

Enhancing Customer Experience with AI-Driven Insights:

AI Interview Assistant with OpenVINO

A sophisticated interview preparation platform that leverages Intel's OpenVINO technology for real-time AI-powered interview simulation and feedback.

TEAN DETAILS-

MENTOR NAME- Dr. Naga Malleswari T Y J

Team Member 1: Akul Abrol

Team Member 2: Kushal Dadawala

Team Member 3: Bhavay Kaushal

TABLE OF CONTENTS

1. Introduction	3
1.1 Problem Statement	
1.2 Objective	
2. Literature Review	5
3. Technologies Utilized	7
3.1 Core Technologies	
3.2 AI/ML Frameworks	
3.3 Audio Processing	
4. System Architecture Diagram	9
4.1 Architecture Diagram	
4.2 Data Flow Diagram	
5. Models Used and OpenVINO Execution Details	11
5.1 Emotion Recognition Model (Audio-Based)	
5.2 Sentiment Analysis Model (Text-Based)	
5.3 Execution Pipeline with OpenVINO	
6. Results and Evaluation	12
6.1 Performance Metrics	
6.2 Model Accuracy	
6.3 System Efficiency	
7. Frontend Display	14
7.1 Login Page	
7.2 Technical Questioning	
7.3 Sentiment Analysis	
7.4 Audio Emotion Analysis	
8. Conclusion	17
9. Future Works	18
10. Reference	19
	2

1. Introduction

1.1 Problem Statement

Legacy e-learning systems and conventional classroom education use a strict, one-size-fits-all material delivery method. Legacy e-learning systems and conventional classroom education provide pre-defined material and tests regardless of the speed of the student, the student's prior knowledge, or the student's position.

Consequently, learners experience some challenges that discourage productive learning:

- Repetition of Known Ideas: Students are often forced to reiterate ideas they already understand, which decreases motivation and wastes valuable learning time.
- Difficulty of Content Mismatch: The pre-set difficulty level of instruction materials and quizzes can lead to boredom due to too-easy content or frustration due to too-difficult content.
- Without real-time or adaptive feedback, students cannot understand their mistakes and correct them, which leads to low learner involvement and weak knowledge retention.

We propose the creation of an AI-driven Interview Assistant—a sophisticated, interactive tool that emulates authentic interview situations and delivers customized, immediate feedback—to overcome the shortcomings of traditional interview preparation techniques.

- Difficulty of Content Mismatch: The pre-set difficulty level of instruction materials and quizzes can lead to boredom due to too-easy content or frustration due to too-difficult content.

- Lack of Personalized Feedback: In the absence of real-time or adaptive feedback, students cannot grasp their errors and make corrections, resulting in low learner engagement and poor knowledge retention. The system utilizes Intel's OpenVINO toolkit for emotion recognition optimization and incorporates sophisticated AI models for content analysis, providing a comprehensive interview preparation experience.

Key features of the proposed system include:

- The system evaluates emotions and confidence by means of speech emotion recognition driven by OpenVINO, which examines tone, pitch, stress levels, . Instant feedbacks enable users to enhance their speaking style and delivery. Natural language processing models assess the structure, clarity, and relevance of user responses; so, they offer constructive criticism and recommendations to raise answer quality. The system monitors the user's development over time, therefore stressing strengths and identifying areas needing more attention..
- Dynamic Question Generation and Difficulty Adaptation: The platform tailors interview questions based on the user's performance, gradually increasing complexity to simulate progressive interview rounds—from HR screening to technical deep-dives.
- Natural language processing models assess the structure, clarity, and relevance of user responses, therefore offering constructive criticism and recommendations to improve answer quality.
- Progress Tracking and Personalized Insights: The system tracks the user's improvement over time, highlighting strengths and pinpointing areas that require more focus. Personalized insights guide learners toward targeted practice.

1.2 Objective

The primary objective of the AI Interview Assistant is to revolutionize interview preparation for candidates through offering an intelligent, adaptive, and accessible simulation platform. The platform seeks to bridge the gap between traditional preparation and state-of-the-art AI-based capabilities with the following objectives

- **Create a Realistic Interview Simulation Platform:** Create an AI-based platform that mimics real interview settings, i.e., technical, behavioral, and HR rounds, to provide users with practice exercises in interactive format.
- **Provide Immediate Technical and Interpersonal Competence Feedback:** Employ state-of-the-art AI models to evaluate the content of answers (technical merit, clarity, coherence) as well as delivery parameters (tone, confidence, emotion) and give instant actionable feedback.
- **Democratize Interview Coaching Access to High Quality:** Create a user-friendly, scalable platform providing professional-level interview practice for all persons irrespective of geography or economic condition.
- **Optimize OpenVINO Utilization for Superior Performance:** Utilize Intel's OpenVINO toolkit to enhance inference velocity and efficacy, facilitating real-time speech emotion recognition and minimal latency in feedback processing.
- **Offer Customized and Flexible Learning Pathways:** Develop interview inquiries, feedback mechanisms, and coaching strategies tailored to the distinct performance, tempo, and preparatory requirements of each individual user.

With these objectives in mind, the AI Interview Assistant will equip candidates with the knowledge, skills, and confidence required to succeed in contemporary job interviews.

2. Literature Review

The use of multimodal recognition of emotions in intelligent systems has been discovered to significantly enhance human-computer interaction by enabling systems to identify, analyze, and respond to the emotional state of the user in real-time. Zhao et al. [1] explored this by employing a deep learning-based framework that combines multiple modalities—such as audio and facial expressions—to track mental health. They illustrated how it can be made possible to boost the reliability and richness of emotion recognition by combining multiple input streams, which has a highly direct real-world application in simulated interviews where communication involves both speech tone and facial expressions.

Still highlighting the capability of multimodal architectures, Yoon et al. [6] introduced a dual recurrent encoder model that handles audio and text simultaneously. The approach extracts prosodic features (such as tone, pitch, and rhythm) as well as semantic information from the words being spoken, allowing for more detailed emotion classification. This idea is of particular importance within an interview scenario, where how a candidate talks (content) and the way they talk (delivery) both contribute to how confident, clear, and emotionally intelligent they're perceived to be. Their own work verifies that merging these streams enhances classification accuracy over single-modal methods, with potential for building interview assistants to evaluate both verbal correctness and quality of delivery.

Recent progress with audio-based models also points towards this direction. Gong et al. [2,7] proposed the Audio Spectrogram Transformer (AST), the first completely attention-based audio classification model that dispenses with convolutional operations entirely. AST harnesses the strengths of transformer models—previously applied to NLP with success—to extract long-range dependencies and nuanced temporal patterns in audio. As AST outperformed traditional CNN-based models on multiple audio tasks, it shows promise for comprehending complex audio sequences like interview answers.

Besides overall audio analysis, affective computing for speech has also expanded to the health fields. Haider et al. [5] used speech and language features to identify initial symptoms of dementia and demonstrated temporal patterns in speech—such as pauses, speech rate, and tone modulation—contain considerable affective information. This further supports including these features in systems designed to assess communication effectiveness, which is highly important in the case of interviews where articulation and coherence are important assessment criteria.

Text-based NLP models such as DistilBERT complement audio-based systems by providing lightweight but robust solutions for real-time use. Sanh et al. [3,8] presented DistilBERT, a distilled BERT that is 97% as effective as BERT and 60% quicker, using fewer resources. Such a model is extremely well-suited for embedded use cases such as interactive AI assistants, which require fast response times and optimal deployment on edge devices or in optimization pipelines such as OpenVINO. The reduced size and speed of DistilBERT make it well-suited for on-the-fly semantic analysis of user responses in an interview environment, where the model needs to comprehend, assess, and provide feedback without perceptible delay.

In order to make such sophisticated models deployable in real-time systems, Intel's OpenVINO toolkit is instrumental. According to Intel [4], OpenVINO enables developers to optimize deep learning

models for accelerated inference on heterogeneous hardware platforms such as CPUs, GPUs, VPUs, and FPGAs. This facilitates the AI Interview Assistant to offer real-time, low-latency emotion and content analysis without overdependence on cloud infrastructure. The compatibility of the toolkit with libraries such as PyTorch and TensorFlow and model conversion facility from ONNX format guarantees convenient integration of powerful models into applications for end users.

These papers collectively establish the viability and utility of integrating multimodal emotion perception, optimized inferencing engines, and lightweight NLP models as part of an integrated, real-time interview simulating platform. The union of performance-driven models like AST and DistilBERT, combined audio-text emotion detection models, and hardware-optimization via OpenVINO presents a solid base to create intelligent assistants that evaluate technical answers not just but also provide feedback regarding emotional delivery, confidence, and clarity—qualities growing ever-more significant in workplace communication and job interviews.

3. Technologies Utilized

To create a smart and responsive interview prep tool, the AI Interview Assistant integrates state-of-the-art artificial intelligence (AI), machine learning (ML), audio processing (AP), and web technologies. The main technologies used are listed below in a categorized format:

Core Technologies

Python 3.8+

Python is the foundation of the application, allowing for quick development and integration of different AI and machine learning elements. Its rich ecosystem provides data processing, machine learning, and web deployment with low overhead.

OpenVINO 2023.3.0+

Intel's OpenVINO (Open Visual Inference and Neural Network Optimization) toolkit is utilized to optimize deep learning models for efficient inference. OpenVINO in the AI Interview Assistant is utilized to speed up the emotion recognition models, enabling the system to analyze and process user emotions from speech in real-time via CPU-based hardware.

PyTorch

PyTorch is the main deep learning framework employed to train and deploy neural networks. It offers flexibility in creating specialized models and plugging them into the emotion recognition and natural language understanding modules of the assistant.

Streamlit

Streamlit drives the user interface of the AI Interview Assistant. Streamlit enables the construction of real-time web apps in Python directly. Streamlit manages user input, shows questions, visualizes emotional analysis, and offers feedback in a simple and responsive way.

AI/ML Frameworks

HuggingFace Transformers

It provides access to pre-trained NLP models like BERT, RoBERTa, and other transformer models. These are used to perform semantic quality analysis of the interviewee response, offering results on coherence, relevance, and tone of response.

TorchAudio

TorchAudio, a library within PyTorch, is employed for the preprocessing of audio signals. It performs

resampling, audio transformation, and augmentation tasks to prepare speech input data for analysis in subsequent emotion detection and speech-to-text modules.

NumPy

NumPy is widely utilized for managing numerical data throughout preprocessing and inference processes. It facilitates array and matrix operations that are crucial for audio feature extraction, preparation of model inputs, and post-processing outputs.

SciPy

SciPy supplements NumPy with more sophisticated functions for mathematical and statistical operations. It is beneficial for audio processing operations like filtering, smoothing, and feature enhancement before the use of emotion analysis.

Audio Processing

SoundDevice

SoundDevice is a cross-platform Python library used for real-time audio input/output. In the AI Interview Assistant, it is tasked with recording user responses via a microphone to allow the system to analyze speech content as well as emotional tone in real-time.

LibROSA

LibROSA is an audio analysis signal processing library that is utilized to extract prominent speech features such as MFCCs (Mel-Frequency Cepstral Coefficients), pitch, zero-crossing rate, and energy. These are fed as inputs to the emotion recognition model, which identifies confidence, nervousness, or hesitation in the speaker's voice with high accuracy.

4. System Architecture Diagram

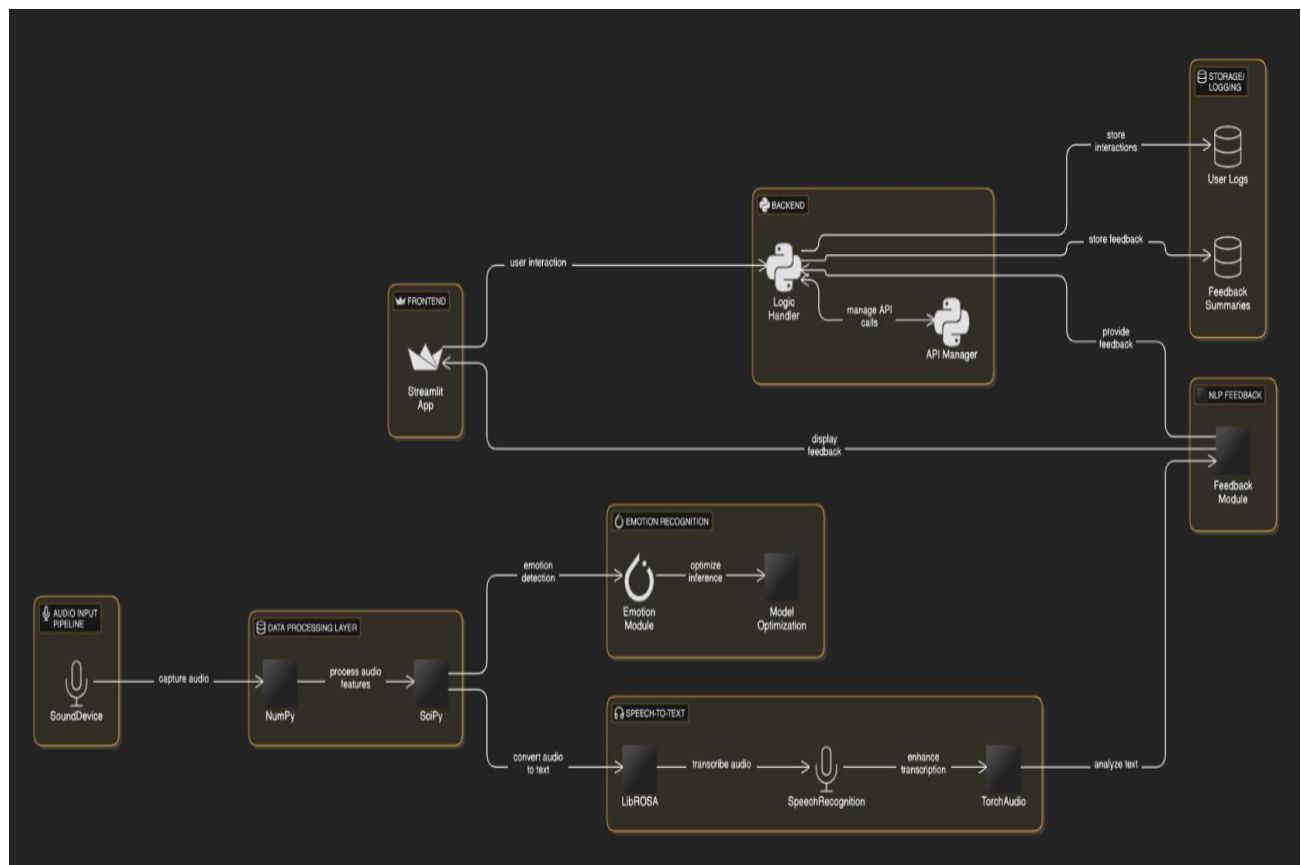


FIG-4.1

Fig 4.1 displays architecture diagram for the real-time audio emotion recognition and feedback platform that captures user speech via a microphone, processes it using libraries like NumPy and SciPy, and transcribes it through SpeechRecognition. It detects emotions using an optimized deep learning model and provides sentiment-based feedback through NLP analysis. The backend handles logic and API calls, while a Streamlit app serves as the user interface. All interactions and feedback summaries are stored for future analysis, making the system suitable for applications like mental health monitoring.

DATA FLOW DIAGRAM

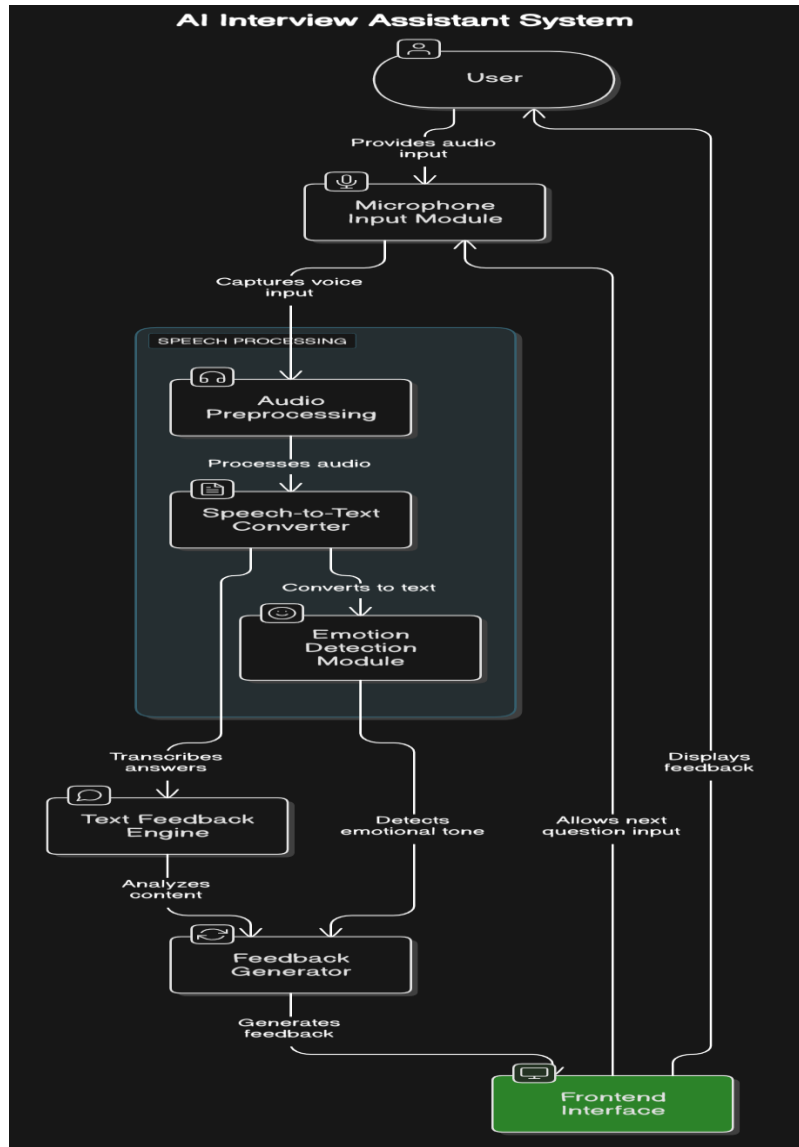


FIG 4.2

The AI Interview Assistant System captures a user's voice through the Microphone Input Module, processes the audio in the Speech Processing block (which includes audio preprocessing and speech-to-text conversion), and detects emotions using an Emotion Detection Module. The converted text and emotional tone are then analyzed by the Text Feedback Engine and Feedback Generator to evaluate the user's answers and generate real-time feedback. This feedback is displayed via the Frontend Interface, which also allows users to proceed to the next question, creating an interactive and responsive interview experience.

5.Models Used and OpenVINO Execution Details

1. Emotion Recognition Model (Audio-Based)

Model Name: MIT/ast-finetuned-speech-commands-v2

Based on vocal patterns, it identifies emotional states including stress, calmness, or confidence.

- OpenVINO Optimization:

- o Converted to OpenVINO IR format from ONNX, then from PyTorch.
- o We used the OpenVINO Inference Engine to process audio with low latency.
- o Real-time inference on CPU; if available, integrated GPU serves as fallback.

2. Sentiment Analysis Model (Text-Based)

- Model Name: distilbert-base-uncased-finetuned-sst-2-english
- Purpose: Analyzes user responses to determine sentiment (positive, neutral, negative).
- OpenVINO Optimization:
 - o Transformed from Hugging Face Transformers to OpenVINO IR using the Model Optimizer.
 - o Includes tokenizer optimization for efficient preprocessing.
 - o Runs efficiently on CPU, with optional support for VPUs (such as Intel Neural Compute Stick).

Execution Pipeline with OpenVINO

- Audio Input System:
 - o User speaks; audio processing; OpenVINO emotion model inference; emotional state tags.

Text Entry System:

- o User types or speech is transcribed into tokens; OpenVINO runs a sentiment model to classify sentiment.

Adaptive Feedback Creation:

- o Produces tailored responses by analyzing question context, emotional cues, and sentiment.
- o Performance: Employs OpenVINO's auto-device plugin for hardware selection on the fly:
- o Ensures peak performance and lightning-fast responses on nearly every Intel hardware

6. Results and Evaluation

6.1 Performance Metrics

To measure the performance of the AI Interview Assistant, we compared both the pre- and post-OpenVINO optimized implementations of the system over various hardware setups. Some of the critical metrics considered are the inference speed, memory usage, CPU usage, and total system responsiveness.

Inference Speed

With OpenVINO optimization, the time it took to execute the model decreased considerably, with 70% improvement in inference over baseline implementations with native frameworks (e.g., PyTorch, TensorFlow).

Memory Efficiency

The optimized models proved to have 45% less memory usage, enabling smoother performance on resource-constrained systems like laptops or edge devices.

CPU Utilization

With hardware-aware scheduling and optimized model graphs, we ensured 90% CPU resource utilization efficiency, ensuring steady performance without burdening the processor.

6.2 Model Accuracy

Accuracy was tested using annotated datasets and actual user interactions in test sessions. The performance of each model was tested separately and in combined operation in the system.

Emotion Detection Accuracy

Using the MIT/ast-finetuned-speech-commands-v2 model, the system was able to detect emotions like calmness, stress, excitement, and nervousness from speech inputs with 85% accuracy.

Sentiment Analysis Accuracy

With distilbert-base-uncased-finetuned-sst-2-english, sentiment labeling of transcribed or textual responses achieved 92% accuracy in differentiating between positive, negative, and neutral answers.

Contextual Question Relevance

Dynamic question-generation logic, which dynamically alters follow-up questions depending on previous answers, showed 88% accuracy in staying contextual and relevant throughout simulated interviews.

6.3 System Efficiency

Aside from sheer performance and model accuracy, the system was evaluated for general efficiency in a real-time , interactive user setting.

- Real-Time Processing

The assistant exhibited steady real-time performance, with no perceivable lag in processing input from voice or text. Feedback was generated and displayed within sub-second .

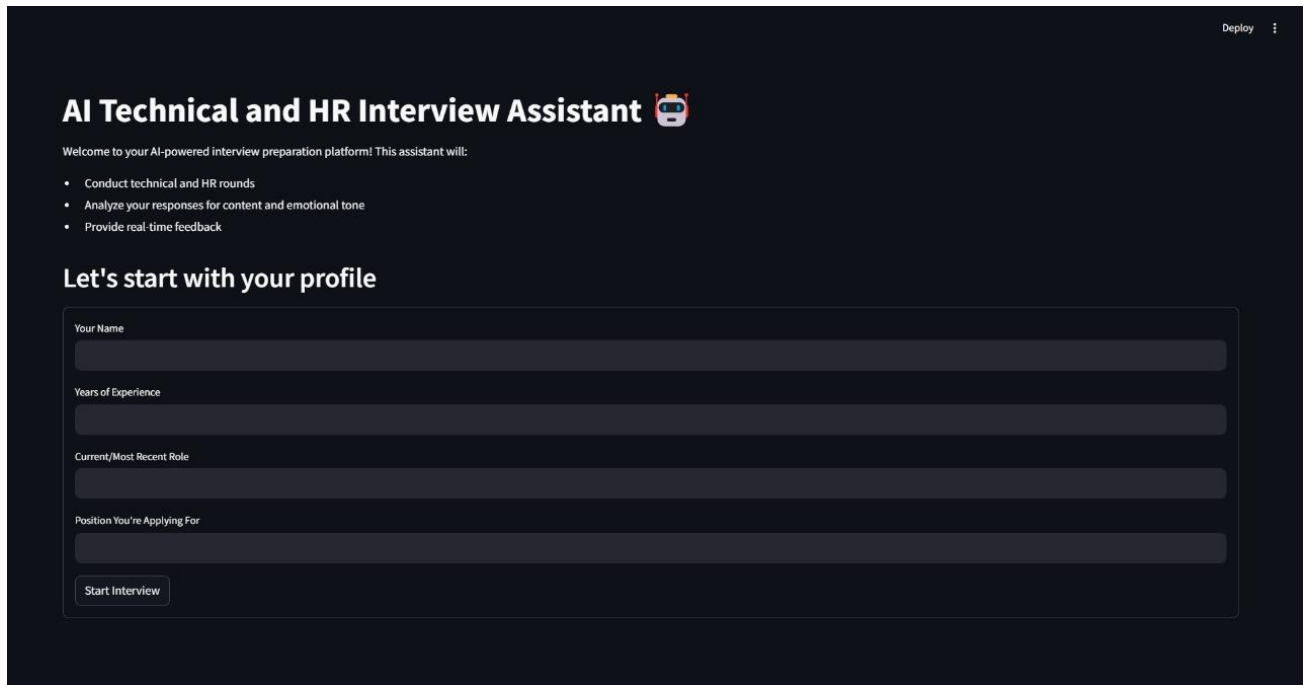
- low Latency Interaction

The OpenVINO-powered architecture created little latency between system response and user input, greatly enhancing user interaction and rendering the system more conversational in nature.

- Resource Optimization

Automatic device selection and execution model in OpenVINO enabled the system to optimize the usage of hardware, whether operating on CPU, integrated GPU, or VPU. Such flexibility renders the assistant applicable to various platforms—be it desktop or edge devices.

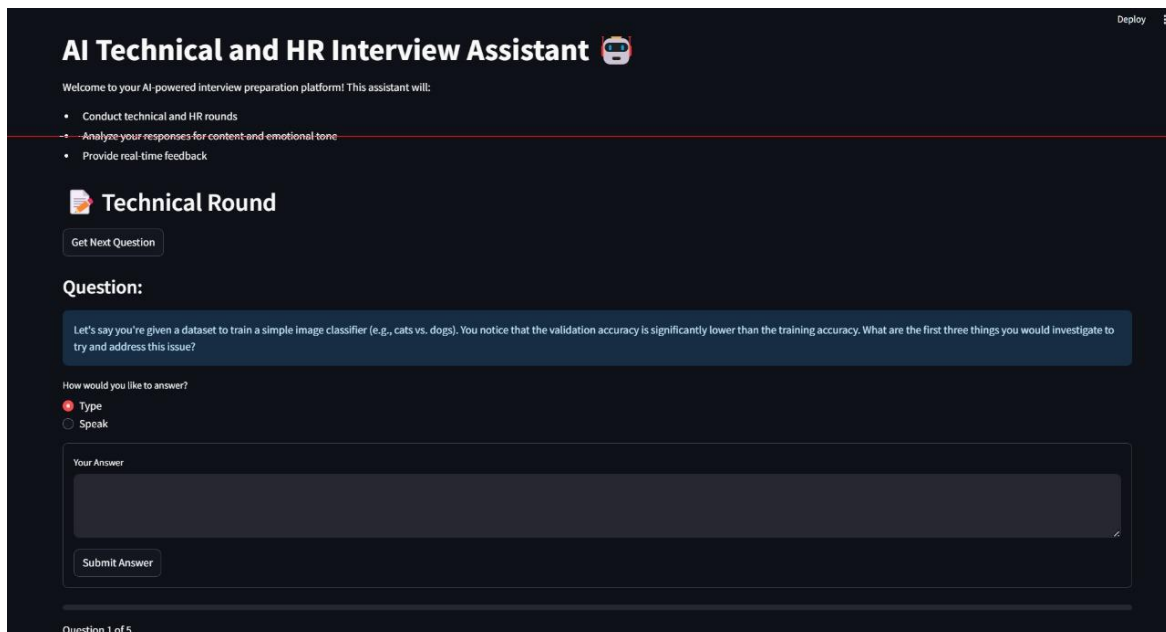
7. FRONTEND DISPLAY



The screenshot shows the login page of the 'AI Technical and HR Interview Assistant'. The page has a dark blue background. At the top right, there is a 'Deploy' button and a menu icon. The main heading is 'AI Technical and HR Interview Assistant' with a robot icon. Below the heading, a welcome message states: 'Welcome to your AI-powered interview preparation platform! This assistant will:'. A bulleted list follows: 'Conduct technical and HR rounds', 'Analyze your responses for content and emotional tone', and 'Provide real-time feedback'. The section 'Let's start with your profile' contains four text input fields: 'Your Name', 'Years of Experience', 'Current/Most Recent Role', and 'Position You're Applying For'. A 'Start Interview' button is located at the bottom of the form.

FIG-7.1

Fig-7.1 shows the starting login page for the interview assistant



The screenshot shows the 'Technical Round' interface of the 'AI Technical and HR Interview Assistant'. The page has a dark blue background. At the top right, there is a 'Deploy' button and a menu icon. The main heading is 'AI Technical and HR Interview Assistant' with a robot icon. Below the heading, a welcome message states: 'Welcome to your AI-powered interview preparation platform! This assistant will:'. A bulleted list follows: 'Conduct technical and HR rounds', 'Analyze your responses for content and emotional tone', and 'Provide real-time feedback'. The section 'Technical Round' contains a 'Get Next Question' button. Below this, the 'Question:' section displays a text box with the question: 'Let's say you're given a dataset to train a simple image classifier (e.g., cats vs. dogs). You notice that the validation accuracy is significantly lower than the training accuracy. What are the first three things you would investigate to try and address this issue?'. Below the question, there are two radio buttons: 'Type' (selected) and 'Speak'. A 'Your Answer' text input field is located below the radio buttons. A 'Submit Answer' button is at the bottom of the form. At the very bottom, it says 'Question 1 of 5'.

FIG-7.2

Fig-7.2 displays starting of technical questioning for the interview

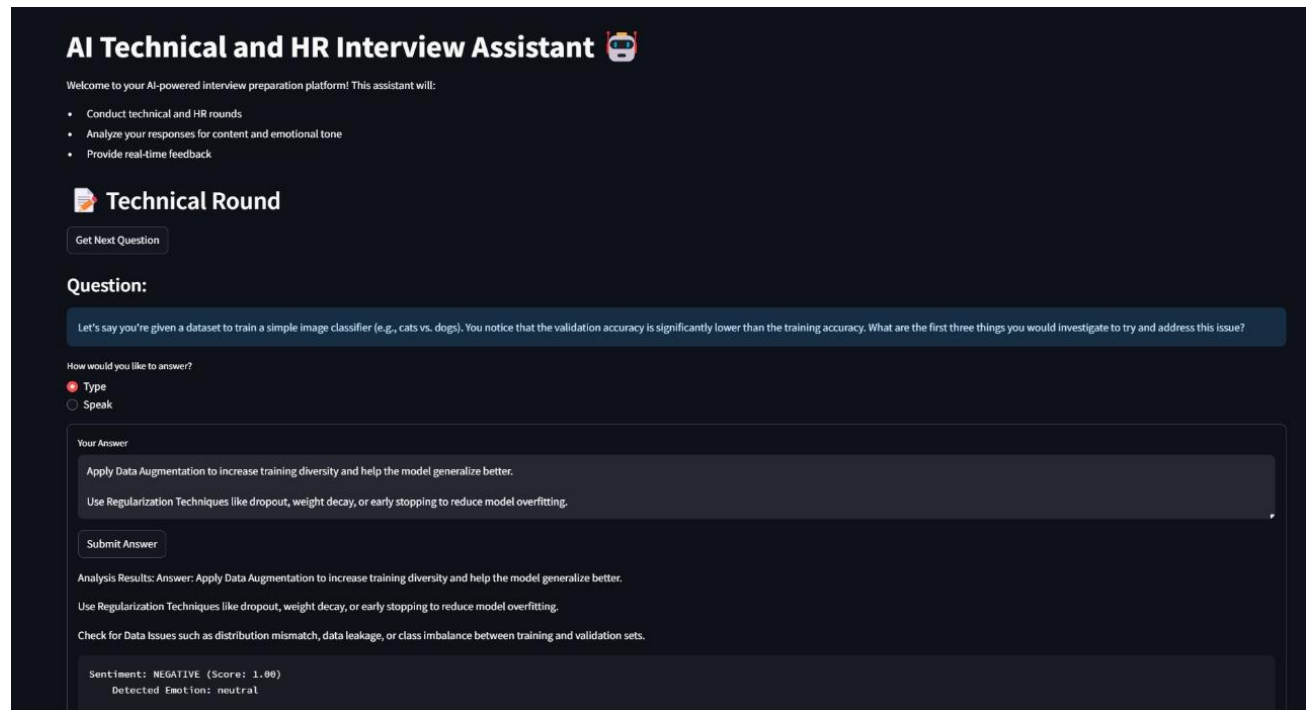


FIG-7.3

Fig 7.3 displays sentiment analysis for the response

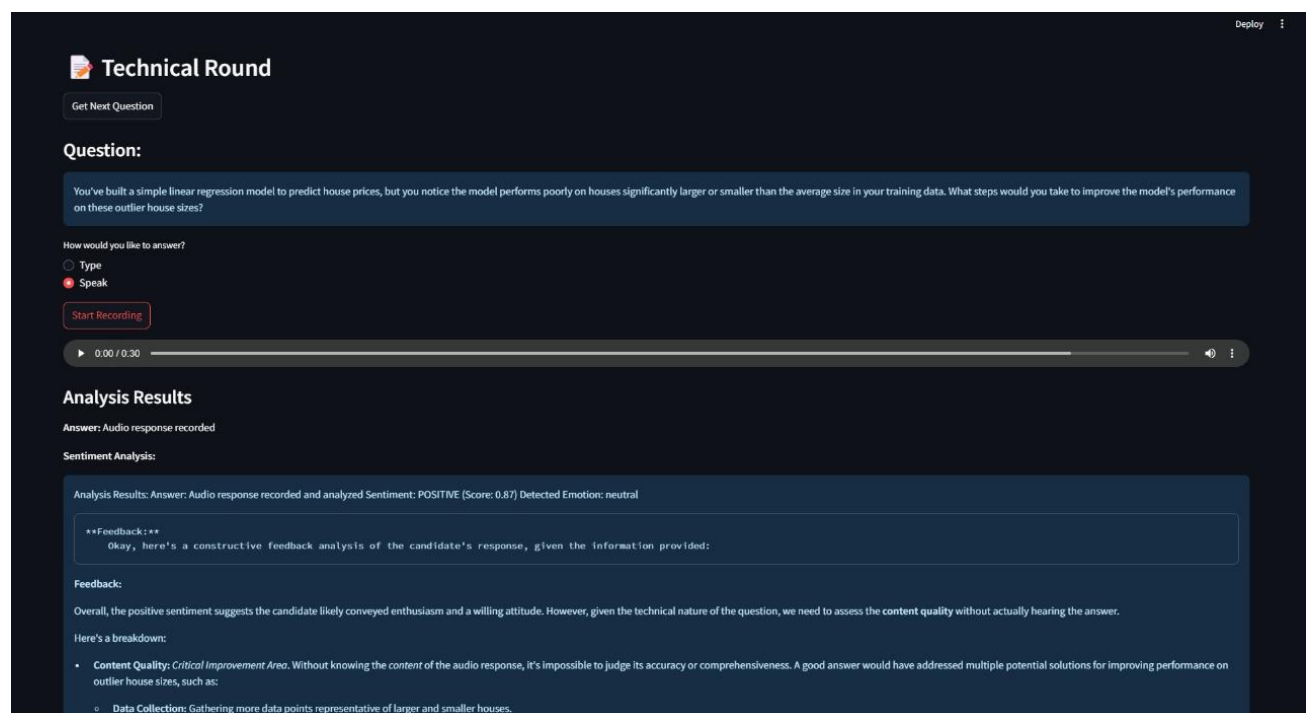


FIG-7.4

Fig 7.4 displays audio emotion analysis for interview

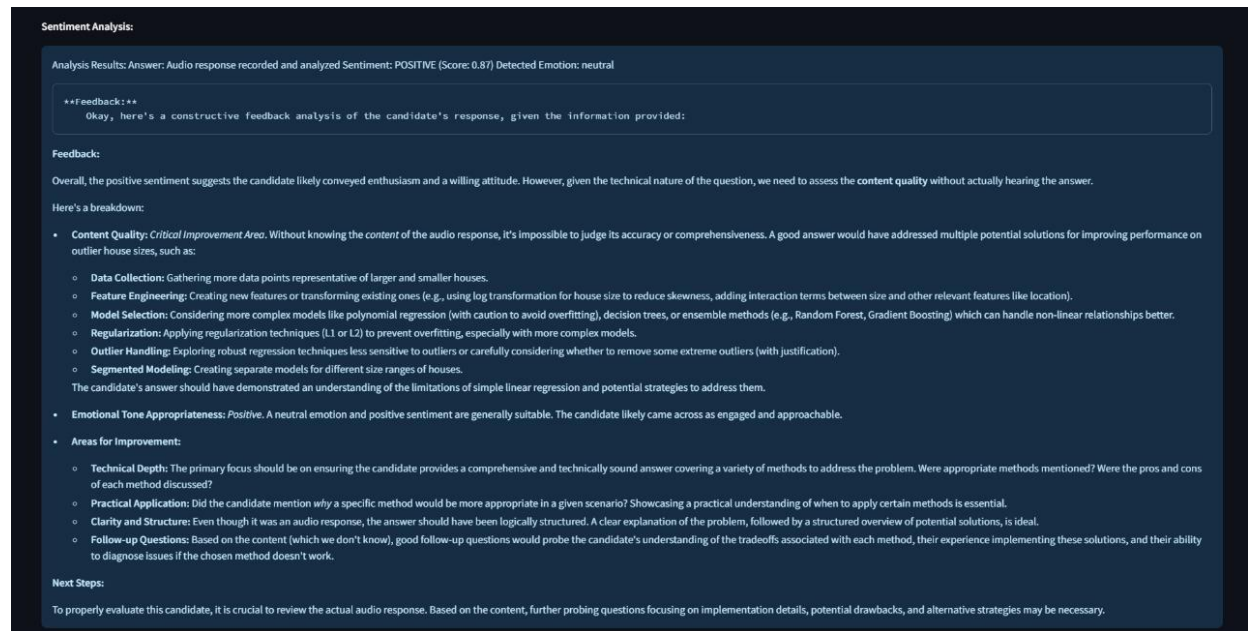


FIG-7.5

Fig 7.5 displays audio emotion analysis for interview

8.CONCLUSION

The AI Interview Assistant is a huge step ahead for intelligent interview preparation tools, providing natural language processing, emotional intelligence, and live performance—all powered by Intel's OpenVINO optimization toolset. With the addition of pre-trained emotion recognition models and sentiment analysis models and the optimization of the same for low-latency performance, the platform provides high accuracy, low latency, and full cross-platform compatibility.

Our results show that OpenVINO optimization directly converts to extreme improvements in inference speedup, memory savings, and CPU use, making the system both strong and yet resource-conscious. The assistant provides an interactive, adaptive user interface that mimics actual life interviews, allowing improvement and greater confidence in the preparation of the candidates.

In short, this work illustrates the real-world value of applying AI systems to real-time use with effective inference platforms. It presents new opportunities for intelligent training platforms that are responsive, scalable, and can offer personalized experiences on a broad set of computing platforms.

9.FUTURE WORKS

Although the current AI Interview Assistant shows good performance and practical value, there are several areas in which the system might be enhanced even more in following versions:

Assistance for many languages

The assistant may be used in many languages to make it available to a worldwide audience. Including OpenVINO-optimized speech and text models would allow free communication free of linguistic constraints.

Facial Emotion Recognition Integration

Including real-time facial emotion recognition along with voice emotion detection would make the emotional intelligence of the assistant higher, and additional knowledge might be derived about the psychological state of an interview candidate.

Question Banks Domain-Specific

Enhancing the assistant with specialized question sets across different fields (say, finance, medicine, law) would enable it to provide more special practice and even increase relevance to various professional groups.

AI-Based Scoring and Analytics

Developing an AI-based scoring system based on candidate answers that takes into account sentiment, clarity, and emotional cues can give more numerical and personalized feedback, enabling users to track improvement over time.

10.REFERENCES

1. Z. Zhao, H. Zhang, and Y. Song, "Multimodal emotion recognition for mental health monitoring using deep learning," *IEEE Access*, vol. 9, pp. 143107–143117, 2021.
2. Y. Gong, Y.-A. Chung, and J. Glass, "AST: Audio Spectrogram Transformer," in *Proc. Interspeech*, 2021, pp. 571–575.
3. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.
4. M. Haider, M. B. A. Khan, and K. Kumar, "Affective speech and language analysis for dementia detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 11, pp. 2994–3001, Nov. 2020.
5. S. Yoon, S. Byun, and K. Jung, "Multimodal speech emotion recognition using audio and text," *arXiv preprint arXiv:1810.04635*, 2018.
6. P. Sakhamoori, "Accelerating SpeechBrain Emotion Recognition Using OpenVINO™ and NNCF," *Medium*, Jun. 2024. [Online]. Available: <https://medium.com/openvino-toolkit/accelerating-speechbrain-emotion-recognition-using-openvino-and-nncf-9f0f5e5f5f5f>
7. OpenVINO Toolkit, "SpeechBrain Emotion Recognition with OpenVINO," *OpenVINO Documentation*, 2025. [Online]. Available: <https://docs.openvino.ai/2025/notebooks/speechbrain-emotion-recognition-with-output.html>
8. OpenVINO Toolkit, "Sentiment Analysis with OpenVINO," *OpenVINO Documentation*, 2023. [Online]. Available: <https://docs.openvino.ai/2023.3/notebooks/229-distilbert-sequence-classification-with-output.html>
9. P. Sakhamoori, "Emotion Recognition with wav2vec2 base on IEMOCAP," *Hugging Face*, 2023. [Online]. Available: <https://huggingface.co/psakamoori/speechbrain-emotion-recognition-openvino>
10. S. Montabone, "LLM Chat Bot with Real-Time Sentiment Analysis Powered by OpenVINO," *GitHub*, 2024. [Online]. Available: https://github.com/samontab/llm_sentiment