

A Survey of AI Techniques for Meeting Summarization and Engagement Analysis in Video Conferencing

Assignment 2 Report

Submitted by Name: Mohammed Omar Abdullah Alqathmi

Lecturer: Dr. Mohammed Anbar

Assignment 2 Report: A Survey of AI Techniques for Meeting Summarization and Engagement Analysis in Video Conferencing

Course Title: Research Methodology in Computer Science (AI800)

Lecturer: Dr. Mohammed Anbar

Weighting: 30%

Hand-out Date: 1 November 2025

Hand-in Date: 25 November 2025

Hand-in Method: Online Submission via E-Learning

Submitted by Name: Mohammed Omar Abdullah Alqathmi

Email: 253001437@st.midocean.edu.km

Student Number: 253001437

CONTENTS

1.0 Introduction	4
2.0 Literature Review	5
2.1 Action-Item-Driven Summarization of Long Meeting Transcripts (Golia & Kalita, 2023)	5
2.2 Investigating Automatic Metrics on Meeting Summarization (Kirstein et al., 2024) ..	5
2.3 Real-time Emotion Recognition in Video Conferencing (Lin, 2024)	5
2.4 Automated Minutes of Meeting Using a Multimodal Approach (Singh et al., 2023) ..	5
2.5 Facial Expression Recognition using Visible and IR (2025)	5
2.6 Recognition of Automatic Student Engagement in Online Classrooms (2025)	5
3.0 Problem Statement	6
3.1 Problem Domain	6
3.2 Problem Description	6
3.3 Significance of the Problem.....	6
3.4 Existing Solutions	6
3.5 Research Gaps	7
4.0 Research Objectives.....	8
4.1 Primary Goal.....	8
4.2 Specific Objectives.....	8
5.0 References	10

A Survey of AI Techniques for Meeting Summarization and Engagement Analysis in Video Conferencing

1.0 INTRODUCTION

The dominant effect of the explosive expansion of online collaboration tools has been the massive increase in the number of virtual meetings using tools like Zoom, Google Meet, and Microsoft Teams. These meetings are recorded because they provide a wealth of information to be extracted to capture the meetings decisions, tasks, and engagement of the participants. As organisations transition to using virtual communication channels, being able to draw actionable intelligence from these meetings becomes paramount for productivity and knowledge management. This report reviews recent literature on AI-based meeting summarisation and engagement analyses, providing an understanding of what methodologies have been used, what is lacking, and how it can be improved.

2.0 LITERATURE REVIEW

2.1 ACTION-ITEM-DRIVEN SUMMARIZATION OF LONG MEETING TRANSCRIPTS (GOLIA & KALITA, 2023)

In this paper, the authors propose a recursive summarization architecture that utilizes a BART transformer-based model to summarize action items from transcripts of meetings. The authors attempt to mitigate memory loss by addressing transcripts in segments distinguished by topic. Their model was evaluated for accuracy with a reported BERTScore of 64.98. Limitations include generalizability across datasets and cost of computation.

2.2 INVESTIGATING AUTOMATIC METRICS ON MEETING SUMMARIZATION (KIRSTEIN ET AL., 2024)

This paper examines meeting summarization metrics and finds that one-third of metric-error combinations either obfuscate or reward error. There was a weak correlation with human judgments that demonstrates existing summarization metrics are not reliable.

2.3 REAL-TIME EMOTION RECOGNITION IN VIDEO CONFERENCING (LIN, 2024)

The authors presented a CNN-LSTM-TCN framework representing a real-time emotion analysis model used with virtual meetings. The authors noted the improvement in performance by applying edge computing around latency. There are privacy concerns associated with the study and but again it is still sensitive to lighting conditions.

2.4 AUTOMATED MINUTES OF MEETING USING A MULTIMODAL APPROACH (SINGH ET AL., 2023)

The authors' system utilizes "Whisper" for transcription and employs "natural language processing" (NLP) for analysis. They achieve high key-point and action-item accuracy using this method. However, there were issues related to performance in noisy environments for meetings or meetings specific to a particular domain.

2.5 FACIAL EXPRESSION RECOGNITION USING VISIBLE AND IR (2025)

This article involved early fusion of visible and infrared imaging, achieving 84–85% recognition accuracy; they noted that hardware requirements and previous research datasets limited scalability to laboratory environments.

2.6 Recognition of Automatic Student Engagement in Online Classrooms (2025)

A multi-scale CNN with attention fusion is applied to distinguish engagement levels resulting in an 82 to 88% accuracy. The limitations concerning are high-resolution videos and privacy issues.

3.0 PROBLEM STATEMENT

3.1 PROBLEM DOMAIN

Automated meeting summarization and action item extraction in virtual collaboration environments.

3.2 PROBLEM DESCRIPTION

Thousands of meetings are recorded every month by various organizations on platforms such as Zoom, Teams and Google Meet. These meetings contain important decisions, actions to be taken and responsibilities assigned. Reviewing these long recordings manually is slow and prone to human errors.

Studies show that employees spend 3–4 hours per week reviewing meeting notes (Singh et al., 2023), and manual extraction often leads to missed action items and unclear follow-ups (Golia & Kalita, 2023). The problem becomes worse for people who miss meetings or new employees who rely on past meeting context.

Meeting transcripts often include irrelevant content and repeated discussions, off topic conversations. Understanding context, speaker turns and flow of conversation is essential for extracting clean summaries and structured action items. These tasks are generally hard to automate.

3.3 SIGNIFICANCE OF THE PROBLEM

“Poor summarization and documentation of meetings lead to 20–30% of action items being forgotten or delayed, resulting in repeated discussions and productivity loss (Golia & Kalita, 2023).”

Saving at least 1 hour per employee in a week through automation results in savings of over 5,200 hours per year in an organization of about 100 employees.

Automated summaries also improve knowledge sharing by helping new or remote team members quickly understand past decisions without watching long recordings.

3.4 EXISTING SOLUTIONS

- Transformer-based Summarization (Golia & Kalita, 2023):
BART with recursive summarization improved meeting action-item extraction, achieving a BERTScore of **64.98** on AML.
- Multimodal Meeting Minutes System (Singh et al., 2023):
Whisper AI + NLP achieved **92% key point** and **87% action item** accuracy, producing minutes **5× faster** than manual writing.
- Evaluation Metrics Analysis (Kirstein et al., 2024):
Showed that many summarization metrics are unreliable, with weak correlation to human judgment ($\alpha \approx 0.79$).

3.5 RESEARCH GAPS

Despite progress, key issues remain:

- High computational requirements:
Models like BART and PEGASUS need powerful GPUs, limiting accessibility.
- Weak generalization:
Models trained on AMI or clean datasets struggle with noisy, real-world meetings (Singh et al., 2023).
- Lack of fair comparison across models:
No consistent benchmarking of common pre-trained models (BART, T5, PEGASUS, LED) using the same dataset.
- Action item extraction is still difficult:
Many models summarize well but fail to extract structured tasks (owner, deadline, action).
- Accessibility gap:
Most research focuses on model design, not on easy, practical deployment for students or small teams.

Because of these gaps, a clear, practical comparison of readily available summarization models is needed.

4.0 RESEARCH OBJECTIVES

4.1 PRIMARY GOAL

Evaluation and comparison of the performance of existing pre-trained AI models for meeting summarization and action item extraction using publicly available datasets,

4.2 SPECIFIC OBJECTIVES

1. **To** identify and prepare suitable public meeting transcript datasets (AMI, ICSI) with proper preprocessing such as speaker normalization and text cleaning.
2. **To** implement summarization using two approaches:
 - **Open-source models** including BART-base, PEGASUS-base, T5-small, LED-base using HuggingFace.
 - **Commercial APIs** including GPT-5, Claude, Gemini with structured prompts.
3. **To** evaluate summarization quality using:
 - ROUGE and BERTScore (where references exist)
 - qualitative review of clarity and correctness
 - action item extraction accuracy
4. **To** implement action item extraction using:
 - prompt-based or fine-tuned sequence labeling (open-source)
 - structured JSON prompting (API models)
5. **To** compare practical aspects of each approach including:
 - computational requirements for open-source models
 - cost, response time, and context limits for API models
 - ease of setup and deployment

6. **To** analyze the impact of different prompting strategies including but not limited to zero-shot, few-shot, chain-of-thought on summarization and action item extraction.
 7. **To** implement a simple demo application (Streamlit/Gradio) that:
 - accepts meeting transcripts
 - allows model/method selection
 - generates summaries and action items
 - displays quality metrics and cost/time info
 - provides downloadable outputs
-

5.0 REFERENCES

1. Golia, L., & Kalita, J. (2023). Action-Item-Driven Summarization of Long Meeting Transcripts. *arXiv preprint arXiv:2312.17581*.
<https://arxiv.org/abs/2312.17581>
 2. Kirstein, F., Wahle, J. P., Ruas, T., & Gipp, B. (2024). Investigating Automatic Metrics on Meeting Summarization. *arXiv preprint arXiv:2404.11124*.
<https://arxiv.org/pdf/2404.11124.pdf>
 3. Lin, W. (2024). The Application of Real-time Emotion Recognition in Video Conferencing. *Journal of Computer Technology and Applied Mathematics*, 1(4), 79-88. <https://www.suaspress.org/ojs/index.php/JCTAM/article/view/v1n4a10>
 4. Singh, A., et al. (2023). Automated Minutes of Meeting Using a Multimodal Approach. *International Journal for Research in Applied Science and Engineering Technology*, 11(12).
<https://www.ijraset.com/research-paper/automated-minutes-of-meeting-using-a-multimodal-approach>
 5. Facial expression recognition using visible and IR by early fusion of deep learning with attention mechanism. (2025). *PeerJ Computer Science*, e-2676.
<https://peerj.com/articles/cs-2676/>
 6. Automatic student engagement recognition in online classrooms based on deep image segmentation and attention mechanisms. (2025). *Traitement du Signal*, 42(4). <https://www.iieta.org/journals/ts/paper/10.18280/ts.420428>
-