

PARTH RRAHUL SHAH

parthrrshah202@gmail.com | +91 9588463603 | India | [Linkedin](#) | [GitHub](#) | [Website](#)

EDUCATION

SCTR's PUNE INSTITUTE OF COMPUTER TECHNOLOGY - SPPU

09/2025 – Present

Bachelor of Technology (B.Tech) in Computer Engineering

Second Year (3rd Sem from Jul 2026)

CGPA - **9.85**

Relevant coursework: Data Structures & Algorithms, Object-Oriented Programming, Discrete Mathematics, Database Management Systems.

PROJECTS

FinFlow - Financial Data Normalization Pipeline

05/2026 – Present

FastAPI • PostgreSQL • pgvector • Sentence Transformers • Python

- Building an end-to-end ETL pipeline normalizing financial data from 5+ Indian bank statement formats (SBI, HDFC, Axis) across 4 file types (PDF, CSV, Excel, text) into a unified 6-field canonical schema; designed as v1 of a Universal Data Ingestion Engine extensible to arbitrary structured data sources.
- Engineered a pgvector-based semantic column-mapping engine using NLP embedding similarity to align 50+ column-name variations with confidence scoring; benchmarking 3 matching approaches on unseen formats.

[Auto Schema Pipeline](#)

02/2026 – 04/2026

Python • PostgreSQL • Gemini API • Streamlit

- Built a zero-configuration system that auto-infers database schemas from CSV files up to 1M+ rows, eliminating manual schema design; deployed live with real users.
- Integrated the Gemini API to generate up to 10 context-aware SQL analytical insights per dataset, surfaced via a real-time Streamlit dashboard with full run-history tracking.

[Log Data Pipeline](#)

01/2026 – 02/2026

Python • Pandas • Matplotlib • Streamlit

- Engineered a modular ETL pipeline transforming 10,000+ server logs into PostgreSQL with automated anomaly flagging, a Streamlit dashboard, and validation from a senior at PICT via LinkedIn.

Confidence Detection for LLMs (LLM Calibration Research) — with Prof. Pankaj Patil

02/2026 – Present

Model Reliability • Probabilistic Machine Learning • LLM Evaluation & Calibration • ECE

- Built a calibration framework measuring Expected Calibration Error on the IndicQA dataset to flag low-confidence LLM outputs before they reach users.

TECHNICAL SKILLS

Programming Languages — C | C++ | Python

Core Concepts — Data Structures & Algorithms | Data Analysis | Object Oriented Programming

Data Engineering — FastAPI | REST APIs | PostgreSQL | ETL/ELT Pipelines

AI & ML — Machine Learning | Natural Language Understanding | NLP | Sentence Transformers | Embeddings & Vector Databases | pgvector | RAG | Deep Learning Fundamentals

Tools — Git & GitHub | VS Code | Docker | Pandas | Streamlit | Matplotlib | Gemini API

ACHIEVEMENTS

Competitive Programming

- LeetCode 200+ problems, peak contest rating 1,426 (all weekly/biweekly contests with upsolving)
- HackerRank 5-star in Problem Solving (C++).

Co-curricular

Highly active across competitions and hackathons like

- Google Big Code
- Microsoft AI Skills Fest & Agents League Hackathon,
- 6 hackathons in first year including PICT InC fest (Pradnya CP, Impetus Hackathon)