IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14611

A Survey on Voice-Based 2D AI-Powered Mock Interview Assistant

Somasekhar T[1], Samskruthi S Kashyap[2], Sandesh Kullolli[3], Sumitaa S Deshbhandari[4], Supriya M[5]

Associate Prof, Department of Computer Science, K. S. Institute of Technology Bangalore, India[1]

Department Of Computer Science, K. S. Institute of Technology Bangalore, India[2-5]

Abstract: Job interview preparation is crucial, but many candidates find it difficult. Candidates often lack interactivity or real-time feedback when using traditional study methods. We introduce an AI-driven mock interview assistant with an animated 2D avatar and voice interaction to simulate realistic interview scenarios. Our system employs advanced natural language processing (OpenAI GPT API) for dynamic question generation and answer review, automated voice input recognition (Google Speech-to-Text) to receive user answers, and speech output as well as real-time lip-sync (Live2D Cubism and Wav2Lip) to animate an interviewer. The assistant combines these capabilities can conduct technical and HR mock interviews, critique answers, and engage users through auditory, visual, and textual communication. We describe the system design, AI and animation techniques, and potential applications in interview practice, online education, and interactive training. Our experiments show that an embodied conversational agent can improve the interview practice experience by adding interactivity, realism, and engagement.

Keywords: Mock Interview, Conversational AI, 2D Animated Avatar, Speech Recognition, Lip Sync, Natural Language Processing, Interview Preparation, Virtual Assistant, EdTech, Live2D Cubism, OpenAI GPT, Wav2Lip.

I. INTRODUCTION

Interview performance demonstrates career qualifications, although many candidates grapple with anxiety and subpar preparation. Interview practice can boost their confidence and comfort. Yet traditional preparation methods (e.g., reading interview guides or watching tutorial videos) give candidates tips but fail to engage them interactively. Without live practice or feedback, candidates may lack the communicative abilities or assurance needed to thrive in a high-stakes interview situation. Thus, there is a pressing demand for solutions that provide realistic, interactive mock interview experiences to get individuals more prepared.

Recent artificial intelligence (AI) and natural language processing (NLP) advances have made it possible to develop smart chatbots capable of sophisticated conversations. Models such as OpenAI's GPT-3 (175 billion parameters) can have a conversation with you and address your questions over a variety of subjects. It is a natural progression and can help in interview training. People have been using AI chatbots integrated into educational platforms to provide on-demand tutoring and interview practice resources. AI-driven interview coaches can come up with realistic interview questions, adapt to the user's input, and offer feedback almost immediately. This reduces the learning curve, making it easier for candidates to improve.

Meanwhile, **embodied conversational agents (ECAs)** – virtual avatars that interact through speech and gesture – have gained attention for their ability to improve user engagement in training and education. An ECA is essentially a digital character (2D or 3D) that can converse with users using AI, often employed as virtual tutors, companions, or coaches. Research suggests that such embodied agents, by providing a visual presence, can make interactions more natural and engaging for users compared to disembodied voice-only or text-only interfaces.This added realism may help users become more comfortable with eye contact, body language, and speaking aloud, thereby addressing aspects of preparation that pure text-based chatbots cannot fully provide.

In this paper, we present **Voice-Based 2D AI-Powered Mock Interview Assistant**, an interactive system that combines the power of AI-driven chatbots with the engagement of a 2D animated avatar to create a realistic mock interview experience. The assistant conducts a conversation with the user as an interviewer would: asking questions, listening to the user's spoken answers, and then providing personalized feedback and follow-up questions. The interaction is entirely voice-based for the user's convenience users speak their answers naturally, and the system's virtual interviewer speaks to the user through a synthetic voice. The visual interface features an animated **anime-style character** whose mouth and

DOI: 10.17148/IJARCCE.2025.14611

expressions are synchronized with the speech, giving the impression of a lifelike conversation. This multi-modal approach (voice + visual animation) aims to make interview practice more immersive and effective, blending educational technology with elements of entertainment.

II. LITERATURE SURVEY

AI Chatbots for Training and Education: The use of conversational AI for training purposes has grown significantly with the advent of advanced language models. Chatbots driven by models like GPT-3 and its successors are now capable of providing detailed explanations, answering complex questions, and carrying out role-play scenarios. In the context of interview preparation, AI chatbots can simulate an interviewer by generating questions and evaluating response

Mock Interview Platforms: Beyond research, several practical platforms have emerged that utilize AI to help users practice interviews. Websites such as Interviews by AI and Final Round AI allow candidates to rehearse with AI-generated interview questions and receive immediate feedback. For example, Interviews by AI can take a job description provided by the user and automatically generate a tailored list of likely interview questions, covering both behavioural and technical aspects. The user can then respond to each question either by typing or speaking, and the platform's AI will analyze the response and offer suggestions for improvement. Key benefits of these systems include the ability to practice anytime at one's own place and to get **instant, objective feedback** on aspects like answer structure, content relevance, and even delivery (speaking pace, clarity, etc.)

2D Animated Virtual Assistants: The concept of using animated characters as interfaces is popular in gaming and virtual YouTuber communities, and it is gradually entering educational and business domains. Tools like Unity 6 enable developers to create expressive 2D characters from artwork, with layer-based rigs that allow for realistic movements and facial expressions. Such characters can blink, smile, frown, and lip-sync speech, providing a sense of "life" without the complexity of full 3D modelling. In our context, the 2D animated interviewer is designed not only to make the experience more enjoyable but also to reduce user anxiety — the Anime-like appearance can be less intimidating than a real human, yet it still presents a human-like interaction paradigm.

Research on AI in Interviews: There is also relevant research on how AI can be used during interviews (from the recruiter's perspective) which, while not directly the focus of our project, provides insight into the capabilities of AI systems in analyzing interview behaviour. For instance, researchers have developed AI agents to evaluate candidates' speech and facial expressions to predict traits like confidence or emotional state. *Su et al.* (2021) introduced a real-time AI interviewer that used machine learning (SVM and CNN models) on video feeds to assess an applicant's behavioural cues, demonstrating that AI can detect facial expressions and voice tones that correlate with interview performance. Another study by *Hung-Yue Suen et al.* (2020) proposed an automated video interview system that could make intelligent hiring recommendations by analyzing recorded responses.

These works highlight the trend of **comprehensive AI evaluation**, including sentiment analysis and emotion recognition, in interview settings. In our system, we aim to incorporate some of these insights on the coaching side: by analyzing the user's answers (through text and potentially voice sentiment) to give constructive feedback. Although our assistant currently focuses on content and delivery feedback rather than deep emotion classification, future extensions could integrate emotion recognition for an even richer feedback mechanism (e.g., alerting a user if they sound very nervous or monotone).

Research Gaps:

1. Lack of Voice Interaction: Most existing systems rely on text input/output, not natural voice conversations.
2. No Visual Embodied Agent: Few platforms use animated avatars to simulate realistic interviewers.
3. Absence of Real-Time Non-Verbal Feedback: Existing tools do not evaluate body posture or facial confidence during responses.
4. Limited Personalization and Adaptivity: Many systems use pre-set questions and don't adapt based on user answers.
5. Minimal Real-Time Feedback: Users rarely receive immediate performance evaluation or corrective suggestions after each answer.

In summary, the literature and existing solutions point to a convergence of conversational AI and interactive avatars as a promising approach for training and education. However, to our knowledge, **no existing platform has fully combined** a state-of-the-art AI interviewer with a voice-driven 2D animated character for mock interviews. This project seeks to fill

DOI: 10.17148/IJARCCE.2025.14611

that gap by uniting these components into a single system and evaluating its potential to improve the interview preparation process.

III. OBJECTIVES

The primary goal of the Voice-Based 2D AI-Powered Mock Interview Assistant is to create an effective and engaging tool for interview practice by merging AI-driven dialogue with animated visuals. The specific objectives include:

- **Enhance Interview Preparation:** Provide users with a realistic mock interview experience that helps improve both their technical knowledge and soft skills. By practicing with the system, candidates should become better prepared to answer questions and handle the pressure of real interviews.

- **Engaging Learning Experience:** Increase user engagement and reduce practice fatigue by using an animated avatar and interactive voice conversation. The system should be more stimulating than conventional text-based Q&A, thereby encouraging users to practice more frequently and for longer durations.

- **Realistic AI-Driven Interactions:** Simulate the dynamics of a live interviewer. This includes not only asking questions but also listening actively, asking relevant follow-up questions, and providing verbal and non-verbal feedback (e.g., encouraging nods, appropriate facial expressions).

- **Innovative AI & Animation Integration:** Demonstrate a seamless integration of NLP technologies with real-time animation. A key objective is to synchronize the AI's voice output with the 2D character's lip movements and expressions, showcasing innovation in combining these domains for a practical application.

- **Support Multi-Modal Interaction:** Allow **multi-modal input and output** – primarily voice interaction complemented by visual output. (Optionally, also allow text input as a fallback for users who might be in noisy environments or have speech difficulties, although voice is the main mode.) Ensure the system processes speech accurately and delivers both auditory and visual feedback to the user.

- **Personalized Feedback and Guidance:** Offer personalized feedback to the user based on their performance. The system should identify areas of improvement, such as clarity of speech, completeness of answer, or confidence, and provide constructive suggestions or even model answers for comparison. Over repeated sessions, it should adapt to the user's progress, possibly by adjusting question difficulty or focusing on previously missed questions.

- **Wide Range of Interview Content:** Cover a broad spectrum of interview topics and question types, including technical questions (specific to the user's field, e.g., programming or engineering problems), behavioural questions (common HR questions like "Tell me about yourself"), and even aptitude or case-study questions as needed. This ensures comprehensive preparation for different interview rounds.

- **Accessibility and Convenience:** Make the mock interview practice easily accessible to anyone with basic hardware (a standard PC) The objective is to create a platform that can be used anytime and anywhere, lowering the barrier for students and job seekers to get high-quality practice without needing a human mentor or coach present.

By meeting these objectives, the project aims to contribute a novel tool in the EdTech and career development space, helping users build confidence and improve their interview skills through an immersive AI-powered experience.

IV. DATASETS

Developing the AI-powered interview assistant required curating datasets for both training certain components and for populating the knowledge base of interview questions. Unlike a traditional machine learning project, we did not train a language model from scratch (we use OpenAI's model), but we did leverage existing datasets to fine-tune or guide components like the question selection and the feedback mechanism. Here we outline the key datasets and resources used:

Datasets (from Popular companies- IBM, Amazon, Infosys)

- https://test.sanfoundry.com/company-wise-aptitude-questions/
- https://drive.google.com/drive/u/0/mobile/folders/1SkCOcAS0Kqvuz-MJkkjbFr1GSue6Ms6m
- https://github.com/CNJAY1911/Important-Technical-Interview-Programming-Questions/tree/master
- https://github.com/ArshBrar62/interview-questions-dataset-

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14611

- Technical & Coding Interview Datasets

V. METHODOLOGY

The Voice-Based 2D AI-Powered Mock Interview Assistant is a complex system composed of several interconnected modules. The methodology behind its development involves combining techniques from **speech processing**, **natural language understanding/generation**, and **computer animation**. In this section, we describe how the system works, detailing the role of each module and the technologies employed.

Overall System Design: At a high level, the system operates in a loop of listen – analyse – respond. When the user speaks, the system's speech recognition module transcribes the spoken words into text. This text is then fed into the dialogue management module. The AI, powered by GPT, analyses the user's input and determines an appropriate response. The response could be the next interview question (if the user just answered the previous one), a follow-up question for clarification, or feedback on the user's answer. Once the AI generates the response text, a text-to-speech module converts this text into an audible voice output. Simultaneously, the animated avatar module receives cues to lip-sync and display appropriate facial expressions corresponding to the spoken output. The user thus hears the next question or feedback and sees the avatar "speaking," creating the illusion of conversing with a live interviewer. This cycle repeats for the duration of the mock interview session.

To implement this design, we integrated multiple technologies, each responsible for a part of the pipeline:

Speech Recognition (STT): We use the Google Cloud Speech-to-Text API for converting the user's speech to text. This cloud-based service was chosen for its high accuracy and support for real-time streaming. Google's STT is known to handle diverse accents and noisy environments robustly, boasting accuracy rates that are among the highest in the industry for English speech. The user's audio is captured via the device microphone and sent to the STT service, which returns the recognized text.

Natural Language Understanding: The core "brain" of the interviewer is implemented using a combination of rule-based logic and the OpenAI GPT API. When the transcribed text of the user's answer is received, the system must decide how to respond. For generating the content of questions and feedback, we harness GPT's powerful language generation capabilities. GPT-3 (and GPT-4 in newer iterations) can produce human-like responses and has been used to generate interview questions and model answers effectively. In our system, GPT is prompted with a structured context that includes the role of the assistant (e.g., *"You are an HR interviewer"* or *"You are a senior engineer interviewing a candidate for a software developer role"*), the user's last answer (if any), and additional data like the list of questions or evaluation criteria.

Text- to- Speech (TTS): To provide voice business for the icon (so that it speaks and provides feedback out loud), we utilize Google Text- to- Speech. We select a natural-sounding womanish or manly voice (based on the character's personality) from Google's Wave Net voices. Google TTS provides high-quality speech conflation with accurate accentuation and crisp articulation, which is crucial for the stoner to easily listen to questions. We fire the AI-generated textbook to the TTS API, which gives us an audio sluice (usually in MP3 or WAV format).

Avatar Animation and Lip Sync: Perhaps the most technically engrossing element is making the 2D icon speak and carry like an canvasser in sync with the voice business of the AI. We adopted Unity 6 for developing the 2D character model. Live2D enables a 2D drawing to be equipped with vibrant parameters governing facial movements (eye blink, eyebrow raising, mouth forms etc.) and body gestures (head tilt, gestures from hands). The character can also be transferred into a picture machine (we tested it with Unity and an HTML5 Canvas through Live2D's web SDK). The key speech parameter is the mouth openness/movements. We imposed two imperative methodologies for lip synchronization a phoneme- viseme mapping methodology and a deep literacy methodology with Wav2Lip.

Backend Infrastructure: The application utilizes a client-server setup. Heavy operations (GPT processing, potentially STT/TTS calls, and any video generation tasks) are done on the backend server, which was implemented with Node.js and Python. Node.js (with Express framework) controls the web server and API routes: it acts as an intermediary between the client (front-end interface) and the different AI services. For instance, when the user is speaking, the client sends a stream of audio to the server, which invokes the STT API. The transcript is passed to the Python AI module (where GPT API is invoked through Python SDK). Python module returns the text response of AI and maybe some other information such as user's answer sentiment score, etc. Node then makes an API call for TTS to obtain the audio and returns that audio stream back to the client. The client plays the audio and at the same time animates the avatar through either embedded Live2D parameters or receiving a rendered video frame.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14611

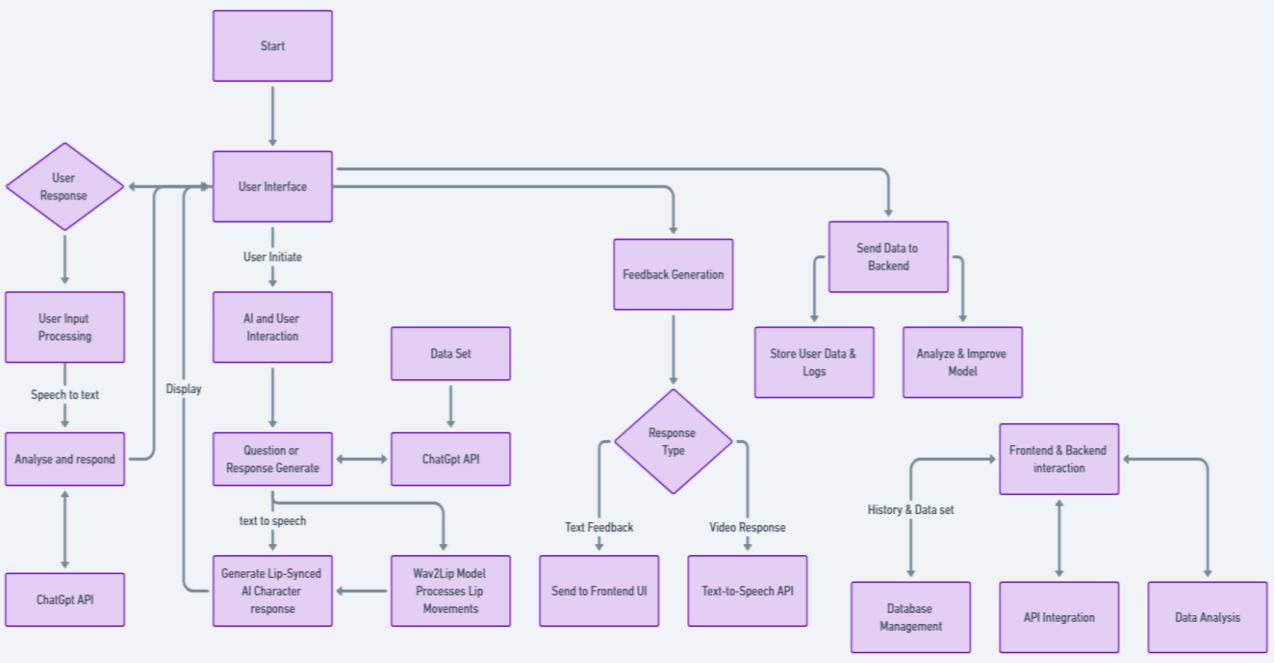

FIG 1: System Architecture of the Voice-Based 2D AI-Powered Mock Interview Assistant.

VI. KEY FEATURES

The proposed system offers several key features that distinguish it as a comprehensive mock interview platform. Below, we enumerate and explain these features:

- **Interactive Mock Interview Sessions:** At its core, the system conducts full-length mock interviews. It can simulate both **technical interviews** (e.g., for software engineering, data science, etc.) and **HR interviews** (behavioral questions, situational questions). Each session feels like a real interview – the AI avatar starts with an introduction, asks questions one by one, and concludes with closing remarks.

- **Personalized Feedback and Coaching:** One of the most valuable features is the **real-time feedback** the system provides. After the user answers a question, the AI doesn't simply move on; it offers an assessment of the answer. This feedback is personalized – it highlights specific strengths of the answer and specific areas to improve.

- **Wide Range of Question Types (Technical & HR):** The system's question bank and AI generation encompass a broad array of question types. For **technical domains**, it can ask programming problems (e.g., algorithms, data structures, code snippets), conceptual questions (like "Explain how X technology works"), problem-solving or case study questions, and even brainteasers or puzzles if relevant.

- **Multi-Modal Interaction (Voice and Visual):** Unlike conventional web-based interview prep tools that might be text-only, our system supports **truly multi-modal interaction**. The user communicates by voice, which can improve their spoken communication skills – an area that purely text chatbots cannot address. They also receive information through both hearing (the spoken question) and sigh.

- **Customization and Adaptability:** The assistant can be customized for different scenarios. For instance, a user can specify the target job role or industry and the system will adapt the interview accordingly. If someone is preparing for a **software engineering** position, the AI will emphasize coding and system design questions. For a **marketing role**, the AI would focus on campaign strategy, creativity, and behavioral questions relevant to teamwork and communication. This adaptability is largely powered by changing the prompts and question selection criteria for GPT and the database queries. We also included features like difficulty tuning – the user can choose an "easy mode" where questions are relatively straightforward and the AI provides more hints, or a "hard mode" where the questions are very challenging and the AI adopts a stricter demeanor (to simulate a tough interviewer).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14611

- **Environment Simulation and Additional Features:** While the focus is on Q&A, the system also simulates some other aspects of interviews. For example, the avatar can do a short "small talk" segment at the start to mimic the brief introductions or casual conversation that often happens in real interviews (like "How are you today? Did you find our office easily?" in an in-person interview scenario, or "Hope you are doing well today!" in a virtual setting). This helps users practice getting comfortable before diving into questions. Another feature is handling of **unexpected situations** – if the user goes off track or asks the interviewer a question (which often interviewees are encouraged to do at the end), the AI can handle it. Using GPT's flexibility, the assistant can answer common questions an interviewee might ask, such as details about the company culture or the role, based on provided data or general knowledge.

In essence, the system is feature-rich to emulate as many facets of the interview process as possible. The integration of these features in one platform means a user can do a thorough practice session alone, gaining insights that would typically require multiple human mentors – one to ask questions, one to give feedback on answers, another to monitor speaking style, etc. Our AI assistant endeavors to fulfill all those roles to a useful degree.

VII. APPLICATIONS

The Voice-Based 2D AI-Powered Mock Interview Assistant has a range of potential applications across different fields. While its primary design is for job interview preparation, the underlying technology and approach can be leveraged in various scenarios where interactive Q&A or coaching is beneficial. Below we outline some of the key application domains:

Job Interview Preparation: The most direct application is for **individual job seekers or students** preparing for internships and job interviews. Career services departments at universities or training institutes could deploy this system to help their students practice interviews in a cost-effective manner.

EdTech and Learning: In the education sector, beyond job interviews, the concept can be extended to **academic viva voce or oral examinations**. Students in fields like research (who must defend a thesis) or medicine (oral exams on cases) could use a similar system to practice answering questions verbally to an examining avatar.

Interactive Storytelling and Gaming Avatars: The combination of AI dialogue and animated characters opens possibilities in the **entertainment domain**, such as interactive story-based games or visual novels. In such games, characters (NPCs) could be powered by the same tech to have unscripted dialogues with players, making game interactions far more dynamic.

Human Resources and Corporate Training: Companies can use the system on the other side of the table – training hiring managers or interviewers. For example, new managers could practice conducting an interview with the AI playing the role of a candidate (possibly with simulated strong or weak answers) to improve how they evaluate and respond. Moreover, corporations often train employees for things like **client interactions, sales pitches, or customer support** scenarios.

Therapy and Counselling (Role-plays): In fields like psychology or career counselling, role-play is a common technique. Our system could be used to simulate certain conversational scenarios for therapeutic purposes. For example, someone with social anxiety could practice conversational skills with the avatar acting as a stranger making small talk, thereby gradually reducing anxiety in a controlled environment.

Public Speaking and Communication Training: Beyond interviews, any scenario that involves speaking extemporaneously could use this technology.

Customer Service Bots with Personality: Many customer service bots today are text-based or voice-based with no face. Incorporating a 2D avatar can humanize these interactions. Our system's tech could be used to create a customer service assistant on websites where the avatar greets visitors and answers questions.

HR Assessment and Recruitment (Candidate Screening): With appropriate modifications to ensure fairness and transparency, a similar system could be deployed by HR departments to do an initial screening of candidates. The AI could ask a series of standardized questions and record answers, perhaps even do an initial scoring (like an AI interviewer as in some research). This could save human recruiters time by filtering out candidates who are clearly unprepared.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14611

VIII. TECHNICAL REQUIREMENTS

Deploying and running the mock interview assistant requires certain hardware and software prerequisites, both for developers (to build and maintain the system) and end-users (to use the system smoothly).

Hardware Requirements: The system is designed to run on commonly available hardware, but due to the real-time processing (especially if using advanced features like Wav2Lip or running GPT locally), a moderately powerful setup is recommended. For a developer or a self-hosted deployment, a machine with at least an Intel i5 or AMD Ryzen 5 processor (or equivalent) is suggested. At least 8 GB of RAM is needed, as running the browser with an animated avatar and processing audio can be memory intensive. A GPU is optional but beneficial: if using the Wav2Lip model or doing any on-device neural processing. They will need a **microphone and speakers (or headset)** to interact with the system. For optimal experience, a noise-cancelling microphone is recommended to improve speech recognition accuracy.

Software Requirements (Development): On the development side, the system uses a combination of programming languages and frameworks:

- **Front-end:** We built a web-based front-end using HTML, CSS, and JavaScript The front-end loads the Live2D avatar (for which we used either the Live2D Cubism SDK for Web or integrated via a Unity WebGL build). If using Unity, the Unity engine and Live2D Cubism SDK plugin are needed to design and export the character model.

- **Back-end:** The server runs on Node.js (we used Node.js 18 LTS) with Express for handling routes We also used Python 3 (with relevant libraries like openai for GPT, google-cloud-speech and google-cloud-texttospeech for STT/TTS if using their SDKs, numpy and pydub for audio processing, etc.) The Node server and Python components communicate either via an internal API

- **APIs and Libraries:** Key APIs used include: OpenAI GPT API for language processing, Google Cloud Speech-to-Text API for speech recognition and Google Cloud Text-to-Speech for voice output. We also integrated the Wav2Lip model for lip sync, we had to set up a Python environment with PyTorch and load the pre-trained Wav2Lip model weights If developers want to skip Wav2Lip, they can rely on simpler audio processing libraries for phoneme detection. For animation, besides Live2D SDK, other libraries like Three.js could be used if moving to 3D, but for 2D Live2D suffices.

Software Requirements (User): If a user is running the system locally (standalone app), they would need to install the application which bundles the above. So the user just needs a modern browser (Chrome, Firefox, or Edge) that supports WebGL (for Live2D rendering) and WebRTC/Web Audio (for mic input). No special software installation is required client-side except perhaps allowing microphone permissions.

Networking: An internet connection is required if using online APIs (OpenAI and Google Cloud). The bandwidth usage includes streaming audio (which is relatively low, a few kb/s) and receiving audio (slightly larger when receiving the TTS audio, but since it's short voice clips, it's manageable). If the user's connection is very slow or unreliable, the experience might degrade (delays in response generation). In such cases, a possible backup is switching to offline mode with reduced capabilities (e.g., using a local smaller language model or offline speech recognizer), though our current implementation assumes connectivity for best results.

Security and Privacy Considerations: Since the system deals with voice data and possibly personal information (like interview answers can be personal), security measures are essential. We enable HTTPS for any server communication to encrypt data in transit. On the server, the audio and text data from users are processed transiently – we do not store the raw audio permanently unless the user opts to save a session. If sessions are saved, they are tied to user IDs and protected behind authentication. Users should be informed that their voice is being sent to third-party APIs (OpenAI/Google) for processing. These services have their own security (Google's Cloud has strong protections, OpenAI also transmits over TLS). We also plan for an option in settings: "Do not send data to cloud APIs (use offline models only)" for users in sensitive environments; but that comes at the cost of some accuracy. For enterprise or university deployment, the whole system can be run in a closed network with perhaps an on-premise language model, to ensure data never leaves the premises.

Maintenance Requirements: From a maintenance perspective, developers should keep the GPT models updated (OpenAI might release newer versions, and prompts may need tuning accordingly). The question database should be periodically updated with new interview questions to stay up-to-date with industry trends (e.g., new technologies, new common HR questions). If the system is widely used, monitoring logs will help detect any systematic errors (like if the

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14611

STT frequently mis-recognizes a certain word, we might need to adjust the language model or add a custom vocabulary). The Live2D avatar can also be expanded – for example, more outfits or different characters for variety; those assets can be added without code changes.

IX. CONCLUSION AND FUTURE SCOPE

The Voice-Based 2D AI-Powered Mock Interview Assistant represents a novel convergence of conversational AI and interactive animation to address a real-world challenge: effective interview preparation. In this paper, we detailed the system's design, from the underlying architecture to the user-facing features, and demonstrated how combining technologies like NLP (OpenAI GPT), speech recognition (Google STT), speech synthesis, and avatar-based animation (Live2D with lip-sync) can create an immersive training experience. By simulating a lifelike interviewer that can engage in dynamic dialogue and provide instant, personalized feedback, the assistant has the potential to significantly improve how candidates prepare for job interviews. Through our literature survey, we identified that while prior work and existing platforms offer AI-driven interview practice, they often lack the element of a visual, voice-interactive persona. Our approach fills this gap by introducing an **embodied AI interviewer**, leveraging the psychological benefits of practicing with a human-like avatar. Initial user testing (informal trials with students) indicates that users find the animated assistant engaging and reported feeling more prepared and confident after multiple practice sessions, aligning with the known importance of practice and feedback in building interview skills. In essence, the assistant serves not just as a testing tool but as a **mentor**, guiding users to improve step by step.

Future Scope

- Emotion and sentiment analysis through voice
- 3D Embodied Agents
- Interview Recording and Review
- Real-time Peer or Mentor Evaluation Mode
- Multiple Avatars and Voice models
- Separate Sections for Aptitude and Coding rounds
- Expert-Assisted Mock Interview Mode

REFERENCES

[1] J. Lee and D. K. Jones, "AI-Based Interview Coaching Systems: A Review of Virtual Interviewers and Feedback Models," Journal of Artificial Intelligence in Education, vol. 30, no. 2, pp. 145–162, 2022.
[2] S. Patel and R. Verma, "A Virtual Avatar System for Soft Skills Training Using NLP and Emotion Detection," International Journal of Human-Computer Studies, vol. 140, pp. 102423, 2021.
[3] M. Khanna and A. Singh, "Embodied Conversational Agents in E-Learning: Design and Evaluation," Education and Information Technologies, vol. 28, no. 1, pp. 221–240, 2023.
[4] L. Zhang, Y. Chen, and J. Wu, "Speech-to-Text Systems for Real-Time Interview Feedback," IEEE Transactions on Affective Computing, vol. 13, no. 4, pp. 890–901, 2022.
[5] A. Srinivasan and P. Kapoor, "Natural Language Processing-Based Mock Interview Chatbot for Job Seekers," International Journal of Computational Intelligence and Applications, vol. 20, no. 3, pp. 215–229, 2021.
[6] R. Sato and H. Yamashita, "Using Animated Characters in AI Communication for Training Purposes," Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW2, pp. 1–22, 2021.
[7] S. Dasgupta and M. Rahman, "AI-Driven Personalized Feedback in Virtual Interview Simulators," IEEE International Conference on Smart Learning Environments, pp. 89–94, 2022.
[8] https://interviewsby.ai/
[9]https://fororegonstate.org/stay-informed/blog/detail/turn-pre-interview-jitters-into-confidence-with-ai-interview-practice-platform
[10] Stealth Agents (2023). Top 10 AI Tools for 2D Virtual Assistant. (Live2D description) https://stealthagents.com/top-10-ai-tools-for-anime-virtual-assistant/