AI-Enhanced Interview Proctoring and Assessment System

Team Members:

Kritika Acharya(THA077BCT024)
Rashna K.C.(THA077BCT039)
Suyog Achary(THA077BCT045)
Yogesh Aryal(THA077BCT046)

Supervised By:

Er. Sudip Rana

Department of Electronics and Computer Engineering
Institute of Engineering, Thapathali Campus

June 21, 2024

Presentation Outline

- Motivation
- Objectives
- Project Scope
- Proposed Methodology
- Expected Results
- Project Applications
- Tentative Timeline
- Estimated Project Expenses
- References

Motivation

- Traditional interviewing and research methods are often resource-intensive, time-consuming, and people-dependent.
- Create a fairer and more consistent recruitment process.
- Integrating AI to judge on their skills and answers rather than subjective impressions.
- Scanning and filtering resumes for further evaluation.
- Conducting interviews and surveys continuously without human intervention which significantly speeds up the process.

Objectives

- To implement advanced AI models like T5, RoBERTa, Wavenet, Whisper, OpenCV for precise assessment of candidate’s technical qualifications, coding responses, and problem-solving abilities.

- To automate the technical interview process, reducing manual effort and ensuring unbiased evaluations.

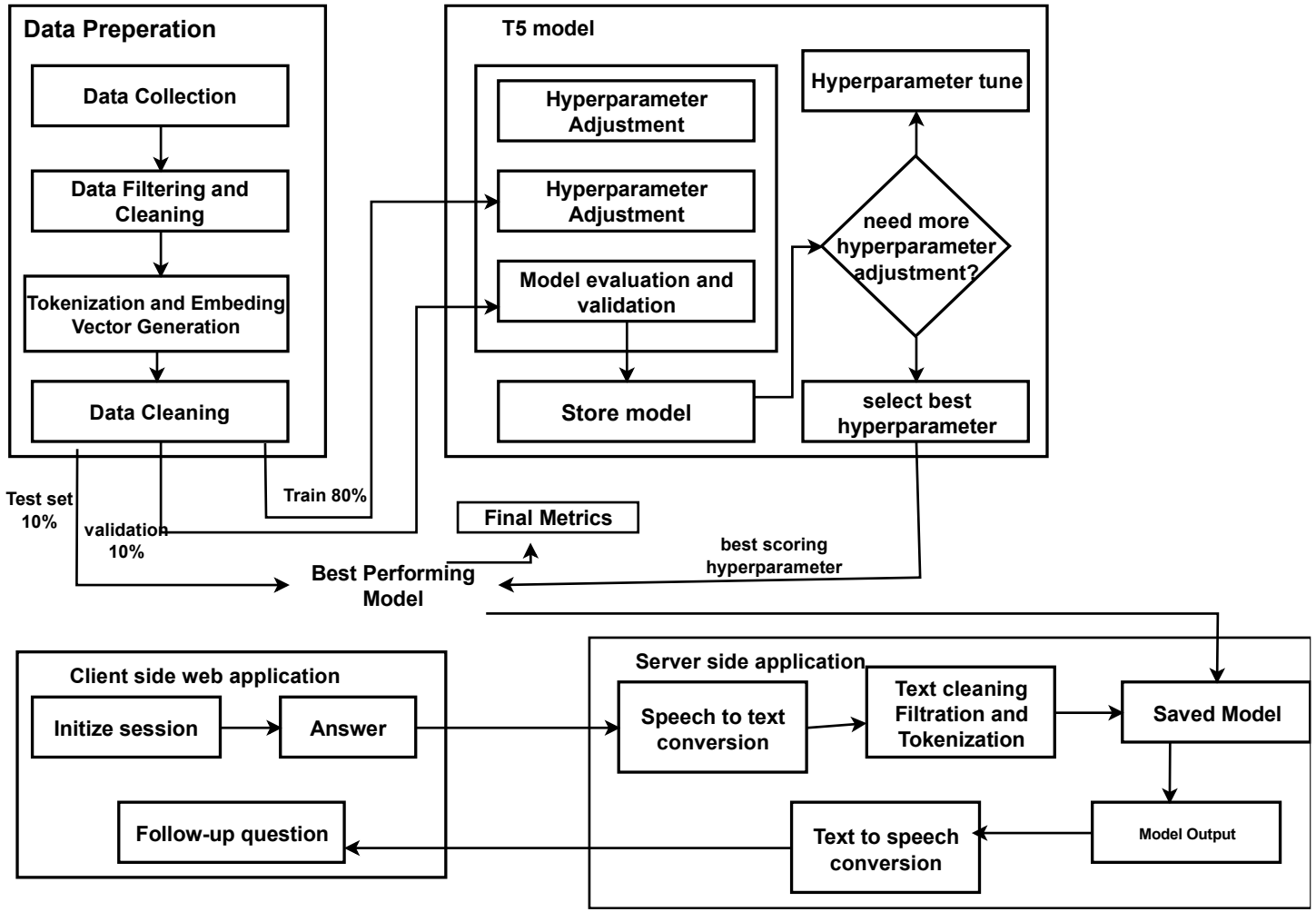
Scope of Project

- Effectively transform the hiring process by taking care of the preliminary screening of candidates, maintaining consistency, and effectively handling high quantities.

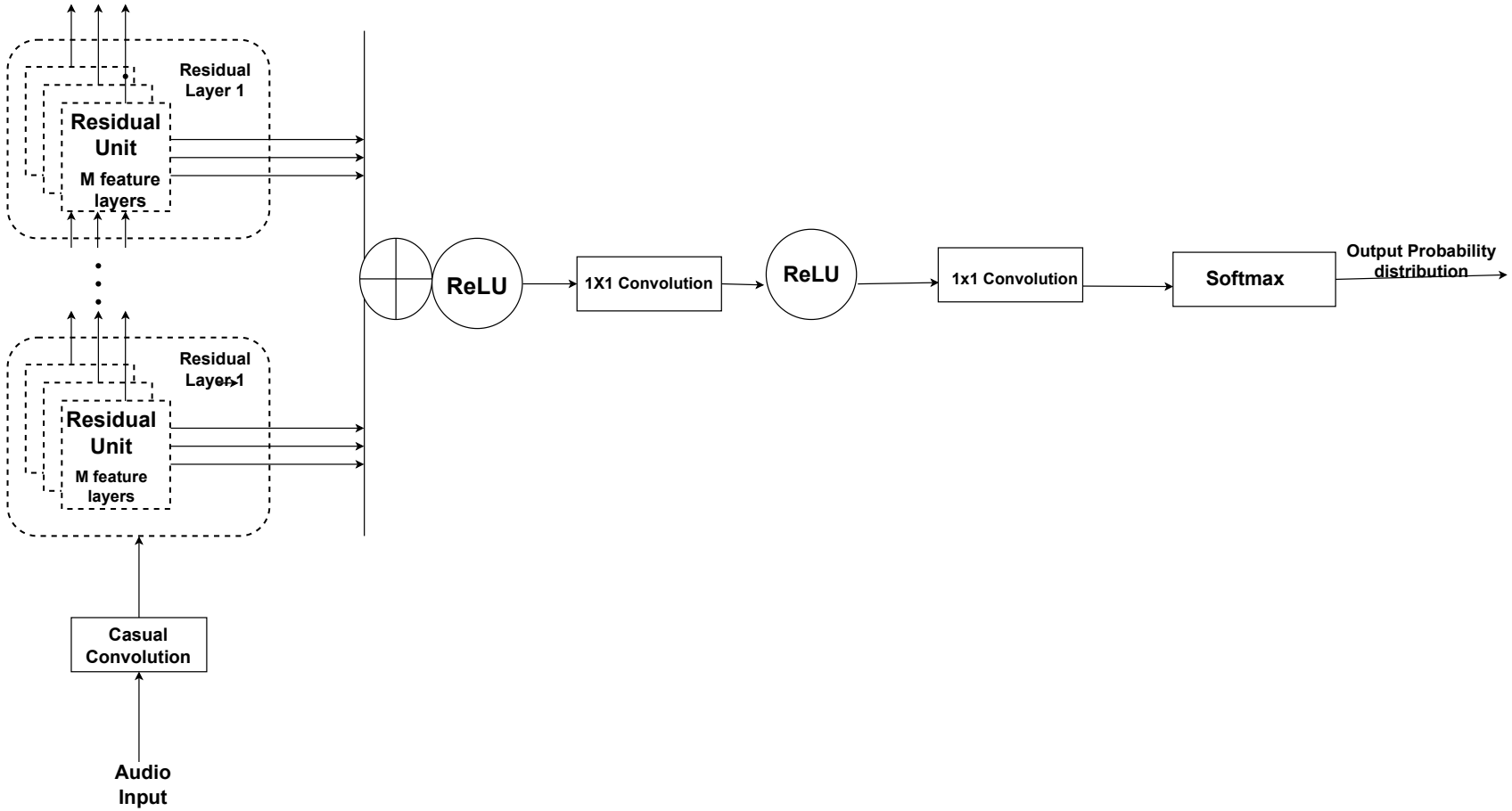
- Ensure a uniform interview process for all candidates minimizing human bias and error.

- Handle large volume of applicants without degradation of performance.

Methodology - [1] (System Block Diagram)



Methodology – [2] (WaveNet Text-to-Speech)



Methodology – [3]
(WaveNet Working principle I)

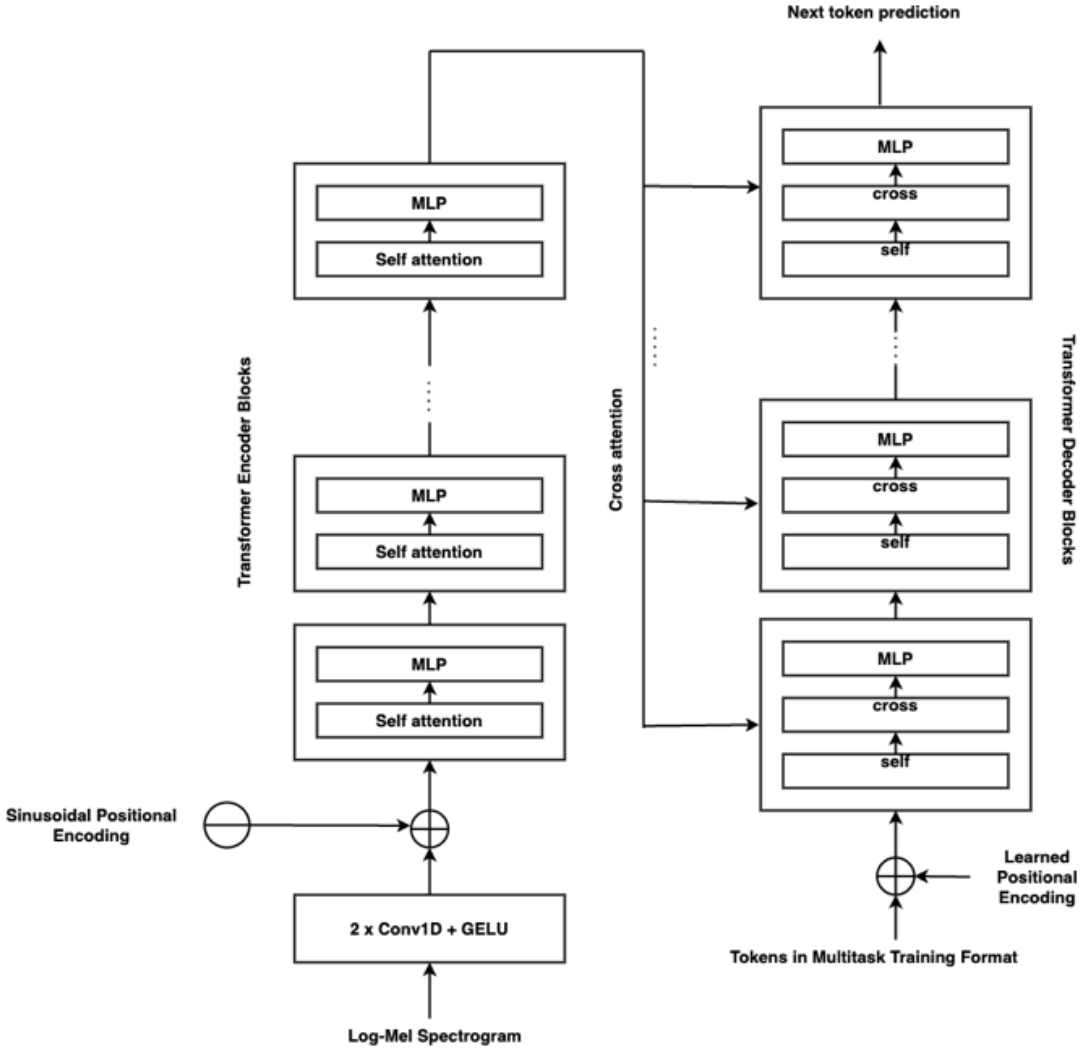
WaveNet will be used for Text-to-Speech conversion during interview process.

- **Autoregressive and Probabilistic Model**: Conditions each audio sample on all previous samples for accuracy.
- **High-Quality Speech Generation**: Produces natural and contextually relevant speech output from text inputs.
- **Linguistic Feature Conditioning**: Enhances speech relevance by conditioning on linguistic features from input text.
- **Speaker Identity Conditioning**: Accurately captures and reproduces characteristics of different speakers.
- **Extensive Training Data**: Trained on thousands of audio samples per second for nuanced learning.

Methodology – [4]
(WaveNet Working principle II)

- **Dilated Convolutions**: Expands receptive fields exponentially, capturing long-range temporal dependencies in audio.
- **Efficient Training**: Handles large audio datasets effectively, optimizing performance.
- **Deep Generative Architecture**: Directly models raw audio waveforms for high-fidelity speech synthesis.

Methodology – [5] (Whisper Speech-to-Text)



Methodology – [6]
(Whisper Working principle I)

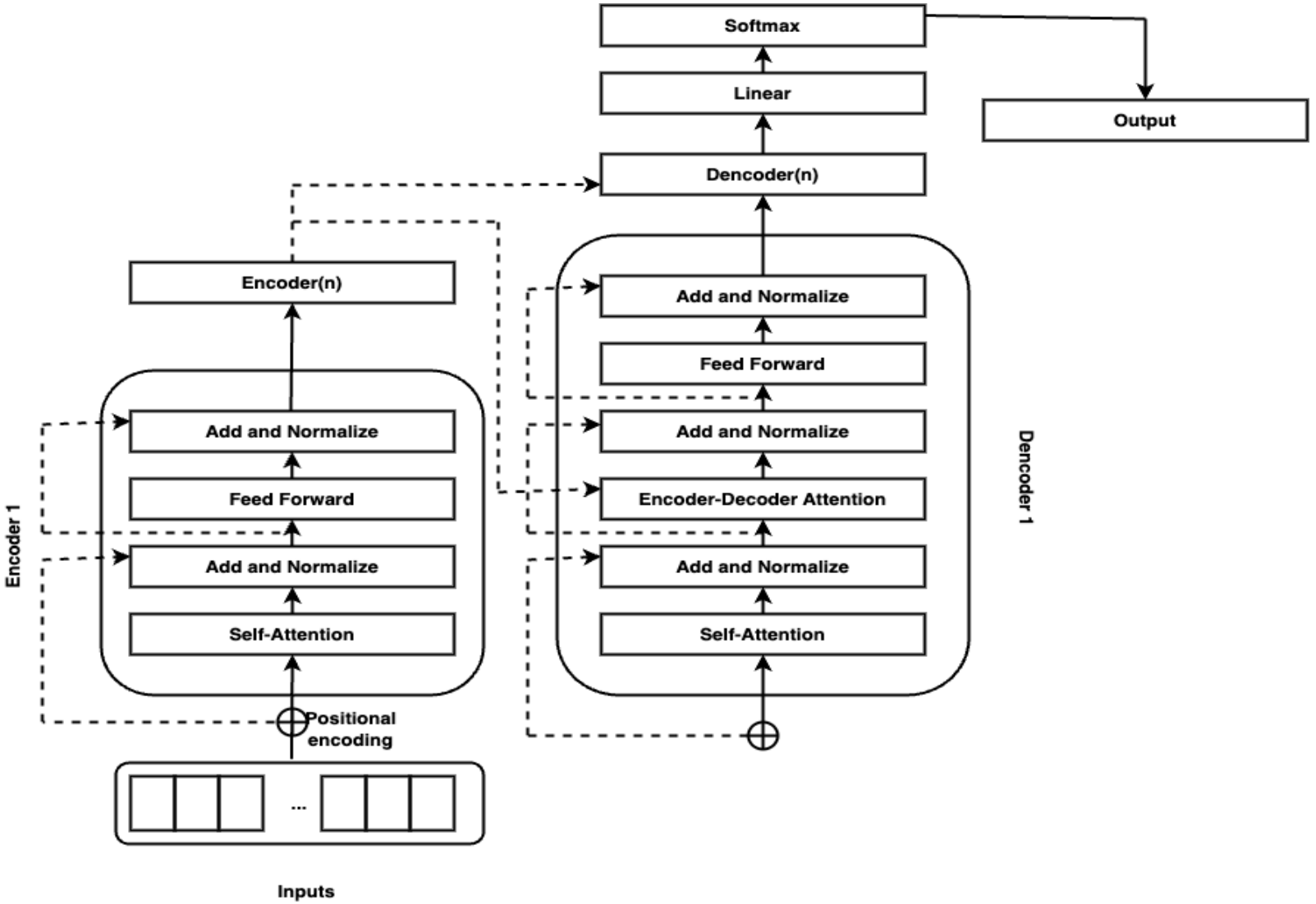
Whisper model will be used for Speech-to-Text conversion during interview process.

- **Innovative Encoder-Decoder**: Efficiently captures speech nuances and produces coherent, accurate text from audio.
- **Self-Attention Mechanism**: Allows sequence parts to communicate, capturing long-range dependencies and contextual relationships.
- **MLP Layers**: Uses feed-forward neural networks to transform and process input audio sequences effectively.
- **Normalization Layers**: Ensures model stability through regulation and normalization processes.
- **Complex Relationship Learning**: Captures long-range dependencies between input elements, enhancing understanding.

Methodology – [7]
(Whisper Working principle II)

- **Cross-Attention Mechanism**: Focuses on specific input parts using contextual information from the encoder.
- **Residual Connections**: Utilizes strengths of all components to create accurate, coherent subtitles.
- **Special Tokens**: Directs tasks like language identification, timestamps, multilingual transcription, and translation.

Methodology – [8]
(T5 for Conditional Generation)



Methodology – [9]
(T5 Working principle I)

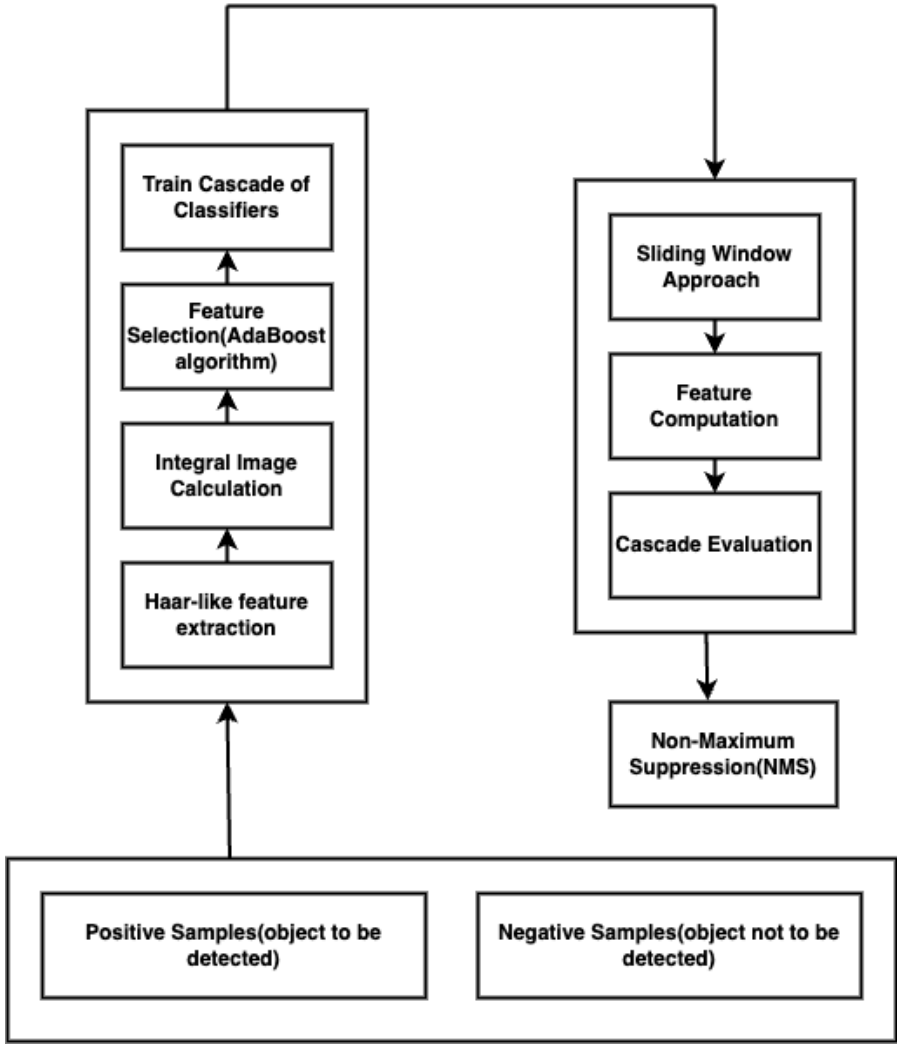
T5 model will be used for generating dynamic questions based on the answer of the interviewee.

- The encoder consists of multiple identical transformer layers, each with a self-attention sub-layer and a feed-forward neural network sub-layer to model the relationships between different parts of the input sequence.
- The decoder also consists of multiple identical transformer layers which also uses self-attention to model output sequence relationships, as well as cross-attention to account for encoder outputs.
- Language modeling head permits the version to carry out language modeling tasks, along with textual content generation, through predicting the maximum probably subsequent token given the preceding context.

Methodology – [10]
(T5 Working principle II)

- The combination of linear transformations followed by a softmax function at the end of the architecture plays a crucial role in converting the model's final hidden states into a probability distribution over the possible output tokens

Methodology – [11] (OpenCV Haar Cascade Model)



Methodology – [12]
(Haar Cascade Working principle)

- OpenCV Haar Cascade model is a machine learning approach widely used for detecting objects, particularly faces and eyes.
- Haar-like features capture essential visual cues like edges, while the integral image enables rapid computation.
- AdaBoost selects the most significant features to create a robust classifier, organized in a multi-stage cascade to enhance detection accuracy.
- Training involves collecting diverse images and extracting features. Detection uses a sliding window approach and Non-Maximum Suppression to refine results.
- In our project, this model ensures real-time, accurate eye detection, crucial for monitoring engagement

Instrumentation – [1] (Hardware Requirements)

- Cloud Computing Resources:

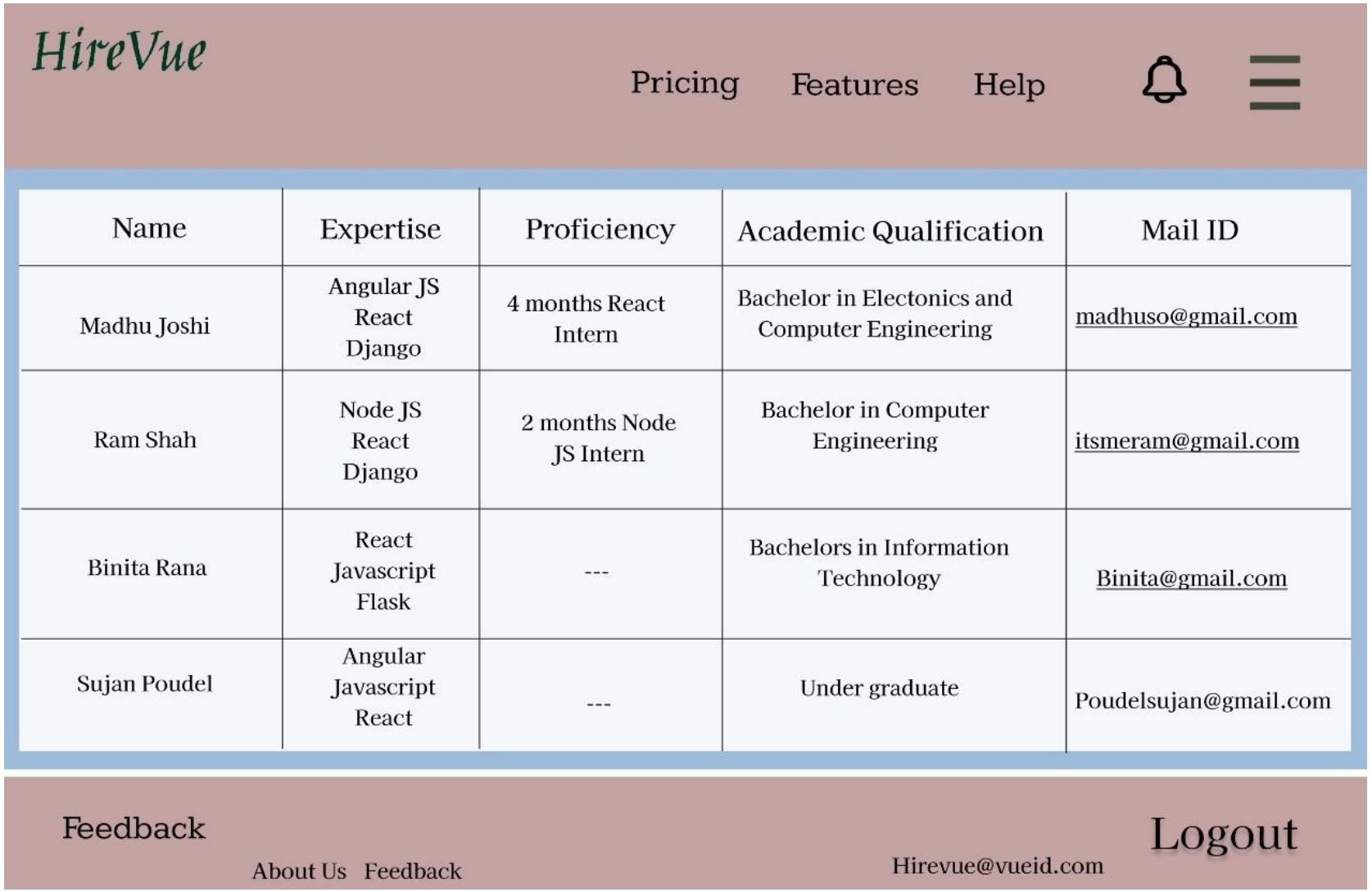
  Kaggle:

Hardware Components	**Details**
CPU	Intel Xeon 2.20 GHz
GPU	NVIDIA T4 x2 GPU
Number of GPUs	2
CUDA cores per GPU	2560
RAM	16 to 30 GB
Memory Bandwidth	320 GB/s

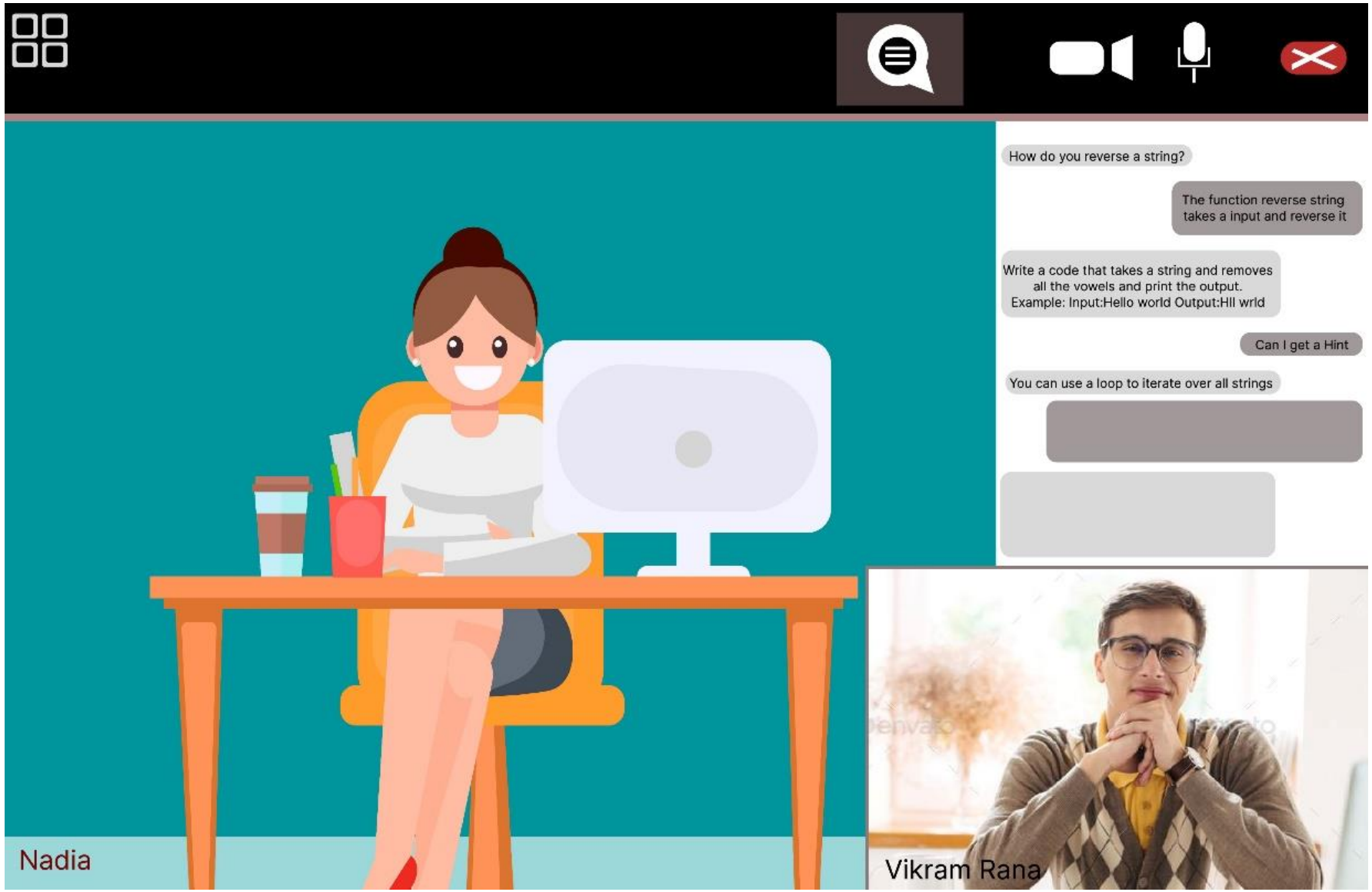
Instrumentation – [2] (Software Requirements)

- Text Editor: Visual Studio Code
- Programming Language: Python
- Version Control : Github
- Natural Language Processing: NLTK, SpaCy
- Framework : PyTorch
- Data Visualization : Matplotlib
- Frontend: HTML, CSS, JavaScript
- Backend : Nodejs , ExpressJs

Expected Results - [1]



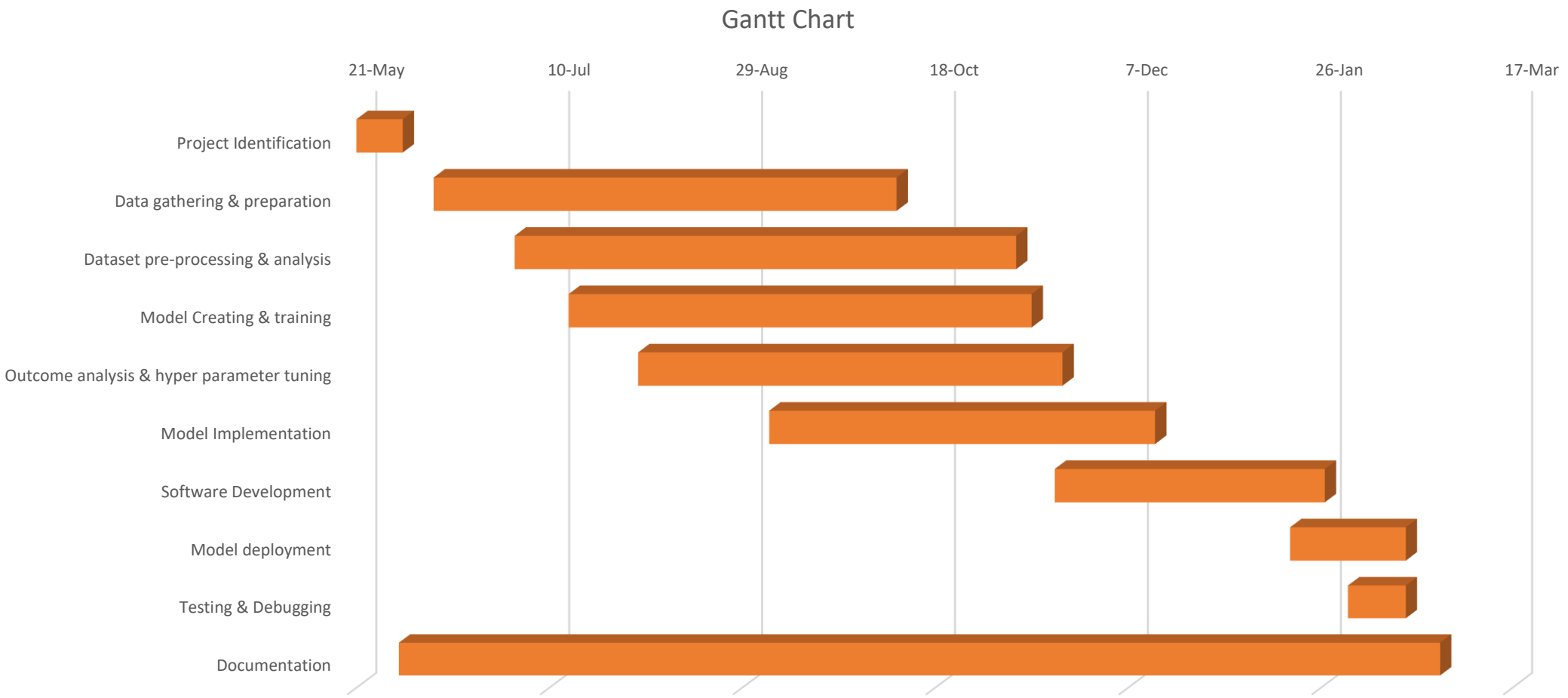
Expected Results - [2]



Project Applications

1. Recruitment and Hiring:
   - Automated Screening
   - Unbiased Evaluations
   - Scalability
2. Business Efficiency:
   - Cost Saving (32.7%)
   - Resource Optimization
3. Educational Sector:
   - Practice Interviews
4. Research:
   - Behavioral studies
   - Data collection.

Tentative Timeline



Estimated Project Expenses

Cost Domain	**Expected Cost**
Cloud Storage	Rs. 3000
Miscellaneous	Rs. 3000
Total	**Rs. 6000**

References – [1]

- [1] F. Chopra and I. Haalad, "Conducting Qualitative Interviews with AI(2023)," Cesifo Working Papers, p. 10666, 2023.
- [2] B. C. Lee and B. Y. Kim, "Development of an AI-based Interview System for Remote Hiring," International Journal of Advanced Research in Engineering and Technology (IJARET), vol. 12, no. 3, pp. 654-663, 2021.
- [3] A. Nigam, R. Pasricha, T. Singh and P. Churi, "A Systematic Review on AI-based Proctoring Systems: Past, Present and Future," Educ Inf Technol, vol. 26, pp. 6421-6445, 2021.
- [4] P. R. S B, M. Agnihotri and D. B. Jayagopi, "Improving Asynchronous Interview Interaction with Follow-up Question Generation," International Institute of Information Technology, vol. 6, pp. 79-89, 2021.

References – [2]

- [5] T. R.K, B. Shinde, N. Rasal and S. Ghorpade, "Smart Interview System using AI Technology," International Research Journal of Modernization in Engineering Technology and Science, vol. 6, no. 2, 2024.
- [6] R. Mengi, H. Ghorpade and A. Kakade, "Fine-tuning T5 and RoBERTa Models for Enhanced Text Summarization and Sentimental Analysis," 2023.
- [7] N. Mulla and P. Gharpure, "Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications, Progress in Artificial Intelligence, vol. 12, pp. 1-32, 2023.

References – [3]

- [8] M. Wng, P. XIe, Y. Du and X. Hu, "T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions," Applied Sciences, vol. 13, no. 12, 2023.
- [9] V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," arXiv preprint arXiv:1609.03499, vol. 12, 2016.