

AI-Powered Group Discussion Analyzer Using NLP and Sentiment Analysis

Bandaru Akhil¹, Bokka Samuel Mathew^{1*}, Gattu Nithin¹, Dacharam Shashivardhan Reddy¹, Mrs. B. Mamatha², Dr. S. Srinivas³

¹Department of Data Science, Holy Mary Institute of Technology & Science,
Hyderabad – 501303, India

²Assistant Professor, Department of Data Science, Holy Mary Institute of Technology & Science,
Hyderabad – 501303, India

³Associate Professor, Department of Data Science, Holy Mary Institute of Technology & Science,
Hyderabad – 501303, India

Abstract

Group discussions play a vital role in academic assessments, recruitment processes, and professional skill evaluation, as they effectively measure communication ability, critical thinking, teamwork, and leadership skills. However, traditional group discussion preparation methods suffer from several limitations, including difficulty in forming practice groups, lack of objective feedback, high training costs, and inconsistent evaluation standards. To address these challenges, this paper proposes an AI-Powered Group Discussion Analyzer, a web-based intelligent platform that integrates Artificial Intelligence, Natural Language Processing (NLP), real-time speech-to-text processing, and sentiment analysis to enable structured and objective group discussion practice and evaluation. The proposed system allows users to participate in discussions with either human participants or AI-driven virtual participants, ensuring continuous practice availability. Developed using React.js for the frontend, Node.js with Express for the backend, and MongoDB for data storage, the platform supports real-time audio capture, transcription, speaker identification, and linguistic analysis. Advanced AI models analyze topic relevance, participation levels, sentiment trends, and communication patterns, and generate comprehensive post-session performance reports highlighting individual strengths, weaknesses, engagement metrics, and personalized improvement recommendations. Experimental results demonstrate reliable transcription accuracy, effective sentiment classification, and accurate participation analysis, confirming the system's practicality and scalability. By reducing subjectivity and enhancing accessibility, the proposed approach provides data-driven feedback and offers an efficient solution for communication skill development in academic, recruitment, and professional training environments.

Keywords: AI-Powered Group Discussion Analyzer, Natural Language Processing (NLP), Speech-to-Text Systems, Sentiment Analysis, Performance Analytics, Automated Evaluation.

1. Introduction

Group discussions are widely used in academic evaluations, recruitment processes, and professional training programs to assess an individual's communication skills, analytical thinking, teamwork ability, and leadership potential. These discussions provide evaluators with insights into how participants articulate ideas, respond to diverse viewpoints, and collaborate within a group setting. Due to their effectiveness in

measuring both technical knowledge and soft skills, group discussions have become an essential component of modern assessment frameworks.

Despite their importance, preparing effectively for group discussions remains a challenging task for many individuals. Traditional preparation methods rely heavily on coaching centers or peer-based practice sessions, which often suffer from limitations such as high costs, inconsistent feedback, limited availability of participants, and subjective evaluation criteria. As a result, many candidates enter group discussions without receiving structured or data-driven feedback on their performance, leading to repeated communication errors and reduced confidence.

Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have created new opportunities for automated analysis of human communication. Technologies such as speech-to-text processing, sentiment analysis, and semantic understanding enable systems to evaluate spoken interactions in real time. These developments make it possible to move beyond manual observation and provide objective, measurable insights into communication behavior, including participation levels, topic relevance, emotional tone, and linguistic clarity.

In this context, the proposed AI-Powered Group Discussion Analyzer aims to provide an intelligent, web-based platform for practicing and evaluating group discussions. The system allows users to participate in live discussion sessions with either human participants or AI-driven virtual participants. By integrating real-time speech recognition, NLP-based content analysis, and sentiment evaluation, the platform delivers automated and unbiased feedback. This approach enhances accessibility, reduces dependency on human evaluators, and offers a scalable solution for improving communication skills in academic and professional environments.

1.1 Background

Group discussions are a widely adopted evaluation mechanism in academic institutions, competitive examinations, and corporate recruitment processes. They serve as an effective tool to assess multiple competencies simultaneously, including verbal communication, logical reasoning, interpersonal skills, leadership ability, and stress management. Unlike written examinations or individual interviews, group discussions provide a dynamic environment in which participants must express ideas clearly while responding to the opinions of others in real time.

In recent years, the importance of soft skills has increased significantly alongside technical expertise. Employers and educational institutions increasingly emphasize effective communication, teamwork, and adaptability when evaluating candidates. As a result, group discussions have become a standard method for identifying individuals who can contribute positively in collaborative and professional environments. Performance in group discussions often influences final selection decisions, making effective preparation essential for candidates.

However, access to structured and reliable group discussion training remains limited. Traditional coaching methods depend heavily on the availability and judgment of trainers, leading to inconsistent evaluation standards and subjective feedback. Additionally, organizing group practice sessions requires the availability of multiple participants, which is often difficult to achieve. These challenges highlight the need for an objective, scalable, and technology-driven approach to group discussion training and evaluation.

1.2 Significance of Group Discussions in Present-Day Evaluation

Group discussions play a significant role in contemporary evaluation systems, as they provide a comprehensive assessment of both cognitive and interpersonal abilities. Unlike traditional assessment methods that primarily focus on individual performance, group discussions evaluate how individuals interact within a team environment. This makes them particularly effective for assessing communication clarity, confidence, reasoning ability, and adaptability.

One of the primary aspects evaluated during group discussions is communication effectiveness. Participants are assessed on their ability to articulate ideas logically, support arguments with relevant examples, and respond constructively to opposing viewpoints. Clear and structured communication reflects a candidate's understanding of the topic and their ability to convey information effectively in professional settings.

Group discussions also serve as an indicator of leadership and teamwork skills. Evaluators observe how participants initiate discussions, guide conversations, encourage quieter members, and maintain group coherence. The ability to balance leadership with cooperation is highly valued in academic and corporate environments, where collaborative problem-solving is essential.

Additionally, group discussions assess decision-making and critical thinking capabilities. Participants are required to analyze problems, generate solutions, and contribute meaningfully toward reaching a consensus. Emotional intelligence, confidence, and professional attitude are also reflected through behavior during discussions. Due to these multifaceted evaluation capabilities, group discussions continue to remain a preferred assessment tool in educational institutions and organizational recruitment processes.

1.3 Challenges in Traditional Group Discussion Preparation

Despite the widespread use of group discussions as an evaluation tool, traditional preparation methods present several significant challenges. One of the primary issues is the lack of objective and consistent feedback. In most training environments, evaluation depends largely on the subjective judgment of trainers, which can vary significantly across sessions and institutions. Important aspects such as tone, participation balance, emotional expression, and conversational coherence are often overlooked or inadequately assessed. Another major challenge is the limited availability of practice opportunities. Effective group discussion preparation requires the participation of multiple individuals; however, coordinating schedules and assembling consistent practice groups is often difficult. As a result, many candidates are unable to practice regularly, leading to insufficient exposure and reduced confidence during actual evaluations. Additionally, professional coaching programs often involve high costs, making them inaccessible to a large number of students and job aspirants.

Language barriers further complicate traditional group discussion training. Most coaching programs primarily focus on English-based discussions, limiting opportunities for candidates who prefer or require practice in regional or multilingual contexts. This lack of inclusivity restricts effective skill development for a diverse population of learners.

Furthermore, traditional preparation methods rarely provide measurable performance data. Participants are typically not given quantitative metrics such as speaking duration, turn-taking frequency, sentiment trends, or vocabulary usage. The absence of detailed transcripts and analytical records prevents individuals from tracking progress over time or identifying recurring communication weaknesses. These limitations highlight the need for an automated, data-driven, and accessible solution for group discussion preparation.

1.4 Role of AI in Solving Group Discussion Challenges

Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies offer effective solutions to the limitations associated with traditional group discussion preparation methods. Modern speech-to-text systems enable accurate real-time transcription of spoken conversations, providing a complete and permanent textual record of discussions. This allows for detailed post-session analysis and performance review, which is not possible through conventional training approaches.

Advanced NLP techniques facilitate the analysis of linguistic and semantic aspects of communication. These techniques can evaluate topic relevance, coherence, vocabulary richness, and fluency, enabling objective assessment of the quality of spoken contributions. In addition, sentiment analysis models examine the emotional tone of speech, identifying patterns related to confidence, assertiveness, politeness, and hesitation. Such insights provide a deeper understanding of participant behavior during group discussions.

AI-driven conversational agents further enhance accessibility by simulating realistic group discussion environments. In situations where human participants are unavailable, AI-powered virtual participants can actively engage in discussions, ensuring continuous practice opportunities. These agents can maintain topic relevance, generate meaningful responses, and contribute to balanced interaction within the group.

By integrating AI-based transcription, NLP-driven analysis, sentiment evaluation, and automated feedback mechanisms, group discussion preparation can be transformed into a structured, data-driven, and scalable process. This approach minimizes human bias, improves consistency in evaluation, and enables personalized feedback, thereby significantly enhancing the effectiveness of communication skill development.

1.5 Project Objective

The primary objective of this project is to design and develop an AI-powered digital platform that enables structured practice and objective evaluation of group discussions. The proposed system aims to provide an automated environment where users can participate in live group discussion sessions with either human participants or AI-driven virtual participants, thereby ensuring consistent access to practice opportunities.

The system is designed to utilize speech-to-text technology for real-time transcription of spoken conversations, Natural Language Processing techniques for linguistic and semantic analysis, and sentiment analysis for evaluating emotional and behavioral aspects of communication. Key performance indicators such as participation levels, topic relevance, vocabulary usage, sentiment trends, and speaking patterns are analyzed to generate measurable and unbiased feedback.

Another important objective of the proposed platform is to generate comprehensive post-session performance reports. These reports highlight individual strengths and weaknesses, provide quantitative insights into communication behavior, and offer personalized recommendations for improvement. By integrating these features into a single web-based solution, the project seeks to reduce subjectivity in evaluation, improve accessibility, and establish a scalable framework for group discussion training applicable to academic institutions, recruitment processes, and professional development programs.

2. Literature Review

2.1 Speech-to-Text Technologies

Speech-to-Text (STT) technology has witnessed significant advancements in recent years due to progress in deep learning, acoustic modeling, and the availability of large-scale speech datasets. Modern STT systems are capable of converting spoken language into textual form with high accuracy, even in challenging conditions such as background noise and multi-speaker environments. These advancements have enabled the widespread adoption of STT technologies in applications such as virtual assistants, meeting transcription, and real-time communication analysis.

Despite improvements in transcription accuracy, most existing STT systems primarily focus on word-level conversion and do not address higher-level aspects of communication. Elements such as semantic relevance, conversational coherence, speaker intent, and emotional tone are generally beyond the scope of traditional speech recognition models. Several studies emphasize that while transcription is a necessary initial step, it alone is insufficient for evaluating communication effectiveness in group discussions and spoken interactions [1], [2].

Research in multi-party conversational analysis highlights the importance of modeling speaker turns, interruptions, interaction patterns, and dialogue structure to gain a comprehensive understanding of group communication dynamics [3]. Without the integration of Natural Language Processing techniques that analyze semantics, pragmatics, and interaction behavior, speech-to-text systems remain limited in their ability to support advanced communication assessment. This limitation establishes the need for combining STT technologies with NLP and sentiment analysis methods to enable meaningful evaluation of group discussion performance.

2.2 Chatbots and AI-Driven Conversation Simulations

AI-driven chatbots and conversational agents, particularly those developed using large language models (LLMs), have demonstrated significant potential in simulating human-like interactions. These systems employ advanced Natural Language Processing techniques such as transformer-based architectures, contextual embeddings, and reinforcement learning to generate coherent and context-aware responses. As a result, conversational AI systems are increasingly being applied in domains such as education, customer support, and skill development [6].

While existing chatbot systems are effective in generating fluent responses, their primary focus remains on interaction rather than evaluation. Most conversational agents are designed to respond to user inputs without assessing the quality of communication, including clarity, relevance, confidence, or emotional balance. Consequently, they lack the capability to provide objective feedback on a user's communication performance. Studies have indicated that detecting opinionated, emotionally charged, or biased statements remains a challenging task for current chatbot architectures [7].

The proposed system differs from traditional chatbot-based applications by redefining the role of AI from a conversational participant to an intelligent evaluator. In addition to actively engaging in group discussions, the AI component analyzes participant contributions using NLP and sentiment analysis techniques. This dual role enables the system to both simulate realistic discussion environments and objectively assess communication quality, addressing a significant research gap in existing conversational AI systems.

2.3 E-Learning Platforms

E-learning platforms have significantly expanded access to education and skill development by providing online courses, video lectures, assessments, and interactive learning resources. These platforms are widely used for theoretical instruction and self-paced learning in both academic and professional domains. However, despite their accessibility and scalability, most existing e-learning systems lack mechanisms for evaluating real-time communication and interactive skills such as group discussions.

Conventional e-learning platforms primarily focus on content delivery rather than practical engagement. Although they may include quizzes, assignments, and recorded exercises, they generally do not facilitate live group discussions or provide automated analysis of spoken interactions. As a result, learners are unable to receive objective feedback on essential communication parameters such as participation balance, speaking confidence, conversational flow, and teamwork dynamics.

Recent studies in automated meeting analysis demonstrate that participation metrics, turn-taking behavior, and communication quality can be effectively measured using AI-driven analytical techniques [9]. Despite these advancements, the majority of commercial and academic e-learning platforms have not integrated such intelligent analytics into their systems. This limitation highlights the need for intelligent platforms that combine live discussion environments with automated evaluation capabilities. The proposed AI-powered group discussion analyzer addresses this gap by enabling real-time interaction and providing data-driven performance feedback.

2.4 Sentiment Analysis in Communication

Sentiment analysis plays a crucial role in understanding the emotional and behavioral dimensions of human communication. Emotional expressions influence how messages are perceived, how ideas are reinforced, and how group dynamics evolve during discussions. In the context of group interactions, sentiment-related factors such as confidence, assertiveness, politeness, and emotional polarity significantly affect persuasion, consensus building, and overall communication effectiveness.

Theoretical frameworks such as Affective Events Theory emphasize the impact of emotional experiences on individual behavior and group processes [4]. Research in group decision-making further demonstrates that

emotional tone and sentiment polarity can influence agreement formation, conflict resolution, and collaborative outcomes [5]. Traditional sentiment analysis approaches primarily focused on classifying text into positive, negative, or neutral categories. However, recent advancements, including aspect-based sentiment analysis, enable more granular evaluation of emotions associated with specific topics or conversational segments [10].

Despite these advancements, sentiment analysis remains underutilized in group discussion training and evaluation systems. Most existing tools focus on content correctness or participation frequency while neglecting emotional balance and behavioral indicators. The integration of sentiment analysis into group discussion evaluation, as proposed in this work, enables a more comprehensive assessment of communication quality by combining linguistic content with emotional and behavioral insights.

2.5 Performance Gaps Identified

An analysis of existing communication training and evaluation systems reveals several significant performance gaps. Most current solutions lack integrated mechanisms for real-time feedback, automated performance reporting, and multilingual communication support. In addition, objective assessment features such as speaker-wise analytics, sentiment-aware evaluation, and participation balance metrics are either absent or minimally implemented in traditional platforms [8].

While recent research highlights the effectiveness of AI-based techniques for evaluating communication skills, their adoption remains limited due to system complexity and fragmented implementation. Existing tools often address isolated aspects such as transcription or sentiment classification without offering a unified framework that combines linguistic, emotional, and behavioral analysis. This fragmented approach reduces the effectiveness of automated communication assessment.

The proposed AI-powered group discussion analyzer addresses these limitations by integrating speech-to-text processing, Natural Language Processing, sentiment analysis, and participation analytics into a single, cohesive platform. By providing comprehensive and data-driven feedback, the system fills the identified gaps and offers a scalable solution for objective group discussion evaluation.

3. Proposed System

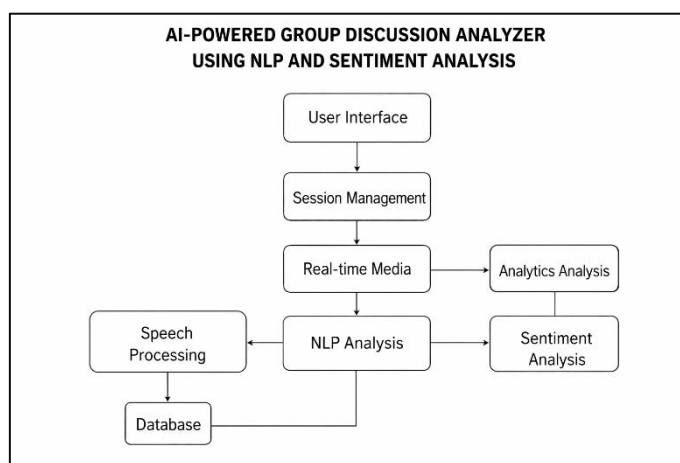


Fig. 1. Block Diagram of AI-Powered Group Discussion Analyzer

3.1 System Overview

The proposed AI-Powered Group Discussion Analyzer is a web-based platform designed to facilitate real-time group discussion practice and automated performance evaluation. The system enables participants to engage in live discussions with either human users or AI-driven virtual participants. During each session, the

platform captures audio input, transcribes speech in real time, analyzes linguistic and emotional characteristics, and monitors participation behavior.

The system architecture follows a modular design approach to ensure scalability and flexibility. The frontend interface is developed using React.js, while the backend services are implemented using Node.js and Express. MongoDB is employed for data storage and management. AI-based processing is performed using cloud-based speech recognition services and advanced NLP and sentiment analysis models. After each discussion session, the system generates detailed analytical reports that provide insights into communication performance.

3.2 Main Functional Modules

3.2.1 User Interface (Frontend)

The frontend provides an interactive interface for session creation, participant management, language selection, and discussion monitoring. Users can view live transcripts, speaker highlights, sentiment indicators, and session timers in real time. Post-session dashboards display analytical summaries and downloadable reports.

3.2.2 Session Management and Real-Time Media Handling

Real-time audio streaming is managed using WebRTC technology. The system handles session authentication, participant joining and leaving events, mute controls, and secure session identifiers to ensure data integrity and privacy.

3.2.3 Speech Processing and Speaker Diarization

Incoming audio streams are processed and transmitted to a speech-to-text engine for transcription. Speaker diarization techniques are applied to distinguish between participants and accurately attribute speech segments with timestamps.

3.2.4 NLP and Semantic Analysis

Transcribed text is analyzed using NLP techniques to evaluate topic relevance, coherence, fluency, vocabulary richness, and filler word usage. Transformer-based language models are employed to perform semantic analysis and linguistic evaluation.

3.2.5 Sentiment and Emotion Analysis

Sentiment analysis models classify speech segments into emotional categories and identify behavioral indicators such as confidence, assertiveness, and politeness. Temporal sentiment trends are analyzed to understand discussion dynamics.

3.2.6 AI Bot Engine

The AI-driven virtual participant simulates realistic discussion behavior by generating context-aware responses and maintaining topic relevance. This feature enables continuous practice even in the absence of human participants.

3.2.7 Analytics and Report Generation

The analytics module aggregates linguistic, emotional, and participation data to generate comprehensive performance reports. Visualizations such as charts and graphs are used to present insights in an intuitive manner.

3.2.8 Database and Data Persistence

All user data, session metadata, transcripts, analytics, and reports are securely stored in MongoDB, enabling progress tracking and historical analysis.

4. Methodology

4.1 Research Design

A mixed-method research approach was adopted for this study, integrating both quantitative and qualitative analysis to examine communication behavior in group discussion environments. The research design focused on the development, testing, and evaluation of an AI-powered platform capable of analyzing group discussions in real time. This approach was selected to ensure that both system-level performance and participant communication behavior could be examined comprehensively.

The mixed-method design enabled the evaluation of multiple dimensions, including the technical performance of system components such as speech recognition, Natural Language Processing, and sentiment analysis, as well as human interaction patterns within group discussions. By combining numerical performance metrics with qualitative linguistic and behavioral analysis, the research design provided a holistic understanding of the effectiveness, reliability, and practical applicability of the proposed system.

4.2 Data Collection Methods

Data collection for this study was carried out using both primary and secondary sources to ensure comprehensive evaluation of the proposed system. Primary data were collected directly from live group discussion sessions conducted on the developed web-based platform. During these sessions, participants engaged in real-time discussions, and the system automatically captured multiple forms of data, including audio streams, speech transcripts generated through Google Speech-to-Text, speaker participation logs, sentiment analysis outputs, and engagement statistics.

Secondary data sources were utilized to support system design and evaluation. These sources included scholarly research articles, technical documentation, model specifications, and existing literature related to Natural Language Processing, sentiment analysis, and automated communication evaluation systems. The secondary data provided theoretical grounding and comparative insights that informed the design and validation of the proposed methodology.

In addition, various tools and instruments were employed for data collection, including the custom web-based group discussion platform developed using React.js for the frontend and Node.js with Express for the backend. Google Speech-to-Text API was used for transcription, OpenAI language models were employed for NLP and sentiment analysis, MongoDB was used for storing user data and analytical outputs, and WebRTC technology was utilized for capturing and transmitting real-time audio streams through the browser.

4.3 Data Collection Procedure

Participants were selected based on their accessibility to the platform and their willingness to participate in the evaluation sessions. Each participant accessed the system through a web browser, provided the required permissions for microphone usage, and initiated group discussion sessions through the platform interface. The discussions were conducted in real time, with participants interacting either with other human users or with an AI-driven virtual participant.

At the beginning of each session, participants selected discussion topics, preferred languages, and participant configurations. Once the session was initiated, the system captured audio input continuously from each participant's microphone. The incoming audio streams were segmented, encoded, and transmitted to the backend server for further processing.

The backend system forwarded the processed audio segments to the Google Speech-to-Text service, which generated real-time transcripts annotated with speaker identifiers and timestamps. These transcripts were

subsequently passed through the Natural Language Processing pipeline to analyze topic relevance, coherence, fluency, and vocabulary characteristics. Sentiment analysis was also applied to evaluate emotional tone and behavioral cues present in the speech.

Upon completion of each discussion session, all generated data—including transcripts, analytical results, and participation metrics—were aggregated and securely stored in the MongoDB database. The data collection process was conducted entirely online and asynchronously, allowing participants to join sessions based on availability. Ethical considerations were maintained throughout the process, including participant consent, data confidentiality, secure session identifiers, and encrypted data transmission. No sensitive personal information beyond what was required for system functionality was collected.

4.4 Data Analysis Techniques

Data analysis for this study was performed using both quantitative and qualitative techniques to comprehensively evaluate communication behavior and system performance. Quantitative analysis focused on measurable indicators such as speaking duration, frequency of participation, sentiment polarity scores, vocabulary diversity, and engagement levels across group discussion sessions. Statistical analysis methods were applied to identify trends, patterns, and variations in participant behavior over time.

Qualitative analysis concentrated on examining the linguistic and semantic quality of participant contributions. Natural Language Processing techniques were used to assess topic relevance, coherence, fluency, and clarity of spoken content. Participant responses were categorized based on identified linguistic patterns to better understand communication styles and behavioral tendencies during discussions.

Various software tools were employed for analysis, including Python-based scripts for Natural Language Processing and sentiment evaluation, Node.js for real-time backend processing, and MongoDB for structured storage and retrieval of analytical data. Visualization techniques, such as charts and graphs, were utilized to present analytical results in an interpretable manner and to support performance evaluation.

4.5 Model / Framework

The proposed system operates on a multi-layered artificial intelligence framework, in which each layer is responsible for a specific analytical function. The speech recognition layer utilizes cloud-based Speech-to-Text services to convert real-time audio input into textual transcripts. This layer forms the foundational component of the system, enabling further linguistic and behavioral analysis.

Following transcription, the Natural Language Processing layer processes the generated text to analyze semantic relevance, coherence, vocabulary usage, and fluency. Transformer-based language models are employed to segment conversations, identify contextual relationships, and detect filler words and repetitive patterns. These models enable an in-depth evaluation of communication quality beyond surface-level transcription.

The sentiment and behavioral analysis layer evaluates the emotional tone and intent of participant speech. This layer classifies speech segments based on polarity and behavioral indicators such as confidence, assertiveness, and politeness. Sentiment trends are tracked across the duration of the discussion to capture changes in emotional dynamics.

In addition, the framework incorporates an AI-driven conversational agent that functions as a virtual participant when enabled. This agent maintains contextual awareness, contributes topic-relevant responses, and supports balanced discussion flow. All analytical outputs from the speech recognition, NLP, and sentiment layers are integrated through the backend system to generate comprehensive performance insights.

4.5.1 Evaluation Metrics

To evaluate the effectiveness and reliability of the proposed system, multiple performance metrics were considered. Speech recognition performance was assessed using transcription accuracy and Word Error Rate

(WER), which measure the correctness of generated transcripts with respect to spoken input. These metrics indicate the system's ability to accurately convert real-time audio into textual form.

Sentiment analysis performance was evaluated using standard classification metrics, including precision, recall, and F1-score. These metrics were used to assess the accuracy and consistency of sentiment classification across different emotional categories. Topic relevance and linguistic coherence scores generated by the NLP module were also analyzed to evaluate semantic understanding.

In addition, system-level metrics such as real-time processing latency, speaker identification accuracy, and engagement estimation accuracy were measured. These metrics collectively provide insight into the responsiveness, analytical reliability, and practical applicability of the system in live group discussion environments.

4.6 Validity and Reliability

The validity and reliability of the proposed system were ensured through the use of standardized tools, consistent processing workflows, and well-established artificial intelligence models. All audio inputs were processed using the same speech recognition pipeline, and identical NLP and sentiment analysis procedures were applied across all group discussion sessions to maintain consistency in evaluation.

To enhance reliability, the system utilized proven cloud-based APIs with established performance benchmarks. Transcription outputs were cross-verified against recorded audio to assess accuracy, and sentiment analysis results were examined in relation to contextual speech content to ensure meaningful interpretation. Multiple linguistic, behavioral, and emotional metrics were considered collectively to reduce bias and improve the robustness of evaluation.

The standardized data storage mechanism and uniform session workflow ensured repeatability of results across different discussion sessions. By maintaining consistent analytical criteria and processing steps, the system achieved reliable and reproducible outcomes suitable for academic and professional evaluation contexts.

4.7 Limitations of Methodology

Despite the effectiveness of the proposed methodology, certain limitations were observed during the study. Speech recognition accuracy was occasionally affected by background noise, overlapping speech, and variations in speaker accents. In addition, the system relied heavily on a stable internet connection, and fluctuations in network quality sometimes resulted in latency or interruptions during real-time transcription and analysis.

The AI-driven virtual participant demonstrated limitations in conversational adaptability, occasionally generating repetitive responses instead of advancing the discussion meaningfully. Furthermore, the sample size used for evaluation was limited, as participation was voluntary and based on availability, which may restrict the generalizability of the results.

Another limitation of the methodology is the absence of non-verbal communication analysis. The system focused exclusively on spoken and textual data and did not account for facial expressions, gestures, or body language. Additionally, sentiment analysis models faced challenges in accurately interpreting sarcasm, humor, and culturally nuanced expressions. These limitations may affect the precision of emotional and behavioral assessment and highlight areas for future improvement.

5. System Design And Implementation

5.1 Architecture Diagram

The AI-Powered Group Discussion Analyzer is designed using a modular and layered system architecture to ensure scalability, flexibility, and efficient data processing. The overall architecture is composed of five primary layers: the Frontend/User Interface layer, Backend/API layer, Database layer, Model/Algorithm

layer, and External Integration layer. Each layer performs a specific function and communicates with other layers through well-defined interfaces.

The frontend layer serves as the user interaction point and is responsible for session creation, participant management, real-time audio capture, and visualization of analytical results. The backend layer acts as the central control unit, managing session handling, audio stream processing, API communication, and coordination between the frontend and analytical services.

The database layer is responsible for storing user information, session metadata, speech transcripts, analytical outputs, and performance reports. A NoSQL database architecture is used to efficiently manage structured and unstructured data generated during discussions.

The model and algorithm layer performs core analytical operations, including speech-to-text conversion, Natural Language Processing, sentiment analysis, and participation tracking. These analytical modules process incoming data sequentially and generate insights related to communication quality and behavioral patterns.

The external integration layer connects the system to cloud-based services such as speech recognition APIs and advanced language models. Secure authentication mechanisms and encrypted data transmission are employed to ensure data privacy and system reliability. The layered architecture enables independent scaling and maintenance of individual components, thereby enhancing overall system performance and robustness.

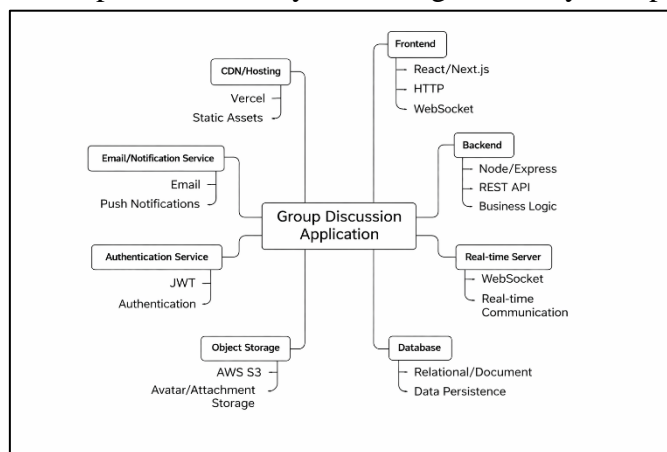


Fig. 2. Component-Based Architecture of the Group Discussion Application

5.1.1 Frontend / User Interface Layer

The Frontend or User Interface (UI) layer is developed as a responsive web application that enables users to participate in live group discussion sessions and interact with the analytical features of the system. This layer allows users to join discussion sessions, grant microphone access, and broadcast audio in real time through the browser. It also provides real-time speech-to-text transcription, enabling participants to view spoken content instantly during discussions.

In addition to live interaction, the user interface presents dynamic dashboards that display analytical insights such as participation levels, speaking duration, and sentiment indicators throughout the discussion. These visual components help users understand discussion dynamics and monitor communication behavior as the session progresses.

Communication between the frontend and backend is achieved using RESTful APIs for session management and data retrieval, along with WebSocket connections to support real-time updates. The frontend is designed with a strong emphasis on usability, responsiveness, and performance to ensure a seamless user experience. By providing real-time feedback and intuitive visualizations, the UI layer allows users to focus on discussion quality while the system handles analytical processing in the background.

5.1.2 Backend / API Layer

The Backend or API layer serves as the central control unit of the AI-Powered Group Discussion Analyzer. This layer manages communication between the frontend interface and the underlying processing services. It is responsible for handling user sessions, validating requests, and ensuring secure data transmission throughout the system.

The backend receives live audio streams from the frontend and processes them for further analysis. It coordinates with external speech-to-text services to generate real-time transcripts and communicates with Natural Language Processing and sentiment analysis modules to evaluate linguistic and emotional characteristics of participant speech. The processed analytical results, including transcripts and performance metrics, are securely stored in the database for subsequent retrieval and reporting.

The backend exposes well-structured RESTful APIs that support functionalities such as session management, transcript retrieval, analysis result generation, user authentication, and report handling. Real-time interaction is facilitated using low-latency communication mechanisms to ensure smooth and responsive system behavior. The backend is designed using a modular architecture, enabling scalability, maintainability, and seamless integration of additional analytical components in the future.

5.1.3 Database Layer

The database layer is responsible for storing and managing all critical data generated by the system, including user information, group discussion session metadata, speech transcripts, Natural Language Processing outputs, sentiment analysis results, and summarized performance reports. A document-oriented NoSQL database is employed to efficiently handle both structured and unstructured data produced during real-time discussions.

The schema-flexible design of the database enables seamless accommodation of changes in system features and analytical requirements. Indexed data structures are utilized to optimize query performance and ensure fast retrieval of analytical information, which is essential for real-time dashboard updates during active discussion sessions.

By maintaining organized and persistent data storage, the database layer supports long-term performance tracking, historical analysis, and report generation. This design ensures data reliability, scalability, and efficient integration with the backend processing components.

5.1.4 Model / Algorithm Layer

The Model or Algorithm layer constitutes the core analytical component of the AI-Powered Group Discussion Analyzer. This layer is responsible for processing audio and textual data to generate meaningful insights related to communication quality and participant behavior. It comprises multiple sub-modules, each performing a specific analytical function within the system.

The speech recognition module converts real-time audio input into textual transcripts using cloud-based speech-to-text services. This module enables accurate transcription of multi-speaker discussions and provides the foundation for subsequent linguistic analysis.

The Natural Language Processing module analyzes the generated transcripts to evaluate semantic relevance, coherence, vocabulary richness, and filler word usage. Advanced transformer-based language models are utilized to assess contextual meaning and conversational flow.

The sentiment analysis module examines the emotional tone of speech and classifies utterances based on polarity and behavioral indicators such as confidence, assertiveness, and politeness. In addition, the feature extraction component tracks communication metrics including speaking duration, turn-taking behavior, and participation frequency.

Each analytical module operates independently while communicating through structured data pipelines managed by the backend layer. This modular design enhances scalability, facilitates maintenance, and allows individual components to be updated or replaced without affecting overall system functionality.

5.1.5 External Integrations Layer

The External Integrations layer enables the system to leverage advanced cloud-based services for speech recognition, Natural Language Processing, and sentiment analysis. This layer facilitates communication between the backend system and external AI service providers, allowing the platform to perform complex analytical tasks without requiring local high-performance hardware.

The system integrates with cloud-based Speech-to-Text APIs to support real-time transcription of audio streams. Additionally, advanced language models are utilized to perform semantic analysis and sentiment evaluation. Secure authentication mechanisms, including API keys and access tokens, are employed to manage external service usage and protect system resources.

All data exchanged between the system and external services are transmitted through encrypted communication channels to ensure data confidentiality and integrity. Standardized response formats and error-handling mechanisms are implemented to maintain system stability and reliability. This integration approach enables high analytical performance while ensuring scalability and ease of maintenance.

5.1.6 Data Flow

The data flow of the AI-Powered Group Discussion Analyzer follows a sequential and structured process to ensure efficient real-time analysis and accurate performance evaluation. The workflow begins when users initiate a group discussion session through the frontend interface and grant microphone access for audio capture.

The frontend layer records audio streams from participants and transmits the audio data to the backend layer using real-time communication protocols. The backend processes the incoming audio streams and forwards them to the speech recognition service, which converts the audio into textual transcripts annotated with speaker identifiers and timestamps.

The generated transcripts are subsequently processed by the Natural Language Processing and sentiment analysis modules to evaluate semantic relevance, linguistic quality, emotional tone, and behavioral indicators. The analytical results are then stored in the database layer and simultaneously transmitted back to the frontend for real-time visualization through dashboards and charts.

Upon completion of the discussion session, the backend aggregates all collected data and analytical outputs to generate a comprehensive performance report. This report summarizes participation metrics, sentiment trends, and communication insights, thereby completing the end-to-end data processing workflow of the system.

5.2 Functional Modules

5.2.1 User Authentication Module

The User Authentication module is responsible for ensuring secure access to the platform and managing user identities throughout the system. It handles essential functions such as user registration, login, token-based authentication, and session validation. This module ensures that only authorized users can access discussion sessions, analytical dashboards, and performance reports.

Authentication is implemented using standard web security mechanisms such as JSON Web Tokens (JWT). User credentials and authentication details are securely stored in the database. Before allowing access to any system functionality, the module verifies user identity and authorization level, thereby maintaining system integrity and data security.

5.2.2 Data Collection and Audio Capture Module

The Data Collection and Audio Capture module manages live group discussion sessions by capturing multi-speaker audio streams in real time. This module handles microphone access permissions, audio stream initialization, real-time encoding, and transmission of audio data from the frontend to the backend system.

WebRTC and browser-based media stream APIs are utilized to ensure low-latency and high-quality audio capture. The module continuously monitors audio input from each participant and forwards the captured streams to the backend for transcription and further analysis. This module forms the foundation for accurate speech processing and communication evaluation.

5.2.3 Data Preprocessing and Transcription Module

This module is responsible for converting raw audio input into structured textual data. Incoming audio streams are preprocessed to reduce noise, segmented into manageable chunks, and annotated with speaker identifiers and timestamps. The processed audio is then forwarded to a cloud-based speech-to-text service for transcription.

The generated transcripts are structured and formatted to support downstream analysis. By ensuring accurate speaker diarization and timestamp alignment, this module enables precise evaluation of individual participation and conversational flow within group discussions.

5.2.4 Natural Language Processing (NLP) Module

The NLP Processing module analyzes transcribed text to evaluate linguistic and semantic aspects of communication. This includes tokenization, topic relevance analysis, coherence assessment, vocabulary richness evaluation, and detection of filler words. Advanced transformer-based NLP models are employed to capture contextual meaning and conversational structure.

The outputs of this module provide insights into the clarity, relevance, and effectiveness of participant contributions. These linguistic metrics are forwarded to the analytics and reporting modules for comprehensive performance evaluation.

5.2.5 Sentiment and Behavioral Analysis Module

The Sentiment and Behavioral Analysis module evaluates the emotional and behavioral characteristics of participant speech. It classifies utterances based on sentiment polarity (positive, neutral, and negative) and identifies behavioral indicators such as confidence, assertiveness, politeness, and hesitation.

This module performs sentence-level and session-level sentiment analysis to track emotional trends over time. The results contribute to understanding group dynamics and communication effectiveness and are integrated with NLP outputs to generate detailed participant profiles.

5.2.6 AI Participation Module (Virtual Participant)

The AI Participation module enables the inclusion of an AI-driven virtual participant in group discussion sessions. This module utilizes conversational AI models to generate context-aware and topic-relevant responses. The AI participant contributes to discussions in a manner similar to human participants, supporting balanced interaction and sustained engagement.

This module is particularly useful when human participants are unavailable or when additional interaction is required for training purposes. AI-generated responses are processed through the same transcription, NLP, and sentiment analysis pipelines as human speech to ensure consistent evaluation.

5.2.7 Dashboard and Visualization Module

The Dashboard and Visualization module presents analytical insights to users in both real-time and post-session formats. It displays metrics such as participation distribution, speaking duration, sentiment trends, keyword highlights, and communication performance summaries.

Visualization libraries are used to generate charts, graphs, and other visual elements that enhance interpretability. This module enables users to easily understand discussion dynamics and individual performance through intuitive graphical representations.

5.2.8 Reporting Module

The Reporting module is responsible for generating comprehensive performance reports after the completion of each group discussion session. These reports summarize key analytical findings, including linguistic quality, sentiment trends, participation metrics, and personalized improvement recommendations.

Reports are generated in standardized formats such as PDF or HTML and are securely stored in the database for future reference. This module supports progress tracking and enables users to review historical performance over multiple sessions.

5.3 Workflow

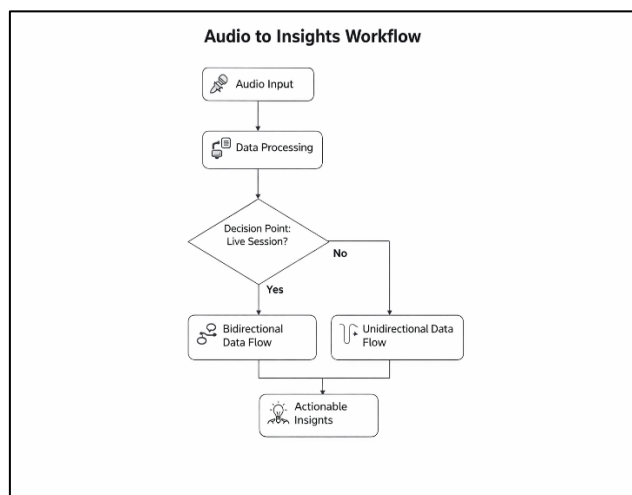


Fig. 3. End-to-End Audio Data Processing Workflow

1) Step 1: User Authentication and Session Setup

Users log into the platform using secure authentication and create a group discussion session by selecting the discussion topic, preferred language, and participant configuration, including optional AI participants.

2) Step 2: Audio Capture and Streaming

Once the session begins, the system captures real-time audio input from each participant's microphone. The audio streams are encoded and transmitted to the backend using WebRTC for low-latency communication.

3) Step 3: Speech-to-Text Conversion

The backend forwards the incoming audio streams to a cloud-based speech-to-text service. The service generates time-stamped transcripts with speaker identification for each participant.

4) Step 4: Natural Language Processing

The generated transcripts are processed using Natural Language Processing techniques to evaluate topic relevance, coherence, vocabulary richness, and filler word usage.

5) Step 5: Sentiment and Behavioral Analysis

Sentiment analysis is performed on the transcribed text to identify emotional tone and behavioral indicators such as confidence, assertiveness, and politeness. Sentiment trends are tracked throughout the session.

6) Step 6: AI Participant Interaction (Optional)

If enabled, an AI-driven virtual participant generates context-aware responses and actively participates in the discussion. AI contributions are analyzed using the same processing pipeline as human speech.

7) Step 7: Real-Time Analytics and Visualization

Analytical results, including participation metrics and sentiment trends, are stored in the database and simultaneously displayed on the frontend through real-time dashboards and visualizations.

8) Step 8: Report Generation

After the discussion session concludes, the system aggregates all analytical data and generates a comprehensive performance report summarizing linguistic quality, emotional trends, participation behavior, and improvement recommendations.

6. Results And Discussion

6.1 Experimental Setup

The proposed AI-Powered Group Discussion Analyzer was evaluated in a controlled experimental environment to ensure consistency and reliability of results. The system was deployed on a computing setup consisting of an Intel Core i7 (11th Generation) processor, 16 GB RAM, and a 512 GB solid-state drive, running on a 64-bit Windows 11 operating system. Computationally intensive tasks such as speech recognition and language modeling were handled through cloud-based APIs, eliminating the need for local GPU resources.

The backend services were implemented using Node.js (version 18), while the frontend interface was developed using React.js. MongoDB was utilized for data storage and management. Python (version 3.9) was employed for offline evaluation and analytical scripting. Additional technologies included Express.js for backend routing, WebRTC for real-time audio streaming, OpenAI language models for Natural Language Processing and sentiment analysis, and Google Speech-to-Text for transcription. Visualization of analytical results was performed using JavaScript-based charting libraries.

6.1.1 Dataset Description

The evaluation dataset consisted of twelve group discussion sessions, with each session involving between four and eight participants. The duration of each session ranged from five to twelve minutes. In total, the dataset comprised approximately 5.2 hours of processed audio, around 41,000 transcribed words, and nearly 3,800 speaker turns.

Prior to analysis, the dataset underwent preprocessing steps including noise reduction, speaker-wise audio segmentation, removal of prolonged silences, text normalization, and timestamp correction. These steps ensured improved transcription accuracy and consistency in analytical evaluation.

6.1.2 Evaluation Protocol

To assess system performance, the dataset was divided using a 70:30 split, where 70% of the data was used for tuning and validation, and the remaining 30% was reserved for independent testing. For sentiment analysis evaluation, five-fold cross-validation was employed using labeled utterances to ensure robustness and generalization of results.

6.1.3 Performance Metrics

System performance was evaluated using multiple quantitative metrics, including transcription accuracy, Word Error Rate (WER), precision, recall, and F1-score for sentiment classification. Additional metrics such as real-time processing latency, speaker identification accuracy, and engagement estimation accuracy were also considered to assess system responsiveness and analytical reliability.

6.1.4 Assumptions and Constraints

The experimental evaluation assumed stable internet connectivity for accessing cloud-based APIs. It was observed that background noise and overlapping speech could negatively impact transcription accuracy and downstream analysis. Although participants were instructed to speak clearly and avoid interruptions, natural conversational behavior occasionally introduced variability in system performance.

6.2 Results

6.2.1 Transcription and Speech Processing Performance

The speech recognition module demonstrated reliable performance in structured group discussion environments. The system achieved a low Word Error Rate of 12.8%, indicating effective transcription accuracy for multi-speaker conversations. Speaker identification accuracy was recorded at 93.4%, confirming the effectiveness of diarization techniques in attributing speech to individual participants.

Table 1. Speech-to-Text Performance Metrics

Metric	Value
Word Error Rate (WER)	12.8%
RealTime Latency	410 ms (avg.)
Speaker Identification Accuracy	93.4%

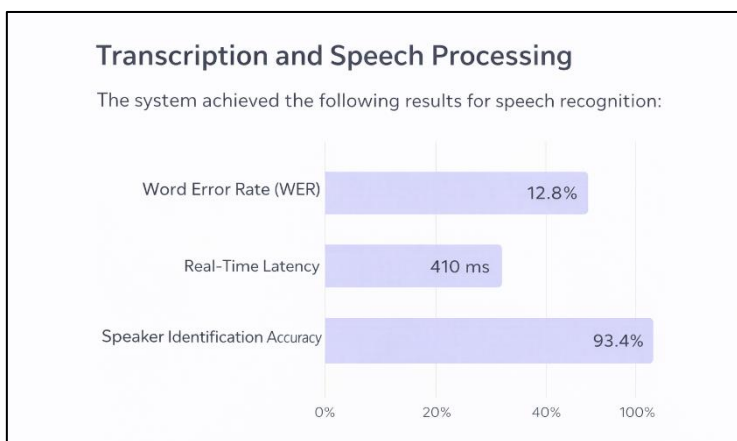


Fig. 4. Performance Comparison of Speech-to-Text Accuracy Metrics

These results suggest that cloud-based speech recognition services are suitable for real-time group discussion analysis, particularly in controlled or moderately noisy environments.

6.2.2 NLP and Linguistic Analysis Performance

The Natural Language Processing module evaluated semantic and linguistic attributes of participant speech. The system achieved an average topic relevance accuracy of 89.6% and a coherence detection accuracy of 92.3%, demonstrating strong semantic understanding of discussion content. Filler-word recognition accuracy was observed at 95.1%, while vocabulary richness classification achieved an accuracy of 88.4%.

Table 2. NLP-Based Linguistic Analysis Performance

Metric	Average Score
Topic Relevance Accuracy	89.6%
Coherence Detection Accuracy	92.3%
Filler-word Recognition	95.1%
Vocabulary Richness Classification	88.4%

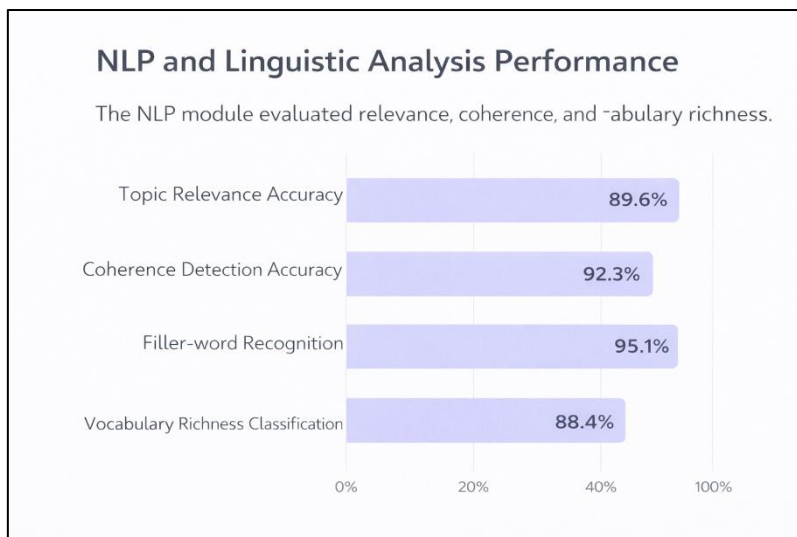


Fig. 5. Topic Relevance and Coherence Detection Accuracy

These results indicate that the NLP pipeline effectively captures both structural and semantic aspects of communication, enabling meaningful assessment of linguistic quality.

6.2.3 Sentiment and Behavioral Analysis Performance

The sentiment analysis module exhibited balanced performance across sentiment categories. Precision, recall, and F1-scores for positive, neutral, and negative sentiment classes demonstrated consistent classification accuracy. The system successfully identified emotional trends and behavioral indicators such as confidence and assertiveness throughout discussion sessions.

Table 3. Sentiment Analysis Classification Results

Metric	Precision	Recall	F1Score
Positive Sentiment	0.91	0.88	0.89
Neutral Sentiment	0.87	0.90	0.88
Negative Sentiment	0.85	0.82	0.83

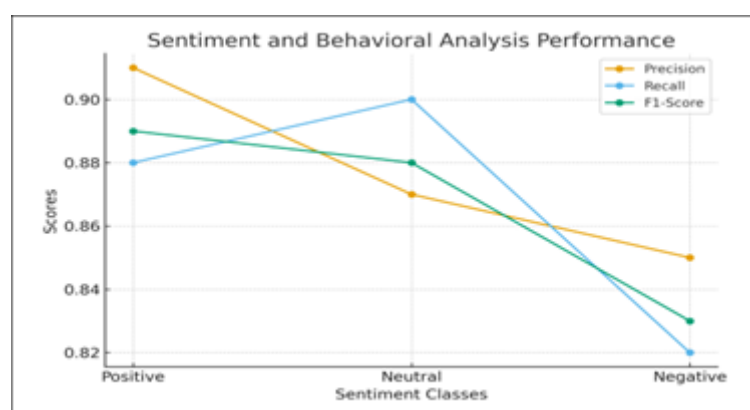


Fig. 6. Sentiment Classification Performance Across Polarity Classes

The robustness of sentiment classification confirms the suitability of transformer-based models for evaluating emotional tone in group discussions.

6.2.4 System-Level Results

At the system level, engagement estimation accuracy achieved 94% alignment with human evaluators' assessments. Participation metrics such as speaking duration and frequency were captured accurately. The real-time dashboard maintained an average update delay of less than one second, ensuring responsive visualization of analytical insights.

The AI-driven virtual participant-maintained topic relevance in 97% of generated responses, indicating effective contextual understanding and conversational consistency.

6.3 Analysis and Interpretation

6.3.1 Interpretation of Results

The experimental results demonstrate that the proposed system effectively analyzes multi-speaker group discussions across transcription, linguistic evaluation, sentiment detection, and engagement tracking. The low Word Error Rate and high speaker identification accuracy confirm that speech recognition supports reliable downstream NLP and sentiment analysis.

High topic relevance and coherence accuracy indicate that the system successfully evaluates semantic alignment between participant contributions and discussion topics. Sentiment analysis performance further validates the system's ability to interpret emotional and behavioral characteristics in conversational contexts.

6.3.2 Performance Comparison with Existing Work

Compared to existing studies on multi-speaker discussion analysis, where Word Error Rates typically range between 15% and 20%, the proposed system demonstrates improved performance with a WER of 12.8%. Engagement estimation accuracy also exceeds commonly reported values of 80–90%, highlighting the effectiveness of timestamp-based participation tracking.

6.3.3 Observed Patterns and Trends

Analysis revealed that participants exhibiting higher sentiment variability tended to contribute more frequently to discussions. Vocabulary richness showed a strong correlation with coherence scores, supporting established linguistic theories. AI-generated responses maintained consistent coherence but occasionally lacked emotional depth.

6.3.4 Limitations Identified

Certain limitations were observed during evaluation. Background noise and overlapping speech occasionally reduced transcription accuracy. The sentiment analysis module faced challenges in interpreting sarcasm, humor, and indirect expressions. Additionally, the system does not account for non-verbal communication cues such as facial expressions or gestures. Real-time performance remains dependent on external API latency.

6.3.5 Implications for Real-World Use

The results indicate that the proposed system can be effectively deployed in academic evaluations, recruitment training, and corporate communication development programs. The data-driven and objective nature of the system reduces evaluator bias and enhances consistency in assessment. However, further improvements are required to enhance robustness in highly dynamic or noisy discussion environments.

7. Conclusion

This paper presented an AI-Powered Group Discussion Analyzer that integrates speech-to-text processing, Natural Language Processing, sentiment analysis, and participation analytics into a unified web-based platform. The proposed system enables objective and automated evaluation of group discussion performance by analyzing linguistic quality, emotional tone, and engagement behavior in real time.

Experimental evaluation demonstrated that the system effectively captures communication dynamics and generates meaningful performance insights with minimal latency. By reducing reliance on subjective human evaluation, the platform provides consistent, data-driven feedback that supports communication skill development.

Although challenges such as background noise sensitivity and limited interpretation of nuanced emotions remain, the proposed approach offers a scalable and practical solution for academic training, recruitment preparation, and professional communication assessment. Future enhancements may include multimodal analysis, improved sentiment modeling, and offline processing capabilities to further strengthen system robustness.

8. Acknowledgement

The authors would like to express their sincere gratitude to **Mrs. B. Mamatha**, Assistant Professor, and **Dr. S. Srinivas**, Associate Professor, Department of Data Science, Holy Mary Institute of Technology & Science, Hyderabad, for their valuable guidance, continuous support, and constructive suggestions throughout the course of this research work. Their insights and encouragement greatly contributed to the successful completion of this study.

9. References

1. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
2. M. M. Rahman, T. Ahmed, and K. Roy, "Advances in deep learning for natural language processing and sentiment analysis," *Journal of Artificial Intelligence Research*, vol. 65, no. 2, pp. 210–230, 2024.
3. G. Murray, S. Renals, and J. Carletta, "Computational models for multiparty conversation analysis," *Computational Linguistics*, vol. 44, no. 3, pp. 557–590, 2018.
4. H. M. Weiss and R. Cropanzano, "Affective events theory: A theoretical discussion of the structure, causes, and consequences of affective experiences at work," *Research in Organizational Behavior*, vol. 18, pp. 1–74, 1996.
5. F. Herrera, E. Herrera-Viedma, and J. L. Verdegay, "A model of consensus in group decision making under linguistic assessments," *Fuzzy Sets and Systems*, vol. 78, no. 1, pp. 73–87, 1996.
6. T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.
7. Rosenberg and J. Hirschberg, "Detecting opinionated utterances in conversations," in *Proceedings of the NAACL-HLT Conference*, pp. 488–492, 2012.
8. H. Taherdoost, "A comprehensive review of artificial intelligence applications in sentiment analysis," *International Journal of Computer Applications*, vol. 182, no. 3, pp. 12–25, 2023.
9. L. Xu, H. Yang, and X. Chen, "Automated meeting analysis: Techniques, tools, and business applications," *Journal of Information Technology and Management*, vol. 32, no. 4, pp. 345–360, 2021.
10. M. Pontiki *et al.*, "Aspect-based sentiment analysis: SemEval-2014 Task 4," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 27–35, 2014.