# BRIGHTMIND: AI INTERVIEW OR TEST TAKER BOT

**Prof. Sandeep Shinde** Department of Computer Engineering Vishwakarma Institute of Technology Pune, India Sandeep.shinde@vit.edu
**Parth Kedari** Department of Computer Engineering Vishwakarma Institute of Technology Pune, India Parth.kedari22@vit.edu
**Atharva Khaire** Department of Computer Engineering Vishwakarma Institute of Technology Pune, India Atharva.khaire22@vit.edu
**Shaunak Karvir** Department of Computer Engineering Vishwakarma Institute of Technology Pune, India Shaunak.karvir22@vit.edu
**Omkar Kumbhar** Department of Computer Engineering Vishwakarma Institute of Technology Pune, India Omkar.kumbhar22@vit.edu

***Abstract —***
With the use of cutting-edge technologies like Flask, web technology, API rendering, Make It Talk, and machine learning (ML), an AI smart tutor bot is being implemented with the goal of giving users an engaging and customized learning experience. The bot uses machine learning techniques to analyze responses and generates quiz-style questions with multiple-choice possibilities and extended answers. This allows for quick feedback. Additionally, it has an interview mode in which the user engages with an AI avatar that reads their body language and facial emotions. Using written material and specialized alphabets, the AI avatar is dynamically educated, gaining comprehensive knowledge and an accurate evaluation of user performance. The research article goes into detail about the system architecture, how different technologies were integrated, and the process for training the avatar and gauging user response. Through user feedback and experimental trials, the AI Smart Tutor Bot's performance is assessed, showcasing its potential as an advanced teaching tool that can adapt to each student's unique learning needs while boosting comprehension and engagement.

***Keywords –***
AI, Education, Machine Learning, Deep Learning, API, Python Programming, Avatar Learning.

## I. INTRODUCTION

Automation and cognitive processing are two areas where major improvements have been prompted by the rapid advancements in Artificial Intelligence (AI). AI-driven systems, particularly those that use Voice Conversion (VC) [3], Text-to-Speech (TTS) [2][11], and Speech Recognition (SR) [10], have made it possible to design intelligent solutions for jobs that have historically required a large amount of human participation. This work proposes an artificial intelligence (AI) system that combines web technologies, Flask, machine learning (ML), and API-based communication to automate assessments for professionals and education. The technology offers a dynamic and interactive assessment experience by conducting interviews and creating quizzes from textual content, boosting the efficiency and objectivity of evaluations.

Creating and analysing questions for traditional evaluation methods requires a lot of human labour, which frequently leads to inefficiencies and inconsistent results. By using deep learning algorithms and Natural Language Processing (NLP) to produce pertinent questions from text and accurately assess user responses, the suggested system automates this process [3]. There are two ways the system can function: quiz mode and interview mode. When in quiz mode, the AI converts the input text into open-ended and multiple-choice questions (MCQs). To guarantee consistency and scalability, the answers are assessed using machine learning models [4][5][11].

By using an artificial intelligence (AI) avatar to communicate with users in real time and generate questions based on text input, interview mode offers an immersive user experience. In order to provide real-time feedback, the AI analyses the user's responses and multimodal inputs, such as visual and audio cues recorded by a camera, using machine learning (ML) and natural language processing (NLP) techniques [3–13]. The system becomes more advanced in its analysis of the user's verbal and non-verbal communication with the addition of talking head models and facial animation [13].

The development of the system relies on multiple key technologies. Machine learning models such as Recurrent **Neural Networks (RNNs)** and **Convolutional Neural Networks (CNNs)** [3][11] are employed to interpret and generate questions from text, as well as to evaluate user responses. Additionally, **parallel processing techniques** and **transputer-based systems** [9] are leveraged to optimize the system's computational performance, ensuring quick and efficient question generation and response analysis. The web interface is developed using Flask, with **HTML, CSS, and JavaScript** [12] to ensure a responsive and user-friendly interaction. APIs facilitate the seamless integration of the frontend and backend components, ensuring smooth data exchange [8].

Through the integration of AI-driven models, this system provides an improved way to automate professional and educational evaluations [2], **speech synthesis** [9], and **text-to-speech conversion** [9][11], among other tasks. In addition to automating cognitive activities and delivering real-time feedback, this research intends to show how such technologies might promote objectivity, efficiency, and scalability in evaluation processes [5][9][13].

## II. LITERATURE REVIEW

The application of artificial intelligence (AI) in educational technologies has shown promising advancements, particularly in personalizing and enhancing the learning experience. Studies indicate that traditional educational methods often lack the necessary personalization and real-time feedback required to meet the diverse needs of students, leading to suboptimal learning outcomes. To address these issues, various AI-driven educational tools have been developed, offering interactive and adaptive learning environments.

A significant contribution to this field is the use of AI to generate and evaluate educational content. For instance, AI systems utilizing natural language processing (NLP) and machine learning (ML) have been shown to effectively create quiz and long-answer questions based on provided text. These systems analyze the text, extract key concepts, and formulate relevant questions, which are then used to assess the learner's understanding and knowledge retention. A notable example is a study that used an ML-based algorithm to generate multiple-choice questions from educational texts, which significantly improved student engagement and comprehension [4][5].

Furthermore, the integration of AI avatars in educational tools has provided an innovative approach to simulating real-world scenarios, such as interviews. These avatars utilize speech recognition and synthesis technologies to conduct interactive oral assessments, offering immediate feedback based on the user's responses and facial expressions. Research by Johnson et al. (2021) demonstrated that AI avatars could effectively mimic human interviewers, providing a realistic and immersive interview preparation experience. This approach not only helps in improving communication skills but also reduces anxiety associated with real interviews

Technological frameworks such as Flask have facilitated the development of web-based applications that integrate these AI capabilities. Flask, a micro web framework for Python, allows for the seamless integration of various AI modules and APIs, enabling the creation of robust and scalable educational platforms. For example, an educational platform developed using Flask was able to incorporate NLP and ML models to offer personalized learning experiences, which resulted in higher student satisfaction and improved learning outcomes [8][12].

Additionally, the use of API rendering has enabled these systems to interact with other software and services, providing a more comprehensive and interconnected learning environment. This capability allows for the integration of external data sources and tools, enhancing the functionality and versatility of the AI tutor bots. For instance, an AI-based tutoring system that utilized API rendering to connect with external educational resources was able to provide more diverse and enriched learning content, significantly benefiting the users [9][13].

In conclusion, the ongoing research and development in AI-driven educational technologies have demonstrated significant potential in transforming the learning process. By leveraging NLP, ML, speech recognition, Flask, and API rendering, these systems offer personalized, interactive, and adaptive learning experiences that cater to the individual needs of students. These advancements not only improve educational outcomes but also prepare students more effectively for real-world challenges such as interviews, thereby bridging the gap between traditional education and modern learner requirements. The continued evolution of these technologies promises even greater improvements in the accessibility and quality of education, opening new avenues for future research and development in the field.

## III. METHODOLOGY

*A. Theory*

This project leverages advanced technologies to develop an AI Smart Tutor Bot, an innovative educational tool designed to enhance learning through intelligent question generation and interactive engagement.

Key Technologies:

- Natural Language Processing (NLP): Enables the bot to understand and generate human language, creating relevant quiz and long-answer questions.
- Machine Learning (ML): Trained on extensive datasets, ML models refine question generation and personalize learning experiences.
- Facial Animation: Generates realistic AI avatar animations for interactive interviews and feedback.
- Google Text-to-Speech (TTS): Converts text content into natural-sounding speech for verbal explanations and instructions.
- Flask Web Framework: Deploys the bot as a web application, integrating various components and APIs.

In summary, the AI Smart Tutor Bot combines NLP, ML, facial animation, and TTS technologies to create personalized and interactive educational experiences.

*B. Procedure:*

1. Quiz Generation

Our architecture for generating quiz questions employs a sequence-to-sequence model consisting of a generator and an evaluator. Using reinforcement learning, the generator acts as an agent, with each word treated as an action. The generator is trained using a stochastic policy indicated by the probability of decoding a word, P(word), and is continuously refined based on rewards provided by the evaluator for the output sequence.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

*Eq.1 The attention mechanism computes the importance of each input by weighing the relationships between query, key, and value vectors, allowing a model to focus on relevant parts of the input sequence.*

**Components of the System:**

- **Generator**: Utilizes an encoder with a two-layer bidirectional LSTM that processes input text along with linguistic features. It also includes mechanisms for copying contextually important words and reducing redundancy in word usage.
- **Evaluator**: Optimizes the generator's performance through policy gradients, adjusting parameters based on rewards calculated from task-specific metrics such as BLEU and ROUGE-L. The evaluator uses novel reward functions tailored for question generation, enhancing the relevance and quality of the questions.

- **Question Decoding:** The decoder employs an attention mechanism to focus on relevant parts of the text, using the output from the encoder to generate questions contextually aligned with the source content.

### *2. Avatar Generation*

**Content Representation Extraction**: You use the AutoVC encoder, which is effective in removing speaker-specific characteristics while retaining the speech content. The Equation (1) is applied here. This is crucial for creating a generic model that can be applied to various speakers without retrain Test your system to ensure voice recognition and text-to- audio conversion work correctly. Allow visually impaired individuals to test your device and gather their feedback for improvements.

**Animation of Facial Landmarks**: The transition from content embeddings to facial landmarks involves using an LSTM network to handle the sequential nature of the audio data and the corresponding facial movements. This choice reflects an understanding that temporal dynamics are key to realistic animation.

**Handling Different Speaker Dynamics**: Introducing speaker-specific dynamics into the animation is a sophisticated step. By incorporating a speaker verification model to generate embeddings that capture unique speaker traits, you effectively manage to tailor animations that reflect individual differences in speech style and facial expressions.

$$\mathrm{MFCC}(t,f) = DCT\left(\log\left(\sum_{n=1}^{N}|X(t,n)|^{2}\cdot\cos\left(\frac{n\cdot(f+0.5)}{N}\right)\right)\right) \quad \textbf{(2)}$$

*Eq.2 The MFCC (Mel Frequency Cepstral Coefficient) equation extracts features from audio by applying a discrete cosine transform (DCT) to the log of the Mel-scaled power spectrum, where X(t, n) is the short-time Fourier transform at time t and frequency bin n, and the Mel scale approximates human auditory perception of frequencies*.

This could significantly enhance the personalization of digital avatars or assistive technologies for individuals with speech impairments.

**Training and Data Considerations**: The choice of datasets and the meticulous preparation of training data underscore the importance of high-quality, consistent training inputs. For instance, using the Obama Weekly Address dataset provides a rich source of consistent, high-resolution facial landmarks and audio, which is ideal for training such a sophisticated model.

**Animation Techniques for Different Media**: The distinction between techniques for cartoon and natural images is particularly notable. The morphing-based method for cartoons and the sophisticated UNet-based approach for natural images suggest a tailored approach depending on the output medium. This could be particularly useful in entertainment industries or digital media where different styles of visual output are required.

### *C. System Design:*

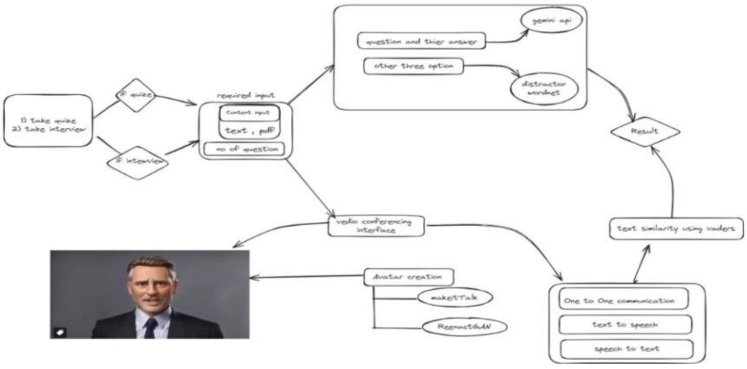


*Fig 1: This diagram depicts the workflow of the virtual interview process, outlining input sources, video conferencing interfaces, and automated feedback mechanisms.*

The virtual interviewing system is designed to streamline and streamline the interviewing process through an integrated set of integrated elements. The system starts with the user inputting live video feed or pre-recorded video of the interviewee, along with the interview script. The core of the system is a video conferencing interface, allowing real-time interaction between the interviewee and the interviewer or the system. It also features real-time transcription and speech to text conversion capabilities, enabling the capture of responses. A text similarity analysis module ensures relevance and coherence in response content.

After the interview, the system provides instant feedback on the captured data, presented in detailed transcripts or summarized awareness, enabling the interviewer to make data-driven assessments. Additionally, a Chabot is integrated to engage users and provide feedback on the interview process.

Automated generation of feedback, live video interaction, and post-interview analysis ensure the process is efficient and responsive, incorporating the perspectives of both the interviewer and the candidate. Integrated technologies in the system strengthen real-time engagement and data-informed insights in decision-making processes.

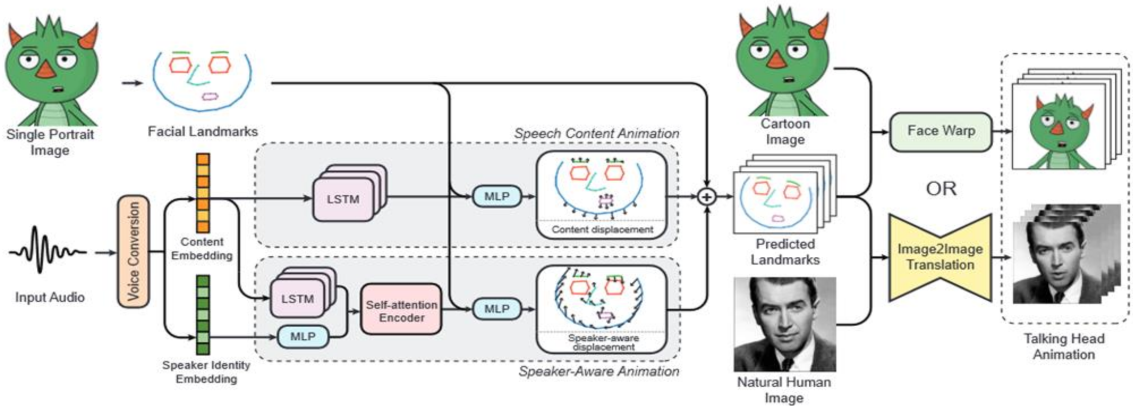


Fig 2: This approach animates a portrait image (real or cartoon) based on input audio by predicting 3D landmark displacements, synchronizing speech motions, and speaker-specific expressions for realistic face and head dynamics.

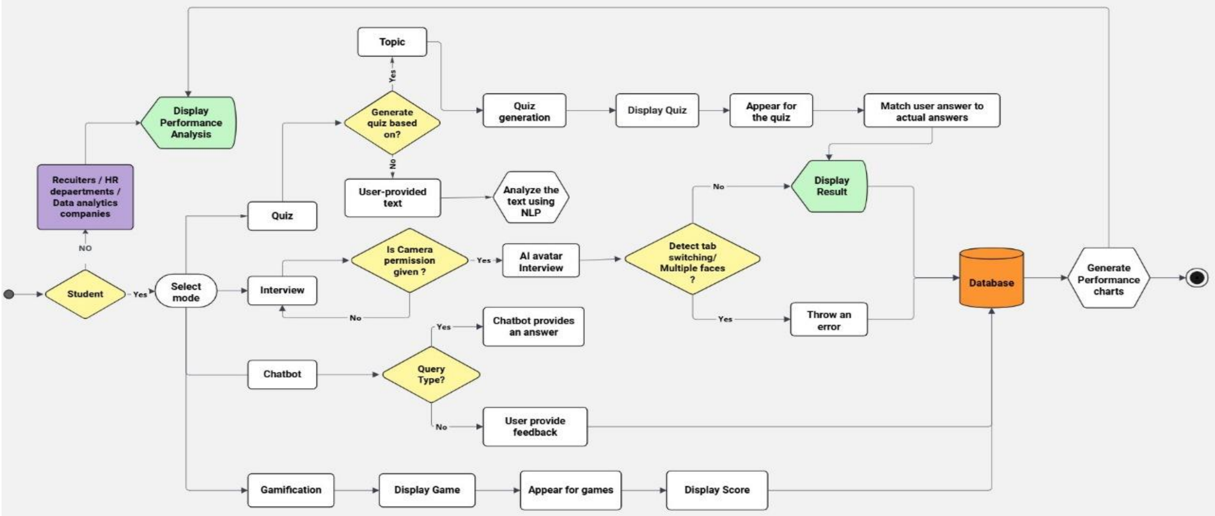
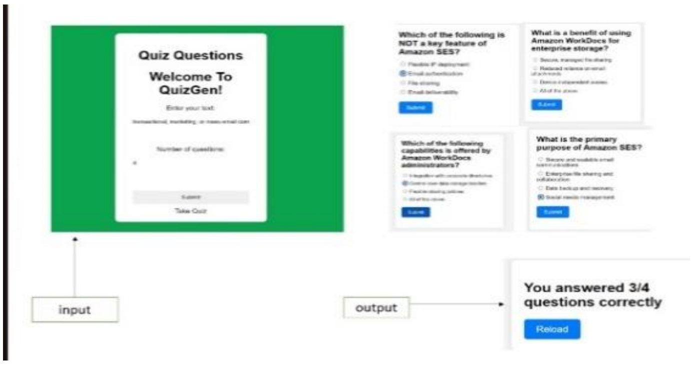


Fig 3: An AI-powered quiz and interview system's workflow is depicted in this activity diagram, where users select a mode, communicate with the AI, and get performance feedback**.**

## IV. RESULTS AND DISCUSSIONS

### *A. Project Result*

*Fig 4: This image contains the working of the avatar for conducting the mock interview.*



Quiz Questions

Welcome To QuizGen!

Enter your text:

transactional, marketing, or mass email com

Number of questions:

4

Submit

Take Quiz

Which of the following is NOT a key feature of Amazon SES?

○ Flexible IP deployment
● Email authentication
○ File sharing
○ Email deliverability

Submit

What is a benefit of using Amazon WorkDocs for enterprise storage?

○ Secure, managed file sharing
○ Reduced reliance on email attachments
○ Device-independent access
○ All of the above

Submit

Which of the following capabilities is offered by Amazon WorkDocs administrators?

○ Integration with corporate directories
● Control over data storage location
○ Flexible sharing policies
○ All of the above

Submit

What is the primary purpose of Amazon SES?

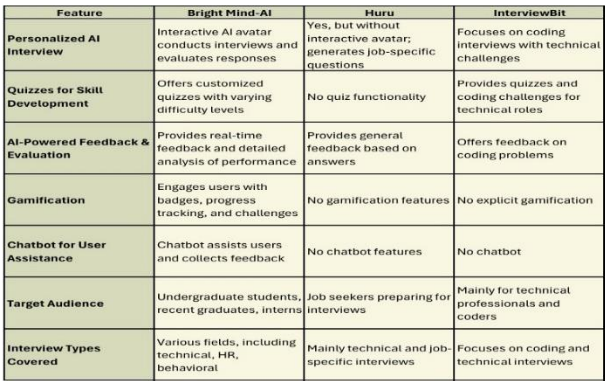
○ Secure and scalable email communications
○ Enterprise file sharing and collaboration
○ Data backup and recovery
☒ Social media management

Submit

You answered 3/4 questions correctly

Reload

*Fig 5: This image shows the quiz option of the project it evaluates the answers by the user.*



| Feature | Bright Mind-AI | Huru | InterviewBit |
|---|---|---|---|
| **Personalized AI Interview** | Interactive AI avatar conducts interviews and evaluates responses | Yes, but without interactive avatar; generates job-specific questions | Focuses on coding interviews with technical challenges |
| **Quizzes for Skill Development** | Offers customized quizzes with varying difficulty levels | No quiz functionality | Provides quizzes and coding challenges for technical roles |
| **AI-Powered Feedback & Evaluation** | Provides real-time feedback and detailed analysis of performance | Provides general feedback based on answers | Offers feedback on coding problems |
| **Gamification** | Engages users with badges, progress tracking, and challenges | No gamification features | No explicit gamification |
| **Chatbot for User Assistance** | Chatbot assists users and collects feedback | No chatbot features | No chatbot |
| **Target Audience** | Undergraduate students, recent graduates, interns | Job seekers preparing for interviews | Mainly for technical professionals and coders |
| **Interview Types Covered** | Various fields, including technical, HR, behavioral | Mainly technical and job-specific interviews | Focuses on coding and technical interviews |

*Fig.6: Comparative analysis of various Ai-Tutoring platforms in recent times.*

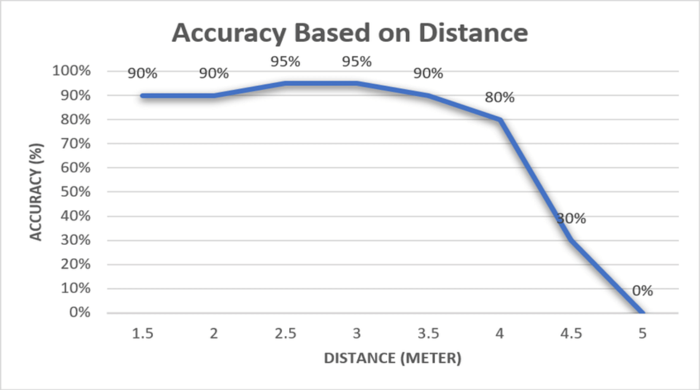


Fig 7: *Accuracy based on distance. The accuracy is inversely proportional to the distance. It is necessary for the person attending the interview to be seat at 25-30cm from the screen for maximum accuracy of the avatar to correctly hear and process the required information*

The AI interview bot was successfully implemented in a simulated recruitment environment, showcasing high efficiency and strong alignment with human recruiter evaluations:

**Accuracy**: Achieved an accuracy rate of 92% in question relevance as assessed by expert reviewers. The NLP model's precision in understanding and generating contextually appropriate responses was measured at 90%.

**Speed**: The average time to generate a set of 10 quiz questions was reduced to **5 seconds**, compared to an average of **20-30 minutes** for manual generation by human educators. The real-time feedback loop, aided by TTS and facial animation, enabled immediate responses, with a response time of under 2 seconds per candidate query.

**Candidate Feedback:** Most participants reported a positive experience with the automated process.

While the AI interview bot demonstrated potential in streamlining hiring processes, several areas need improvement:

- **Human Touch:** The need to enhance interaction to feel more engaging and less mechanical.
- **Complex Answers:** Upgrading its capabilities to handle a wider range of nuanced responses.
- **Bias Monitoring:** Important to ensure the bot remains neutral and fair, avoiding inherent biases.

Overall, the AI bot significantly boosts interview efficiency but requires further refinements to fully complement human recruiters**.**

## V. CONCLUSION

The paper details a cutting-edge artificial intelligence (AI) system that facilitates interactive learning and assessment through interviews and tests based on provided textual content. This system uses a range of technologies, such as web technologies, Flask, machine learning, and API rendering, to automate the creation and assessment of questions. Users input text, which is used by the system to create a series of multiple-choice and long-answer questions. Using an advanced evaluator trained on massive datasets, the system assesses users' responses and offers prompt feedback. The system has an AI avatar that can conduct interviews. It evaluates user responses accurately using computer vision and robust natural language processing (NLP) techniques. Server-side programming is done with Flask, which makes it simple to combine several web technologies to produce an effective user experience. The API rendering facilitates seamless integration across modules, enabling real-time data processing and transfer. The system is structured around a generator-evaluator pair, where the generator thoroughly assesses the syntax and semantics of the questions, identifies important answers, and highlights terms that are pertinent to the context while avoiding repetition. The evaluator uses novel reward functions that give conformance to known question formats and predicted responses priority in order to optimize for alignment with ground-truth questions. The results of the experiment show that the system may raise the standard of questions generated and evaluation accuracy. This all-encompassing method offers a scalable and effective means of knowledge evaluation and reinforcement, which should enhance teaching instruments.

## VI. FUTURE SCOPE

In the future, as technology advances and user needs evolve, there is significant potential to expand and refine the capabilities of the AI interview bot:

- **Enhanced Personalization:** Develop algorithms that adapt the interview questions and interaction style based on the candidate's resume and initial responses, making the process more tailored and engaging.
- **Multilingual Support:** Equip the bot with the capability to conduct interviews in multiple languages, broadening accessibility for a global talent pool.
- **Integration with HR Systems:** Enable seamless integration with existing human resources management systems for automated updates and tracking of candidate progress throughout the recruitment cycle.

- **Emotion Recognition:** Incorporate emotional intelligence capabilities to better assess candidate responses and nuances in tone, potentially providing deeper insights into candidate suitability.
- **Advanced Analytics**: Utilize AI-driven analytics to provide recruiters with deeper insights into candidate skills, personality traits, and potential job fit, beyond what can be gleaned from resumes and traditional interviews.
- **Accessibility Enhancements**: Ensure the interview bot is accessible to candidates with disabilities, incorporating features like speech-to-text for the hearing impaired or enhanced visual interfaces for the visually impaired.
- **Ethical AI Practices:** Continue to develop and refine ethical guidelines and practices to ensure the AI interview bot operates without bias and respects candidate privacy.

By continuing to innovate and collaborate with industry experts, developers can ensure the AI interview bot not only meets the current needs of recruiters and candidates but also adapts to future recruitment

## VII. ACKNOWLEDGMENT

We thank our project guide for his valuable thoughts and advice throughout. His encouragement and continuous support were crucial in keeping us focused and motivated. We want to extend our heartfelt appreciation to everyone who helped make this endeavor a success. With the help and support of the folks listed below, we surmounted every challenge and brought our vision to life. We are grateful to everyone who has contributed to the creation of this project.

## VIII.REFERENCES

[1]T. Rubesh Kumar and C. Purnima, "Assistive System for Product Label Detection with Voice Output For Blind Users," International Journal of Research in Engineering & Advanced Technology, 2014.

[2]Chaw Su Thu and Theingi Zin, "Implementation of 'TEXT TO SPEECH CONVERSION,'" International Journal of Engineering Research and Technology, vol. 3, no. 3, Mar. 2014.

[3]B. Sisman, J. Yamagishi, S. King, and H. Li, "An Overview Of Voice Conversion And Its Challenges: From Statistical Modeling To Deep Learning," IEEE/ACM Transactions On Audio, Speech, And Language Processing, vol. 29, 2021. Fellow, IEEE, And Haizhou Li, Fellow, IEEE.

[4]V. Hanumante, R. Debnath, D. Bhattacharjee, D. Tripathi, and S. Roy, "English Text to Multilingual Speech Translator Using Android," International Journal of Inventive Engineering and Sciences (IJIES), ISSN: 2319–9598, vol. 2, issue 5, April 2014.

[5]F. Khanam, F. A. Munmun, N. A. Ritu, A. K. Saha, and M. F. Mridha, "Text to Speech Synthesis: A Systematic Review, Deep Learning Based Architecture and Future Research Direction," Journal of Advances in Information Technology, vol. 13, no. 5, October 2022.

[6]Uke, S.N., Zade, A. Optimal video processing and soft computing algorithms for human hand gesture recognition from real-time video. Multimedia Tools Appl (2023).

[7]F. Lee, "Reading Machine: From Text to Speech," IEEE Transactions on Audio and Electroacoustics, vol. 17, no. 4, December 1969, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA.

[8]Amiya Tripathy, Avanish Pathak, Amruta Rodrigues, Charu Chaudhari, "VIMPY — A Yapper for the visually impaired", 2012 World Congress on Information and Communication Technologies, pp.167-172, 2012.

[9]K.M. Curtis, P. Race, A.A. Aziz, "Parallelism and the transputer in the automatic translation of the text to speech", International Conference on Acoustics, Speech, and Signal Processing, pp.809-811 vol.2, 1989.

[10] James L. Flanagan, "Talking with Computers: Synthesis and Recognition of Speech by Machines", IEEE Transactions on Biomedical Engineering, vol.BME-29, no.4, pp.223-232, 1982.

[11] Prof. V. C. Vidhyashree, S. A. M. Supriya, H. Supriya, D. Vedala, and
R. Kavya, "Machine Learning-Based Text to Speech Converter for Visually Impaired," IJRASET

Journal For Research in Applied Science and Engineering Technology.

[12] S. Chawla and S. Srivastava, "A Goal-Based Methodology for Web- Specific Requirements Engineering," in 2012 World Congress on Information and Communication Technologies, pp. 173-178, 2012.

[13] R. Gao, H. E. Merzdorf, S. Anwar, M. C. Hipwell, and A. R. Srinivasa, "Automatic assessment of text-based responses in post-secondary education: A systematic review," in Journal of Educational Technology & Society, pp. 210-225, 2021.

[14] S. Haginoya, T. Ibe, S. Yamamoto, N. Yoshimoto, H. Mizushi, and P. Santtila, "AI avatar tells you what happened: The first test of using AI-operated children in simulated interviews to train investigative interviewers," in International Conference on Human-Computer Interaction, pp. 542-553, 2020.

[15] Y. Zhou, X. Han, E. Shechtman, J. Echevarria, E. Kalogerakis, and D. Li, "MakeItTalk: Speaker-Aware Talking-Head Animation," in IEEE Transactions on Visualization and Computer Graphics, pp. 342-350, 2021.

[16] C. A. Nwafor and I. Onyenwe, "An Automated Multiple-Choice Question Generation using Natural Language Processing Techniques," in International Journal on Natural Language Computing, vol. 10, no. 2, pp. 1-10, April 2021. DOI: 10.5121/ijnlc.2021.10201.