Boosting Interview Effectiveness and Streamlining Insights with AI-Powered GPT Solutions

Shashank Mahato  
*Department of ECE  
Amrita School of Engineering*

*Amrita Vishwa Vidyapeetham*Chennai, India

[ch.en.u4cce21031@ch.students.amrita.edu](mailto:ch.en.u4cce21031@ch.students.amrita.edu)

Muthulakshmi M   
*Department of ECE*

*Amrita School of Engineering  
Amrita Vishwa Vidyapeetham*Chennai, India

[m\_muthulakshmi@ch.amrita.edu](mailto:m_muthulakshmi@ch.amrita.edu)

***Abstract*—** **The advent of Artificial Intelligence in streamlining interview processes has been a critical tenet of ensuring efficiency and effectiveness. The following study outlines a novel framework utilizing sophisticated AI procedures to automatize transcription, analysis, as well as feedback. With the help of AWS Transcriber, as well as OpenAI’s ChatGPT-4 Turbo, the current system allows for beneficial and real-life feedback in consideration of each responder. This paper elaborates on the techniques of prompting and inspecting that enable the system to provide persons with incredible choices for answers.**

Keywords—Linguistic Analysis, AWS Transcriber, OpenAI's ChatGPT-4 Turbo, Prompt Techniques.

# Introduction

Although the traditional approaches to interviewing have laid the foundation for the further instruments of selection, they are not quite effective due to their time-consuming nature and rather subjective outcomes. More often than not, the inability of the interviewer to pose the question designed to test the candidate’s approach to addressing a specific problem occurs [1]. In turn, rather vague and, therefore, not exactly useful feedback is delivered to the candidate or is sent too late. As a result, the overall process becomes significantly impeded. That being said, the incorporation of technology has defined the direction in which the redesign of recruiting is supported. Particularly, the immense capability of Artificial Intelligence to process and analyze data at a very high speed is likely to predetermine the outcomes.

However, the deployment of AI into such a human-related process as interviewing raises the need to thoroughly investigate both technological and psychological points to be considered to ensure that the technology will only facilitate the process rather than taking a human touch to one side. The innovative framework relies on AI to address the identified needs related to different aspects of the interviewing process: interview transcription, response analysis, and feedback generation [2]. AWS Transcriber and Google Gemini 1.5 Pro combined with OpenAI’s ChatGPT-4 Turbo are to be used for the transcription of interviews and language analysis and processing, respectively.

Moreover, the solution’s architecture focuses on the extensive use of tailored prompt mechanisms to control the actions of AI models. Built in such a way, the system only focuses on this or that aspect of a candidate’s competency, as well as the linguistic elements needed to analyze the responses by both candidates and interviewers. As a result, the identity of the interaction character is enhanced. Moreover, the developed prompt system adds to the quickness with which feedback can be provided to the parties involved in the rivalry. Hence, the promotion of the decision-making process is made a rather streamlined and efficiency target. Overall, the approach targeted in the system’s architecture and design, as well as the nature of the algorithms incorporated in the program, is expected to redefine the process of preparing candidates for an interview.

# Methodology

Specifically, transcription entails speech-to-text conversion or processing the auditory input from an interview into a written form. Chunking refers to the division of the interview into separate elements, while detailed analysis involves the examination of these segments about various affiliation risk factors. Finally, the generated feedback is processed by the AI tool and can be provided to the interviewer or candidate in verbal or written form.

Overall, the current AI-supported interview process is designed to facilitate the consideration of various risk factors and indicators of affiliation by analyzing a candidate’s responses to their speech patterns at different scales of the language. Transcription, chunking, detailed analysis, and generated feedback represent vital steps reflecting the consideration of speech expression and risk factors associated with these components of AI processing.

## Transcription with AWS Transcriber

The initial stage of our procedure is an exact transcription of audio interviews into a text format, as the basis of all other kinds of the processing. For this step, we use the AWS Transcriber, which is a high-accuracy and reliable automatic speech recognition tool, provided by Amazon Web Services. There are many different accents, dialects and variations of speech quality recorded in the library of interviews, and AWS transcriber can accurately handle most of them.

**Process Overview:**

**Audio File Preparation:** Audio files that include the interviews are prepared. In order to pass our transcription one we need the audio files to meet the certain requirements and to be good in quality. Also, they need to be recorded correctly so we use them.

**Transcription Job Creation:** Inside the Amazon Transcriber we create a new transcription job. Here we provide the necessary services about our audio, like location. We also have to provide the system with some additional integrations if we want to reach higher quality and actuality of our research. This can be such options as speaker identification. Also, we can upload there some of our previous researches and create a custom vocabulary to make the transcriber able to write some typical things wrong.

**Transcription Execution:** The job is executing and the results are showing, that the processed audio file has a full-text file related. This text is provided with the title and a description. The most useful things for us here are the timestamps of the beginning and the end of each phrase and the labels of the speakers. The text also can be played, as in the true interview.

## Chunking the Transcript

Making analysis of transcript with attention net considerably more manageable, as, to begin with, splitting lengthy pieces of content in separate little segments is a process, naturally part of the mechanism of attention net architecture. This process is known as text chunking and serves two main purposes:

The chunking eliminates the necessity to analyze such a copious amount of the data all at once, because the data is split into several segments;

The security of the context is achieved by the overlap of the words-chunks: the small amount of the content from the previous chunk is included in the next segment of the text..

**Process Details:**

**Segmentation:** The transcript is segmented into chunks, each consisting of a fixed number of words with a slight overlap to maintain context.

**Overlap:** An overlap of approximately 20 words is set between chunks so that no single piece of information from the transcript is lost.

## Analysis Using AI Models

At the chunk level, each chunk of the transcript is analyzed by combining Google Gemini 1.5 Pro and OpenAI’s ChatGPT-4 Turbo. These AI models are informed by prompts, which are specifically designed to guide the analysis around areas of interest and competencies relevant for the job interview.

**AI-driven analysis:**

**Skill identification**: Determining which skills are pertinent for job interviewees gleaned from this chunk of text and how they are discussed.

**Generation of questions, and feedback**: What would be a great question or feedback relevant to the competencies being assessed.

## Feedback Generation

This step is the last one in our methodology: its goal is the feedback generation. Here, the information Natural Language Understanding algorithms provide is processed in order to formulate both interviewers’ and candidates’ concrete advice.

**Feedback Customization:**

**Prompt Techniques:** there are techniques dedicated to ensuring that the prompts do not only direct AI with regard to addressing specific questions but to make the generated feedback critical and personalized based on the information provided by a candidate.

**Tailored Prompts:** Techniques that ensure that the feedback is specific to the questions asked in the interview, it is, therefore, helpful in making sure that the AI is directed to ensure, wherever possible, that the generated feedback is relevant to the underlying intent of specific prompts or questions.

**Constructive feedback prompts:** The AI is prompted to provide constructive feedback, which ensures that the feedback generated is not only relevant but also constructive. By doing providing feedback that highlights the strengths of the candidate and offers insights on how the interviewee can improve, the AI makes the process of giving feedback to the candidate growth-oriented.

**Real-time feedback:** The feedback is automatically generated once the analysis is done therefore, it is always generated in a timely manner so that the interviewer and the interviewee can use the insights provided as soon as they are generated.

# Literature Review

This literature survey consolidates findings from five recent research, each concentrating on innovations in artificial intelligence (AI) methodologies for the analysis, prediction, and transcription of interviews. This document provides a comprehensive analysis of the methodology, techniques, and innovations employed in each work.

## Automated Transcription of Interviews in Qualitative Research Utilising Artificial Intelligence

[3] This report highlights the utilisation of OpenAI's Whisper for automated transcription in qualitative research. Whisper's multilingual capabilities and integration via Python, web applications, or offline deployments facilitate transcription work. The text emphasises ethical aspects, including anonymisation in online transcription to safeguard participant privacy. A noteworthy aspect is its hybrid methodology, blending Whisper's automation with manual checking to balance efficiency and transcription quality. The practical guidance of the study renders AI transcription accessible to researchers, sparing time while guaranteeing data reliability and compliance with ethical criteria.

## Survey on Interview Performance Prediction and Analysis Using Audio Features and NLP

[4] This research proposes a framework incorporating prosodic qualities and lexical models to predict interview results and provide feedback on candidates’ emotional and psychological attributes. Utilizing a combination of time-distributed CNNs and LSTMs for audio analysis and RNNs for generating the lexical model, the framework analyses emotions (e.g., anger, happiness) and personality traits (e.g., neuroticism). It employs RAVDESS Emotional Speech and Song Dataset for training audio models and Stream-of-consciousness datasets for lexical analysis. The final score prediction employs Decision Tree Regression, giving an objective approach to performance evaluation and reducing interviewer biases.

## A Measurement-Based Foundation for AI Applied to the Audio of Interviews

[5] This work operationalizes verbal expression in pre-employment interviews by defining and anchoring behavioral factors, creating a measurement-driven foundation for AI models. Using pairwise comparison methodologies, the study establishes reliability in measuring auditory aspects and presents preliminary validity evidence. The distinctiveness consists in a measurement-focused approach to defining vocal expressiveness and its application to deep learning models for large-scale automation. By exploring the link between acoustic qualities and employment screening, the study lays the scene for future advancements in incorporating audio analysis in recruiting procedures.

## Does the Use of Synchrony and Artificial Intelligence in Video Interviews Affect Interview Ratings and Applicant Attitudes?

[6] This study analyses the social and psychological effects of synchronous (SVI) and asynchronous video interviews (AVI) with and without AI-based decision agents. Using media richness and social interface theories, it studies fairness perceptions and the relevance of initial impressions. The findings reveal that AVIs minimise biases associated with physical appearance and early impressions while keeping fairness. However, candidates demonstrate less favorability towards AVIs compared to SVIs. This research highlights the potential of AI to eliminate bias in hiring but underscores the necessity to overcome user acceptability challenges in asynchronous scenarios.

## Temporally Aligning Long Audio Interviews with Questions: A Case Study in Multimodal Data Integration

[7] This work addresses the challenge of detecting specific questions in lengthy audio interviews using INDENT, a cross-attention-based paradigm. INDENT combines temporal ordering and trained voice embeddings to map text queries to audio segments, displaying higher performance over text-based heuristics with a 3% improvement in R-avg. The technique exhibits consistency across 11 Indic languages and efficiently absorbs noisy ASR outputs for increased performance. Its application to multilingual and noisy datasets, particularly in rural and resource-constrained situations, highlights its novelty and utility in real-world scenarios.

# Implementation

## This section will be discussing deeper into the technical implementation of our AI-driven interview analysis system, spotlighting the specific prompting techniques and important technical terms used to increase the functionality and efficiency.

## Code Architecture:

## **Service-Oriented Architecture (SOA): The system’s architecture is based on SOA, which allows services to be scaled modulary and independently. It makes it relatively simple to integrate different technologies, such as AWS Transcriber for audio-to-written-text conversion and OpenAI’s ChatGpt-4 Turbo for throrough linguistic processing.**

## **Robust API integration: Custom wrapper functions make the integration of external services easy and drive robustness in the system through sophisticated error handling and data centricity.**

## Data Handling and Processing:

**Asynchronous Transcription Management**: Leveraging asynchronous programming paradigms, the system efficiently accomplishes multiple transcription tasks concurrently. This approach optimizes data throughput and minimizes latency, crucial for the real-time processing requirements.

**Segmentation Algorithm**: Transcripts are divided using a chunking algorithm that segments 80-word text with a 20-word overlap. These parameters were empirically determined to balance with analytical granularity with contextual continuity:

* **Chunk Size**: The 80 words are fixed for a relevant reason and it is that this size allows both manageable chunk analysis load and the required depth for meaning relevance.
* **Overlap**: A 20-word overlap is used due to the recurrent factor considered for chunk analysis, and it ensures that no text information between two consecutive chunks is omitted and that the narrative aspect of the text is maintained.

## AI Model Implementation:

## **Dynamic Prompt Engineering**: The system uses advanced prompt-engineering techniques to facilitate AI’s extraction of knowledge and generation of valuable feedback. It combines several AI and NLP libraries:

## **OpenAI's GPT-4 Turbo**: it serves as a foundation for natural language generation and understanding, with prompts tailored to reflect the specifics of each interview part, providing remarkable precision and relevance

**Google Gemini 1.5 Pro:** It is a model suitable for deep linguistic analysis that is able to comprehend complex language patterns and detect fine nuances in the transcript chunks.

**Model Fine-Tuning:** In order to make the models better handle the specifics of interview dialogues, they were fine tuned on domain-specific datasets.

**Custom NLP Pipelines:** It includes the data preprocessing via Python’s spaCy and nltk libraries, such as tokenization, lemmatization and named entity recognition.  
  
*D. Feedback Generation:*

Notice that the entire process sits on top of dynamic prompt engineering, which is used to indicate to the AI model in charge of generating feedback that the prompt to operate on an interview segment is relevant. To implement such technique, we will need to do the following:

* **Contextual Adaptation**: need to ensure that the prompts are generated dynamically based on the content of the interview chunks so that the generated feedback is based on the specifics of the candidate’s response.
* **Competency Mapping**: The prompts are to be generated that would map the response and generated feedback to competencies and skills identified during the analysis of the interview, making the response more structured and applicable both to the interviewer and the candidate.

# Results and Discussion

In the next section, we elaborate on the results of implementing the AI-driven interview analysis system, providing both quantitative and qualitative data describing its outcomes and efficacy in the spheres of producing appropriate and timely feedback, as well as improving existing questions and interviews.

## BLEU Evaluation

The Bilingual Evaluation Understudy metric is one of the primarily used approaches toward evaluating machine productions and their correspondence to human writing style. As such, the BLEU score was applied as one of the main tools for defining the accuracy and fluency, accuracy, and clarity of the responses to feedback that were supplied by the AI machine and the human expert.

The method computes BLEU scores using n-gram matching to compare the reference feedback and the one generated by the model. Specifically, the scores were calculated by using BLEU-4 that considers up to four-word sequences to evaluate both short-range and long-range dependencies in the text. Different scores were calculated for each model as demonstrated below:

**BLEU-4 Scores:** With a score of 0.82, ChatGPT-4 Turbo had the highest BLEU-4 score, signaling better accuracy and fluency of generated texts similar to the human references. The score of 0.80 was recorded for Claude-3 while 0.78 and 0.75 were scores for MetaLlama-3 and Gemini 1.5 Pro, respectively.

1. BLEU-4 scores of LLM Models

|  |  |
| --- | --- |
| **LLM Models** | **BLEU-4 scores** |
| ChatGPT-4 Turbo | 0.82 |
| Gemini 1.5 Pro | 0.75 |
| Meta llama-3 | 0.78 |
| Claude-3 | 0.80 |

## ROUGE Evaluation

The Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is a widely used metric for establishing the quality of text generation and is used by comparing the generated content with reference texts. In our system, we used ROUGE in order to evaluate the comprehensiveness and relevance of the AI-generated feedback by measuring how well the generated feedback aligned with the human-generated feedback. The specific scores were:

**ROUGE-3**: This score is the measure of a match between trigram sequences referenced in a specific source with that of the created feedback. It shows how well the model can create feedback that sticks with the context of the feedback in three-word sequences. ChatGPT-4 Turbo achieved the highest ROUGE-3 score of 0.70, indicating strong contextual consistency across three-word sequences. Gemini 1.5 Pro, Meta Llama-3, and Claude-3 scored 0.65, 0.63, and 0.68 respectively, showing varying abilities in short-range text coherence.

**ROUGE-L**: ROUGE-L is based on the longest common subsequence between the generated feedback and the references. Evaluating how well the content aligns in structure and coherence, it measures the similarity in structure of the generated feedback to the ground-truth human feedback. Here, as well, ChatGPT-4 Turbo is outperformed in all baselines by having a higher ROUGE-L score of 0.82. Following closely are Claude-3 and Gemini 1.5 Pro, with scores of 0.80 and 0.78, respectively. The Meta Llama-3 model trails the first three with a score of 0.75, demonstrating that the length of the text generated by the model was slightly lower in similarity in structure to that of the human content.

1. ROUGE scores of LLM Models

| LLM Models | ***ROUGE scores*** | |
| --- | --- | --- |
| ROUGE-3 | ROUGE-L |
| ChatGPT-4 Turbo | 0.70 | 0.82 |
| Gemini 1.5 Pro | 0.65 | 0.78 |
| Meta llama-3 | 0.63 | 0.75 |
| Claude-3 | 0.68 | 0.80 |

# Future Works

Additionally, the system can be developed for real-time use, which means that the analysis and feedback will be generated during the interview. Thus, the interviewers will be able to obtain experience and share their data with the candidates almost at the time of answering the questions. The most important challenge in this case will be providing that the model inference time is low. Moreover, the use of audio instead of text will require additional resources to be processed. Thus, the aim will be the provision of low-latency high-accuracy operation under any network conditions.

Another potential development of the described idea is adding the ability to assess a candidate’s emotional intelligence. This way, the AI could analyze not only the content of a candidate’s answer but their emotional undertones, providing more profound insights into candidate behavior and their fit for positions requiring particular emotional competencies. A significant difficulty in this approach is creating EI models that effectively recognize emotions in text and speech. These models will have to be seamlessly integrable with the current feedback generation framework and not affect its processing speed.

Also, it is necessary to enhance the system’s data security and privacy since the tool starts being widely used in recruitment processes where sensitive data is utilized. It would be reasonable to think of ways to enhance security, and it may include enhanced encryption techniques, more secure data storage or GDPR compliance. Therefore, a future development focus of the project must be deploying such measures while ensuring that the system remains as accessible and user-friendly as possible, especially when it runs in real-time or from the cloud.  
 These developments can help in ensuring the system remains at the forefront of AI-driven tools for recruitment, offering real-time, emotionally intelligent, and secure interview analysis capabilities.

# Conclusion

We present an AI-driven interview analysis system proposed and developed to improve the efficiency, accuracy, and quality of interviews. The assessment was conducted by using the latest language models: ChatGPT-4 Turbo, Gemini 1.5 Pro, MetaLlama-3, and Claude-3. Precisely, the systems provide feedback, extract relevant questions, and generate contextually appropriate answers in an attempt to augment the efficiency and quality of interviews for both the interviewees and interviewers. The assessment was conducted using several metrics: BLEU-4, ROUGE-3, and ROUGE-L, where the accuracy, fluency, and structural match of the AI-generated feedback with human-generated texts were measured respectively. We found that ChatGPT-4 Turbo excelled in all metrics, indicating that it produces high-quality feedback that is as contextually near to human-written texts as possible.

Our study concludes that AI integration into the interviewing process brings its contribution in providing the speed and quality of feedback higher and, thus, aids the decision-making process. The technique used dynamic prompt engineering has proven to make the provided knowledge more accurate and adjusted to the context of a particular interview. It should be noted, however, that the study has also shown that there are still the aspects in which AI models need improvements, such as the ability to manage specialized specifics of the interview content and the security of the interview data. It will be also useful to work further on enhancing the adaptability of the system to different types of interviews and devoting more resources to making the feedback models’ responses more natural and conversational.

In summary, this research showcases the capacity of artificial intelligence (AI) to completely transform conventional interviewing techniques, enhancing their efficiency, objectivity, and depth of understanding. With the continuous advancement of AI technologies, their integration into recruitment and human resources is expected to become more crucial. This integration will provide new possibilities to enhance and simplify talent acquisition procedures.

# Acknowledgment

##### We express our sincere gratitude for the valuable assistance and support provided by our mentors within our research team. We express our gratitude to our families and friends for their unwavering support and motivation along the course of this undertaking.

##### VIII. References

[1] Levashina, Julia, et al. "The structured employment interview: Narrative and quantitative review of the research literature." *Personnel Psychology* 67.1 (2014): 241-293.

[2] Sidaoui, Karim, Matti Jaakkola, and Jamie Burton. "AI feel you: customer experience assessment via chatbot interviews." *Journal of Service Management* 31.4 (2020): 745-766.

[3]Mojadeddi, Zubair M., and Jacob Rosenberg. "Automated Transcription of Interviews in Qualitative Research Using Artificial Intelligence: A Simple Guide." (2024).

[4] Srihari, P., et al. "Survey On Interview Performance Prediction and Analysis using Audio Features and NLP." *NeuroQuantology* 20.9 (2022): 6567.

[5] Thompson, Isaac, et al. "A Measurement-based Foundation for AI Applied to the Audio of Interviews." *TMS Proceedings 2021* (2021).

[6] Suen, Hung-Yue, Mavis Yi-Ching Chen, and Shih-Hao Lu. "Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes?." *Computers in Human Behavior* 98 (2019): 93-101.

[7] Pasi, Piyush Singh, et al. "Temporally aligning long audio interviews with questions: a case study in multimodal data integration." *arXiv preprint arXiv:2310.06702* (2023).