International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

AI-Powered Question Management in Conferences

Sahana Sharma M[1], Bibek Mukherjee[2], Ramya Ganesh Shastri[3], Shiva Prasad [4], Akash A Gaddi[5],

[1] Assistant Professor, Department of CSE-AI, Dayananda Sagar Academy of Technology and Management, Bangalore – 82.

[2,3,4,5] Student, Department of CSE-AI, Dayananda Sagar Academy of Technology and Management, Bangalore – 82.

Abstract- Effective management of audience questions are crucial for enhancing speaker-audience interaction in conferences. However, this process is often hindered by irrelevant or repetitive questions and a lack of participation from introverted attendees. This paper presents an **AI-Powered Question Management System** designed to address these challenges. Leveraging the **DistilBERT** model, enhanced with a **Question-Topic Relevance Prediction Model**, the system predicts whether a question is relevant to the discussion topic using deep learning and NLP techniques. This model combines relevance scoring with semantic similarity matching to deliver highly accurate predictions, ensuring only pertinent questions reach the speaker.

The backend, implemented using Node.js and MongoDB, ensures robust scalability and seamless performance, while the React.js-based frontend delivers user-friendly interfaces tailored to speakers and listeners. Conference speakers benefit from real-time access to prioritized, topic-relevant questions, enabling more focused and impactful discussions. Listeners, on the other hand, can submit queries anonymously, fostering inclusivity and encouraging participation from all attendees.

The Question-Topic Relevance Prediction Model integrates key processes: fine-tuned DistilBERT for binary classification, curriculum learning for optimized training, and similarity-based scoring using topic-specific terms. The model achieves relevance predictions by combining model-generated probabilities and similarity metrics, ensuring a balanced evaluation. Experimental evaluations demonstrate significant improvements in question handling efficiency, including reduced redundancy and enhanced user satisfaction. The proposed solution establishes a scalable, intelligent framework for equitable and effective conference question management, setting a new benchmark in audience engagement technologies.

Index Terms- AI-powered question AI-powered question management, DistilBERT, conference interaction, anonymous submissions, question prioritization, Natural Language Processing (NLP), semantic similarity scoring, audience engagement, scalable question filtering, question-topic relevance prediction.

I. INTRODUCTION

In modern conferences, where effective communication and interaction are pivotal, managing audience questions efficiently remains a significant challenge. Traditional approaches, such as open-floor questioning or manual curation, are often riddled with inefficiencies. These include repeated or irrelevant queries that consume valuable time and the hesitancy of certain individuals, particularly introverts, to voice their concerns in a public setting. Such barriers can disrupt the flow of meaningful discussions, leaving critical questions unaddressed and limiting audience engagement. Furthermore, the lack of an automated system to prioritize questions based on their relevance or popularity often leads to missed opportunities for impactful discussions.

To address these multifaceted challenges, this paper introduces an AI-Powered Question Management System. This innovative platform integrates state-of-the-art natural language processing (NLP) capabilities using the DistilBERT model to intelligently filter and prioritize questions. The training pipeline fine-tunes the model using a combination of advanced techniques such as curriculum learning and contrastive learning to enhance its ability to classify question-topic relevance accurately. By identifying and removing redundant queries, the system ensures that speakers focus only on unique and relevant questions. The inclusion of a priority mechanism, which ranks questions based on how many attendees share similar concerns, further streamlines the interaction.

A notable feature of the system is its anonymous submission option, which fosters inclusivity by allowing participants to contribute without fear of judgment. This is particularly beneficial for introverts or individuals who may feel intimidated by public speaking. Additionally, the system incorporates a topic terms generator powered by NLP techniques like TF-IDF analysis, key phrase extraction, and Word2Vec similarity modelling to dynamically enrich and evaluate the relevance of questions to specific topics. The architecture's modularity further ensures scalability and efficiency, with components such as training pipelines, prediction pipelines, and an API interface working seamlessly together.

The system's robust architecture is powered by Node.js and MongoDB in the backend, ensuring scalability and real-time performance. The React-based frontend provides tailored functionalities for speakers and listeners. Speakers gain access to a live dashboard displaying filtered and prioritized questions, while listeners interact with an intuitive interface to submit their queries. By combining advanced relevance scoring mechanisms—including DistilBERT-based model scoring and topic term similarity matching—the system achieves a balanced decision-making process that optimizes engagement.

By addressing these issues, the proposed system revolutionizes conference question management. It not only enhances the speaker-audience interaction but also ensures equitable participation, setting a new benchmark for effective communication in conferences. With its combination of cutting-edge NLP techniques, real-time processing, and inclusive design, this system transforms how questions are handled, making conferences more engaging, inclusive, and impactful.

II. RELATED WORK

The management of audience interactions in conferences has evolved significantly, with platforms like Slido and Pigeonhole Live enabling Q&A, polls, and surveys. However, these systems rely on manual curation of questions, which can be inefficient for large-scale events. Additionally, they lack advanced natural language processing (NLP) capabilities to filter irrelevant or redundant questions, often leading to suboptimal interactions.

Recent advancements in NLP, especially transformer-based models like BERT and DistilBERT, have opened new possibilities for improving question management. These models excel at tasks like semantic similarity detection and classification, making them ideal for filtering and prioritizing conference questions. DistilBERT offers a lightweight, computationally efficient alternative to BERT, suitable for real-time applications.

While previous studies have demonstrated the potential of NLP in question answering and text summarization, direct applications in conference settings often require significant customization. Multilingual support is also a common limitation in existing systems. Some platforms incorporate NLP for specific tasks, such as automated chatbots or AI-driven analytics, but they often lack the capability to efficiently handle redundancy and prioritize relevant questions.

In contrast, the proposed system integrates advanced NLP capabilities into a real-time conference question management platform. It leverages DistilBERT for intelligent question filtering and prioritization. The system's architecture includes a training pipeline, prediction pipeline, API interface, and topic terms generator, ensuring scalability and efficiency. The training pipeline uses data augmentation techniques to enhance the model's understanding of diverse question formats.

The prediction pipeline fine-tunes DistilBERT to classify questions as relevant or irrelevant, with additional term matching based on TF-IDF and Word2Vec embeddings to refine relevance scores. The topic terms generator creates domain-specific keywords to improve similarity calculations, enabling the model to recognize exact and partial matches.

The system also supports anonymous submissions, fostering inclusivity, and uses a weighted combination of model scores and similarity scores to prioritize questions effectively. While it currently focuses on single-language conferences, the system can be easily adapted for multilingual environments. Its modular structure ensures flexibility and scalability, allowing integration with existing platforms.

By leveraging advanced NLP, efficient computational frameworks, and user-centric design, this system bridges the gap between engagement-focused platforms and NLP-driven technologies. It addresses the limitations of existing systems and sets a new standard for real-time question management in conferences. Its ability to handle large-scale events, filter redundant questions, and prioritize

relevant ones marks a significant advancement in the field. The use of frameworks like TensorFlow, PyTorch, Flask, Node.js, and MongoDB ensures a robust, scalable solution capable of adapting to diverse conference needs.

III. SYSTEM DESIGN AND ARCHITECTURE

The design and architecture of the AI-powered question management system are carefully crafted to address the specific challenges faced during conferences. The system consists of three main layers: the frontend, backend, and database, with the integration of the DistilBERT model. Each layer is designed to ensure scalability, efficiency, and inclusivity while providing a seamless user experience. The system is highly modular, allowing for easy updates and adaptations as new technologies and requirements emerge.

1. Frontend (React.js)

The frontend acts as the user interface, enabling interaction between attendees and speakers. Built using **React.js**, it ensures a responsive and intuitive design, catering to both desktop and mobile users. The frontend is optimized to handle dynamic interactions with minimal latency, ensuring that listeners and speakers can seamlessly interact in real-time. Upon logging in, users select their roles—listener or speaker. Based on this selection, the platform provides tailored functionalities:

- **Listeners:** The interface allows users to type questions, select anonymity options, and submit queries.
- **Speakers:** The dashboard displays a curated list of questions ranked by relevance. The design minimizes distractions and ensures that only high-priority questions are shown, enabling speakers to focus on meaningful discussions. Speakers can also interact with the system by sorting questions based on their priority, selecting topics to address. The dynamic nature of the interface ensures that it remains responsive even during large-scale conferences with multiple participants.

  Real-time updates are facilitated using **React's state management libraries**, ensuring a smooth and dynamic user experience. The platform also integrates **WebSocket technology** for low-latency communication between the frontend and backend.

2. Backend (Node.js)

The backend serves as the system's engine, processing data and managing API endpoints. Built on **Node.js**, it handles incoming questions, applies natural language processing (NLP), and routes data between the frontend and database. The backend leverages asynchronous capabilities to handle multiple requests simultaneously, making it ideal for large conferences with thousands of participants. Key functionalities include:

- **Real-Time Data Flow:** Using **WebSocket technology**, the backend ensures instant updates to both listeners and speakers. For instance, when a listener submits a question, it immediately appears on the speaker's dashboard after processing. This low-latency communication is crucial for maintaining a fluid and uninterrupted interaction during live conferences.
- **NLP Integration:** The backend integrates the **DistilBERT model** for filtering and prioritizing questions. Each incoming question is analyzed for relevance to the session topic, ensuring that only pertinent queries are forwarded to the speakers.

3. Database (MongoDB)

The database stores user data, questions, and session details. **MongoDB**, a NoSQL database, is chosen for its flexibility and scalability. The database structure includes collections for: **User Data:** Contains information about listeners and speakers, including their roles and session details. This data is dynamically updated as users interact with the system, ensuring that the correct data is always available for processing. This collection is essential for managing multiple sessions simultaneously and ensuring that questions are associated with the correct session and speaker. Indexes are created on frequently queried fields, such as **session IDs** and **relevancy scores**, to ensure quick data retrieval.

4. AI Component (DistilBERT)

DistilBERT, a distilled version of the BERT model, is the core **NLP engine** of the system. It processes questions in three stages:

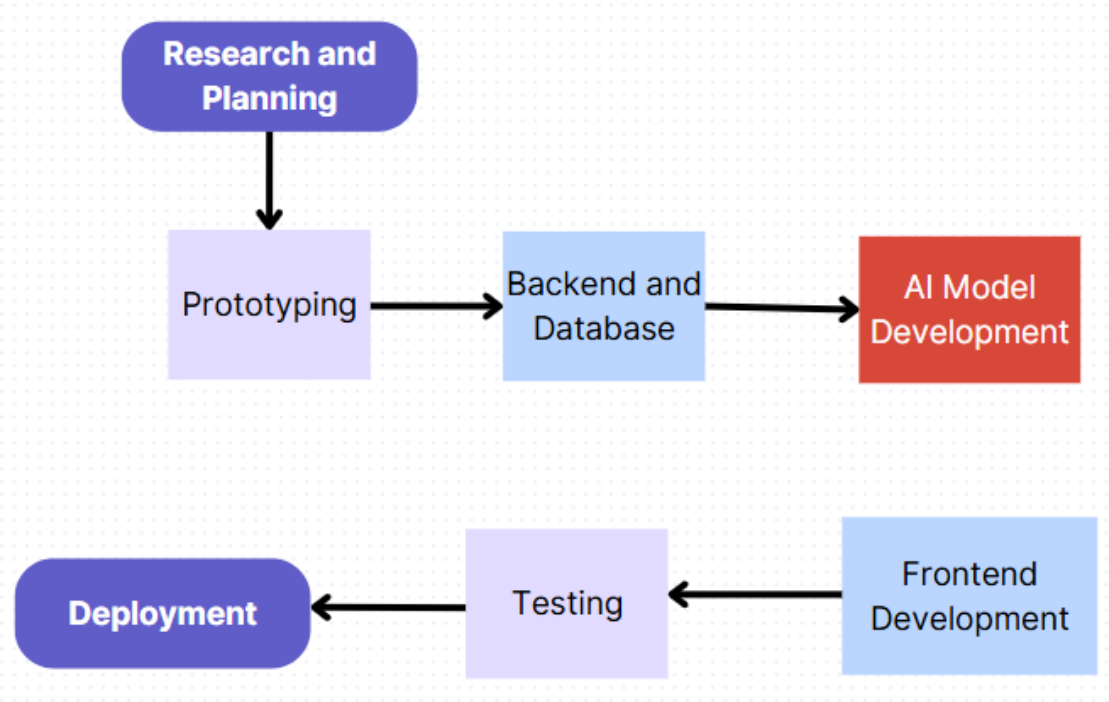
- **Question Filtering:** Evaluates the semantic similarity between each question and the session topic. Irrelevant questions are filtered out to maintain focus on the subject matter of the conference.
- **Priority Ranking:** Assigns scores to questions based on factors such as relevance and frequency. Frequently asked questions are given higher priority, ensuring that common concerns are addressed first. The system uses a weighted combination of **model scores** from DistilBERT and **similarity scores** based on term matching to generate an overall score for each question, which is then used to rank questions.
- The **AI model** is fine-tuned for this specific task, ensuring that it can accurately determine the relevance and priority of each question. Using **transformers** and frameworks like **PyTorch** and **TensorFlow**, the system can efficiently classify questions while maintaining scalability and low computational overhead.

5. Scalability and Reliability

This robust architecture ensures that the AI-powered question management system effectively addresses the challenges of traditional conference interactions, providing a seamless and inclusive experience for all participants.

The system's modular design allows for future enhancements, such as the integration of multilingual support or the inclusion of advanced analytics to provide insights into audience engagement.

IV. Workflow Overview



The system follows a structured workflow to ensure efficient processing, from research to deployment. This iterative approach allows for continuous improvement and alignment with user needs.

1. Research and Planning

- **Objective:** Conduct a study of existing solutions and user needs.
- **Activities:** Research technologies like DistilBERT for NLP, MongoDB for database storage, and React for frontend to ensure scalability, efficiency, and inclusivity. Gather requirements for speaker and listener features, focusing on user interaction patterns. Define architecture for integrating AI, frontend, and backend with real-time updates via WebSocket and database replication. Explore datasets for training the AI model, focusing on optimizing question filtering, moderation, and ranking.

2. Prototyping

- **Objective:** Create a small-scale prototype to validate initial ideas.
- **Activities:** Develop a mockup for frontend login features and role-based interfaces for speakers and listeners. Implement a basic question filtration algorithm using DistilBERT for relevance filtering. Test initial database interactions for storing questions and metadata like priority. Outcome: Gather user feedback and assess real-time performance with light usage.

3. AI Model Development

- **Objective:** Develop and train the core AI model for question filtering and prioritization.
- **Activities:** Train the DistilBERT model on relevant datasets for topic relevance, deduplication, and prioritization. Implement techniques for deduplication and privacy handling, and explore data augmentation strategies. Fine-tune the model, evaluating its accuracy using precision, recall, and F1 score.

4. Backend and Database

- **Objective:** Set up backend infrastructure and database.
- **Activities:** Use Node.js to build APIs for user roles and question submissions, focusing on asynchronous processing. Design a MongoDB schema for storing questions, interactions, and metadata with indexes for fast queries. Implement load balancing and database replication for scalability and high availability during conferences.

5. Frontend Development

- **Objective:** Build an interactive user interface using React.
- **Activities:** Implement role-based login for speakers (question dashboard) and listeners (anonymous question submission). Ensure responsive design optimized for various devices and screen sizes. Integrate frontend with backend APIs using WebSocket technology for real-time updates.

6. Testing

- **Objective:** Ensure reliability, accuracy, and usability.
- **Activities:** Conduct unit tests on components (AI, frontend, backend). Perform end-to-end testing to validate the entire workflow, from question submission to response. Test scalability under simulated conference conditions with load testing. Collect feedback from beta users to refine the user experience and interface.

7. Deployment

- **Objective:** Launch the application for real-world conference use.
- **Activities:** Deploy the frontend on a web hosting platform and backend on cloud infrastructure (e.g., AWS, Google Cloud). Implement monitoring tools for performance tracking and usage analytics. Provide training and documentation for conference organizers on managing roles, moderating questions, and using the AI model

V. Implementation Details

The AI-powered question management system is designed with cutting-edge technologies across the frontend, backend, and database layers to provide a seamless, real-time user experience while maintaining scalability and efficiency.

1. Frontend Implementation (React.js)

The frontend, built with React.js, is designed for modularity, performance, and user experience.

- **Role-Based Login:**
  Users log in as either listeners or speakers, with the UI tailored to their respective roles, ensuring relevant features like anonymous question submission or question moderation are available.
- **Real-Time Updates:**
  React's state management libraries (e.g., Redux) handle real-time updates, allowing listeners' questions to appear instantly on speakers' dashboards.
- **Responsive Design:**
  Optimized for desktop and mobile devices, using CSS frameworks like Material-UI to ensure a consistent, fluid experience.
- **Integration with WebSocket:**
  WebSocket enables real-time communication, ensuring low-latency interactions for a smooth conference experience.

2. Backend Implementation (Node.js)

The backend, powered by Node.js, is responsible for API management, NLP processing, and real-time communication.

- **API Development:**
  RESTful APIs manage question submission, filtration, and retrieval, ensuring secure interactions through JWT authentication.
- **NLP Integration (DistilBERT Model):**
  The backend integrates the DistilBERT model for question filtering, redundancy elimination, and priority ranking, optimized for speed.
- **Real-Time Communication:**
  WebSocket ensures that updates from listeners are instantly reflected in the speaker's dashboard.
- **Scalability:**
  Node.js handles multiple concurrent requests efficiently, crucial for large-scale conferences.
- **Middleware for NLP Operations:**
  A middleware bridge ensures clean separation between data handling and NLP processes, maintaining modularity.

3. Database Implementation (MongoDB)

MongoDB powers the database layer, providing flexibility, scalability, and performance for user data, questions, and session information.

- **Schema Design:**
  Collections for users, sessions, and questions are indexed for fast retrieval, optimizing query performance during live interactions.

- **Data Handling:**
  The MongoDB aggregation framework groups similar questions to reduce redundancy, delivering concise question lists to speakers.

- **Scalability and Replication:**
  MongoDB's distributed architecture and replication ensure high availability and fault tolerance for large-scale conferences.

- **Real-Time Syncing:**
  Real-time syncing between the database and frontend ensures both speakers and listeners are always up to date.

4. Integration of DistilBERT Model

- The **DistilBERT model** is integrated into the system to enhance its NLP capabilities for efficient question processing. It is used to predict whether a given question is relevant to a specific topic, leveraging a combination of deep learning and natural language processing (NLP) techniques.
- Architecture Overview: The system is designed with four main components: the Training Pipeline (train.py), the Prediction Pipeline (predict.py), the API Interface (app.py), and the Topic Terms Generator (predict.py). Each component plays a crucial role in the processing and prediction tasks.

⇒ Training Pipeline (train.py)

The **Training Pipeline** starts by loading a dataset from a CSV file, which contains questions, topics, and relevance labels. This dataset is split into training (70%) and validation (30%) sets. The training data is further augmented using techniques like synonym replacement (via WordNet) and question reformulation (e.g., transforming "what" to "could you explain"). The **model architecture** uses the base DistilBERT model, which is fine-tuned for binary classification, specifically predicting whether a question is relevant or not to a given topic. The input format consists of a combination of the question and the topic: "Question: [question_text] Topic: [topic_text]." During the **training process**, the input data is tokenized, and the relevance labels are converted to integers (0 for not relevant, 1 for relevant). The model is trained using five epochs with a batch size of 8. The learning rate uses a warmup schedule, and weight decay is set to 0.01. Training features like **curriculum learning** (starting with easier examples), **contrastive learning**, and **evaluation during training** are employed. The best model checkpoint is saved during training to ensure optimal performance.

⇒ Prediction Pipeline (predict.py)

First, the model is loaded, and the relevance score is calculated by passing the tokenized input through the model. A **similarity score** is then calculated by comparing the question to a set of topic terms. This score helps to assess the relevance of the question based on partial and exact matches of terms. The **final score** is computed as a weighted combination of the model score (70% weight) and the similarity score (30% weight).

⇒ Topic Terms Generation

The **Topic Terms Generator** uses multiple NLP techniques to generate relevant terms for each topic. **TF-IDF analysis** helps to identify the most important terms in the dataset. **Key phrase extraction** is performed using spaCy, which extracts noun chunks and filters them by relevance and length. Additionally, **Word2Vec similarity** is used to identify semantically similar terms by training Word2Vec on a domain-specific corpus.

⇒ API Interface (app.py)

The **API Interface** provides REST API endpoints, including a **health check** endpoint (GET /) and a **prediction endpoint** (POST /predict). The prediction endpoint receives a JSON request containing a question and a topic, and returns a prediction

indicating whether the question is relevant to the topic. Model Score (70% of Final Score): The model score, which accounts for 70% of the final score, is computed by tokenizing the input text (a combination of the question and topic) and passing it through the DistilBERT model. The output logits are processed through softmax to generate a probability, which reflects how relevant the model believes the question is to the topic. A score of 0.0 means the model is certain the question is not relevant, while a score of 1.0 indicates full relevance. A score around 0.5 indicates uncertainty. Similarity Score (30% of Final Score): The **similarity score** represents 30% of the final score and is calculated based on how many topic terms match the question. **Exact matches** contribute fully to the score, while **partial matches** (where only parts of the term appear in the question) contribute half the weight. The final similarity score is a weighted average of these matches, with a maximum score of 1.0.

Final Score Calculation: The final score is calculated using a dynamic weighting approach based on the similarity score. The logic follows these rules:

1. If the similarity score is high (> 0.5):

The final score is set to the maximum of the model score and 0.7. This ensures that strong term matches can override model uncertainty.

If both scores are very low (model_score < 0.1 and similarity_score < 0.1):

The final score is set to 0.0001. This indicates definitive irrelevance

In all other cases:

The final score is calculated as: final_score = (0.7 model_score) + (0.3 similarity_score)

This gives more weight to the model's prediction (70%) while still considering term matching (30%).

Decision Making: The system determines relevance based on the final score, with special handling for edge cases:

A score of 0.0001 indicates strong confidence in irrelevance

A score ≥ 0.7 when similarity is high indicates strong confidence in relevance

Scores between these extremes represent varying degrees of confidence

Similarity Calculation: The system now uses an enhanced similarity calculation that:

1. Gives special handling to "What is X?" questions:

Direct term matches return a score of 1.0

Partial term matches return a score of 0.9

2. For other questions:

Calculates exact matches (full term found in question)

Calculates topic keyword matches (individual words found)

Weights exact matches more heavily than partial matches

⇒ Factors Affecting Scores

1. Model Score Influences:
   - Question phrasing
   - Topic relevance in training data
   - Semantic understanding
   - Context recognition

2. Similarity Score Influences:
   - Number of matching terms
   - Quality of topic terms
   - Word overlap

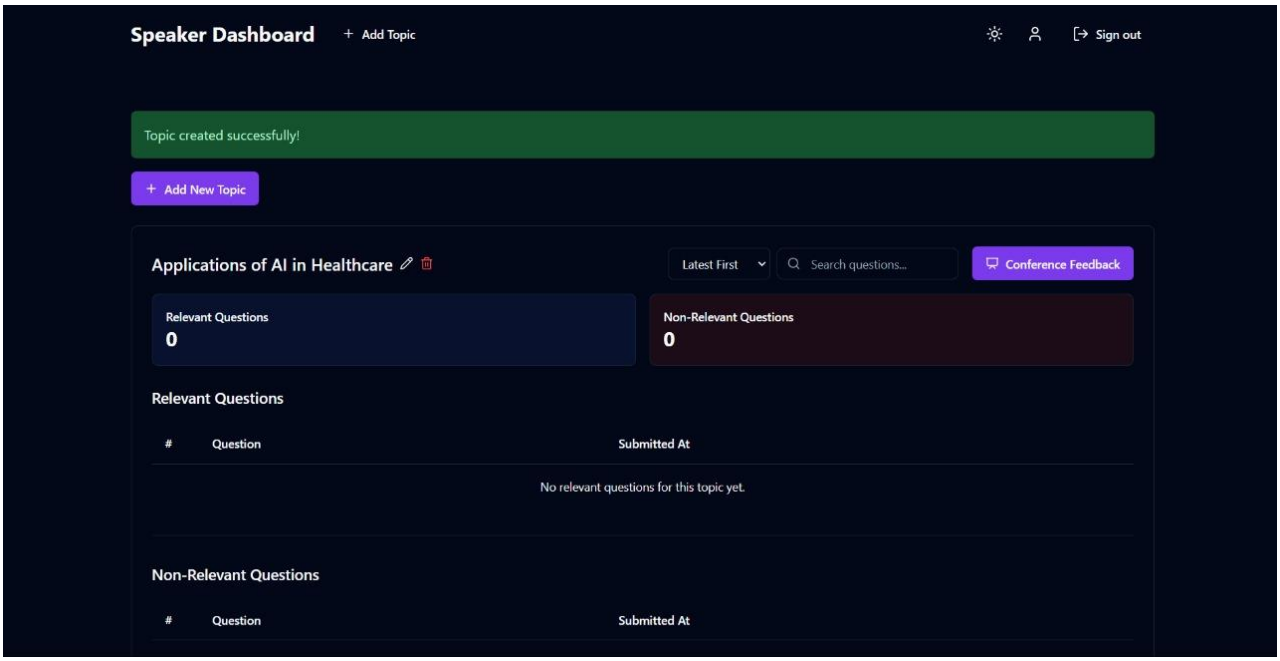
5. System-wide Performance Enhancements

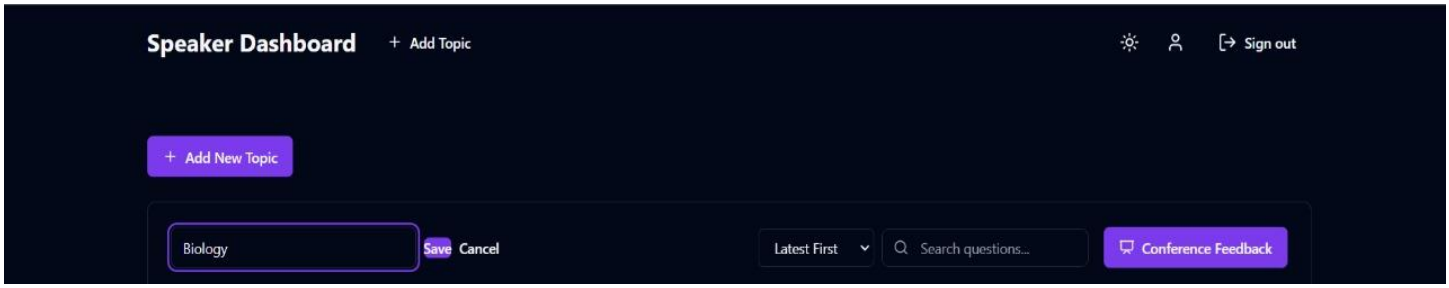
The system focuses on scalability, reliability, and performance.

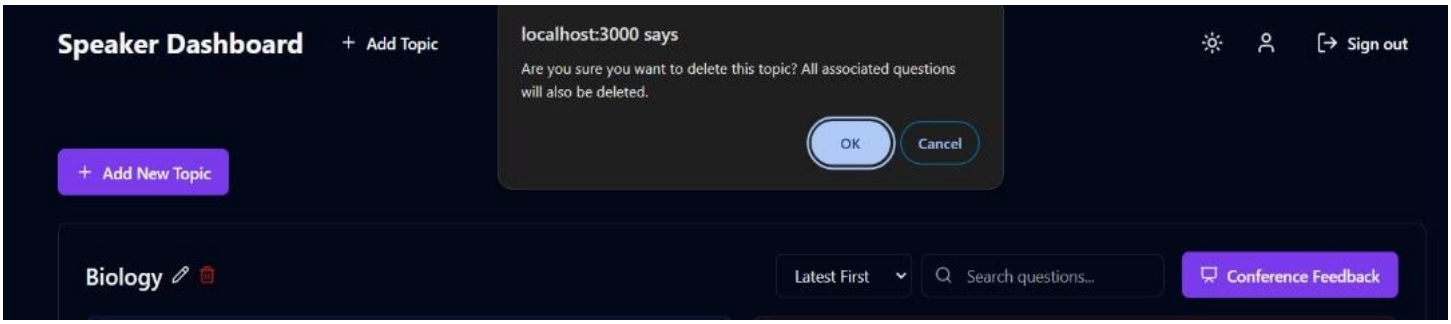
- **Load Balancing and Redundancy:**
  Load balancing distributes requests across multiple servers, ensuring performance consistency during high-traffic periods.
- **Caching and Data Retrieval:**
  Caching strategies minimize database load and optimize retrieval speeds, improving response times.
- **Continuous Monitoring:**
  Real-time monitoring tools track key performance metrics, enabling proactive issue resolution and system optimization.

VI. RESULTS

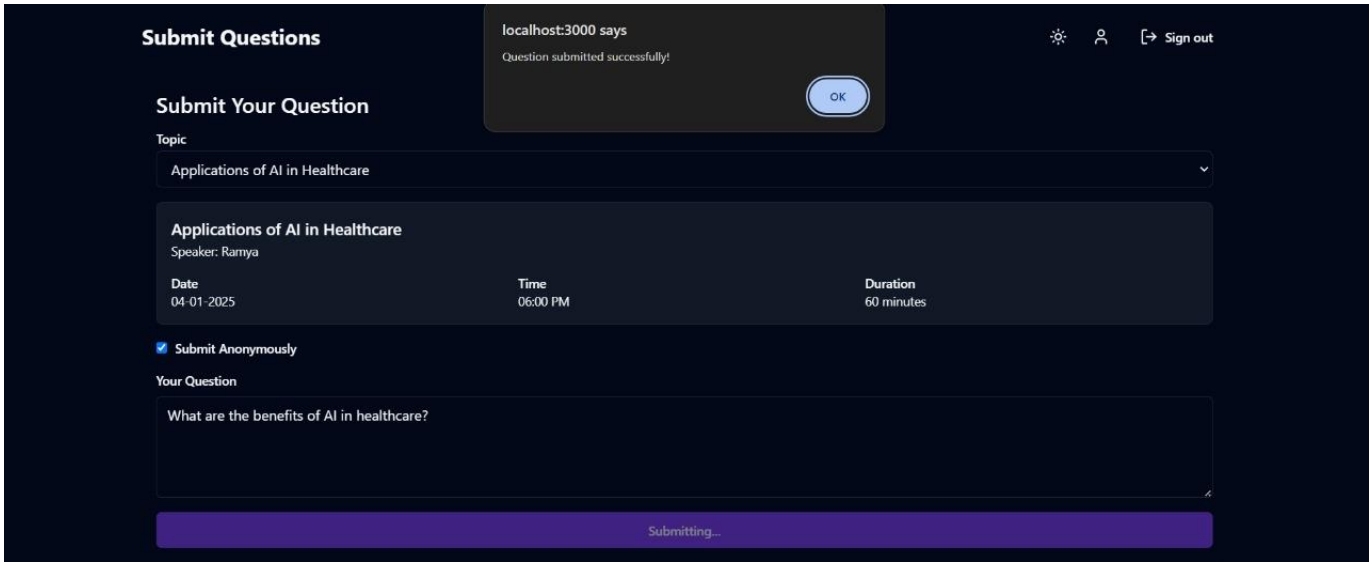
Edit Topic Name, and Topic Delete Feature available to the Speaker

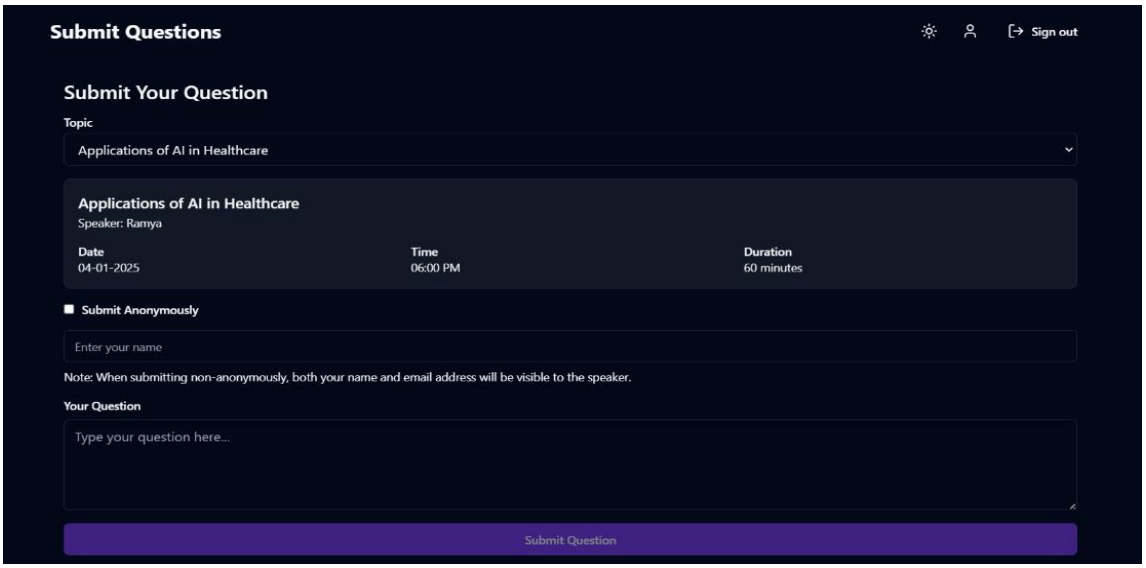




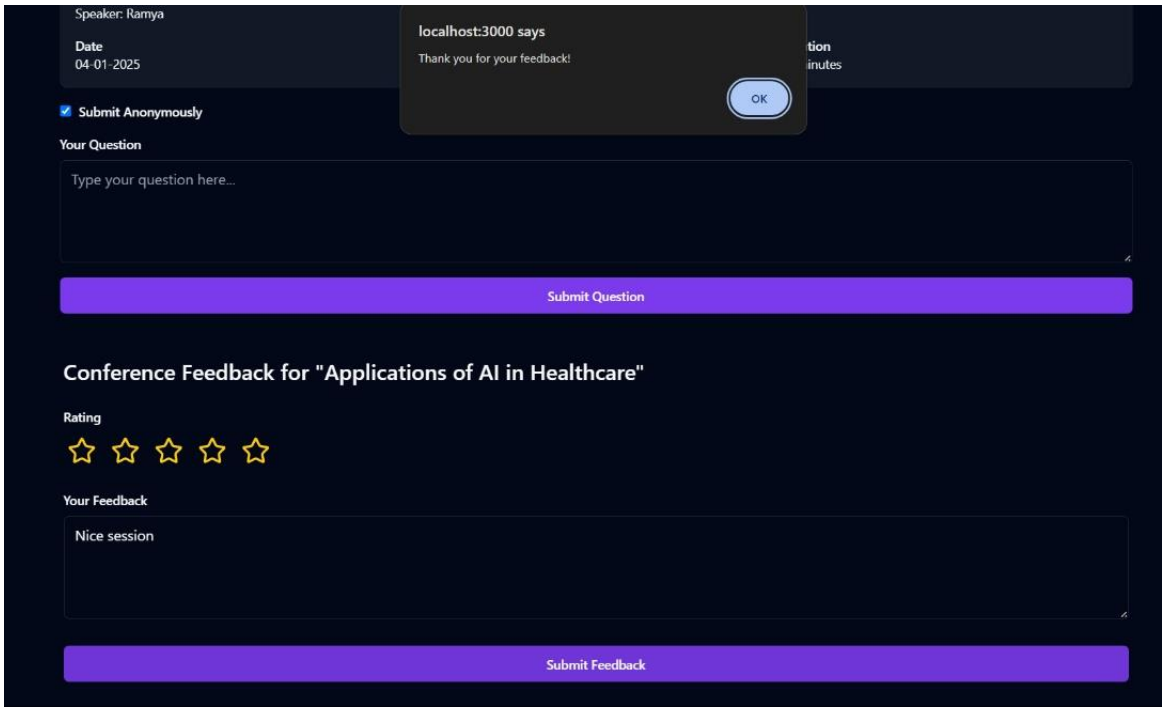


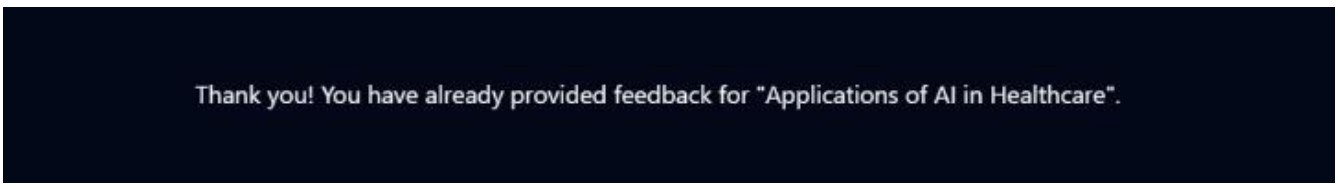
Page visible to the Attendee. They have a choice to choose a topic and submit the questions anonymously or non-anonymously.



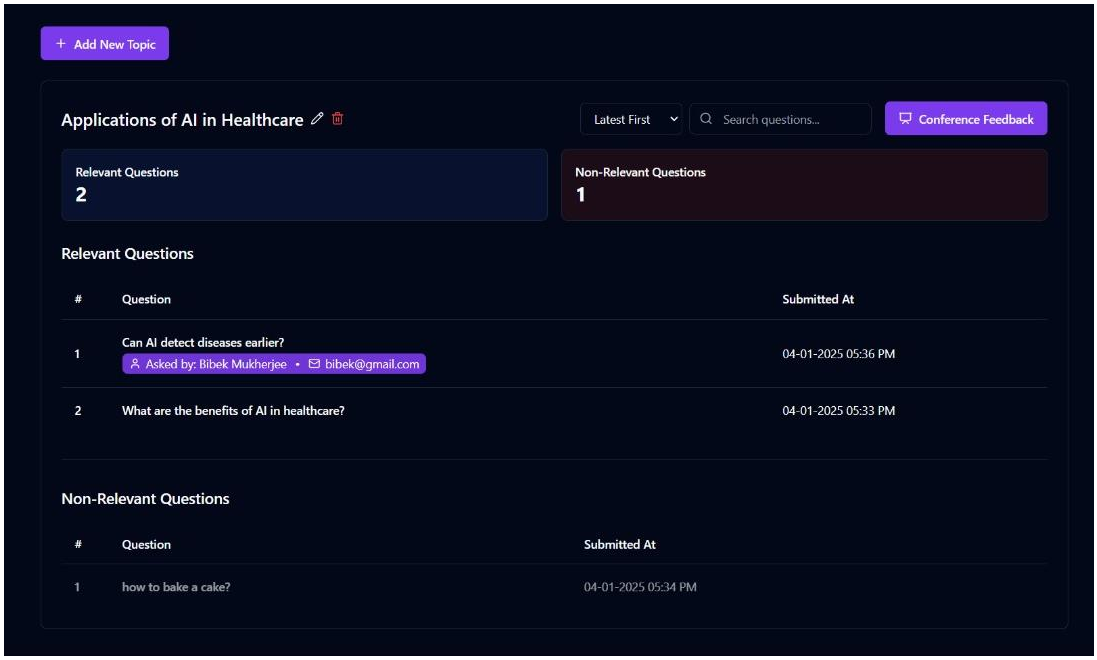


The attendees are given the option to give anonymous feedback about the session. They are permitted to provide feedback on a specific topic only once.

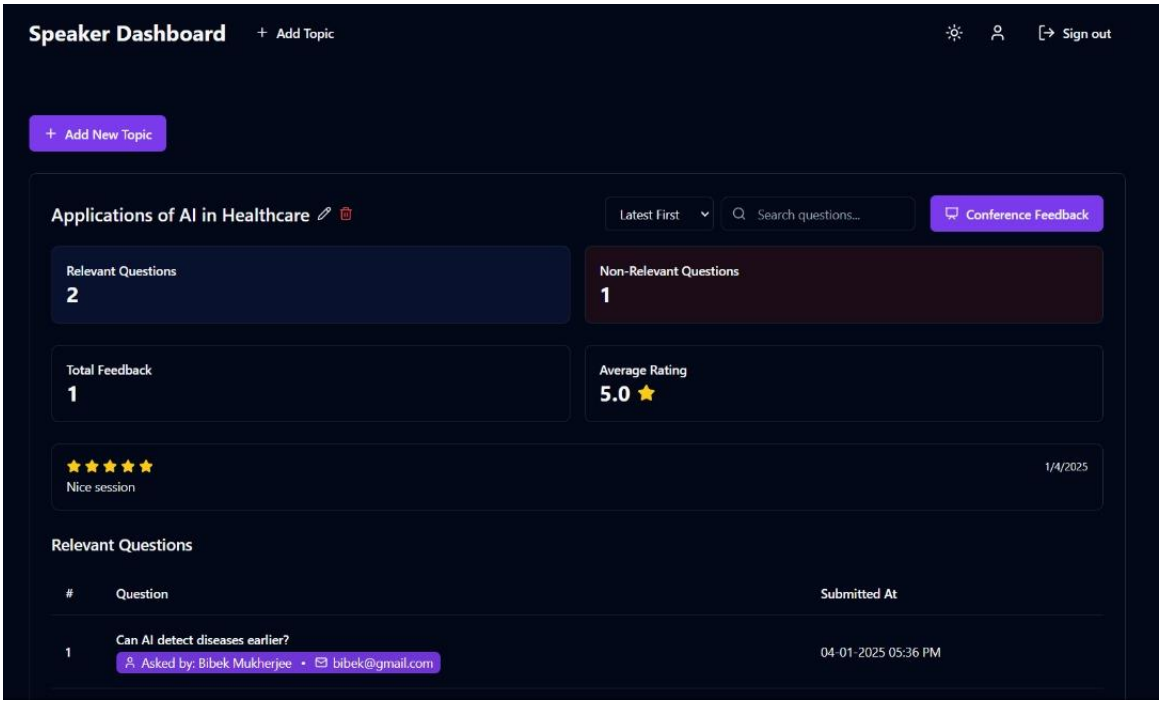




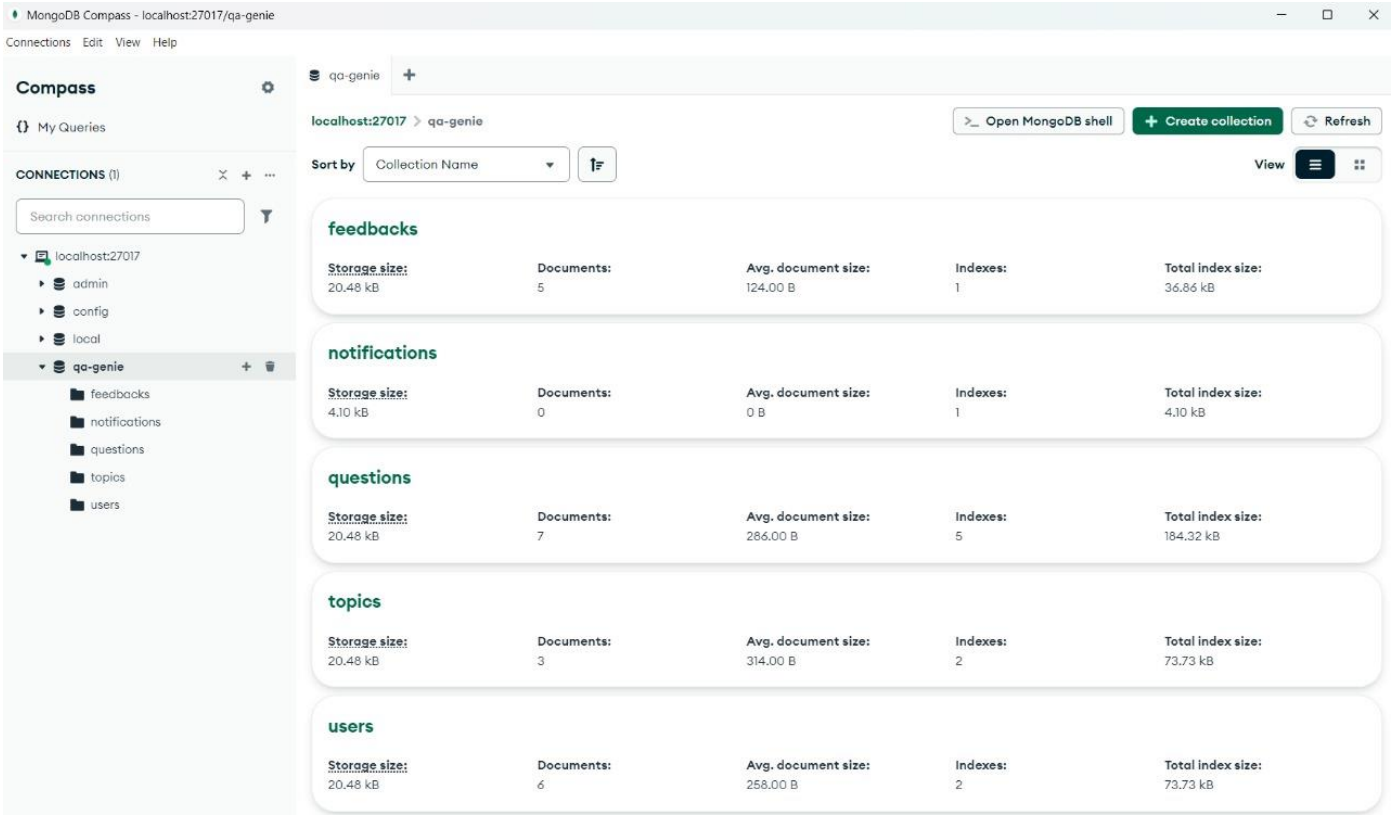
The visuals of Speaker's dashboard. The first question has been submitted anonymously, while the second question has been submitted with their name, while it automatically fetches their email id.



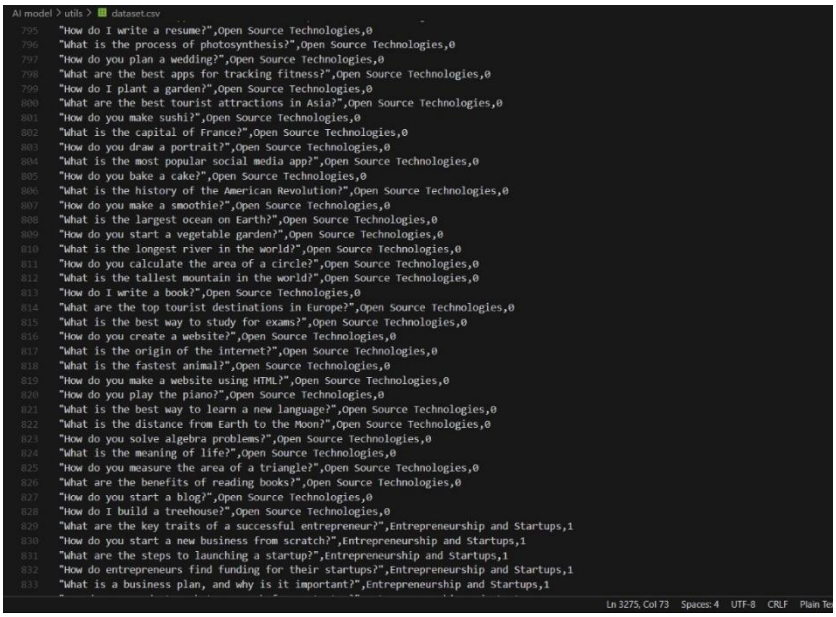
The feedback submitted by the Attendees is visible to the Speaker.



Backend stored data in MongoDB



A few examples of the datasets used to train the AI Model



AI model > utils > dataset.csv

```
"How can businesses build a diverse pipeline of talent?",Diversity and Inclusion in the Workplace,1
"How can organizations address gender inequality in the workplace?",Diversity and Inclusion in the Workplace,1
"What are the challenges of creating an inclusive workplace in multinational companies?",Diversity and Inclusion in the Workplace,1
"How can employers accommodate the needs of diverse religious groups?",Diversity and Inclusion in the Workplace,1
"What are the benefits of promoting mental health awareness as part of diversity?",Diversity and Inclusion in the Workplace,1
"How can organizations create inclusive learning and development programs?",Diversity and Inclusion in the Workplace,1
"What role do employee networks play in advancing diversity and inclusion?",Diversity and Inclusion in the Workplace,1
"How can organizations encourage open conversations about diversity?",Diversity and Inclusion in the Workplace,1
"How can companies track the diversity of their workforce?",Diversity and Inclusion in the Workplace,1
"What are some ways to prevent tokenism in diversity and inclusion efforts?",Diversity and Inclusion in the Workplace,1
"What impact does diversity and inclusion have on employee engagement?",Diversity and Inclusion in the Workplace,1
"How can leaders promote diversity and inclusion through their actions?",Diversity and Inclusion in the Workplace,1
"How can diversity and inclusion initiatives support career growth for employees?",Diversity and Inclusion in the Workplace,1
"Why is allyship important in promoting diversity and inclusion?",Diversity and Inclusion in the Workplace,1
"How can companies incorporate diversity and inclusion into their marketing strategies?",Diversity and Inclusion in the Workplace,1
"How can organizations use data to improve diversity and inclusion efforts?",Diversity and Inclusion in the Workplace,1
"What role does training play in fostering inclusion in the workplace?",Diversity and Inclusion in the Workplace,1
"How can organizations address pay equity as part of diversity and inclusion?",Diversity and Inclusion in the Workplace,1
"How can remote work affect diversity and inclusion in the workplace?",Diversity and Inclusion in the Workplace,1
"How can businesses adapt diversity and inclusion programs in the post-pandemic world?",Diversity and Inclusion in the Workplace,1
"What is the capital of France?",Diversity and Inclusion in the Workplace,0
"How do you make a chocolate cake?",Diversity and Inclusion in the Workplace,0
"Who invented the light bulb?",Diversity and Inclusion in the Workplace,0
"What are the different types of pasta?",Diversity and Inclusion in the Workplace,0
"How does a car engine work?",Diversity and Inclusion in the Workplace,0
"What is the longest river in the world?",Diversity and Inclusion in the Workplace,0
"How do I start a vegetable garden?",Diversity and Inclusion in the Workplace,0
"What is the process of photosynthesis?",Diversity and Inclusion in the Workplace,0
"How do you calculate the area of a triangle?",Diversity and Inclusion in the Workplace,0
"What is the difference between a crocodile and an alligator?",Diversity and Inclusion in the Workplace,0
"How do you cook rice?",Diversity and Inclusion in the Workplace,0
"What is the capital of Japan?",Diversity and Inclusion in the Workplace,0
"What is the distance from the Earth to the Sun?",Diversity and Inclusion in the Workplace,0
"How do you set up a Wi-Fi network?",Diversity and Inclusion in the Workplace,0
"What are the top tourist destinations in Europe?",Diversity and Inclusion in the Workplace,0
"What is quantum physics?",Diversity and Inclusion in the Workplace,0
"How do you use a smartphone?",Diversity and Inclusion in the Workplace,0
"What is the human digestive system?",Diversity and Inclusion in the Workplace,0
"How do I write a resume?",Diversity and Inclusion in the Workplace,0
```

AI Model Testing

```
Testing:
Question: What is cell?
Topic: Biology
Expected: Relevant
Actual Result: {
  "confidence": 69.99,
  "model_path": "C:\\Users\\bibek\\Desktop\\AI model\\model\\model\\relevance_model",
  "question": "What is cell?",
  "result": "Relevant",
  "score": 0.6999,
  "topic": "Biology"
}


Testing:
Question: Why is my name Ankita?
Topic: Biology
Expected: Not Relevant
Actual Result: {
  "confidence": 99.99,
  "model_path": "C:\\Users\\bibek\\Desktop\\AI model\\model\\model\\relevance_model",
  "question": "Why is my name Ankita?",
  "result": "Not Relevant",
  "score": 0.0001,
  "topic": "Biology"
}


Testing:
Question: What's the population of Tokyo?
Topic: Public Health in a Post-Pandemic World
Expected: Not Relevant
Actual Result: {
  "confidence": 99.99,
  "model_path": "C:\\Users\\bibek\\Desktop\\AI model\\model\\model\\relevance_model",
  "question": "What's the population of Tokyo?",
  "result": "Not Relevant",
  "score": 0.0001,
  "topic": "Public Health in a Post-Pandemic World"
```

Detailed AI Analysis while submitting questions

```
Question: Why is my name Ankita?
Topic: Biology
Topic terms: set()
Initial model score: 0.00017290526011493057
Similarity score: 0.0
Final score: 0.0001
127.0.0.1 - - [04/Jan/2025 17:48:39] "POST /predict HTTP/1.1" 200 -

DEBUG - predict.py:
Question: What's the population of Tokyo?
Topic: Public Health in a Post-Pandemic World
Topic terms: {'global', 'global health organizations', 'priorities', 'access', 'nations', 'change', 'improve', 'the future', 'community heal
th workers', 'countries', 'disparities', 'prolonged social distancing', 'social determinants', 'burden', 'practices', 'policy', 'technology'
, 'play post', 'covid 19', 'post-pandemic public health', 'successful post-pandemic recovery', 'post-pandemic recovery', 'what measures', 'c
ooperation', 'vulnerable populations', 'trust', 'the lessons', 'the role', 'recovery', 'vaccination', 'long term', 'pandemic public', 'effec
ts', 'shortage', 'the best practices', 'collaboration', 'covid-19', 'health systems', 'post', 'health misinformation', 'the burden', 'the po
st-pandemic recovery', 'surveillance', 'term', 'future', 'the continuity', 'non-communicable diseases', 'crisis', 'workers', 'telemedicine',
 '19', 'we', 'services', 'system', 'social inequalities', 'public health post', 'what strategies', 'systems', 'pandemic recovery', 'climate'
, 'post-pandemic', 'pandemic public health', 'healthcare', 'the economic impacts', 'communication', 'the challenges', 'a pandemic', 'communi
ty', 'the covid-19', 'world', 'innovation', 'vaccination coverage', 'future health crises', 'ensure', 'mental', 'role does', 'role', 'public
 health responses', 'responses', 'public health communication', 'crises', 'professionals', 'what', 'public health agencies', 'what role', 'd
oes', 'organizations', 'economy', 'the long-term effects', 'the aftermath', 'health post', 'agencies', 'changes', 'era', 'aftermath', 'vacci
ne hesitancy', 'management', 'public health', 'climate change', 'the key challenges', 'disease', 'the context', 'post pandemic world', 'ment
al health professionals', 'continuity', 'public trust', 'post-pandemic lifestyle changes', 'global health', 'future pandemics', 'impact', 'c
ovid 19 pandemic', 'distancing', 'address', 'pandemics', 'education', 'international health collaboration', 'public health infrastructure',
'governments', 'misinformation', 'implications', 'the covid-19 pandemic', 'pandemic', 'hesitancy', 'lessons', 'pandemic world', 'public heal
th surveillance', 'populations', 'considerations', 'health', 'lifestyle', 'post-pandemic healthcare', 'infrastructure', 'policies', 'the key
 policies', 'preparedness', 'play', 'strategies', 'long', 'coverage', 'determinants', 'a post-pandemic world', 'global collaboration', 'heal
thcare access', 'diseases', 'health disparities', 'context', 'public health policy', 'equitable healthcare access', 'challenges', 'pandemic
healthcare', 'analytics', 'terms', 'play post pandemic', 'social', 'workforce', 'impacts', 'the era', 'underserved populations', 'pandemic p
reparedness', 'the future priorities', 'the impacts', 'essential health services', 'public', 'inequalities', 'mental health', 'hospitals', '
health post pandemic', 'healthcare systems', 'covid', 'data', 'health responses', 'global healthcare systems', 'measures', 'post pandemic',
'post pandemic recovery', 'survivors', 'the impact', 'vaccine', 'the ethical considerations'}
Initial model score: 0.00010603429836919531
Exact matches: 1
Topic keywords: 21
Final similarity score: 0.00558659217877095
Similarity score: 0.00558659217877095
Final score: 0.0001
127.0.0.1 - - [04/Jan/2025 17:48:39] "POST /predict HTTP/1.1" 200 -
```

```
DEBUG - predict.py:
Question: What is cell?
Topic: Biology
Topic terms: set()
Initial model score: 0.9998337030410767
Similarity score: 0.0
Final score: 0.6998835921287536
127.0.0.1 - - [04/Jan/2025 17:48:38] "POST /predict HTTP/1.1" 200 -
```

Snapshots of few of the questions used for accuracy analysis and their analysis results

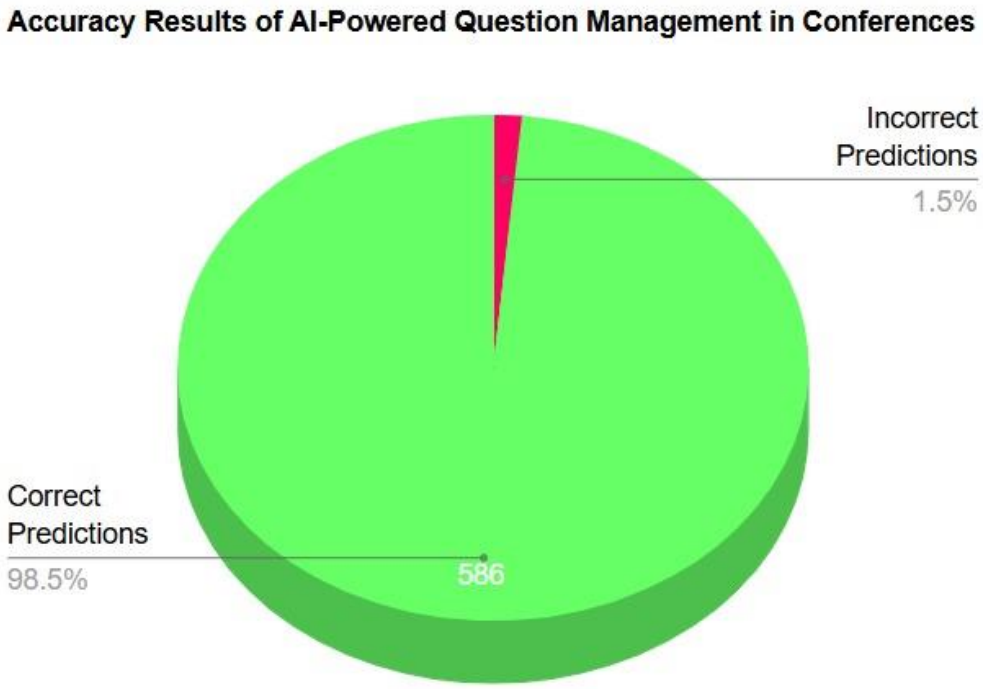
test_results_1000.csv

```
"Question","Topic","Expected","Predicted","Score","IsCorrect"
"What is the periodic table?","Chemistry","Relevant","Relevant","0.6997","True"
"How do chemical bonds work?","Chemistry","Relevant","Relevant","0.6999","True"
"What is a catalyst?","Chemistry","Relevant","Relevant","0.6999","True"
"How to make a sandwich?","Chemistry","Not Relevant","Not Relevant","0.0001","True"
"What is pH balance?","Chemistry","Relevant","Relevant","0.6999","True"
"What's the best smartphone?","Chemistry","Not Relevant","Not Relevant","0.0001","True"
"How do chemical reactions work?","Chemistry","Relevant","Relevant","0.6999","True"
"How to play chess?","Chemistry","Not Relevant","Not Relevant","0.0001","True"
"What is DNA?","Biology","Relevant","Relevant","0.6999","True"
"How does photosynthesis work?","Biology","Relevant","Relevant","0.6999","True"
"What is cellular respiration?","Biology","Relevant","Relevant","0.6999","True"
"How to make coffee?","Biology","Not Relevant","Not Relevant","0.0001","True"
"What is mitosis?","Biology","Relevant","Relevant","0.6999","True"
"What's the best vacation spot?","Biology","Not Relevant","Not Relevant","0.0001","True"
"How do enzymes function?","Biology","Relevant","Relevant","0.6999","True"
"How to learn piano?","Biology","Not Relevant","Not Relevant","0.0001","True"
"What are the barriers to gender equality in leadership?","Gender Equality in Leadership Roles","Relevant","Relevant","0.6999","True"
"How can organizations promote women leaders?","Gender Equality in Leadership Roles","Relevant","Relevant","0.6999","True"
"What is the impact of gender quotas?","Gender Equality in Leadership Roles","Relevant","Relevant","0.6999","True"
"How to bake bread?","Gender Equality in Leadership Roles","Not Relevant","Not Relevant","0.0001","True"
"How to address workplace gender bias?","Gender Equality in Leadership Roles","Relevant","Relevant","0.6999","True"
"What's the best gaming console?","Gender Equality in Leadership Roles","Not Relevant","Not Relevant","0.0001","True"
"How does mentoring support gender equality?","Gender Equality in Leadership Roles","Relevant","Relevant","0.6999","True"
"How to make smoothies?","Gender Equality in Leadership Roles","Not Relevant","Not Relevant","0.0001","True"
"What are the ethical concerns in AI?","Ethical Considerations in Technology","Relevant","Relevant","0.6999","True"
"How can we ensure AI fairness?","Ethical Considerations in Technology","Relevant","Relevant","0.6994","True"
"What is algorithmic bias?","Ethical Considerations in Technology","Relevant","Relevant","0.6998","True"
"How to make pizza dough?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"How to protect user privacy?","Ethical Considerations in Technology","Relevant","Relevant","0.6999","True"
"What's the best movie of 2024?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"What is responsible AI development?","Ethical Considerations in Technology","Relevant","Relevant","0.699","True"
"How to grow herbs?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"How does technology enhance learning?","EdTech: Revolutionizing Education","Relevant","Relevant","0.7062","True"
"What is adaptive learning technology?","EdTech: Revolutionizing Education","Relevant","Relevant","0.7062","True"
```

```
"How to learn guitar?","Internet of Things (IoT) in Modern Applications","Not Relevant","Not Relevant","0.0001","True"
"What are IoT security protocols?","Internet of Things (IoT) in Modern Applications","Relevant","Relevant","0.71","True"
"What's the best coffee maker?","Internet of Things (IoT) in Modern Applications","Not Relevant","Not Relevant","0.0001","True"
"What is responsible AI development?","Ethical Considerations in Technology","Relevant","Relevant","0.699","True"
"How to ensure ethical data collection?","Ethical Considerations in Technology","Relevant","Relevant","0.6999","True"
"How to bake cookies?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"What are privacy-preserving techniques?","Ethical Considerations in Technology","Relevant","Relevant","0.6999","True"
"What's the best movie of 2024?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"How does bias affect AI systems?","Ethical Considerations in Technology","Relevant","Relevant","0.6999","True"
"What is algorithmic accountability?","Ethical Considerations in Technology","Relevant","Relevant","0.6998","True"
"How to make a smoothie?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"What are digital ethics guidelines?","Ethical Considerations in Technology","Relevant","Relevant","0.6999","True"
"What's the best gaming console?","Ethical Considerations in Technology","Not Relevant","Not Relevant","0.0001","True"
"What are the key cybersecurity trends for 2024?","Cybersecurity Trends and Challenges","Relevant","Relevant","0.7146","True"
"How can companies prepare for emerging cyber threats?","Cybersecurity Trends and Challenges","Relevant","Relevant","0.7073","True"
"What are the challenges in protecting cloud environments?","Cybersecurity Trends and Challenges","Relevant","Relevant","0.7102","True"
"How to bake a cake?","Cybersecurity Trends and Challenges","Not Relevant","Not Relevant","0.0001","True"
"Can blockchain improve cybersecurity?","Cybersecurity Trends and Challenges","Relevant","Relevant","0.7117","True"
"What is the capital of France?","Cybersecurity Trends and Challenges","Not Relevant","Not Relevant","0.0001","True"
"How do firewalls help in cybersecurity?","Cybersecurity Trends and Challenges","Relevant","Relevant","0.7087","True"
"How to grow plants?","Cybersecurity Trends and Challenges","Not Relevant","Not Relevant","0.0001","True"
"What are the applications of IoT in smart cities?","Internet of Things (IoT) in Modern Applications","Relevant","Relevant","0.7133","True"
"How does IoT improve home automation?","Internet of Things (IoT) in Modern Applications","Relevant","Relevant","0.7167","True"
"What are IoT-based solutions for healthcare?","Internet of Things (IoT) in Modern Applications","Relevant","Relevant","0.7133","True"
"How to make coffee?","Internet of Things (IoT) in Modern Applications","Not Relevant","Not Relevant","0.0001","True"
"What role does IoT play in autonomous vehicles?","Internet of Things (IoT) in Modern Applications","Relevant","Relevant","0.715","True"
"What's the best recipe for pasta?","Internet of Things (IoT) in Modern Applications","Not Relevant","Not Relevant","0.0001","True"
"How does IoT enable smart grid technologies?","Internet of Things (IoT) in Modern Applications","Relevant","Relevant","0.7184","True"
"How to learn piano?","Internet of Things (IoT) in Modern Applications","Not Relevant","Not Relevant","0.0001","True"
"What is blockchain technology?","Blockchain and Cryptocurrencies","Relevant","Relevant","0.9999","True"
"How do cryptocurrencies work?","Blockchain and Cryptocurrencies","Relevant","Relevant","0.7038","True"
"What are the benefits of blockchain in finance?","Blockchain and Cryptocurrencies","Relevant","Relevant","0.7114","True"
"How to train a dog?","Blockchain and Cryptocurrencies","Not Relevant","Not Relevant","0.0001","True"
"What is Bitcoin mining?","Blockchain and Cryptocurrencies","Relevant","Relevant","0.7056","True"
"What's the best vacation spot?","Blockchain and Cryptocurrencies","Not Relevant","Not Relevant","0.0001","True"
```

Accuracy of the Project

```
Final Results:
Total Tests: 595
Correct Predictions: 586
Accuracy: 98.4873949579832%
Relevant Questions: 426
Non-Relevant Questions: 169
```



VII. CHALLENGES AND LIMITATIONS

The AI-powered question management system, while innovative and efficient, faces several challenges and limitations that need to be addressed for optimal performance and user satisfaction.

1. Dependency on Internet Connectivity: The system relies heavily on stable internet connections for real-time updates and processing. In scenarios where connectivity is limited or unreliable, the system's performance may degrade, leading to delayed updates and a suboptimal user experience.

2. Lack of Multilingual Capabilities: The system currently does not support multilingual processing, limiting its applicability in conferences with non-English-speaking participants or where questions may be submitted in multiple languages. This can result in reduced accessibility for a segment of the audience, potentially excluding valuable insights and queries from diverse participants.

3. Complexity in Handling High Traffic During Peak Periods: As conferences progress, especially during Q&A sessions, the volume of incoming questions can surge significantly. Managing this spike in traffic without degrading the system's responsiveness or overloading the backend servers requires advanced load-balancing techniques and efficient resource allocation, which may still struggle under extreme conditions.

4. NLP Model Biases: The DistilBERT model, like many machine learning models, may inherit biases from the training data. These biases could inadvertently prioritize or exclude certain types of questions, affecting the fairness and inclusivity of the system.

5. Redundancy Detection Limitations: The system's ability to group similar questions relies on semantic similarity detection, which may not always capture nuanced differences between questions. This could result in some unique questions being incorrectly grouped or redundant questions being missed.

6. Integration Complexity with Virtual Platforms: For hybrid or virtual conferences, integrating the system with platforms like Zoom or Microsoft Teams may require significant customization. Ensuring seamless interaction across platforms adds to the development complexity.

7. Privacy and Data Security Concerns: Handling user data, especially anonymous submissions, requires robust security measures to prevent unauthorized access or data breaches. Ensuring compliance with privacy regulations like GDPR adds another layer of complexity.

8. Real-Time Synchronization Challenges: While WebSocket technology ensures real-time communication, the system's dependency on synchronized updates between frontend, backend, and database can be disrupted by network fluctuations. In cases of temporary network failure, users may experience data inconsistencies, where submitted questions or status updates are not immediately reflected across all user interfaces.

9. Ethical Concerns in AI-based Question Ranking: The use of AI for question prioritization raises ethical concerns about transparency and accountability. Since the DistilBERT model is trained on historical datasets, questions with certain wordings or themes might unintentionally be prioritized over others, potentially excluding niche but relevant queries. Transparency in how questions are ranked and clearer explanations of the AI's decision-making process could help mitigate these concerns.

VIII. COMPARISON WITH EXISTING SYSTEM

	EXISTING SYSTEM	PROPOSED SYSTEM
1. Question Filtering and Prioritization	Tools like Slido and Pigeonhole Live allow users to upvote or prioritize questions manually. However, they lack sophisticated AI-driven mechanisms to automatically filter irrelevant or redundant questions.	Uses the DistilBERT NLP model for automated question filtering based on semantic similarity to session topics.
2. Inclusivity for Introverts	While existing platforms allow anonymous question submissions, they often do not actively encourage participation from introverted or hesitant attendees. The lack of anonymity in some tools may discourage engagement.	Strongly emphasizes anonymity as a core feature. Listeners can submit questions anonymously, fostering a comfortable environment for participation, particularly for introverted attendees.
3. Real-Time Processing	Real-time updates in existing tools are often limited to displaying new questions or votes without real-time NLP-based filtering or prioritization.	Implements WebSocket-based real-time updates, ensuring that questions are instantly filtered, grouped, and ranked. Speakers receive a curated list of questions dynamically, enhancing session flow.
4. Role-Based Features	Most systems do not differentiate features based on user roles. Moderators or speakers often share the same interface as attendees, leading to potential inefficiencies.	Offers role-specific features: listeners can submit and track questions, while speakers receive curated dashboards displaying prioritized questions. This tailored approach improves usability and focus for all participants.
5. AI Integration	Few existing platforms leverage advanced NLP or AI techniques for question processing. Filtering and redundancy detection are primarily manual or rudimentary.	Leverages DistilBERT for semantic analysis, ensuring precise filtering and prioritization of questions. This reduces the cognitive load on speakers and moderators while enhancing session quality.
6. Scalability	Scalability is often constrained by reliance on manual moderation and limited backend architecture. Managing large-scale conferences can lead to inefficiencies.	Built using Node.js and MongoDB, the system's architecture is optimized for handling thousands of simultaneous users. Asynchronous programming and distributed database capabilities ensure scalability.

7. Cost Efficiency	Subscription-based models of tools like Slido may involve high recurring costs, especially for premium features.	While initial implementation costs may be higher due to AI integration, the long-term operational costs are optimized by automating key functionalities, reducing dependency on human moderators.
8. Customizability	Limited customization options for tailoring workflows to specific conference needs. Most tools offer fixed feature sets.	Highly customizable with modular components. Features like sentiment analysis or analytics can be added as needed, making it adaptable to diverse use cases.

IX. FUTURE ENHANCEMENTS

To ensure the continuous evolution and adaptability of the AI-powered question management system, several potential enhancements can be implemented. These improvements aim to address current limitations, expand functionalities, and ensure the system remains relevant in diverse use cases.

1. Sentiment Analysis Integration: Incorporating sentiment analysis can provide speakers with insights into the audience's emotional tone. By analysing the sentiment of submitted questions, the system can highlight concerns or enthusiasm, allowing speakers to tailor their responses accordingly.

2. Offline Mode for Question Submission: Developing an offline mode would allow attendees to submit questions even without internet connectivity. These questions could be queued and processed once the connection is restored, ensuring uninterrupted participation.

3. Multimodal Input Options: Enable attendees to submit questions through voice or image inputs. For example:

- Voice-to-text conversion for verbal questions.
- Image-based submission for hand-written or diagrammatic queries, processed using OCR (Optical Character Recognition) and NLP. This feature would make the system more inclusive for differently-abled participants.

4. Enhanced Multilingual Support: Expand the system’s multilingual capabilities to include:

- Context-aware translations that retain the intended meaning of technical or domain-specific terms.
- Regional dialect recognition for more accurate processing. This would ensure greater accuracy and inclusivity for global audiences.

5. AI-Driven Personalization: Introduce AI-based personalization features for attendees, such as:

- Suggested questions based on the session’s topic and past submissions.
- Tailored recommendations for follow-up sessions or related content.

6. Gamification for Audience Engagement: Incorporate gamified elements to encourage participation, such as:

- Points or badges for submitting questions or voting on queries.
- Leaderboards for active participants. These features can make the interaction more engaging and enjoyable.

7. Improved Scalability for Massive Events: Enhance the system's architecture to support events with tens of thousands of participants. This could involve:

- Load balancing across multiple servers.
- Advanced caching mechanisms to reduce latency.

8. Feedback Loop for Continuous Improvement: Incorporate a feedback mechanism where users (both speakers and attendees) can provide suggestions for system improvement. This feedback can be analysed to prioritize future updates.

9. Integration with Social Media Platforms for Broader Engagement: Enable integration with social media platforms to capture a wider range of audience interactions. Attendees could submit questions via social media channels like Twitter, and these could be filtered, processed, and integrated into the session in real-time.

X. Conclusion

The AI-powered question management system presents a transformative approach to enhancing interactions in conferences. By leveraging advanced technologies such as DistilBERT for NLP, real-time processing, and multilingual support, the system addresses significant challenges like question redundancy, irrelevance, and barriers to participation. It fosters inclusivity by enabling anonymous and multilingual submissions, creating a platform that accommodates diverse audiences and encourages active engagement.

The system's modular architecture, comprising React.js for the frontend, Node.js for the backend, and MongoDB for data storage, ensures scalability and efficiency for conferences of all sizes. Tailored features for listeners and speakers optimize the user experience, while real-time updates and prioritization ensure that the most relevant questions are addressed.

Despite its strengths, the system acknowledges limitations such as dependency on internet connectivity and biases in NLP models. Proposed enhancements, including sentiment analysis for emotional tone insights, offline submission functionality, and advanced analytics, promise to improve its capabilities. Multimodal inputs, such as voice and image-based question submissions, could further increase inclusivity for differently-abled participants.

In conclusion, this project not only solves existing gaps in conference question management but also sets a precedent for how AI and NLP can revolutionize audience-speaker interactions. With continuous improvements and adaptability, the system can become an indispensable tool for conferences, webinars, and hybrid events, fostering meaningful dialogues and enhancing overall session effectiveness.

XI. Acknowledgement

The success of the AI-Powered Question Management in Conferences system is attributed to the collective efforts and contributions of numerous individuals and organizations. We express our heartfelt gratitude to the Department of Computer Science – Artificial Intelligence Engineering, Dayananda Sagar Academy of Technology and Management for their unwavering support and provision of resources, which were instrumental in the development of this project. The guidance and mentorship from Ms. Sahana Sharma played a crucial role in shaping the core functionalities of the system and addressing challenges effectively.

We are deeply indebted to the open-source community for providing robust tools and frameworks, including the DistilBERT model, React.js, MongoDB, and Node.js, which served as the backbone of our project. These technologies enabled us to seamlessly implement features like real-time question filtering, anonymous submissions, multilingual support, and future enhancements like sentiment analysis and offline submission functionality.

Special thanks to the conference organizers and attendees who inspired the need for a platform to improve question management, fostering inclusivity and efficiency in live discussions. Their insights and feedback throughout the testing phases were invaluable in fine-tuning the system to meet practical needs.

We also extend our appreciation to our peers and collaborators for their encouragement, constructive criticism, and innovative ideas, which added depth to this endeavour. Lastly, we are grateful for the support from our families and friends, whose patience and encouragement sustained us during this journey.

XII. REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding (Conference Paper)," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, Association for Computational Linguistics, Minneapolis, MN, 2019, pp. 4171–4186.
[2] H. Chen, P. Xu, and J. Liu, "Building interactive applications with React (Book style)," in *Modern Web Development Techniques*, 1st ed., A. Miller, Ed. San Francisco, CA: O'Reilly Media, 2021, pp. 102–148.
[3] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text (Journal article)," in *Computer Vision and Natural Language Processing*, vol. 12, no. 4, pp. 1–9, Jan. 2016.
[4] J. Mongo, "MongoDB architecture and use cases (Book style)," in *Database Management Systems*, 3rd ed., M. A. Stonebraker, Ed. Boston, MA: Addison-Wesley, 2020, ch. 6, pp. 301–345.
[5] R. Dahl, "Node.js: Building scalable network applications (Book style)," in *Server-Side JavaScript Programming*, 2nd ed., vol. 5, J. Brown, Ed. New York: McGraw-Hill, 2018, pp. 201–250.
[6] T. Wolf, V. Sanh, J. Chaumond, and C. Delangue, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, vol. 1, Association for Computational Linguistics, Hong Kong, 2019, pp. 3645–3650.
[7] S. Bird, E. Loper, and E. Klein, *Natural Language Processing with Python*, 1st ed. Sebastopol, CA: O'Reilly Media, 2009.
[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Mar. 2003.
[9] M. Al-Rfou, B. Choi, R. V. Kleinberg, and A. R. Trivedi, "Character-level Convolutional Networks for Text Classification," in *Proceedings of the 29th International Conference on Neural Information Processing Systems*, vol. 1, 2016, pp. 1–9.
[10] M. Ruder, "Universal Transformers," in *Proceedings of the 2019 Conference on Machine Learning*, vol. 1, Association for Computing Machinery, 2019, pp. 1836–1845.
[11] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems*, vol. 1, 2013, pp. 3111–3119.
[12] H. Lin, X. Zhao, and Z. Lu, "Topic modeling in text mining using TF-IDF and LDA," *Journal of Computational Linguistics*, vol. 8, pp. 134–148, May 2022.