Interactive GD Platform

Team members

Sanghamitra Menon

Sanoy Boby

Susan Mathew

Vijay K V

Guide

Ms. Jomina John

Contents

- Problem definition
- Purpose & need
- Project objective
- Literature survey
- Methodology
- Architecture diagram
- use case diagram
- Modules
- Each module in detail

- Assumptions
- Work breakdown & responsibilites
- Hardware & software requirements
- Time line
- Budget
- Risk & challenges
- Expected output
- Conclusion
- References

Problem definition

Effective communication and collaboration skills are vital in academic and professional settings, yet many lack opportunities for structured practice. There is a clear need for a dedicated platform that allows users to practice group discussions, receive constructive feedback, and build confidence, ultimately enhancing their communication skills for collaborative challenges.

Purpose & need

The purpose of this project is to create a platform that enhances group discussions through multiple conversational models, allowing individuals to practice their speaking skills in a real-life simulation. This tool fosters communication and critical thinking without interruptions, making it valuable for interview preparation and helping candidates excel in group discussions.

Project objective

The platform aims to provide individuals with a structured environment to practice and improve their group discussion skills. It will facilitate engaging conversations, offer real-time feedback, and connect users with diverse perspectives, fostering effective communication and collaboration.

Literature survey

PAPER	ADVANTAGES	DISADVANTAGES
Ul Haq et al. [1] Natural Language Processing Approach to Evaluate Real-Time Flexibility of Ideas to Support Collaborative Creative Process (2024).	• NLP can analyze conversations in real-time and providing instant feedback	• NLP struggles to fully understand and emotional tone or cultural context
Halvoník et al. [2] Large Language Models and Rule-Based Approaches in Domain-Specific Communication (2024).	• LLMs can be fine-tuned for specific tasks and domains, improving their performance in particular areas • LLMs can generate varied viewpoints, enriching discussions by presenting multiple angles on a topic.	• LLMs depend on high-quality and diverse training data for effective performance. • For real-time updates, LLMs require continuous internet access to deliver accurate information.

| PAPER	ADVANTAGES	DISADVANTAGES
Olofintuyi et al. [1] A Survey of Variants of Round Robin CPU Scheduling Algorithms (2020).	• Equal turn-taking ensures all LLMs contribute.	• Turn-based delays can slow down the interaction.
Topsakal et al. [1] Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast (2023).	• Optimised memory compononets to keep track of conversations	• Limited size of buffers

<u>Methodology</u>

- The platform simulates group discussions by integrating human users with two AI participants powered by Gemini and LLaMA models.

- Animated avatars with real-time lip synchronization enhance participant realism

- Real-time speech processing utilizes WebKit API for bidirectional speech-text conversion

- Periodic facial engagement analysis evaluates eye contact and orientation through OpenCV-processed screenshots.

- The Qwen model performs content evaluation, scoring topic adherence and argument coherence.

- MongoDB's architecture efficiently stores user profiles and session-specific interaction data.

- Post-session feedback combines facial engagement metrics and content evaluation in visual dashboards.

Architecture diagram
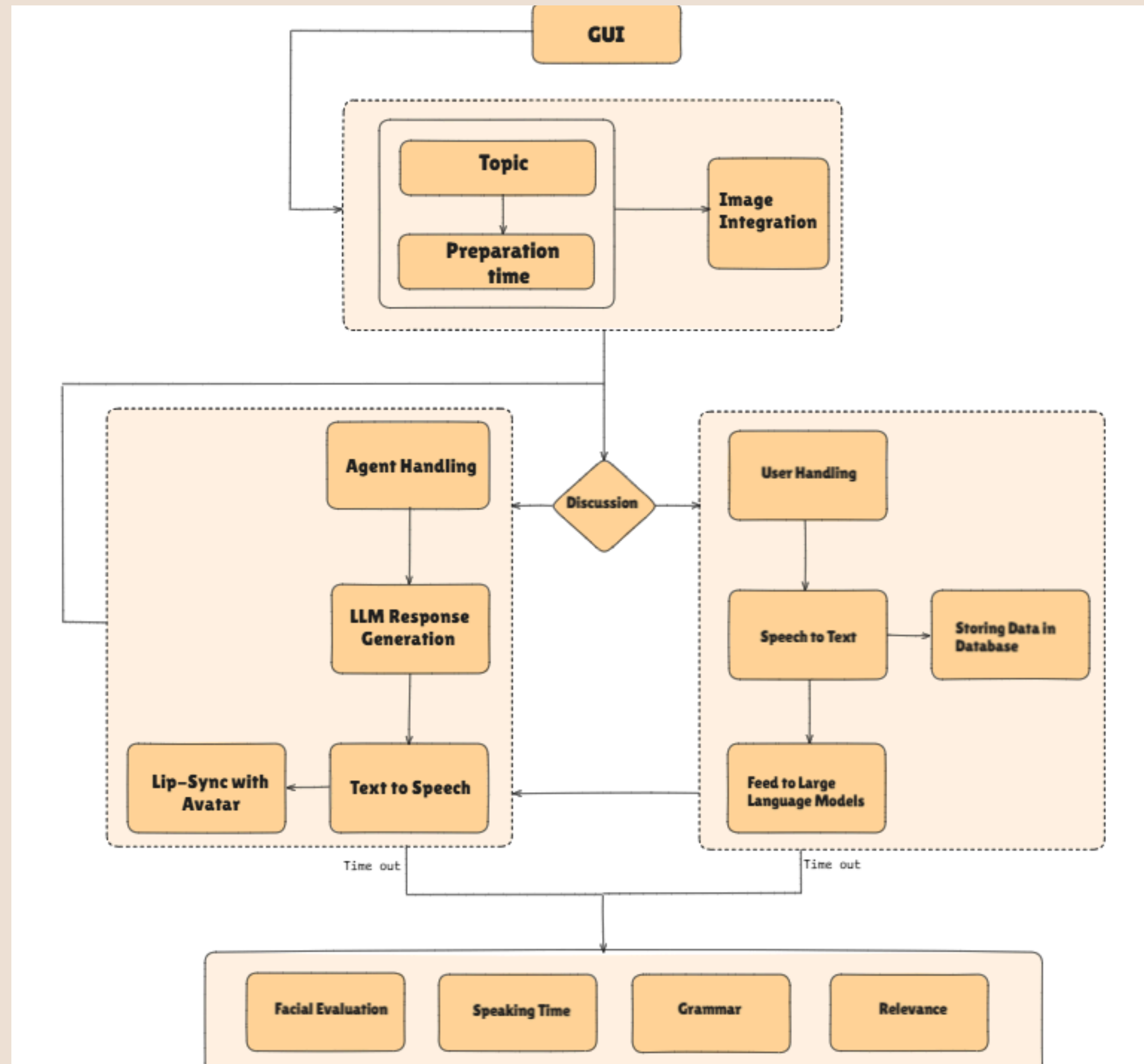
Use case diagram
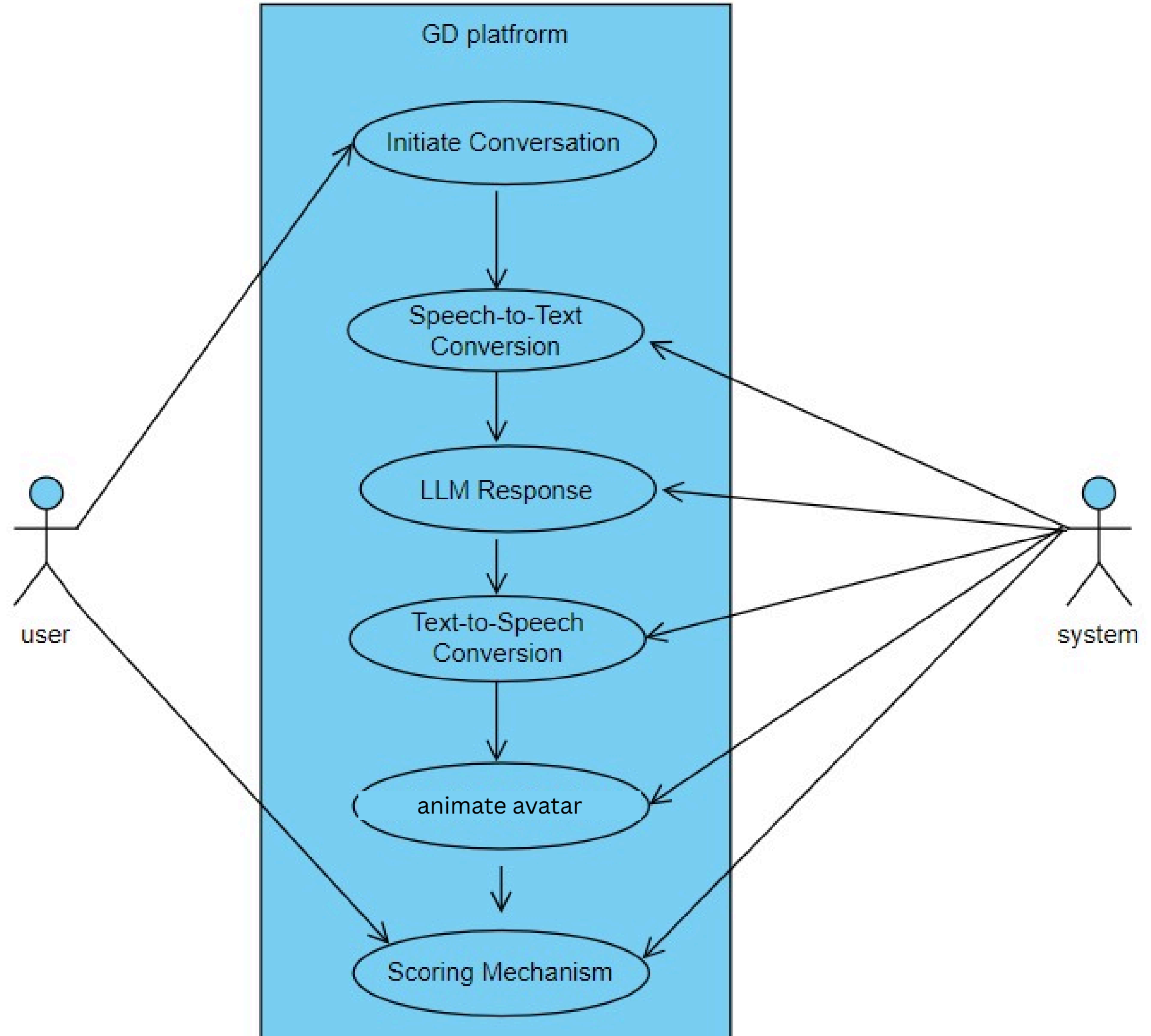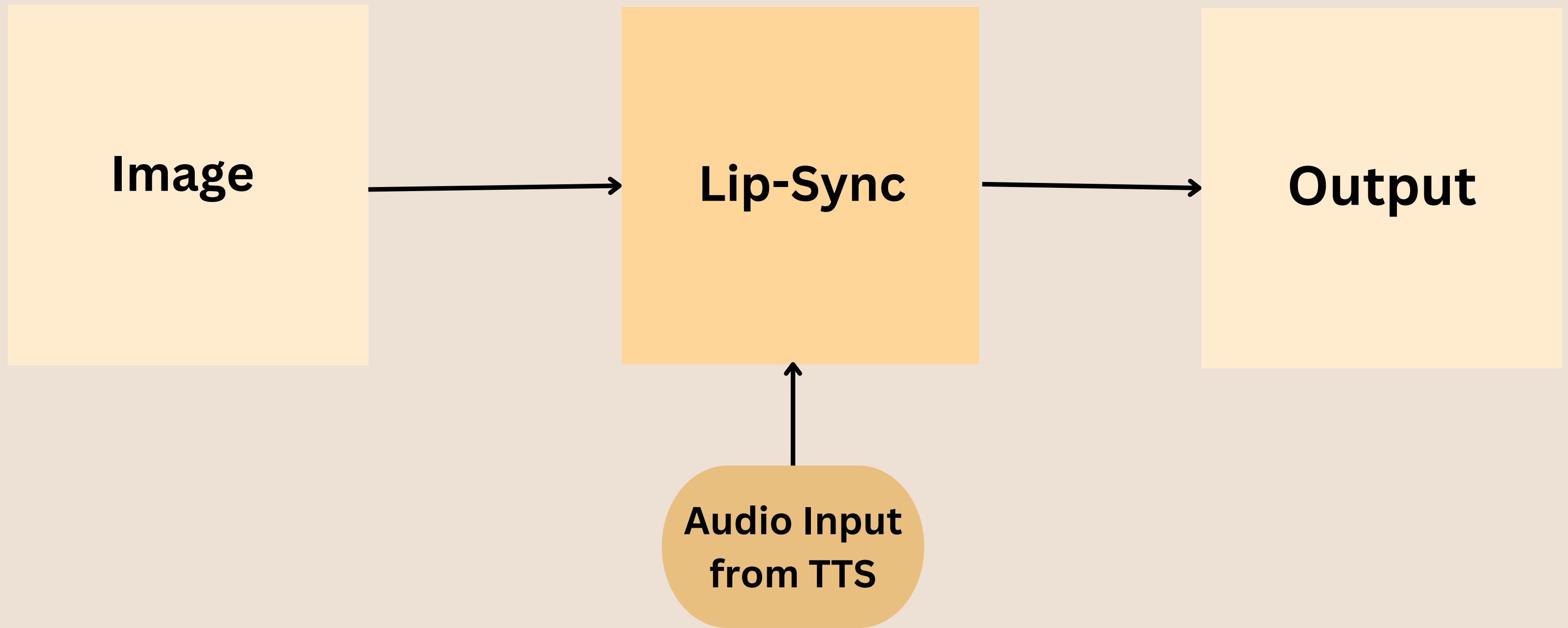
Modules

Model handling

Evaluation

Animation integration

Speech-text module

Animation Integration

- Wav2lip is an open-source project that uses AI to animate images

- It utilizes deep learning techniques like GANs to synchronize a person's lip movements with an image, enabling realistic lip-sync animations.

Animation Integration process

15

- It recieves audio and static image as input

- Wav2lip modifies the face of the static image (keeping the background and other facial features constant) to lip-sync the movements with the speech audio.

- After Avatarify generates video frames based on the lip-synced audio, stream these frames back to the frontend in real time.

# **Speech-to-Text (SST) and Text-to-Speech (TTS)**

- User Speaks: User presses a button to talk.

- STT Conversion: Transcribes speech to text.

- LLM Response: LLM generates a reply.

- TTS Conversion: Converts text to speech.

- Synthesized speech: Played back to the user.
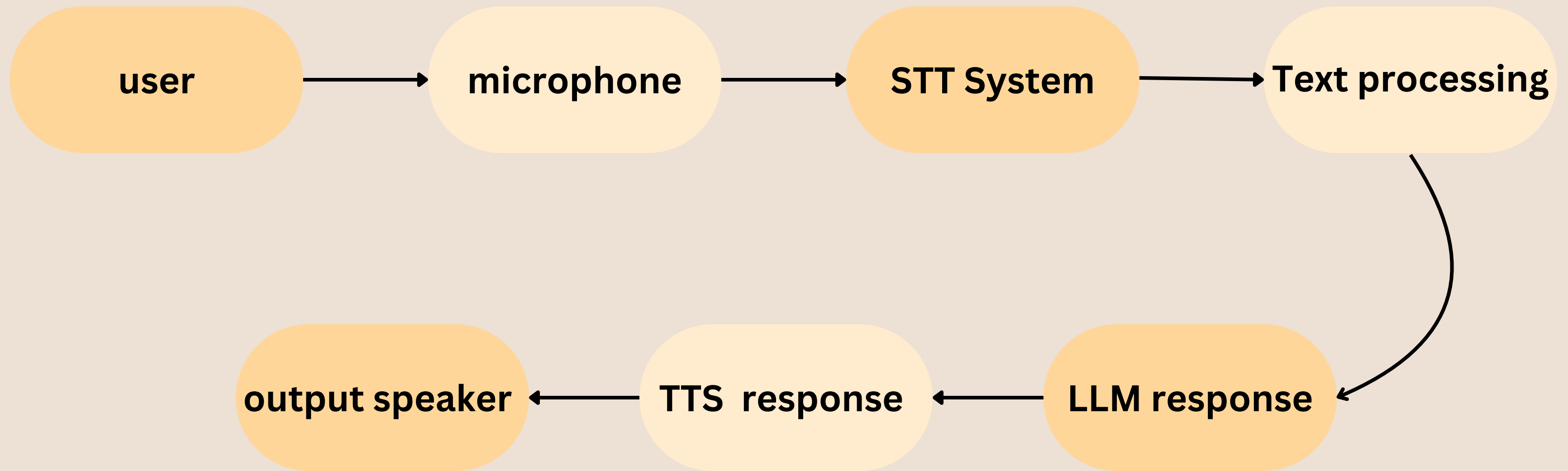
SST and TTS process

- Speech Recognition Toolkit: WebKit Speech Recognition API is used to transcribe user speech into text for processing by the LLM.
- Speech-to-Text Processing: When the user presses "Talk," the Speech Recognition Toolkit captures and transcribes the speech into text, which is then processed by the LLM to generate a response.
- Text-to-Speech Output: The LLM-generated response is converted into speech using Text-to-Speech, played back to the user with distinct voices assigned to each LLM for realistic interaction.

Model handling process

Select the right models → Load the models → Manage conversation history → Handle multiple models
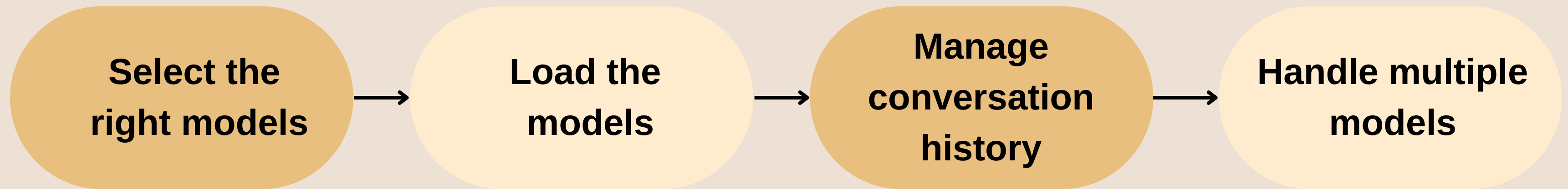
Select the right models

Decide which models will be used in the platform for generating conversation responses.

Gemini

- Gemini's advanced reasoning capabilities enable meaningful and context-aware contributions to group discussions.

- It's multimodal abilities allow seamless integration of text, images, and audio, enhancing interactive dialogue experiences.

LLaMA

- LLaMA (Large Language Model Meta AI) is an open-source, large-scale language model developed by Meta, designed to understand and generate human-like text across multiple applications.
- It is useful for a wide range of tasks, such as natural language processing, conversation generation, content creation, and enhancing AI-driven systems with scalable and efficient performance.

Load the models

- Ensure the models are properly loaded and ready to generate responses.
- Gemini is loaded to generate detailed, context-aware responses, effectively balancing counterarguments and supporting discussions.
- LLaMA, when loaded, delivers flexible and scalable responses, ensuring seamless integration into conversations while providing both supportive and challenging points.
- The models are initialized to seamlessly integrate and respond based on user input and conversation history.

Managing Conversation History

An array is used to store LLM responses to prevent repetition and maintain context-aware replies throughout the discussion.

Allow User Contributions

Users can contribute thoughts only after a model completes its point, maintaining a smooth flow of conversation.

Evaluation

Topic relevance

Grammar

scoring module

Facial Evaluation

Speaking Time
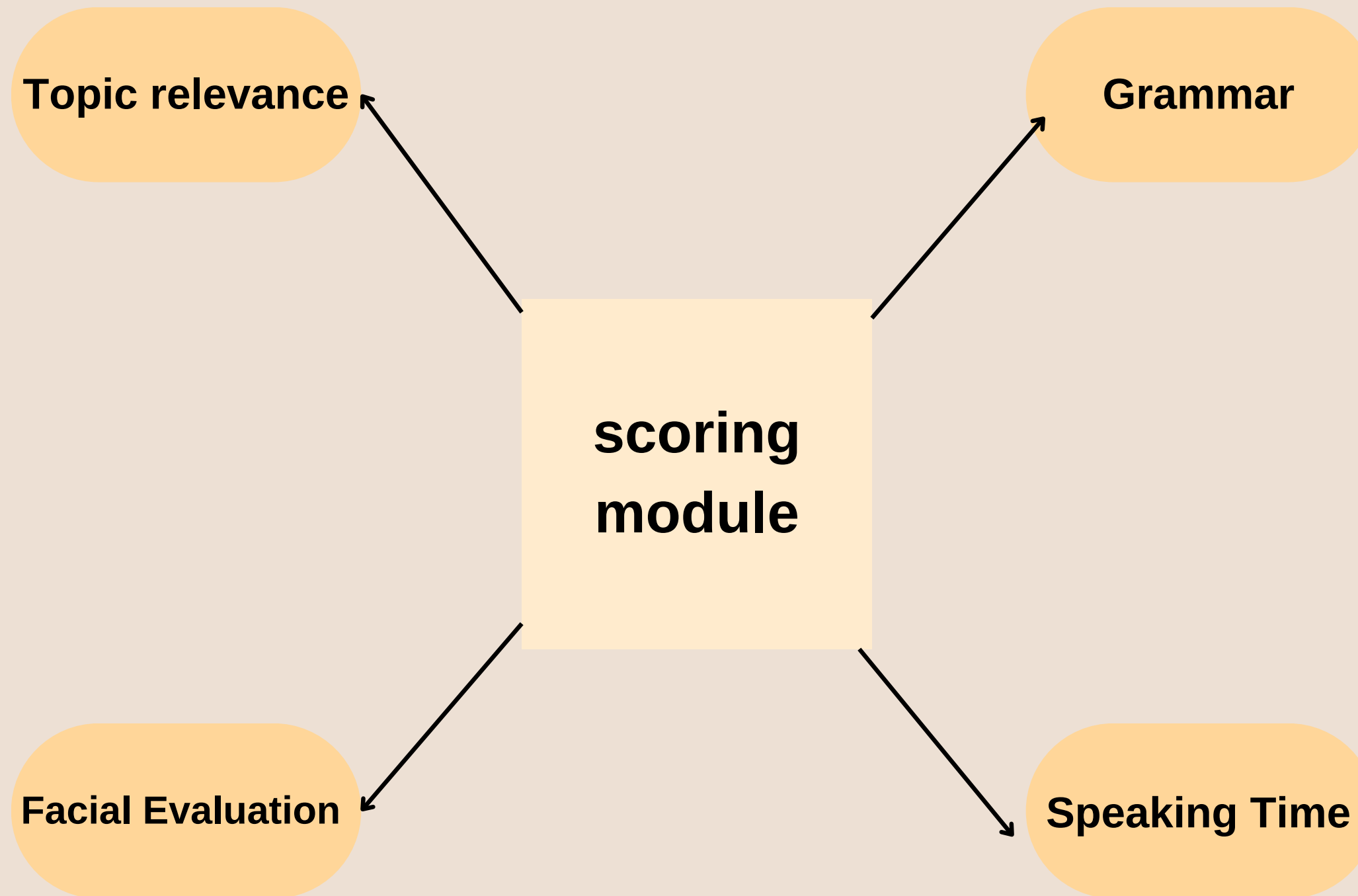
Topic Relevance

- Speech evaluation implemented using the Qwen language model.
- Qwen provides insights in three areas:

    - Analysis, Key Points Covered, and Missing Points.

- Scoring based on:

    - Depth of Analysis (0.8), Relevance (0.85), Structure (0.7).

- Generates an Overall Score: 7.8/10.
- Delivers a summary and improvement suggestions to enhance speech quality.

Grammar

- Evaluation is performed on transcribed text only (no audio analysis).

- Text analysis is done using:

1. TextStat – for readability, complexity, and structural scoring.
2. LanguageTool (Python) – for grammar, spelling, and style checks.

Speaking Time

- User speaking time is extracted using state management during the Group Discussion (GD).
- Tracks when a user starts and stops speaking within the total GD duration.
- Final speaking score is calculated using the formula:
    (User's Speaking Time / Total GD Time) × 100

Work breakdown & responsibilities

Sangamitra Menon

STT and TTS

Speech-to-Text and Text-to-Speech modules enable seamless voice interaction and real-time AI-driven dialogue simulation.

Susan Mathew

Model handling

The system picks and loads the best AI models to keep the conversation smooth and on-topic.

Sanoy Boby

Animation

Avatars are animated with moving lips that match the speech, making the discussion feel more real.
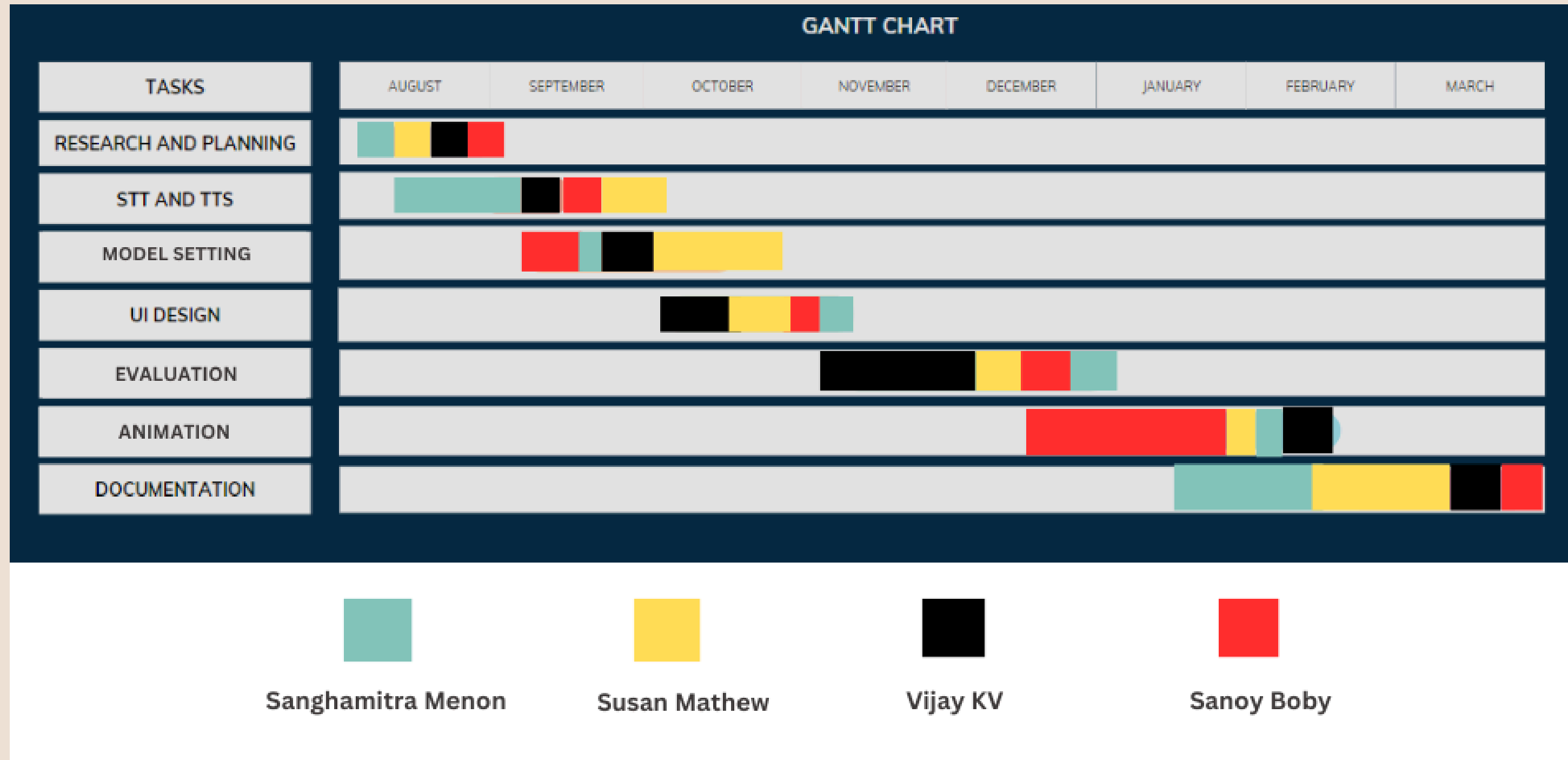
Vijay KV

Evaluation

The evaluation system assesses topic relevance, grammar, facial engagement, and speaking time to generate performance feedback.

Assumptions

- User Consent is given for facial and voice data analysis.
- System Requirements like GPU and internet are met.
- User Readiness to engage with AI-driven feedback.
- Model Availability and stability of LLM APIs is ensured.

Gantt chart

GANTT CHART

TASKS	AUGUST	SEPTEMBER	OCTOBER	NOVEMBER	DECEMBER	JANUARY	FEBRUARY	MARCH
RESEARCH AND PLANNING								
STT AND TTS								
MODEL SETTING								
UI DESIGN								
EVALUATION								
ANIMATION								
DOCUMENTATION								

Sanghamitra Menon   Susan Mathew   Vijay KV   Sanoy Boby

30

Requirements

Hardware

- **Intel Core i5, 8GB RAM**
- **4GB GPU**
- **NVIDIA GeForce GTX**
- **1650 - graphics card**
- **OS: Windows 10 64-bit,**
- **Ubuntu**

Software

- **Development environment: Visual Studio code**
- **Libraries: React,tailwind,openpose**
- **Frameworks: Flask**
- **Database: MongoDB**

Budget

| SL NO	ITEMS	AMOUNT
1	Hosting Service	2000
2	Cloud Database	1000
3	Avatarify (Premium)	1000
4	Publishing	2000
TOTAL:		6000

Risk & Challenges

- Performance and Scalability: Processing large amounts of data efficiently and quickly to generate video elements could pose performance challenges, especially when dealing with diverse video lengths.

- Requires beta-testing to uncover edge cases and bugs.

<u>Outputs</u>
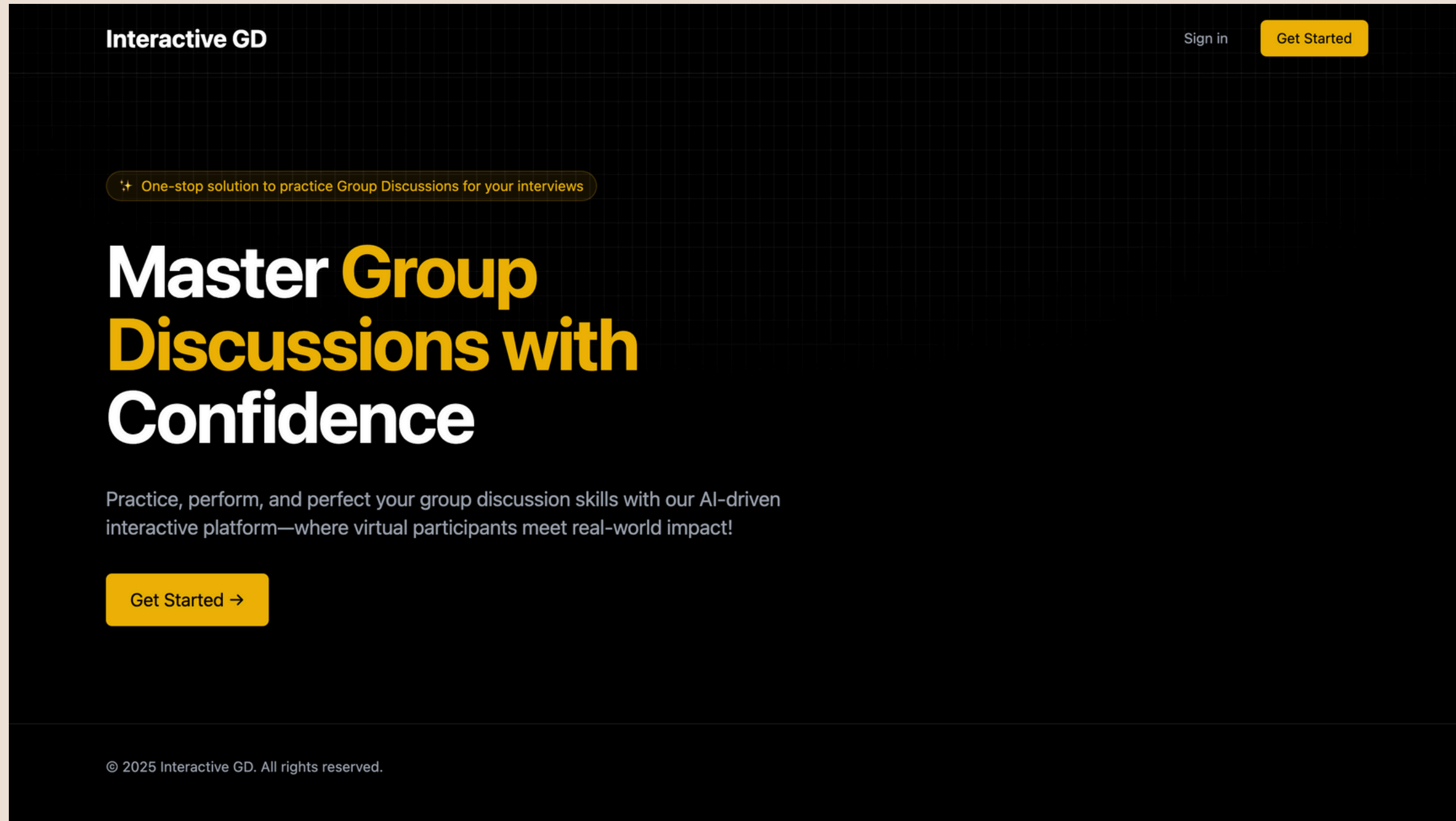
Welcome Back

Sign in to continue to Interactive GD

Sign in with Google
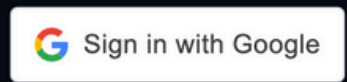
Return to Home

GD Topic
Impact of Artificial Intelligence on Job Markets
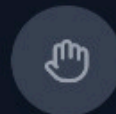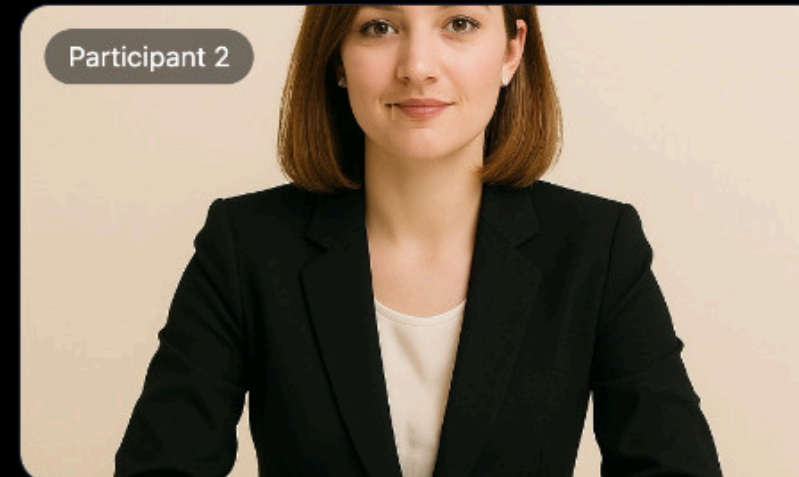
Guidelines for the Group Discussion:

1. Listen actively and respect others' opinions

2. Speak clearly and concisely

3. Support your arguments with relevant examples

4. Encourage quieter participants to contribute

5. Stay on topic and avoid personal attacks

6. Summarize key points towards the end

36

Interactive GD

VIJAY KV Rajagiri

Welcome back, VIJAY KV Rajagiri!

Your GD skills are improving. Keep practicing!

Start New GD Session

Speaking Time Analysis
Based on your 2 GD sessions

🎤 Speaking Time

0%

Average speaking time in sessions

Ideal range   20-30%

Evaluating your results

Please wait while we process your data...

⚡ Speaking Score

1/10

Based on optimal speaking time

In a 4-minute session, you speak for approximately:

0:00

out of 4:00 total session time

Scoring Formula

Your score is calculated based on how close your speaking time is to the ideal range (20-30% of session time).

Score Breakdown: Time Calculation:

Interactive GD

Welcome back, VIJAY KV Rajagiri!

Your GD skills are improving. Keep practicing!

Start New GD Session

Overall Performance

55%

Topic Coverage
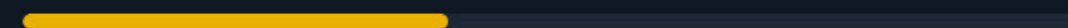
Score: 50%

Key Points Covered
- Impact on junior developers
- Use of LLMs in work

Missing Points
- Broader job market trends
- AI's role in job creation
- Impact on various industries

Detailed Analysis

Depth of Analysis
Score: 40%

Relevance
Score: 60%

Structure

39

Summary

The speech provides a limited view of AI's impact on the job market, focusing narrowly on junior developers and LLMs. It lacks depth, broader relevance, and a structured approach.

Suggestions for Improvement

- Expand the discussion to include various job sectors and roles
- Provide a clear introduction and structure to guide the audience
- Explore both the positive and negative impacts of AI on jobs

Screenshot Evaluation Results

Screenshot Summary

Total Screenshots: 8

Valid Screenshots: 3

Attention Metrics

Eyes Closed Count: 1

Head Turned Count: 3

Scores

Eye Score: 88%

Face Score: 63%

Average Score: 75%

40

Speaking Time Analysis
Based on your 2 GD sessions

🎤 Speaking Time

0%

Average speaking time in sessions

Ideal range   20-30%

⭐ Speaking Score

1/10

Based on optimal speaking time

In a 4-minute session, you speak for approximately:

0:00

out of 4:00 total session time

Scoring Formula

Your score is calculated based on how close your speaking time is to the ideal range (20-30% of session time).

Score Breakdown:

• 20-30% = 10 points (Optimal)
• 15-20% or 30-35% = 7 points (Good)
• 10-15% or 35-40% = 5 points (Fair)
• 5-10% or 40-45% = 3 points (Needs improvement)
• <5% or >45% = 1 point (Significant improvement needed)

Time Calculation:

For a 4-minute (240 seconds) session:

• Optimal: 48-72 seconds
• Good: 36-48 or 72-84 seconds
• Fair: 24-36 or 84-96 seconds
• Needs improvement: 12-24 or 96-108 seconds
• Significant improvement needed: <12 or >108 seconds

41

Conclusion

The Group Discussion Practicing Platform provides a realistic, AI-powered environment to improve communication skills through interactive sessions with LLM avatars. With integrated speech processing, facial evaluation and real-time feedback, the system delivers personalized insights that enhance performance and readiness for real-world GD scenarios.

Future references

- VR/AR Integration for immersive, real-life GD simulations.
- Multilingual Support to cater to diverse user backgrounds.
- Emotion Detection to analyze tone, expression, and mood.
- Academic & Corporate Use as a soft-skill training tool.

References

1. Ul Haq, I., Pifarré, M., & Fraca, E. (2024). Natural Language Processing Approach to Evaluate Real-Time Flexibility of Ideas to Support Collaborative Creative Process. International Journal of Emerging Technologies in Learning (iJET), 19(05), pp. 93–107. https://doi.org/10.3991/ijet.v19i05.47465
2. D. Halvoník and J. Kapusta, "Large Language Models and Rule-Based Approaches in Domain-Specific Communication," in IEEE Access, vol. 12, pp. 107046-107058, 2024, doi: 10.1109/ACCESS.2024.3436902.
3. Olofintuyi, S.S., Omotehinwa, T.O. and Owotogbe, J.S., 2020. A survey of variants of round robin CPU scheduling algorithms. FUDMA JOURNAL OF SCIENCES, 4(4), pp.526-546.
4. Topsakal, Oguzhan, and Tahir Cetin Akinci. "Creating large language model applications utilizing langchain: A primer on developing llm apps fast." International Conference on Applied Engineering and Natural Sciences. Vol. 1. No. 1. 2023.

THANK YOU