# 113學年度大學部專題海報展

國立清華大學 資訊工程學系
National Tsing Hua University Department of Computer Science

# Enhancing Meeting Quality with NLP: From Speech-to-Text to Summarization and Analysis

組員:梅可親、陳祐瑞、姚東廷、黎秉諺、簡雅雯、邱宇婕

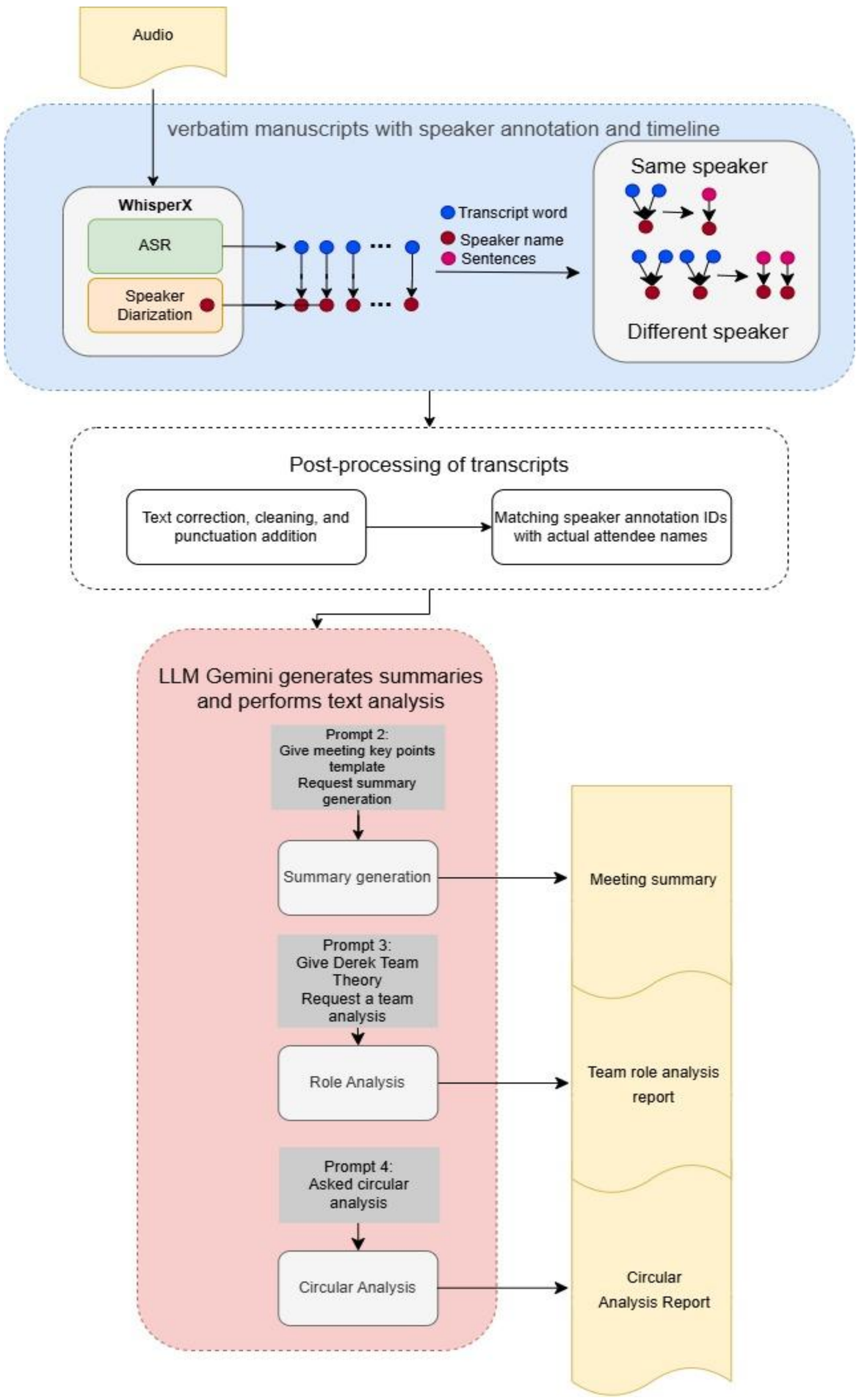
## Abstract

Inefficient meetings often lead to the challenge of organizing records. This study developed a system integrating WhisperX and large language models to transcribe audio, summarize content, and analyze team dynamics, helping users improve performance by understanding member roles and team strengths.

## Motivation

1. Meetings are often inefficient.
2. Few participants actively engage.
3. Verbatim transcription is time-consuming and labor-intensive.
4. Verbatim drafts don't enhance team capabilities effectively.

## Framework



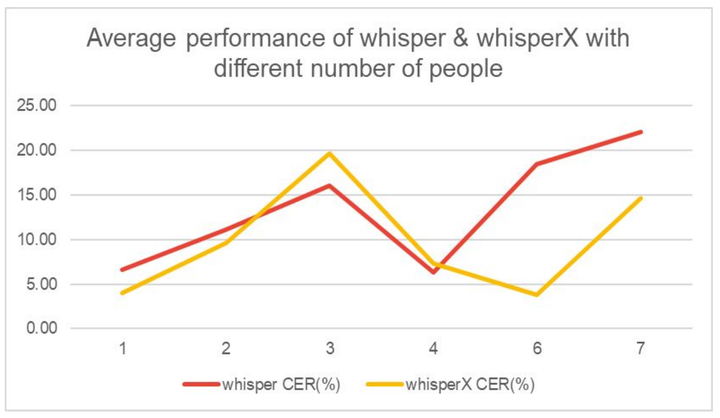

## Experiments

**Automatic Speech Recognition:** Obtain the correct text through manual correction of verbatim manuscripts, use Character Error Rate (CER) for comparison, and obtain data related to error types during the correction process.

*CER = (substitutions+deletions+insertions) / reference text*

**Summarization Models:** The transcribed meeting transcripts are input into the model for testing, followed by manual evaluation by the meeting participants to determine whether the key points of the meeting were successfully summarized.

## Result




CER with time

| Length of time | whisper CER(%) | whisperX CER(%) | CER difference(%) whisper-whisperX |
|---|---|---|---|
| under 10 minutes | 17.82 | 15.65 | 2.17 |
| Within 10 to 30 minutes | 12.86 | 11.43 | 1.43 |
| More than 30 minutes | 14.35 | 3.56 | 10.79 |

We tested summarization models including PEGASUS, TextRank, and LexRank, on our group's meeting transcripts. After initial tests, TextRank and LexRank were quickly eliminated. PEGASUS also performed poorly on multiple transcripts and was eventually discarded.

| Benchmark | Gemini Flash | ClaudeHaiku | GPT-3.5 Turbo |
|---|---|---|---|
| MMLU | 77.9 | 73.8 | 69.8 |
| GPQA | 38.6 | 35.7 | 30.8 |
| DROP | 79.7 | 78.4 | 70.2 |

## Conclusion

**Automatic Speech Recognition:**

- WhisperX resolves repetition issues and improves multi-speaker video stability.

**Summarization Model**:

- Non-LLM models performed poorly due to the conversational, informal nature of transcripts.
- Gemini model provided the best quality, handling repetition and enhancing linguistic diversity.

**Future Outlook**:

- Explore faster speaker annotation models beyond pyannote.
- Refine Gemini for varied transcripts.
- Strengthen prompt engineering and error analysis for precise, efficient summaries.

**Project Demo**