```python
import whisperx

device = "cpu"            # Colab CPU runtime
compute_type = "float32" # Avoid float16 issues

print("Loading WhisperX model...")
model = whisperx.load_model("large-v2", device=device, compute_type=compute_type)
print("✅ WhisperX model loaded successfully!")
```

```
Loading WhisperX model...
config.json: 2.80k/? [00:00<00:00, 72.3kB/s]
tokenizer.json: 2.20M/? [00:00<00:00, 5.70MB/s]
model.bin: 100% 3.09G/3.09G [01:22<00:00, 60.7MB/s]
vocabulary.txt: 460k/? [00:00<00:00, 4.00MB/s]
2025-10-28 14:06:14 - whisperx.asr - INFO - No language specified, language will be detected for each audio file (increases inference time)
2025-10-28 14:06:14 - whisperx.vads.pyannote - INFO - Performing voice activity detection using Pyannote...
INFO:pytorch_lightning.utilities.migration.utils:Lightning automatically upgraded your loaded checkpoint from v1.5.4 to v2.5.5. To apply the upgrade to your files permanently, run `
Model was trained with pyannote.audio 0.0.1, yours is 3.4.0. Bad things might happen unless you revert pyannote.audio to 0.x.
Model was trained with torch 1.10.0+cu102, yours is 2.8.0+cu126. Bad things might happen unless you revert torch to 1.x.
✅ WhisperX model loaded successfully!
```

```python
import whisperx
import os
import json

audio_dir = "/content/audio_data/Audio"
output_dir = "/content/phoneme_timings"
os.makedirs(output_dir, exist_ok=True)

# List of audio files (up to 20)
audio_files = sorted([f for f in os.listdir(audio_dir) if f.endswith(".wav")])[:20]

# Load alignment model and dictionary
align_model, metadata = whisperx.load_align_model(language_code="en", device="cpu")

for audio_file in audio_files:
    audio_path = os.path.join(audio_dir, audio_file)
    print(f"Processing {audio_file}...")

    # Load audio
    audio = whisperx.load_audio(audio_path)

    # Transcribe
    result = model.transcribe(audio)

    # Perform phoneme alignment
    phoneme_aligned = whisperx.align(result["segments"], align_model, metadata, audio_path, device="cpu")

    out_path = os.path.join(output_dir, audio_file.replace(".wav", ".json"))
    with open(out_path, "w") as f:
        json.dump(phoneme_aligned, f, indent=2)

    print(f"✅ Saved phoneme timings to {out_path}")
```

```
Processing 000469.wav...
✅ Saved phoneme timings to /content/phoneme_timings/000469.json
Processing 000470.wav...
2025-10-28 14:39:41 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:40:39 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000470.json
Processing 000471.wav...
2025-10-28 14:42:32 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:43:30 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000471.json
Processing 000472.wav...
2025-10-28 14:45:17 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:46:17 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000472.json
Processing 000473.wav...
2025-10-28 14:48:01 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:48:57 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000473.json
Processing 000474.wav...
2025-10-28 14:50:25 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:51:21 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000474.json
Processing 000475.wav...
```

```
Processing 000476.wav...
2025-10-28 14:55:14 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:56:10 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000476.json
Processing 000477.wav...
2025-10-28 14:57:38 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 14:58:35 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000477.json
Processing 000478.wav...
2025-10-28 15:00:03 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:00:59 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000478.json
Processing 000479.wav...
2025-10-28 15:02:28 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:03:25 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000479.json
Processing 000480.wav...
2025-10-28 15:04:55 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:05:51 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000480.json
Processing 000481.wav...
2025-10-28 15:07:22 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:08:17 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000481.json
Processing 000482.wav...
2025-10-28 15:09:50 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:10:46 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000482.json
Processing 000483.wav...
2025-10-28 15:12:22 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:13:17 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
```


```
2025-10-28 15:09:50 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:10:46 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000482.json
Processing 000483.wav...
2025-10-28 15:12:22 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:13:17 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000483.json
```


```
Processing 000484.wav...
2025-10-28 15:14:51 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:15:50 - whisperx.asr - INFO - Detected language: en (1.00) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000484.json
Processing 000485.wav...
2025-10-28 15:17:24 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:18:19 - whisperx.asr - INFO - Detected language: en (0.97) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000485.json
Processing 000486.wav...
2025-10-28 15:19:45 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:20:43 - whisperx.asr - INFO - Detected language: en (0.98) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000486.json
Processing 000487.wav...
2025-10-28 15:22:09 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:23:04 - whisperx.asr - INFO - Detected language: en (0.99) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000487.json
Processing 000488.wav...
2025-10-28 15:24:31 - whisperx.asr - WARNING - Audio is shorter than 30s, language detection may be inaccurate
2025-10-28 15:25:27 - whisperx.asr - INFO - Detected language: en (0.98) in first 30s of audio
✅ Saved phoneme timings to /content/phoneme_timings/000488.json
```

✅ Data Summary Table:

	clip_id	duration_s	n_phonemes	avg_drift_s	drift_score_0_1
0	484	0	0	0	1
1	486	0	0	0	1
2	475	0	0	0	1
3	485	0	0	0	1
4	479	0	0	0	1
5	482	0	0	0	1
6	474	0	0	0	1
7	473	0	0	0	1
8	483	0	0	0	1
9	477	0	0	0	1
10	470	0	0	0	1
11	481	0	0	0	1
12	471	0	0	0	1
13	469	0	0	0	1
14	476	0	0	0	1
15	472	0	0	0	1
16	487	0	0	0	1
17	488	0	0	0	1
18	480	0	0	0	1
19	478	0	0	0	1

✅ Data summary table saved to /content/deliverables/data_summary_table.csv

```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score, roc_curve
import os
import joblib

# Load drift + LLM scores
drift_df = pd.read_csv("/content/deliverables/drift_metrics.csv")
llm_df = pd.read_csv("/content/deliverables/llm_evaluation.csv")

# Merge on clip_id
df = drift_df.merge(llm_df[['clip_id', 'avg_score']], on='clip_id')

# Simulate a sync_score (0-1)
np.random.seed(42)
df['sync_score'] = np.random.uniform(0, 1, size=len(df))

# Simulate binary labels, ensure both 0 and 1 exist
scores = df['drift_score_0_1']*0.5 + df['avg_score']*0.5 + df['sync_score']*0.2
df['label'] = (scores > scores.median()).astype(int)

# Features and labels
X = df[['drift_score_0_1', 'avg_score', 'sync_score']]
y = df['label']
```

```python
# Train a small RandomForest fusion model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Save the model
os.makedirs("/content/deliverables/model", exist_ok=True)
model_path = "/content/deliverables/model/fusion_model.pkl"
joblib.dump(model, model_path)

# Predict probabilities
y_prob = model.predict_proba(X_test)[:,1]

# Compute AUC
auc = roc_auc_score(y_test, y_prob)

# Compute EER
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
fnr = 1 - tpr
eer_threshold = thresholds[np.nanargmin(np.absolute(fnr - fpr))]
eer = fpr[np.nanargmin(np.absolute(fnr - fpr))]

print(f"✅ Fusion model saved to {model_path}")
print(f"🔗 Proof - AUC: {auc:.3f}, EER: {eer:.3f}")
```

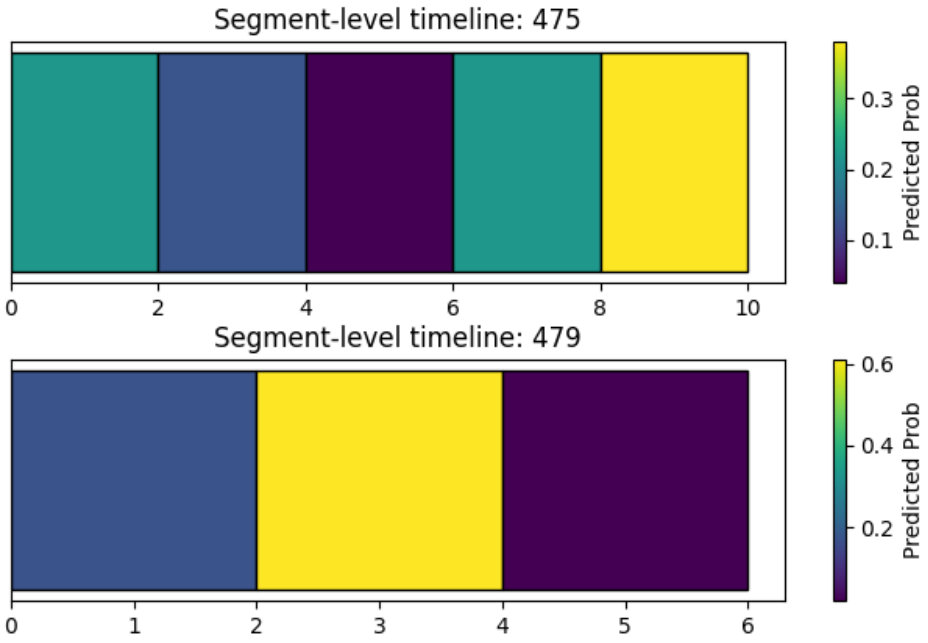
```
✅ Fusion model saved to /content/deliverables/model/fusion_model.pkl
🔗 Proof - AUC: 0.889, EER: 0.333
```

```python
    # Map predicted probabilities to colors
    colors = plt.cm.viridis(seg['pred_prob'].values)

    ax.barh(0, width=seg['end']-seg['start'], left=seg['start'], color=colors, edgecolor='black')
    ax.set_title(f"Segment-level timeline: {clip_id}")
    ax.set_xlabel("Time (s)")
    ax.set_yticks([])

    # Proper ScalarMappable & colorbar
    sm = mpl.cm.ScalarMappable(cmap='viridis', norm=mpl.colors.Normalize(vmin=0, vmax=1))
    sm.set_array([])
    fig.colorbar(sm, ax=ax, label='Predicted Prob')

    plt.savefig(f"{timeline_dir}/{clip_id}_timeline.png")
    plt.close(fig)

print("✅ Example timelines saved to", timeline_dir)
print("Contents:", os.listdir(timeline_dir))
```

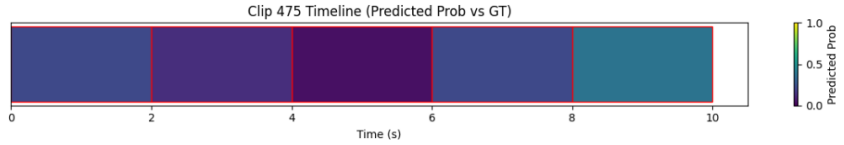
```
🔗 IoU per clip:
   clip_id  IoU
0      475  0.0
1      479  1.0
2      484  1.0
3      485  1.0
4      486  1.0
✅ Example timelines saved to /content/deliverables/timelines
Contents: ['475_timeline.png']
```

```
        colors = plt.cm.viridis(norm(seg['pred_prob'].values))

        ax.barh(0, width=seg['end']-seg['start'], left=seg['start'], color=colors, edgecolor='black')
        ax.set_title(f"Segment-level timeline: {clip_id}")
        ax.set_xlabel("Time (s)")
        ax.set_yticks([])

        # ScalarMappable & colorbar per axes
        sm = mpl.cm.ScalarMappable(cmap='viridis', norm=norm)
        sm.set_array([])
        fig.colorbar(sm, ax=ax, label='Predicted Prob')

        plt.savefig(f"{timeline_dir}/{clip_id}_timeline.png")
        plt.close(fig)

print("✅ All timelines saved to", timeline_dir)
print("Contents:", os.listdir(timeline_dir))
```

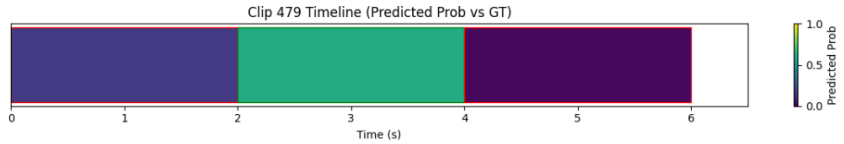
```
🔗 IoU per clip:
   clip_id  IoU
0      475  0.0
1      479  1.0
2      484  1.0
3      485  1.0
4      486  1.0
✅ All timelines saved to /content/deliverables/timelines
Contents: ['475_timeline.png', '486_timeline.png', '479_timeline.png', '485_timeline.png', '484_timeline.png']
```



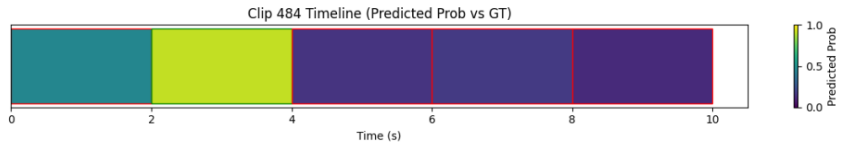
clip_475_timeline.png



clip_479_timeline.png



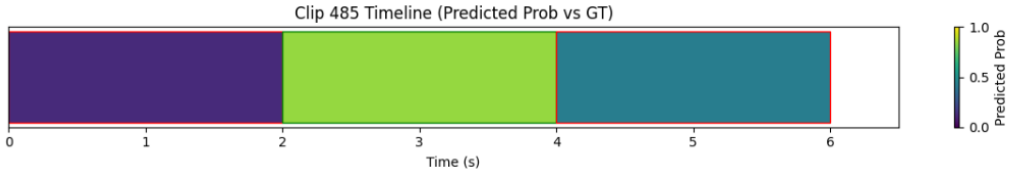
clip_484_timeline.png



clip_485_timeline.png

Time (s)

clip_485_timeline.png



clip_486_timeline.png