

AI-Powered Classroom Assistant using Multimodal Learning

Aiswarya Anil, Anagha Anil, Hima Rose George

Saintgits Group of Institutions

June 24, 2025

Introduction

- AI-powered assistants support personalized learning in classrooms.
- Integrates voice, text, and image-based student inputs.
- Improves engagement and query resolution using multimodal inputs.

Input Modalities Handled

- ① Text-Based Queries
 - User types questions for quick topic explanation
- ② Speech-Based Queries
 - Student speaks into the mic for voice-driven interaction
- ③ Image-Based Queries
 - Students submit photos or diagrams for context-aware answers

Challenges in Development

- Handling diverse and noisy classroom inputs
- Ensuring fast response time for each modality
- Combining NLP, speech, and vision models efficiently
- Building lightweight solutions for real-time deployment

Core Functional Modules

- ① Visual Question Answering
 - Extracts text meaning from uploaded images or diagrams
- ② Speech-to-Text Conversion
 - Uses pretrained models to process voice inputs
- ③ Natural Language QA
 - Processes typed and transcribed queries using transformers

Model Optimization with OpenVINO

- Used Intel's OpenVINO toolkit for runtime acceleration
- Optimized transformer-based QA for faster inference
- Reduced latency without loss of accuracy
- Ideal for edge or low-resource classroom environments

Evaluation Metrics

- Accuracy: Correctly answered student queries across modules
- Precision: Relevant answers to specific, unstructured inputs
- Recall: Number of correctly processed queries among all types
- F1 Score: Weighted performance of precision and recall

Conclusion

- AI assistants can revolutionize smart classroom learning
- Combining text, speech, and vision improves accessibility
- OpenVINO brings real-time performance to multimodal AI

Questions?

