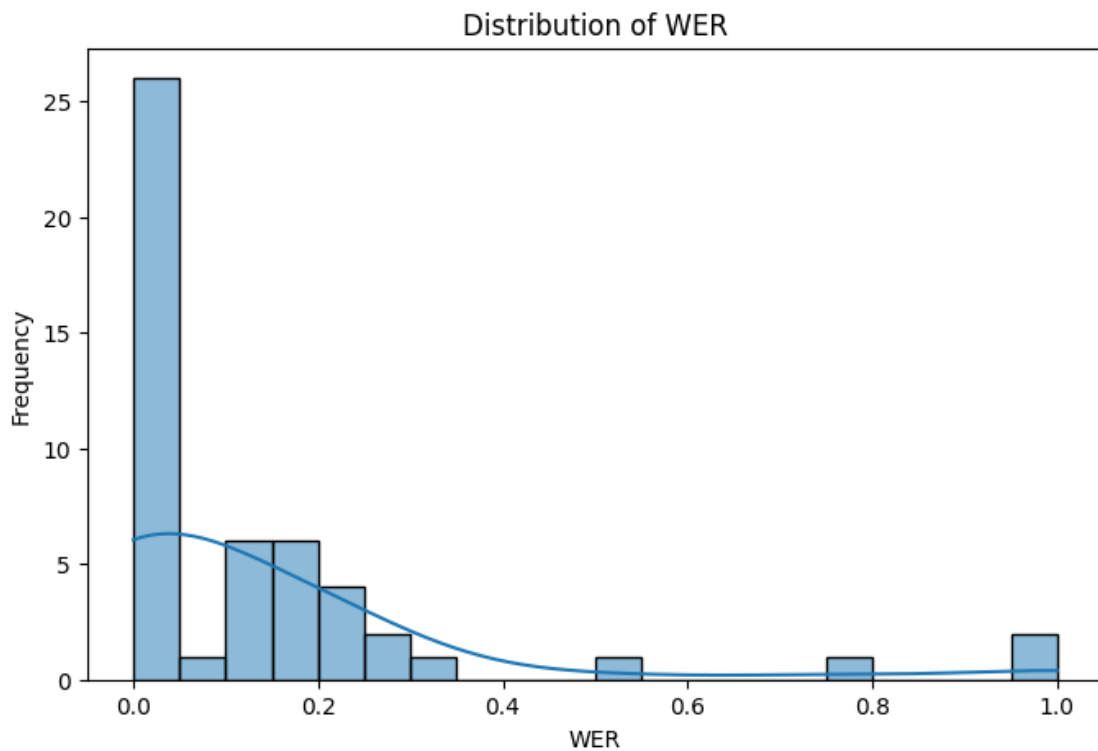
```

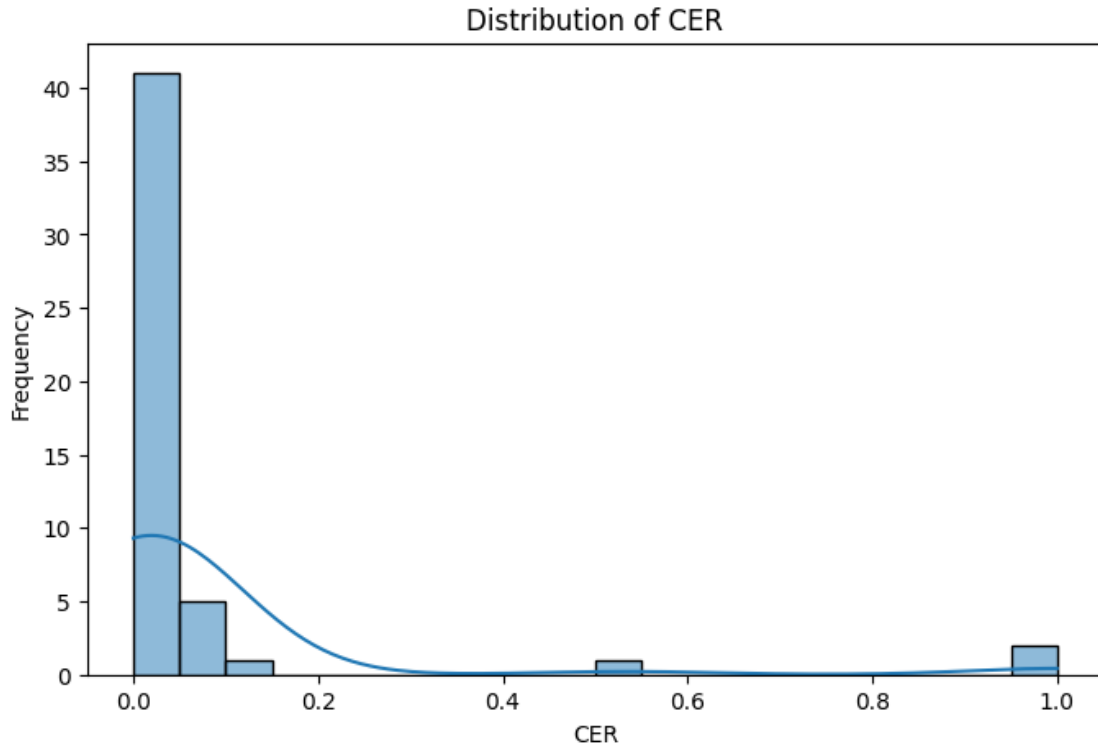
```

# Plot WER distribution
plt.figure(figsize=(8,5))
sns.histplot(subset['wer'], bins=20, kde=True)
plt.title('Distribution of WER')
plt.xlabel('WER')
plt.ylabel('Frequency')
plt.show()

# Plot CER distribution
plt.figure(figsize=(8, 5))
sns.histplot(subset['cer'], bins=20, kde=True)
plt.title('Distribution of CER')
plt.xlabel('CER')
plt.ylabel('Frequency')
plt.show()

```





Summary:

1. **Most transcriptions have low CER and WER** • The majority of the data points are concentrated around low error rates (close to 0).
 - This suggests that the ASR (Automatic Speech Recognition) model performs well on most of the dataset.
2. **Long-Tail Distribution (Presence of Outliers)** • There are some high CER and WER values, creating a long tail towards the right.
 - This means that some transcriptions are significantly worse than others, possibly due to:
 - Poor audio quality
 - Strong accents
 - Background noise
 - Rare words not well represented in the training set
3. **CER is lower than WER on average** • This is expected because CER measures errors at the character level, whereas WER measures errors at the word level.
 - A small character-level mistake (e.g., missing a single letter) may not necessarily cause a word-level mistake.

Step 5: Process and Compare AI vs Human Transcriptions

```
[29]: # Load dataset
df = pd.read_csv('/content/drive/MyDrive/Colab_Notebooks/M2M Tech/Mozilla_
↳Common Voice/ai_transcripts_with_metrics.csv')

# Rename columns for clarity
```

```

df.rename(columns={'sentence': 'ground_truth'}, inplace=True)

# Compute WER & CER for AI
df['WER_AI'] = df.apply(lambda row: wer(row['ground_truth'],
    ↪row['ai_transcript']), axis=1)
df['CER_AI'] = df.apply(lambda row: cer(row['ground_truth'],
    ↪row['ai_transcript']), axis=1)

# Assign WER & CER = 0 for human reference (since it's the ground truth)
df['WER_Human'] = 0
df['CER_Human'] = 0

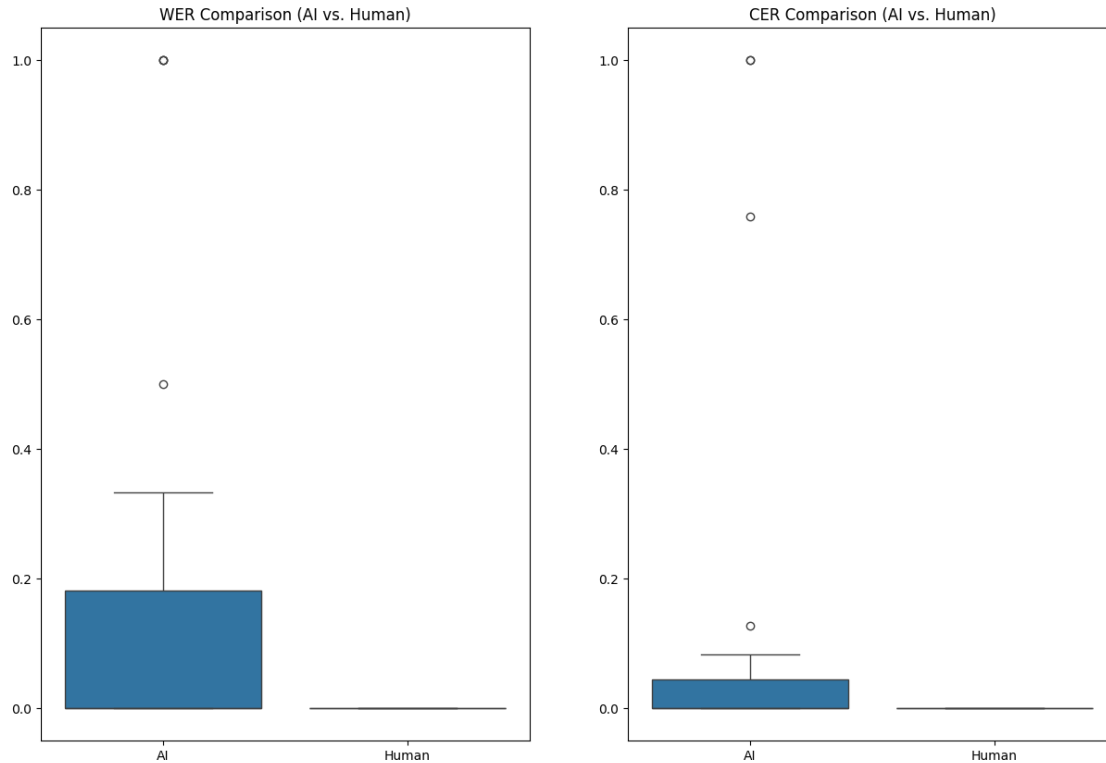
# Plot boxplots to visualize the difference
plt.figure(figsize=(15, 10))

# WER Comparison
plt.subplot(1, 2, 1)
sns.boxplot(data=df[['WER_AI', 'WER_Human']])
plt.xticks([0, 1], ['AI', 'Human'])
plt.title("WER Comparison (AI vs. Human)")

# CER Comparison
plt.subplot(1, 2, 2)
sns.boxplot(data=df[['CER_AI', 'CER_Human']])
plt.xticks([0, 1], ['AI', 'Human'])
plt.title("CER Comparison (AI vs. Human)")

plt.show()

```



Summary of the Graphs (WER & CER Distribution)

1. Most AI-generated transcriptions have low error rates, with WER(Word Error Rate) and CER(Character Error Rate) Concentrated around 0.

This means AI performs well on most samples

2. A smaller portion of cases has extremely high WER(nearly 1.0).

These could be outliers, possibly due to noisy audio, uncommon words or poor AI recognition in certain scenarios.

3. The distribution is right-skewed (more low-error cases, fewer high-error cases).

AI is mostly accurate than WER but fails significantly on some difficult samples.

4. CER is lower than WER, indicating that AI tend to make more word-level mistakes rather than character-level mistakes.

Example: "The cat ran fast" -> "The cats run fast" (presents word-level error but few character changes: The tense of the verb has changed from past to present, and the subject has changed from singular to plural. WER is more sensitive to these types of errors than CER.)

```
[62]: data = df[['WER_AI', 'CER_AI', 'WER_Human', 'CER_Human']]
      data.head(10)
```



```
[62]:
```

	WER_AI	CER_AI	WER_Human	CER_Human
0	0.181818	0.044776	0	0
1	0.100000	0.063492	0	0
2	0.166667	0.036145	0	0
3	1.000000	0.758621	0	0
4	0.000000	0.000000	0	0
5	0.230769	0.044776	0	0
6	0.000000	0.000000	0	0
7	0.100000	0.033898	0	0
8	0.000000	0.000000	0	0
9	0.285714	0.040000	0	0

```
[61]: numeric_df = df[['WER_AI', 'CER_AI', 'WER_Human', 'CER_Human']]
corr_matrix = numeric_df.corr()

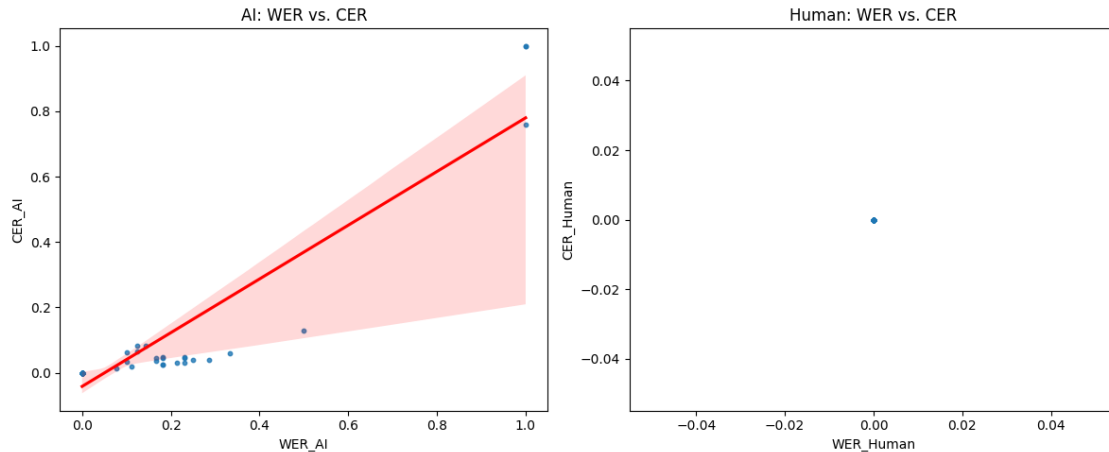
fig = px.imshow(
    corr_matrix,
    text_auto=True, # Show correlation values
    color_continuous_scale=px.colors.sequential.RdBu, # Use a valid colorscale
    labels=dict(x="Metrics", y="Metrics", color="Correlation"),
    title="Correlation Heatmap"
)
fig.update_xaxes(tickangle=45, tickfont=dict(size=12))
fig.update_yaxes(tickfont=dict(size=12))
fig.show()

# --- 2. Scatterplots with Regression Lines using Seaborn --- #
fig, axes = plt.subplots(1, 2, figsize=(12, 5))

sns.regplot(x=df["WER_AI"], y=df["CER_AI"], ax=axes[0], scatter_kws={'s':10},
    ↪line_kws={'color':'red'})
axes[0].set_title("AI: WER vs. CER")

sns.regplot(x=df["WER_Human"], y=df["CER_Human"], ax=axes[1], scatter_kws={'s':
    ↪10}, line_kws={'color':'blue'})
axes[1].set_title("Human: WER vs. CER")

plt.tight_layout()
plt.show()
```



Step 6: HeatMap, Correlation Matrix and Table Comparison

```
[57]: import plotly.express as px

fig = px.scatter_matrix(df[['WER_AI', 'CER_AI', 'WER_Human', 'CER_Human']],
                        dimensions=['WER_AI', 'CER_AI', 'WER_Human', 'CER_Human'],
                        title="Correlation Scatter Matrix",
                        color_continuous_scale=px.colors.sequential.RdBu)

fig.update_layout(width=800, height=800)
fig.show()
```

Summary of the Correlation Scatter Matrix

Strong positive correlation between WER_AI and CER_AI:

The top-left plot shows that as AI word error rate increases, character error rate tends to increase as well. This makes logical sense as more word errors would typically coincide with more character errors.

Limited range for human metrics:

Both WER_Human and CER_Human appear to be clustered near zero, suggesting that human performance has consistently low error rates compared to AI.

Little correlation between human and AI performance:

The plots comparing WER_AI vs WER_Human and CER_AI vs CER_Human don't show strong patterns, indicating that AI errors don't strongly predict human errors or vice versa. Bimodal distribution in AI metrics: Both AI metrics (WER_AI and CER_AI) show some interesting clustering - there are points near zero (perfect performance) and another cluster at higher values (around 0.3-0.5 for WER and around 0.1 for CER), suggesting the AI system might perform very well on some tasks but struggle significantly on others.

Human performance consistency:

Human metrics are mostly concentrated near zero with very few outliers, suggesting high consistency in human performance across different inputs or tasks.

```
[32]: from IPython.display import display, HTML

# Create a table for side-by-side comparison
comparison_table = df[['ground_truth', 'ai_transcript']]

# Rename the columns for clarity
comparison_table = comparison_table.rename(columns={'ground_truth': 'Human_↵
↵Translation', 'ai_transcript': 'AI Translation'})

# Convert the table to HTML
human_table_html = comparison_table[['Human Translation']].to_html()
ai_table_html = comparison_table[['AI Translation']].to_html()

# Display the tables side-by-side using HTML
display(HTML(f"<table><tr><td>{human_table_html}</td><td>{ai_table_html}</td></
↵tr></table>"))
```

<IPython.core.display.HTML object>

```
[ ]: !sudo apt-get update
!sudo apt-get install texlive-xetex pandoc
```

```
0% [Working]          Get:1 https://cloud.r-project.org/bin/linux/ubuntu
jammy-cran40/ InRelease [3,632 B]
0% [Connecting to archive.ubuntu.com] [Waiting for headers] [1 InRelease 3,632
Get:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease [1,581 B]
0% [Connecting to archive.ubuntu.com] [Waiting for headers] [1 InRelease 3,632
0% [Connecting to archive.ubuntu.com] [Waiting for headers] [Connecting to r2u.
0% [Connecting to archive.ubuntu.com (185.125.190.81)] [Waiting for headers] [C
Get:3 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Get:4 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ Packages [69.9
kB]
Get:5 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
Packages [1,378 kB]
Hit:6 http://archive.ubuntu.com/ubuntu jammy InRelease
Get:7 https://r2u.stat.illinois.edu/ubuntu jammy InRelease [6,555 B]
Get:8 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
Get:9 https://r2u.stat.illinois.edu/ubuntu jammy/main all Packages [8,768 kB]
Get:10 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages
[1,236 kB]
Hit:11 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:12 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy
```

```

InRelease
Hit:13 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Get:14 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages
[2,692 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Get:16 https://r2u.stat.illinois.edu/ubuntu jammy/main amd64 Packages [2,680 kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [3,000
kB]
Get:18 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages
[1,535 kB]
Fetched 21.8 MB in 3s (7,327 kB/s)
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository
'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does not seem to provide
it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3
  libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
  libgs9 libgs9-common libidn12 libijs-0.35 libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53
  libtexluajit2 libwoff1 libzip-0-13 lmodern pandoc-data poppler-data
  preview-latex-style rake ruby ruby-net-telnet ruby-rubygems ruby-webrick
  ruby-xmlrpc ruby3.0 rubygems-integration tclutils teckit tex-common tex-gyre
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-extra texlive-latex-recommended texlive-pictures
  texlive-plain-generic tipa xfonts-encodings xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java
  texlive-luatex pandoc-citeproc context wkhtmltopdf librsvg2-bin groff ghc
  nodejs php python libjs-mathjax libjs-katex citation-style-language-styles
  poppler-utils ghostscript fonts-japanese-mincho | fonts-ipafont-mincho
  fonts-japanese-gothic | fonts-ipafont-gothic fonts-arphic-ukai
  fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper gv
  | postscript-viewer perl-tk xpdf | pdf-viewer xzdec
  texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
  icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
  texlive-latex-extra-doc texlive-latex-recommended-doc texlive-pstricks
  dot2tex prerex texlive-pictures-doc vprerex default-jre-headless tipa-doc
The following NEW packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3

```

```

libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
libgs9 libgs9-common libidn12 libijs-0.35 libjbig2dec0 libkpathsea6
libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53
libtexluajit2 libwoff1 libzip-0-13 lmodern pandoc pandoc-data poppler-data
preview-latex-style rake ruby ruby-net-telnet ruby-rubygems ruby-webrick
ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-common tex-gyre
texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra texlive-latex-recommended texlive-pictures
texlive-plain-generic texlive-xetex tipa xfonts-encodings xfonts-utils
0 upgraded, 58 newly installed, 0 to remove and 35 not upgraded.
Need to get 202 MB of archives.
After this operation, 728 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all
1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1
[2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all
0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17
[33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all
20200910-1 [6,367 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common
all 9.55.0~dfsg1-0ubuntu5.10 [752 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64
1.38-4ubuntu1 [60.0 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64
0.35-15build2 [16.5 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64
0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64
9.55.0~dfsg1-0ubuntu5.10 [5,031 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libkpathsea6
amd64 2021.20210626.59705-1ubuntu0.2 [60.4 kB]
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libwoff1 amd64
1.0.2-1build4 [45.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/universe amd64 dvisvgm amd64
2.13.1-1 [1,221 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-lmodern all
2.004.5-6.1 [4,532 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-noto-mono all
20201225-1build1 [397 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-texgyre all
20180621-3.1 [10.2 MB]
Get:17 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:18 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcmarm-
gfm0.29.0.gfm.3 amd64 0.29.0.gfm.3-3 [115 kB]

```

Get:19 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcmark-gfm-extensions0.29.0.gfm.3 amd64 0.29.0.gfm.3-3 [25.1 kB]
Get:20 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-parent-java all 43-1 [10.8 kB]
Get:21 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-logging-java all 1.2-2 [60.3 kB]
Get:22 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libfontenc1 amd64 1:1.1.4-1build3 [14.7 kB]
Get:23 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libptexenc1 amd64 2021.20210626.59705-1ubuntu0.2 [39.1 kB]
Get:24 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rubygems-integration all 1.18 [5,336 B]
Get:25 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby3.0 amd64 3.0.2-7ubuntu2.8 [50.1 kB]
Get:26 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-rubygems all 3.3.5-2 [228 kB]
Get:27 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby amd64 1:3.0~exp1 [5,100 B]
Get:28 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rake all 13.0.6-2 [61.7 kB]
Get:29 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-net-telnet all 0.1.1-2 [12.6 kB]
Get:30 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-webrick all 1.7.0-3ubuntu0.1 [52.1 kB]
Get:31 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-xmlrpc all 0.3.2-1ubuntu0.1 [24.9 kB]
Get:32 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libruby3.0 amd64 3.0.2-7ubuntu2.8 [5,113 kB]
Get:33 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libsynchronet2 amd64 2021.20210626.59705-1ubuntu0.2 [55.6 kB]
Get:34 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libteckit0 amd64 2.5.11+ds1-1 [421 kB]
Get:35 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexlua53 amd64 2021.20210626.59705-1ubuntu0.2 [120 kB]
Get:36 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexluajit2 amd64 2021.20210626.59705-1ubuntu0.2 [267 kB]
Get:37 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libzip-0-13 amd64 0.13.72+dfsg.1-1.1 [27.0 kB]
Get:38 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-encodings all 1:1.0.5-0ubuntu2 [578 kB]
Get:39 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-utils amd64 1:7.7+6build2 [94.6 kB]
Get:40 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 lmodern all 2.004.5-6.1 [9,471 kB]
Get:41 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 pandoc-data all 2.9.2.1-3ubuntu2 [81.8 kB]
Get:42 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 pandoc amd64 2.9.2.1-3ubuntu2 [20.3 MB]

```

Get:43 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.2 [9,860 kB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:55 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:56 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:57 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:58 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 202 MB in 9s (23.5 MB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line 78,
<> line 58.)
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 125044 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Selecting previously unselected package fonts-lato.

```

```

Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
Unpacking fonts-lato (2.0-2.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
Unpacking poppler-data (0.4.11-1) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.17_all.deb ...
Unpacking tex-common (6.17) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
Unpacking fonts-urw-base35 (20200910-1) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.10_all.deb ...
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.10) ...
Selecting previously unselected package libidn12:amd64.
Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15build2) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.10_amd64.deb ...
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.10) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build4) ...
Selecting previously unselected package dvisvgm.
Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
Unpacking dvisvgm (2.13.1-1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
Unpacking fonts-lmodern (2.004.5-6.1) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
Unpacking fonts-noto-mono (20201225-1build1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...
Unpacking fonts-texgyre (20180621-3.1) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...

```



```

Selecting previously unselected package libcmark-gfm0.29.0.gfm.3:amd64.
Preparing to unpack .../17-libcmark-gfm0.29.0.gfm.3_0.29.0.gfm.3-3_amd64.deb ...
Unpacking libcmark-gfm0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Selecting previously unselected package libcmark-gfm-
extensions0.29.0.gfm.3:amd64.
Preparing to unpack .../18-libcmark-gfm-
extensions0.29.0.gfm.3_0.29.0.gfm.3-3_amd64.deb ...
Unpacking libcmark-gfm-extensions0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../19-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../20-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../21-libfontenc1_1%3a1.1.4-1build3_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../22-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../23-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...
Selecting previously unselected package ruby3.0.
Preparing to unpack .../24-ruby3.0_3.0.2-7ubuntu2.8_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.8) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../25-ruby-rubygems_3.3.5-2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../26-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../27-rake_13.0.6-2_all.deb ...
Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../28-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../29-ruby-webrick_1.7.0-3ubuntu0.1_all.deb ...
Unpacking ruby-webrick (1.7.0-3ubuntu0.1) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../30-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../31-libruby3.0_3.0.2-7ubuntu2.8_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.8) ...

```

```

Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../32-libsyntax2_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libsyntax2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../33-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../34-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
.../35-libtexluajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../36-libzip-0-13_0.13.7+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.7+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../37-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../38-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../39-lmodern_2.004.5-6.1_all.deb ...
Unpacking lmodern (2.004.5-6.1) ...
Selecting previously unselected package pandoc-data.
Preparing to unpack .../40-pandoc-data_2.9.2.1-3ubuntu2_all.deb ...
Unpacking pandoc-data (2.9.2.1-3ubuntu2) ...
Selecting previously unselected package pandoc.
Preparing to unpack .../41-pandoc_2.9.2.1-3ubuntu2_amd64.deb ...
Unpacking pandoc (2.9.2.1-3ubuntu2) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../42-preview-latex-style_12.2-1ubuntu1_all.deb ...
Unpacking preview-latex-style (12.2-1ubuntu1) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../43-t1utils_1.41-4build2_amd64.deb ...
Unpacking t1utils (1.41-4build2) ...
Selecting previously unselected package teckit.
Preparing to unpack .../44-teckit_2.5.11+ds1-1_amd64.deb ...
Unpacking teckit (2.5.11+ds1-1) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../45-tex-gyre_20180621-3.1_all.deb ...
Unpacking tex-gyre (20180621-3.1) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../46-texlive-
binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb ...

```

```

Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../47-texlive-base_2021.20220204-1_all.deb ...
Unpacking texlive-base (2021.20220204-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../48-texlive-fonts-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-fonts-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../49-texlive-latex-base_2021.20220204-1_all.deb ...
Unpacking texlive-latex-base (2021.20220204-1) ...
Selecting previously unselected package libfontbox-java.
Preparing to unpack .../50-libfontbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libfontbox-java (1:1.8.16-2) ...
Selecting previously unselected package libpdfbox-java.
Preparing to unpack .../51-libpdfbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libpdfbox-java (1:1.8.16-2) ...
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack .../52-texlive-latex-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-latex-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../53-texlive-pictures_2021.20220204-1_all.deb ...
Unpacking texlive-pictures (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../54-texlive-latex-extra_2021.20220204-1_all.deb ...
Unpacking texlive-latex-extra (2021.20220204-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../55-texlive-plain-generic_2021.20220204-1_all.deb ...
Unpacking texlive-plain-generic (2021.20220204-1) ...
Selecting previously unselected package tipa.
Preparing to unpack .../56-tipa_2%3a1.3-21_all.deb ...
Unpacking tipa (2:1.3-21) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../57-texlive-xetex_2021.20220204-1_all.deb ...
Unpacking texlive-xetex (2021.20220204-1) ...
Setting up fonts-lato (2.0-2.1) ...
Setting up fonts-noto-mono (20201225-1build1) ...
Setting up libwoff1:amd64 (1.0.2-1build4) ...
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libijs-0.35:amd64 (0.35-15build2) ...
Setting up libtexluaajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libfontbox-java (1:1.8.16-2) ...
Setting up rubygems-integration (1.18) ...
Setting up libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Setting up fonts-urw-base35 (20200910-1) ...
Setting up poppler-data (0.4.11-1) ...
Setting up tex-common (6.17) ...
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based

```

```

frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line
78.)
debconf: falling back to frontend: Readline
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-1build3) ...
Setting up libjbig2dec0:amd64 (0.19-3build2) ...
Setting up libteckit0:amd64 (2.5.11+ds1-1) ...
Setting up libapache-pom-java (18-1) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up t1utils (1.41-4build2) ...
Setting up libidn12:amd64 (1.38-4ubuntu1) ...
Setting up fonts-texgyre (20180621-3.1) ...
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up ruby-webrick (1.7.0-3ubuntu0.1) ...
Setting up libcmark-gfm0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Setting up fonts-lmodern (2.004.5-6.1) ...
Setting up libcmark-gfm-extensions0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Setting up pandoc-data (2.9.2.1-3ubuntu2) ...
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Setting up libsynchronet2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.10) ...
Setting up teckit (2.5.11+ds1-1) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.10) ...
Setting up preview-latex-style (12.2-1ubuntu1) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.13.1-1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up pandoc (2.9.2.1-3ubuntu2) ...
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) ...
Setting up texlive-base (2021.20220204-1) ...
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.

```

```

tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line
78.)
debconf: falling back to frontend: Readline
Setting up tex-gyre (20180621-3.1) ...
Setting up texlive-plain-generic (2021.20220204-1) ...
Setting up texlive-latex-base (2021.20220204-1) ...
Setting up texlive-latex-recommended (2021.20220204-1) ...
Setting up texlive-pictures (2021.20220204-1) ...
Setting up texlive-fonts-recommended (2021.20220204-1) ...
Setting up tipa (2:1.3-21) ...
Setting up texlive-latex-extra (2021.20220204-1) ...
Setting up texlive-xetex (2021.20220204-1) ...
Setting up rake (13.0.6-2) ...
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.8) ...
Setting up ruby3.0 (3.0.2-7ubuntu2.8) ...
Setting up ruby (1:3.0~exp1) ...
Setting up ruby-rubygems (3.3.5-2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for mailcap (3.70+nmulubuntu1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.8) ...
/sbin/ldconfig.real: /usr/local/lib/libumf.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libhwloc.so.15 is not a symbolic link

```

```
/sbin/ldconfig.real: /usr/local/lib/libtcm_debug.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link
```

```
Processing triggers for tex-common (6.17) ...
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line
78.)
debconf: falling back to frontend: Readline
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.
    This may take some time...
```

```
[ ]: !jupyter nbconvert --to pdf "/content/drive/MyDrive/Colab_Notebooks/M2M Tech/
↳Capstone 1: Data Analysis of speech-to-text accuracy.ipynb"
```