



Tech Architects



# NOVA Intelligence

MULTIMODAL AI ASSISTANT FOR  
CLASSROOMS



# The Problem

Modern classrooms face a critical gap in delivering personalized, real-time learning experiences that adapt to the unique needs and behaviors of each student. Traditional teaching methods often lack:

- Dynamic responsiveness to student queries across text, speech, and visuals.
- Tools to support latecomers or absent students with contextual class summaries.
- Mechanisms to detect confusion or disengagement in real-time.
- Language localization, limiting accessibility for diverse learners.
- Proactive learning support, like personalized Q&A or revision prompts.
- Continuous student performance tracking through automated portfolios.

These limitations result in passive learning, reduced retention, and a lack of equity in educational support across diverse classrooms.



# Vision & Mission

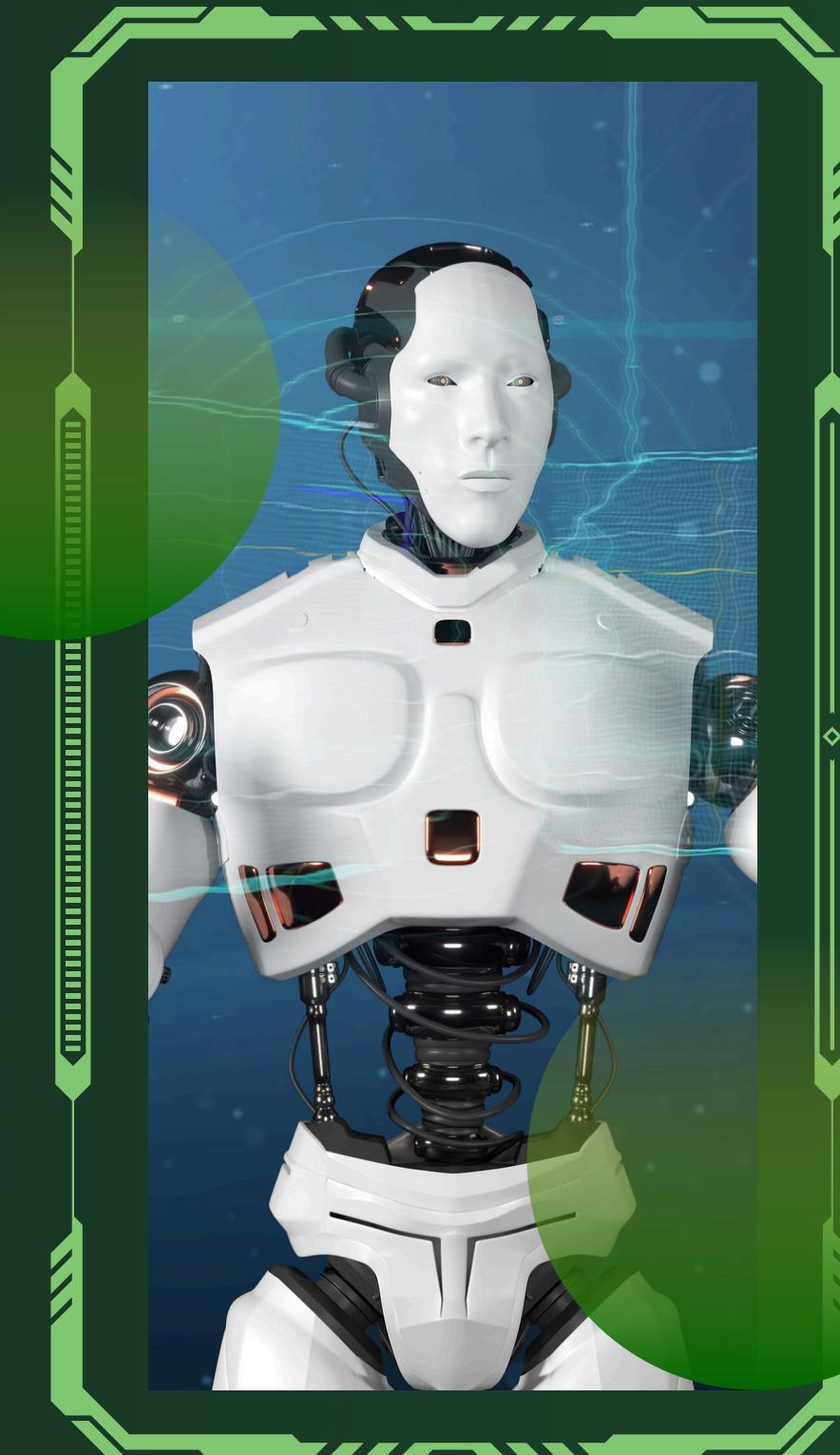
## Vision -----

We envision a future where classrooms are adaptive, inclusive, and emotionally intelligent — a space where AI assists in creating personalized learning journeys for every student, regardless of pace, background, or learning style.

## Mission -----

NOVA's mission is to augment teaching with AI that can:

- Comprehend and respond to multimodal student inputs
- Summarize, translate, and personalize classroom content
- Detect emotional cues and engagement levels
- Foster an active, supportive, and intelligent classroom ecosystem



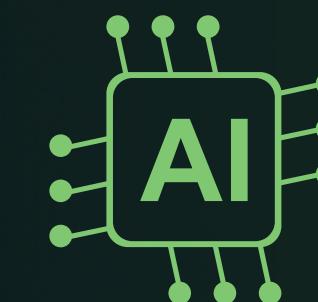
# Feature Overview



-  Text QnA: Ask questions and get answers using a local LLM
-  Image Captioning: Upload images and receive descriptive captions generated by the BLIP model.
-  Speech-to-Text: Record or upload audio and convert it to text for question input using speech recognition
-  Image QnA: Ask questions about uploaded images; the system uses image captions as context for the LLM.
-  Offline Operation: All models run locally after initial setup, so no internet connection is required for core features.



# Future Enhancements



## Voice Note Transcription

# Convert lectures into accurate, clean notes instantly

## Student Performance Portfolio

## Dynamic learning profiles evolving with progress

# Multilingual Translation

**Supports diverse language needs seamlessly**

# AI-Generated Q&A

Instant revision aid with smart question creation

## Emotion & Confusion Detection

Tracks engagement using advanced vision A

## Class Note Summarization

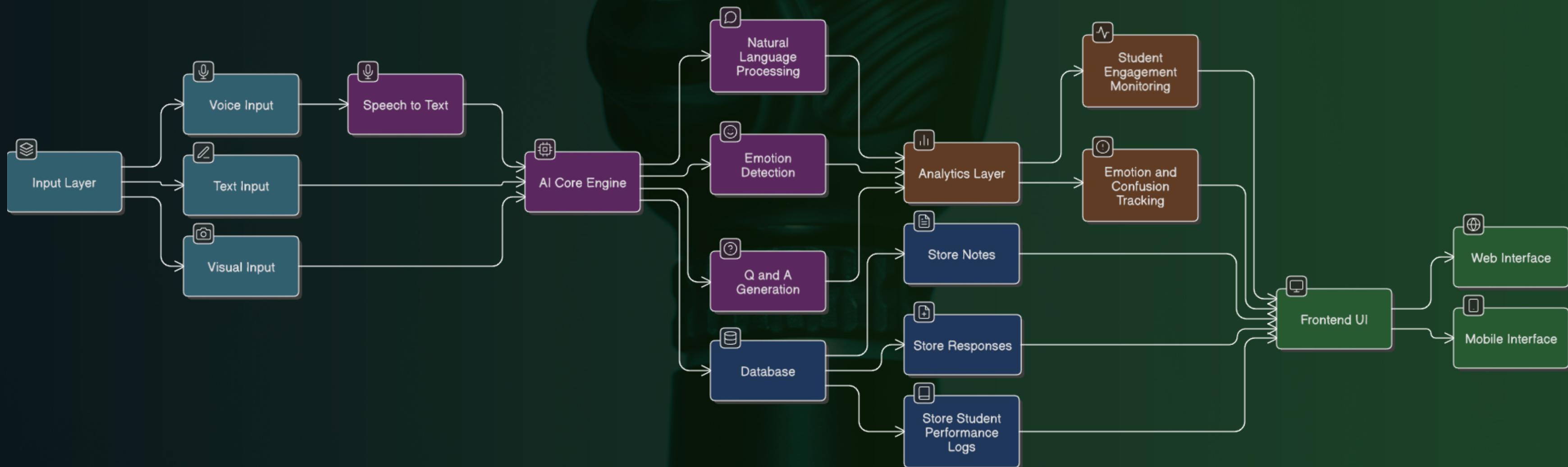
**Summarizes key points for quick review or catch-up**

# Real-Time Interactive Quizzes

**Proactive participation through instant quizzes**



# System Architecture





# Technology Stack

## 🗣️ Speech Processing

- Whisper (OpenAI) – Robust real-time speech-to-text, effective in noisy classrooms
- Wav2Vec2 (Meta) – Fine-tuned ASR for precise voice transcription with low latency

## 🧠 Natural Language Processing

- Phi-2 (Microsoft) – Lightweight, efficient language model used for reasoning, Q&A generation, and instruction following
- BERT (Google) – Deep contextual model for understanding student queries and refining classroom transcripts
- Hugging Face Transformers – Framework for fine-tuning and deploying transformer-based NLP models

## 👁️ Computer Vision

- OpenCV – Frame capture, face detection, and image preprocessing
- FER+ / AffectNet – Emotion recognition using facial expressions for confusion/disengagement tracking

## 🚀 Deployment Infrastructure

- Intel NPU / GPU – Hardware-accelerated processing for real-time inference
- OpenVINO Toolkit – Optimized runtime for deploying models like Phi-2 and Wav2Vec2 at the edge
- ONNX Runtime – Ensures model portability and cross-framework compatibility



# Aditya khandelwal

VOICE RECOGNITION AND CONVERSION TO TEXT

# Team

## Dhruv Bhadhotiya

LLM AND IMAGE CAPTIONING

## Deepti Yadav

TEAM LEAD AND VISION

## Kajal Chaudhary

FRONTEND AND DATABASE DESIGN

## Madhavika Sisodia

FRONTEND AND DATABASE DESIGN





# Thank You!