Multimodal Classroom AI Assistant using OpenVINO

Submitted by Kewal Thacker and Siddharth Subramanian

Introduction to the Problem Statement

The landscape of education is undergoing a rapid digital transformation. Traditional classroom setups often fall short in meeting the varied needs of learners, especially in a post-pandemic world that demands hybrid, asynchronous, and personalized learning experiences. Students frequently face difficulties such as:

- Inability to attend or revisit live lectures

- Overload of information without adequate summarization

- Limited opportunities for individual doubt resolution

- Lack of engaging visual aids or supplementary content

Simultaneously, educators are stretched thin, managing both content delivery and individual queries, which reduces the quality of teaching. There is a clear and urgent need for a system that can enhance the teaching-learning process using intelligent automation and artificial intelligence.

## Project Overview and the Problem It Aims to Solve

This project presents a comprehensive solution in the form of a multimodal Classroom AI Assistant. Designed with both students and educators in mind, this assistant leverages state-of-the-art AI models optimized with OpenVINO to deliver a powerful educational tool that transforms videos into digestible and interactive learning experiences.

The assistant takes a recorded lecture video or YouTube link as input and outputs:

- A complete lecture transcript

- Concise and well-structured notes generated using the Qwen 2.5 language model

- Relevant Google search results for extended reading

- Curated Google Images and additional educational YouTube videos

- An education-focused chatbot powered by Qwen for real-time doubt clarification

By automating and enhancing content delivery, this assistant aims to improve knowledge retention, increase accessibility, and enable on-demand learning for students across diverse educational backgrounds.

Technical Architecture and Pipeline Details
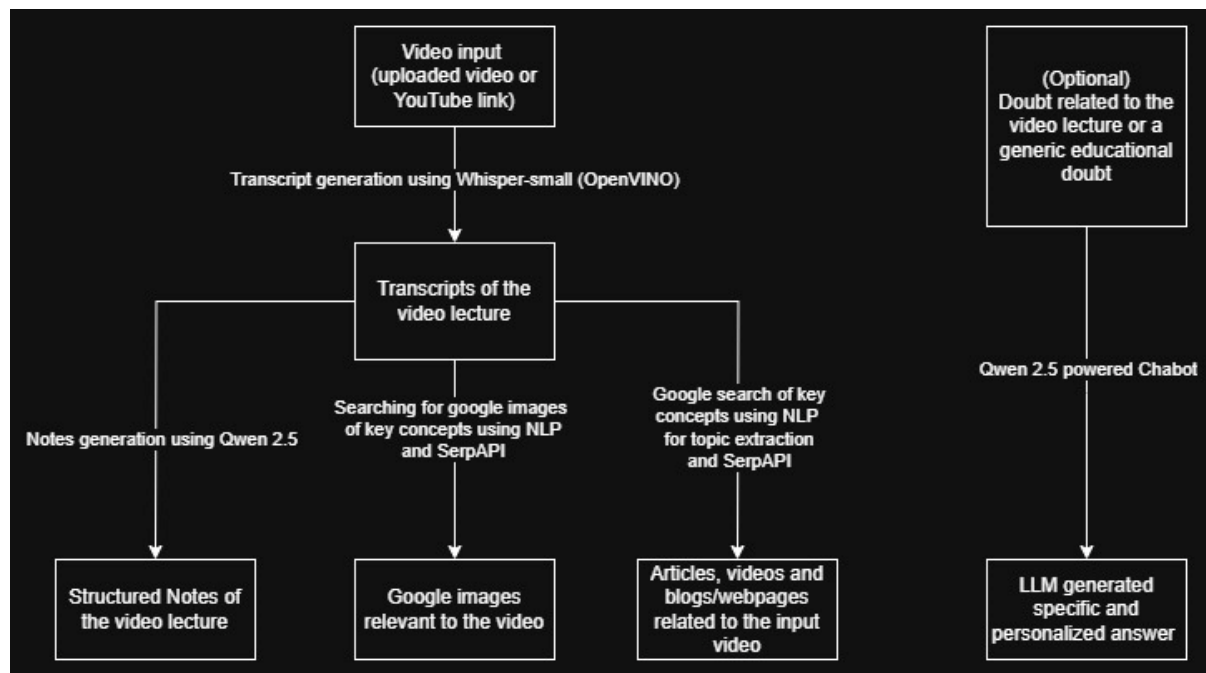
The system consists of a modular pipeline with several AI-powered components. All heavy AI inference tasks are accelerated using the OpenVINO Toolkit for low latency and efficient performance.

Input Module

- Accepts either a video file uploaded by the user or a YouTube link
- Uses yt_dlp to download and extract content from YouTube
- Processes video for audio extraction using moviepy or ffmpeg

Transcription Module

- Uses the pre-trained Whisper small model for speech-to-text conversion
- The Whisper model is converted into OpenVINO IR format for optimized inference
- Outputs a time-aligned, punctuated transcript of the lecture

Note Generation Module

- The full transcript is passed to the Qwen 2.5 large language model for summarization
- Qwen generates structured, concise lecture notes
- OpenVINO optimizes Qwen's inference, enabling fast processing even for lengthy inputs

Search and Visual Reference Module

- Key topics from the transcript are extracted using simple NLP heuristics
- These topics are used to fetch:
    - Google search results using SerpAPI
    - Related YouTube videos for deeper learning
    - Google Images to visually supplement the concepts

Educational Chatbot Module

- Uses the Qwen LLM fine-tuned for educational dialogue and knowledge retrieval
- Provides intelligent and context-aware responses to student queries
- Deployed with OpenVINO backend for seamless and efficient real-time interaction

User Interface

- A unified and intuitive Gradio-based interface hosts all components
- Allows seamless interaction with the assistant from video upload to chatbot communication

Note : The UI has provisions to access all individual components of the pipeline separately, so they can be used generally instead of specifically for videos as a part of the complete pipeline. All the models are individually optimized for educational purposes and hence give appropriate results.

Technologies and Libraries Used

- **OpenVINO Toolkit**: Optimization and acceleration of AI models
- **Whisper-small**: Speech-to-text transcription
- **Qwen 2.5 LLM**: Note generation and chatbot capabilities
- **yt_dlp**: Downloading and extracting YouTube videos
- **SerpAPI**: Google and YouTube search integration
- **Gradio**: Interactive user interface
- **Python (NumPy, requests, re)**: Backend scripting and processing

Instructions for Use

1. Run the application from the Jupyter Notebook or deploy it as a Python web app

2. Choose the video input method: upload a local file or paste a YouTube link

3. Wait for the system to extract audio and generate a full transcript

4. View automatically generated lecture notes created by Qwen

5. Explore topic-wise Google search results, related YouTube videos, and image references

6. Use the chatbot to ask educational questions related to the video or general topics

7. Optionally, download the output material for offline use

Use Cases and Benefits

For Students:

- Review full transcripts and notes for better retention

- Visual learners benefit from images and related videos

- Clarify doubts instantly using the chatbot

- Perfect for asynchronous learning and self-paced study

For Educators:

- Automatically generate study materials and reference content

- Share videos knowing students can review and learn autonomously

- Focus more on teaching by offloading repetitive queries to the chatbot

For Institutions:

- Integrate with e-learning platforms to enhance the digital classroom

- Improve accessibility for students with special needs

- Offer scalable support for large classroom sizes without compromising individual attention

Future Scope and Potential Improvements

- Add multilingual support with localized language models

- Integrate with LMS platforms such as Moodle, Google Classroom, or Canvas

- Add emotion and engagement detection using video-based facial analysis

- Enable real-time classroom summarization during live lectures

- Improve notes by extracting diagrams or slides using OCR and image analysis

- Generate quizzes and flashcards based on lecture content for self-assessment

- Offer cloud-based and offline deployment options for various educational setups

- Expand chatbot capabilities to support peer learning and collaborative Q&A

Limitations

- Model inference, while accelerated by OpenVINO, can still be time-consuming for long videos

- Performance may vary depending on the quality of audio input

- Google and YouTube search accuracy depends on API limits and relevance

Conclusion

The Classroom AI Assistant project showcases a powerful application of multimodal AI, using OpenVINO to build an accessible, intelligent, and user-friendly educational tool. By converting passive video lectures into interactive learning sessions, the assistant empowers students to learn effectively at their own pace and enables educators to scale their teaching efforts. With future enhancements and integrations, this system has the potential to become a cornerstone of smart digital education.

Note : UI screenshots and demonstrations are included in the GitHub repo's README doc as well as the demo video.

Teammate Contribution

Both teammates (Kewal Thacker and Siddharth Subramanian) worked on all aspects of the project parallelly. We shared our observations and training results with each other to help each other at every step of the project since this method allowed us to try more things and was more efficient than splitting the work modularly.