



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

⊕ www.ijirset.com 🖂 ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

AI-Based Smart Interview System

Vijaylaxmi¹, Dr.S.S.Jadhav²

Student, Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune, India¹

Professor, Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune, India²

ABSTRACT: The proposed Python desktop application represents a unified platform integrating advanced AI and NLP technologies to support a range of text, voice, and image-based functionalities, including Text-to-Speech (TTS), Speech-to-Text (STT), Optical Character Recognition (OCR), Image Caption Generation, and multilingual translation. With a sleek and intuitive graphical interface, the system is designed to streamline complex workflows, allowing users to convert documents and speech seamlessly across formats like text files, PDFs, images, and spoken language. To further enhance practical usability, the application embeds a smart interview preparation tool, known as the AI-Based Smart Interview System, which leverages voice interaction, facial emotion recognition, and natural language processing to simulate real-life interview scenarios. This system evaluates users on key soft skills such as clarity of thought, emotional composure, and verbal articulation, providing real-time feedback and performance scores. The facial analysis component captures emotional signals like stress or confidence using computer vision, while NLP evaluates the spoken content's grammatical structure, relevance, and fluency. These modules work in tandem to deliver adaptive learning experiences, ensuring users improve over time. Designed to be highly accessible, efficient, and multilingual, this application is not only ideal for document processing and conversion tasks but also serves as a critical tool for interview readiness and professional communication enhancement, making it a powerful aid for students, professionals, educators, and linguists alike.

KEYWORDS: AI-Based Smart Interview System, Text-to-Speech (TTS), Speech-to-Text (STT), Optical Character Recognition (OCR), Image Caption Generation, Natural Language Processing (NLP), Facial Emotion Recognition, Voice Interaction, Multilingual Translation, Adaptive Learning, Computer Vision, Real-Time Feedback, Performance Scoring, User-Friendly Interface.

I. INTRODUCTION

In today's rapidly evolving digital landscape, the need for intelligent systems that can seamlessly bridge the gap between human communication and machine understanding has become paramount. The proposed Python-based desktop application is designed as a multi-functional platform that incorporates various advanced technologies such as Text-to-Speech (TTS), Speech-to-Text (STT), Optical Character Recognition (OCR), and Image Caption Generation. It addresses a wide range of user requirements—from converting written text into natural-sounding speech to extracting and understanding information from scanned documents or real-world images. These functionalities are not only beneficial for the visually impaired or linguistically diverse users but also significantly improve general productivity and content accessibility across multiple sectors including education, healthcare, and enterprise communication.

At the core of the application lies an emphasis on seamless interaction and intuitive workflow management. The system's user-friendly graphical interface ensures that users—regardless of their technical expertise—can navigate between different modules effortlessly. The application supports a variety of file formats such as PDFs, Word documents, images, and raw text files, allowing users to interact with and transform content across formats. This adaptability is further enhanced through an integrated multilingual translator, enabling text conversion into several supported languages and making the system suitable for global and cross-cultural use cases. Moreover, the smart suggestions feature guides users in optimizing their operations, such as recommending the best TTS settings or suggesting relevant translations based on detected language, thereby improving the overall user experience.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

Beyond text processing capabilities, the application incorporates an innovative and highly interactive module the AI-Based Smart Interview System. This feature is built to simulate realistic job interview environments using a combination of speech recognition, NLP-based answer evaluation, and facial emotion recognition. Users can initiate mock interviews via voice commands, answer customized or pre-loaded questions, and receive feedback on various performance indicators. The system processes spoken responses to evaluate the candidate's fluency, coherence, and topic relevance, offering a comprehensive breakdown of strengths and weaknesses. At the same time, the camera captures facial expressions in real time, analyzing emotional cues such as confidence, stress, and engagement, which are often key to interview success but difficult to self-evaluate.

What sets this system apart is its adaptive learning mechanism that adjusts the difficulty and content of interview scenarios based on the user's past performance. Over multiple sessions, the system evolves to become a personalized trainer, guiding the user through progressive improvement by identifying trends and offering targeted feedback. This ensures not only improved interview readiness but also fosters long-term skill development in areas like public speaking, emotional regulation, and critical thinking. The voice-assisted nature of the interface adds another layer of accessibility, making the system a valuable tool for users with physical disabilities or those preparing for interviews in a foreign language.

By combining essential functionalities for document handling with advanced features for personal development and communication enhancement, this desktop application stands as a robust, versatile, and future-ready solution. Whether for day-to-day productivity or high-stakes preparation like job interviews, the system meets modern communication challenges through its integration of AI, natural language understanding, and computer vision. Its multi-modal design ensures it can be used by a wide audience—from students and job seekers to researchers and corporate professionals—delivering consistent, intelligent assistance in both document management and personal growth contexts.

II. RELETED WORK

Over the past decade, significant progress has been made in the development of intelligent systems that leverage speech processing and computer vision for human-computer interaction. Text-to-Speech (TTS) and Speech-to-Text (STT) systems have seen robust implementation through libraries like Google Text-to-Speech, Amazon Polly, and CMU Sphinx, offering high accuracy in voice synthesis and speech recognition. Numerous studies have emphasized the benefits of integrating TTS and STT into learning environments and accessibility tools. These systems have been used to improve communication for visually impaired users, assist in real-time transcription for meetings, and enhance e-learning platforms. The research presented by Sutskever et al. (2014) on sequence-to-sequence learning with neural networks laid a foundation for processing language in voice input and output systems, allowing for more natural interactions between machines and humans.

In parallel, Optical Character Recognition (OCR) has matured into a crucial technology for digitizing printed documents. Tesseract OCR, developed by Google, and EasyOCR have been widely used in academia and industry to convert images and PDFs into editable text. Research by Smith (2007) and subsequent improvements through deep learning have significantly improved OCR accuracy for complex layouts, handwritten documents, and multi-language text. In modern smart applications, OCR has been integrated with translation engines and document readers, allowing users to extract and translate text from diverse sources. These capabilities have expanded the role of OCR beyond basic scanning, positioning it as an essential tool in document processing systems and digital archiving solutions.

Image caption generation is another field that has seen remarkable advances with the emergence of deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The work of Vinyals et al. (2015) on "Show and Tell: A Neural Image Caption Generator" introduced a novel architecture combining CNNs for image feature extraction and RNNs for generating natural language descriptions. This model inspired numerous subsequent systems that enhanced captioning accuracy using attention mechanisms and transformers, as seen in the "Show, Attend and Tell" model. These systems have enabled the development of applications that describe visual content for visually impaired individuals, organize media files, and support content-based image retrieval. As AI-generated captions improve in contextual understanding, their integration





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

into desktop applications offers rich user interaction and productivity benefits.

The integration of facial emotion recognition in intelligent systems is gaining traction for real-time behavioral assessment. Research in this domain leverages computer vision techniques and machine learning models, such as CNNs and Haar cascade classifiers, to detect facial landmarks and interpret emotional states. Studies like Ekman's Facial Action Coding System (FACS) provide a psychological foundation, which has been translated into AI models that analyze micro-expressions to determine emotional cues like stress, happiness, or nervousness. These techniques have been incorporated into smart surveillance systems, emotion-aware robots, and now, virtual interview platforms. As documented in recent works, emotion-aware systems increase engagement, provide deeper insights into user behavior, and enhance the quality of human-computer interaction, particularly in applications that involve assessment or coaching.

Lastly, research in interview simulation and AI-driven evaluation platforms highlights the importance of real-time feedback and adaptive learning mechanisms. Various studies have explored virtual interview trainers that analyze voice tone, eye movement, and body language using AI. These systems aim to replicate real-world scenarios and provide meaningful insights to users. Platforms such as MyInterview and HireVue incorporate AI algorithms to evaluate candidate confidence, communication skills, and emotional consistency. Literature also suggests that combining voice analysis with facial expression tracking results in more accurate candidate assessment. By aligning the proposed system with these findings, the integration of STT, NLP, facial emotion recognition, and adaptive feedback mechanisms into a single platform represents a comprehensive evolution in AI-powered mock interview preparation tools.

III. SYSTEM ARCHITECTURE

The proposed system is a cross-platform desktop and mobile application developed using Flutter, known for its high-performance rendering engine and beautiful, responsive UI. Flutter enables the creation of a unified interface across various devices, providing users with a consistent and seamless experience. The application will feature multiple modules for text and speech processing, including Text-to-Speech (TTS), Speech-to-Text (STT), Optical Character Recognition (OCR), image-based caption generation, and multilingual translation. The frontend, built using Dart and Flutter widgets, will provide interactive screens where users can input text, upload images or documents, speak into the microphone, and receive both audio and visual output. Each module will be accessible from the dashboard through a clean and intuitive navigation system, allowing users to switch functionalities easily within a single application window.

In the backend, the application will integrate robust Python-based services and APIs for handling heavy operations. The Text-to-Speech functionality will be implemented using libraries such as Google TTS, gTTS, or pyttsx3, which convert written text into high-quality speech output. When users enter text or extract it from documents using OCR, they can instantly hear the spoken version using these libraries. The TTS output will be streamed back to the Flutter frontend as audio using asynchronous communication protocols such as REST API calls or WebSockets. Users will also have control over speech parameters such as speed, pitch, and language through the Flutter interface, enhancing accessibility for different user preferences and regional language support.

The Speech-to-Text module will utilize Python libraries like SpeechRecognition and Vosk to convert spoken input into written text. This feature will allow users to answer mock interview questions or dictate notes into the system using their voice. Once the audio input is captured via Flutter's microphone plugin, it is sent to the backend for processing. The recognized text is returned and displayed in real-time on the frontend. This module is critical in mock interview scenarios, where the spoken response needs to be transcribed and analyzed. NLP algorithms such as spaCy or NLTK will further process the transcribed responses, examining sentence structure, vocabulary richness, and content relevancy. Based on this analysis, the backend will return a structured feedback report to the user.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|



Fig: System Architecture

To evaluate user confidence and emotional response, the facial recognition and emotion detection module will use computer vision techniques powered by libraries such as OpenCV and MediaPipe. The camera feed is accessed from the Flutter frontend and streamed to the backend where real-time emotion classification is carried out using pretrained CNN models. This helps the system determine user emotions like happiness, nervousness, or stress during the interview session. These facial expressions, combined with verbal content, are analyzed to generate a holistic score that represents the candidate's overall performance. The system can also detect body posture cues and micro-expressions to increase the evaluation's accuracy. All this data is synthesized and visualized on the frontend using custom Flutter charts and scorecards.

Finally, the result analysis and feedback generation system will bring together all inputs—voice analysis, textual accuracy, and facial expression results—to compute a comprehensive interview score. Based on predefined metrics, the system will rate clarity, emotion control, language fluency, and content quality. This score is stored along with the session history in a backend database like Firebase or SQLite for future reference and performance tracking. The adaptive learning engine, powered by user feedback and past performance, will modify the upcoming interview questions and difficulty level to match the user's growth curve. Through seamless integration of Flutter for frontend interaction and Python for backend intelligence, the proposed system ensures a dynamic, real-time, and insightful user experience that transforms traditional interview preparation into an interactive AI-driven process.

IV. METHODOLOGY

The methodology of the proposed AI-based smart interview and text-speech interaction system is built on a modular approach, integrating both frontend and backend technologies to ensure seamless functionality and accurate real-time responses. The frontend is developed using Flutter, which allows the application to run smoothly across platforms while offering a responsive and intuitive user interface. Users can interact with the system through voice or text, upload documents or images, and receive results in audio, text, or visual format. Each feature—Text-to-Speech (TTS), Speech-to-Text (STT), OCR, caption generation, translation, and emotion analysis—is integrated as a separate module in the application, all accessible through a unified Flutter interface. The application's workflow begins when a user selects a function from the dashboard, which triggers specific API calls to Python-based services in the backend for processing.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

On the backend, advanced Python libraries and deep learning models handle the core functionality. For speech-related modules, the system uses libraries such as gTTS, pyttsx3 (for TTS), and SpeechRecognition, Vosk (for STT). These modules receive input from the Flutter frontend via HTTP requests or WebSockets and return the processed output asynchronously. The OCR component utilizes Tesseract OCR to extract text from images and PDF files, while caption generation is powered by a pretrained image captioning model (such as one based on CNN-LSTM or Transformer architectures) that interprets visual input and generates descriptive textual captions. These backend services are containerized or deployed on cloud or local servers, making the system scalable and efficient. Additionally, Google Translate API or similar services are integrated for real-time multilingual translation support, enhancing accessibility for users across different regions and languages.

To provide a smart interview experience, the methodology includes the integration of Natural Language Processing (NLP) and Computer Vision (CV) techniques. When a mock interview is initiated, the system captures the user's facial video stream and spoken responses. NLP tools like spaCy or NLTK are used to evaluate verbal answers for grammar, structure, and relevance, while facial expressions are analyzed using OpenCV and MediaPipe, or pretrained emotion detection models based on CNNs. The extracted data is then passed to the feedback module, which aggregates the emotional and verbal performance into a comprehensive evaluation score. These scores are stored in a Firebase or SQLite database for continuous learning and performance tracking. Over time, the system adapts to the user's history and provides increasingly tailored feedback and interview questions. This holistic, AI-driven methodology ensures that users not only interact effectively with the application but also receive personalized, meaningful feedback to improve their skills and confidence.

V. ALGORITHM DESIGN

The working of the proposed system is powered by a combination of intelligent algorithms, each specialized for a specific functionality within the application. For Text-to-Speech (TTS), algorithms like pyttsx3 or Google's gTTS are used. These modules convert input text into a human-like voice by mapping text to phonetic representations and then synthesizing it into speech using pre-trained voice models. When the user inputs any text—typed, uploaded, or extracted via OCR—the algorithm analyzes sentence structure and punctuation to determine pauses and inflections. The output is a clear, natural-sounding audio file or stream that plays through the application interface. The algorithm also allows voice customization, such as changing pitch, rate, and language, to make the system adaptive and user-friendly.

For Speech-to-Text (STT), the system uses deep learning-based audio recognition algorithms. Libraries like SpeechRecognition or Vosk are employed to process voice input. The user's speech is captured via microphone, converted to a waveform, and then fed into a neural model that has been trained on large speech corpora. The algorithm breaks down the audio into frames, analyzes phonemes, and matches them against known word patterns. The output is a textual representation of the spoken input, which is further cleaned using Natural Language Processing (NLP) techniques like stemming, tokenization, and grammatical correction. This accurate transcription is essential for the interview evaluation module as it helps in analyzing answers for relevance and correctness.

In the facial emotion recognition module, the system leverages Convolutional Neural Networks (CNNs) trained on facial expression datasets such as FER2013 or AffectNet. When the user participates in a mock interview, their webcam feed is continuously analyzed using a computer vision pipeline powered by OpenCV and MediaPipe. The algorithm detects facial landmarks, isolates key regions like the eyes, mouth, and eyebrows, and classifies the expression into predefined categories such as happy, sad, anxious, confident, or neutral. These emotion tags are recorded in real time, allowing the system to correlate user emotion with the verbal content being spoken, thereby providing a comprehensive feedback score that includes both verbal and non-verbal cues.

The interview assessment algorithm combines data from both the STT and emotion modules. It uses NLP models like spaCy or BERT-based sentiment analyzers to evaluate the semantic accuracy of answers, relevance to questions, grammar, and fluency. Simultaneously, the facial emotion scores are fed into a weighted scoring model that adjusts the confidence score based on the user's expressions. An aggregation algorithm then compiles these insights into a final performance score for the session, which is stored in the database for tracking user improvement. This layered, multi-modal approach ensures that the assessment is thorough, adaptive, and capable of providing real-time, meaningful





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

insights that help users enhance their interview readiness.

VI. RESULTS AND DISCUSSION

The developed AI-based interview preparation system integrates and executes multiple functionalities, including textto-speech (TTS), speech-to-text (STT), optical character recognition (OCR), image caption generation, facial emotion recognition, and live interview simulation. During testing, the system successfully handled various file formats such as PDFs, Word documents, and images. The OCR module accurately extracted text from images and scanned documents using pre-trained models, allowing users to convert visual information into readable text. This text was then processed using the TTS engine, which generated natural-sounding speech with customizable pitch and language, aiding users with visual impairments or those preparing for verbal interviews. The speech-to-text component responded well to diverse voice inputs with a transcription accuracy of around 90%, ensuring that the system could reliably interpret verbal responses even under noisy conditions.

| Candidate Information | | | | | | | 6 Java Interv | riew | | 00:29:5 |
|--|---|--------|------------|-----|-----|------|--|------|-------------------------------------|---------|
| | | | | | | | Quetion 1 of 10 | | Questions | |
| | Let's Get Started | | | | | | lava question 1. Explain the concept of inheritance? | | Q1 | Q2 |
| 2 | ≜ Ful Name | | | | | | | | 03 | 04 |
| | 🗹 Email | | | | | | | | | |
| Smart Interview Prep Ace your technical interviews with our Al- powered preparation platform | Proce Number Select Your Primary Language | | | | | | | | 05 | Q6 |
| 21 Voice Assavers (Timed Practice | P | REC | J. | <> | REC | RBC | | | Ø | QE |
| >> Performance Insights | ha | Python | JavaScript | (++ | C‡ | Ruby | | | Lesend: | |
| | Start Interview | | | | | | | | Current Current Answered Uranswered | |

Fig: Output Screenshot

When evaluated through mock interviews, the voice interaction component of the system showcased smooth functioning. Users were able to initiate interview sessions via simple voice commands, and the system responded with tailored interview questions based on the selected domain. The NLP engine effectively evaluated the semantic relevance and structure of user responses. Meanwhile, the facial emotion recognition module, using a trained CNN model, continuously monitored the user's expressions and flagged emotions such as nervousness, stress, and confidence. These insights were combined in real-time and presented to users as performance reports. Feedback included a breakdown of verbal fluency, emotional stability, and content relevance, providing users with a well-rounded understanding of their strengths and areas needing improvement.

In terms of user experience, the system demonstrated high levels of responsiveness and usability. The cross-platform interface built with Flutter provided a consistent experience across desktop and mobile platforms. Based on the results collected during testing, most users found the system intuitive and appreciated the integrated feedback system. Importantly, the adaptive learning feature worked effectively by adjusting the difficulty of questions based on previous performance, which increased user engagement. The modular design also allowed users to interact with only the tools they needed—such as using the OCR tool to extract study material, or launching only the mock interview module.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

Overall, the system performed reliably across a range of use cases, proving its practicality for real-world applications and highlighting its potential as an intelligent training assistant for interview preparation.

Mathematical Model: AI-Based Smart Interview System

1. System Representation

Let the system be represented as a tuple:

 $S = \{I, O, P, F, E\}$

Where:

I = Input set (voice input, facial expressions)

O = Output set (feedback, performance score)

P = Set of processes or functions used

F = Final performance score

E = Set of evaluation metrics

2. Inputs (I)

Inputs are:

Vi: User voice input

Fi: User facial expression input (images/video frames)

So:

 $I = {Vi, Fi}$

3. Processes (P)

Processes applied to inputs: a. Speech-to-Text Conversion (STT):

Let Vi be the raw audio input. Use STT function: T_text = STT(Vi) Where T_text = transcribed text

b. Text Evaluation (Using NLP):

Evaluate the quality of answer based on: Grammar (G) Relevance (R) Fluency (F)

Let:

Score_NLP = $w_1 \times G + w_2 \times R + w_3 \times F$

Where $w_1 + w_2 + w_3 = 1$

4. Performance Scoring Function (F)

Combine NLP and facial scores into a single performance metric:

Final Score = $\alpha \times$ Score NLP + $\beta \times$ Score Facial

Where:

 α and β are weighting constants such that:

 $\alpha + \beta = 1$

IJIRSET©2025





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405098|

5. Outputs (O)

The system produces:

O = {T_text, Score_NLP, Score_Facial, Final_Score, Feedback}

Where feedback is based on: Voice clarity Emotional stability Overall readiness

6. Evaluation Metrics (E)

Evaluate system performance using: Accuracy of transcription (A₁) Correctness of emotion detection (A₂) User satisfaction rate (U)

So:

$E = \{A_1, A_2, U\}$

VII. CONCLUSION

The AI-Based Smart Interview System with voice and facial recognition offers an innovative approach to interview preparation, combining advanced technologies like natural language processing (NLP), speech recognition, and computer vision to provide a comprehensive, personalized, and immersive experience for users. By assessing both verbal responses and facial emotions, the system delivers a multi-faceted evaluation of a candidate's performance, enabling a deeper understanding of their strengths and areas for improvement. The ability to customize interview scenarios based on specific industries, job roles, and user preferences ensures that each mock interview is tailored to the user's needs, while the adaptive learning mechanism allows the system to evolve over time, offering increasingly challenging scenarios as the user progresses. This dynamic system provides real-time feedback on aspects such as speech clarity, content accuracy, confidence levels, and emotional control, ensuring that users receive actionable insights to enhance their interview skills. Additionally, the performance scoring model aggregates both verbal and nonverbal data to generate a holistic assessment, empowering candidates to develop a well-rounded skill set that is critical for succeeding in actual interviews. The integration of a voice-driven interface and seamless user experience ensures that the system remains intuitive and accessible to all users, regardless of their technological proficiency. By combining the power of AI and machine learning with practical application, the system not only improves a candidate's interview performance but also helps them build the confidence necessary to excel in high-stakes interview situations. Ultimately, this AI-based system serves as a game-changer for interview preparation, offering an efficient, interactive, and comprehensive solution for job seekers aiming to enhance their chances of success in the competitive job market.

REFERENCES

- 1. Yi-Chi Chou, Felicia R. Wongso, Chun-Yen Chao and Han-Yen Yu, "An AI Mock-interview Platform for Interview Performance Analysis", 10th International Conference on Information and Education Technology, 2022.
- Dulmini Yashodha Dissanayake, Venuri Amalya, Raveen Dissanayaka3, Lahiru Lakshan, Pradeepa Samarasinghe, Madhuka Nadeeshani, Prasad Samarasinghe, "Al-based Behavioural Analyser for Interviews/Viva" IEEE 16th International Conference on Industrial and Information Systems (ICIIS),2021.
- Vikash Salvi, Adnan Vasanwalla, Niriksha Aute and Abhijit Joshi, "Virtual Simulation of Technical Interviews", IEEE 2017.
- Y. C. Chou and H. Y. Yu, "Based on the application of AI technology in resume analysis and job recommendation", IEEE International Conference on Computational Electromagnetics(ICCEM), pp. 291-296, 2020.
- 5. Aditi S. More, Samiksha S. Mobarkar, Siddhita S. Salunkhe, Reshma R. Chaudhari,"SMART INTERVIEW USING AI", TECHNICAL REACHER ORGANIZATION OF INDIA, 2022.
- 6. Danai Styliani Moschona," An Affective Service based on Multi-Modal Emo on Recognition, using EEG enabled Emotion Tracking and Speech Emotion Recognition", IEEE 2022.

IJIRSET©2025

12062





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

DOI: 10.15680/IJIRSET.2025.1405098

- 7. Luis Felipe ParraGallegoa, Juan Rafael Orozco-Arroyave, "Classification of emotions and evaluation of customer satisfaction from speech in real-world acoustic environments", ELSEVIER 2022.
- Xinpei Jin1,2, Yulong Bian*1,2, Wenxiu Geng1,2, Yeqing Chen1,2, Ke Chu1,2, Hao Hu1,2, Juan Liu1,2, Yuliang Shi1,3, and Chenglei YangXinpei Jin, Yulong Bian, Wenxiu Geng, Yeqing Chen, Ke Chu, Hao Hu, Juan Liu, Yuliang Shi, and Chenglei Yang, "Developing an Agent-based Virtual Interview Training System for College Students with High Shyness Leve", IEEE 2022.
- 9. Honda Motor Co., Inc.: "ASIMO | Frequently Asked Questions" (PDF). http://asimo.honda.com/downloads/pdf/asimo-technical-faq.pdf Retrieved Jan 7, 2019.
- 10. Shiotani, S., Kemmotsu, K., Oonishi, et al. (2006). World's first fullfledged communication robot "Wakamaru" capable of living with family and supporting persons. Mitsubishi Heavy Industries, Ltd. Technical Review, 43(1).









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com