ShikshaNetra — Technical Summary Report

(Problem Statement PS2 – Mentor Scoring AI)

1. <u>Problem Statement</u>

Mentor evaluation in educational institutions is currently subjective, inconsistent, and difficult to scale. Human evaluators vary widely in judgment, struggle to measure technical correctness, cannot track behavioral cues (eye contact, emotion, gestures), and often introduce accent or language bias.

- Standardized scoring
- Fairness across accents/languages
- Objective measurement of clarity, engagement, technical depth
- Personalized, actionable feedback
- Automated analysis of multimodal teaching behavior

---

2. <u>Approach & AI Components</u>

ShikshaNetra introduces a **multimodal scoring pipeline** combining audio, video, and text analysis with an LLM-powered feedback generator.

A. Audio Analysis

- Whisper V3 (speech-to-text)
- Librosa (clarity, fluency, pauses, fillers)
- Speaker diarization + energy profile (PyAnnote)
- Confidence & tonal stability scoring

B. Video Analysis

- OpenCV frame extraction
- MediaPipe for face landmarks, gesture intensity, eye-contact tracking
- Emotion recognition (FER / DeepFace)
- Engagement scoring based on visual cues

C. Text & NLP Analysis

- BERT/Transformer-based topic relevance
- Technical depth evaluation (concept correctness, coverage)
- Interaction index (questions asked, inclusiveness, structure)

D. Bias Correction Layer

- Accent normalization using SpeechBrain LID
- Pitch & linguistic bias compensation
- Demographic fairness calibration
  This ensures fair scoring for teachers across regions, genders, and language styles.

E. Scoring Engine

A weighted fusion engine generates 0–100 scores for 5 key metrics (Clarity, Confidence, Engagement, Technical Depth, Interaction).

F. Generative AI Feedback Layer

LLM (GPT-5/Gemini-2.5-pro) generates:

- Personalized improvement tips
- Teaching style explanation
- Factual accuracy check

- Multilingual summaries
- Title + hashtag generation

3. Technical Architecture Overview

1. **User uploads a teaching video**
2. **Preprocessing:** audio extraction, frame extraction, speech-to-text
3. **Multi-Modal AI Analysis:**

- Audio → clarity, confidence
- Video → emotion, gestures, engagement
- Text → technical depth, topic coverage

4. **Scoring & Feature Fusion**
5. **Bias & Fairness Normalization**
6. **LLM Feedback Layer**
7. **Dashboard Output (Next.js + Chart.js)**

Backend: **FastAPI, PyTorch, MongoDB**
Frontend: **Next.js, Tailwind, Chart.js, Vercel**

---

4. Challenges & Mitigations

Challenge	Mitigation Strategy
Accent/language bias	Demographic calibration + linguistic normalization
Subjective scoring	Objective multimodal metrics + fairness layer
Technical correctness validation	Transformer-based topic & concept analysis
Real-time performance	Optimized preprocessing + async pipeline
Privacy of mentor data	Local processing + encrypted storage
Model stability across subjects	Fine-tuned multilingual datasets

---

5. <u>Roadmap to Final Build</u>

Phase 1 – Core AI Foundation

- Audio → Video → Text pipeline
- Early scoring engine + Whisper + Emotion AI

Phase 2 – Fairness & Transparency Layer

- Accent/gender/language bias correction
- Explainable AI metrics

Phase 3 – Generative AI Feedback

- Personalized suggestions
- Multilingual summary generation
- Teaching style detection

Phase 4 – Advanced AI Coaching

- Conversational AI coach
- Session-wise emotional engagement timeline

Phase 5 – Deployment & Scaling

- Next.js dashboard
- Cloud hosting (Vercel + FastAPI backend)
- Organization onboarding