Monthly Progress Report

Project ID: P077 **Date: 30 Sept 2025**
Project Title: AI-Driven System for Adaptive Learning Platform
Student names (with Roll no): Vaibhav Yadav (142201015) & Shreyas Vadthya (112201038)
Mentor/s: Dr. Sreenath Vijaykumar

Work done in the current month (Limit to maximum 2 pages):

Cumulative Project Progress Report

We've made solid progress in setting up the foundation of our project and have now moved into a fully implemented backend system. This report provides a cumulative summary of all work done, from initial research and planning through to the current operational status of the application.

1. Domain Research & Analysis

- Looked into the GATE and CAT exam structures, since these form the core focus of the project.
- Collected and studied previous years' question papers to gain practical insights that will guide the design of useful and relevant features.

2. System Design & Architecture

- Created a detailed ER diagram, based on the exam structures, to make sure our database design fits real-world needs.
- Finalized the technology stack after carefully comparing options, with an emphasis on speed, scalability, and modern best practices: **FastAPI**, **PostgreSQL**, **ChromaDB**, and the **Groq API** for LLM access.

3. Implementation & Development

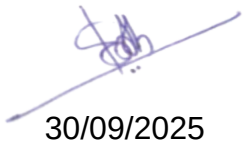
- **Data Pipeline Automation**
  - Developed and implemented a robust data pipeline in Python to process unstructured Previous Year Question (PYQ) papers.
  - **PDF Parsing:** Created a sophisticated script (1_parse_pdfs.py) that intelligently extracts questions, options, and answers from various complex PDF formats.
  - **Data Structuring:** The raw, unstructured data is now automatically converted into clean, machine-readable JSON files, organized section-wise (CAT_QA.json, etc.).
  - **Vector Database Creation:** Implemented a script (2_build_vector_db.py) that uses the sentence-transformers library to convert the structured JSON data into a high-performance, searchable vector database using ChromaDB. This forms the core knowledge base for our AI.
- **AI Question Generation Engine**

- The core AI feature of the application, the **Retrieval-Augmented Generation (RAG) service**, has been fully implemented in rag_service.py.
- **LLM Integration:** Successfully integrated with the high-speed Groq API to leverage the llama-3.1-8b-instant model for generating new questions.
- **Full Exam Generation:** The service can now generate a complete and unique mock exam that adheres to the specific structure of the CAT exam, including the correct number of MCQs and TITA questions for each section (VARC, DILR, QA).
- **Rate-Limit Handling:** Implemented a sequential, section-by-section generation process with built-in delays to respect API rate limits, ensuring stable and reliable performance.

- **Live Backend Application & API**
  - Developed a complete and runnable backend server using **FastAPI**.
  - **User Database:** Set up a **PostgreSQL** database to securely store all user and subscription data.
  - **Multi-User Authentication:** Implemented a secure, token-based (JWT) authentication system that handles multi-user registration and login.
  - **Role-Based Access Control:** The system now supports different user roles, with dedicated API endpoints for regular users and a protected endpoint for administrators.
- **4. Current Status & Next Steps**
  - The Main part of backend is complete and fully operational, have to add more minor features
  - The entire user workflow—from registration and login to generating an AI-powered exam—is testable through the interactive API documentation (/docs).
  - Report system and storing user exam results and calculating exam report and AI-summary to give feedback.

Github Link: https://github.com/The-Vaibhav-Yadav/OELP

Feedback from the mentor/s:



30/09/2025

Signature of the mentor/s with Date