AI Mock Interview Simulator — Technical Documentation

1. Introduction

The **AI Mock Interview Simulator** provides students with a real-time, intelligent mock interview experience using AI-driven question generation, voice-based interaction, animated interviewer avatars, and automated evaluation. It aims to reduce the gap between theoretical preparation and actual interview performance by offering a scalable, adaptive, and accessible platform.
The Round-2 prototype (70% complete) includes a functioning AI pipeline, integrated frontend–backend architecture, cloud deployment, persona-based media components, and structured report generation.

---

2. Problem Statement

Students often lack expert guidance, structured mock interviews, and personalized feedback. Existing platforms do not offer **multimodal interaction**, **dynamic question generation**, or **AI-based evaluation**.
This system solves these limitations by:

- Conducting human-like interviews using avatar personas
- Generating personalized questions
- Evaluating responses for technical, communication, and behavioural skills
- Producing detailed improvement reports
- Supporting cloud inference now and offline inference in Round-3

---

3. System Architecture Overview

Frontend (React + Tailwind)

- Responsive UI for live interview, dashboard, and reports
- Camera/mic integration
- Interviewer avatar (male/female)
- Speech recognition & persona-based TTS output

Backend (FastAPI)

- APIs for question generation, evaluation, scoring, and report creation
- OpenRouter AI integration
- Secure session-based communication

- PDF report generation

AI Pipeline

Cloud (Current): Claude, Mixtral, Amazon Nova Lite
Used for: generating questions, evaluating answers, summarizing reports
Offline (Round-3): Llama-3-8B / Mistral-7B

- Local inference
- Vector DB for skill-based questions
- Improved scoring via rubric-based fine-tuning

Database (SQLite)

Stores transcripts, scores, and session metadata.

Deployment

Frontend on **Vercel**, backend on **Render/Railway**, with cloud build and downloadable reports.
(*Architecture diagram available in original document*)

---

4. Development Summary

Week 1 — Setup & Architecture

- Built architecture workflow, backend structure, database schema
- Created UI wireframes & landing page
- Integrated resume parsing
- Repository + documentation prepared
  Output: Architecture + environment ready

Week 2 — AI Pipeline Integration

- Connected OpenRouter models
- Implemented question → answer → evaluation flow
- Added fallback logic and internal testing
- Created API documentation
  Output: Functional AI-enabled backend

Week 3 — Frontend Integration

- Connected frontend with backend
- Added interviewer personas (male/female/bossy female)
- Enabled avatar videos, TTS, STT, transitions

- Added automatic PDF report generator
- Completed cloud deployment
  Output: Fully working prototype with end-to-end flow

---

5. Challenges & Mitigations

- **Audio/video sync issues:** resolved using state-based media hooks
- **Cloud rate limits:** caching + fallback templates
- **Persona mismatch:** improved persona selection logic
- **High prompt cost:** optimized prompts (<900 tokens)

---

6. Roadmap (Round-3)

Planned enhancements include:

- Full offline LLM inference
- Advanced scoring algorithms
- Enhanced personas & UI animations
- Analytics dashboard
- Facial expression & sentiment scoring
  Goal: Deliver a polished, offline-capable simulator for IIT Bombay.

---

7. Conclusion

The team has delivered a strong and scalable 70%-complete prototype featuring real-time AI interviews, persona-based interactions, automated evaluations, and structured PDF reporting. The architecture is robust, cloud-tested, and ready for the final offline enhancement phase. Round-3 will focus on accuracy, optimization, and complete on-device operation to ensure a high-quality final product.