

Pranay Shaurya

Bhopal, Madhya Pradesh, India

✉ Pranayshaurya.pro@gmail.com — [in linkedin.com/in/pranay-shaurya-06106b252](https://www.linkedin.com/in/pranay-shaurya-06106b252) — [🔊 Pranay-Shaurya](#) — [📱 9334192924](#)

PROFILE SUMMARY

Final-year B.Tech Computer Science student specializing in Generative AI systems and backend engineering. Experienced in building Retrieval-Augmented Generation (RAG) pipelines using LangChain, vector databases, and LLM APIs on Google Cloud. Strong in Python and SQL with hands-on exposure to Vertex AI and Gemini. Passionate about designing scalable AI-powered applications in production-oriented environments.

EDUCATION

Vellore Institute of Technology – Bachelor of Technology, CSE (Health Informatics)
CGPA: 8.16/10.0

Sep 2022 – 2026

TECHNICAL SKILLS

Programming: Python, SQL, C++

Generative AI: RAG Pipelines, LangChain, ChromaDB, Prompt Engineering

LLM Platforms: Gemini, Vertex AI (Google Cloud)

Machine Learning: Scikit-learn, TensorFlow

Databases: MySQL, PostgreSQL, Vector Databases

Cloud & Tools: Google Cloud Platform, Git, Jupyter Lab

EXPERIENCE

Generative AI Virtual Internship – Google Cloud

2025

- Built Generative AI applications using **Gemini and Vertex AI** on Google Cloud.
- Implemented prompt engineering techniques to enhance contextual response quality.
- Developed end-to-end AI workflows including embeddings generation, model invocation, and response handling.
- Gained practical exposure to scalable cloud-based LLM deployment and evaluation strategies.

PROJECTS

Gen-AI Document QA System (RAG-Based Architecture) - *Python, LangChain, ChromaDB, Gemini API*
[GitHub Repository Link](#)

- Designed and implemented an end-to-end Retrieval-Augmented Generation (RAG) pipeline.
- Built document ingestion module for chunking, embedding generation, and vector indexing.
- Integrated LLM API with Top-K similarity retrieval for contextual answer generation.
- Structured modular architecture separating ingestion, retrieval, and generation layers.
- Improved retrieval relevance through chunk-size and embedding parameter optimization.

Retail Sales Analytics System - *MySQL, Python (Pandas)*

- Designed relational schema for customer, product, and transaction datasets.
- Developed 25+ analytical SQL queries using JOINs, CTEs, and window functions.
- Implemented revenue, cohort, and Pareto (80/20) analysis for business insights.
- Performed data cleaning and ETL transformations using Pandas.

Alzheimer Disease Detection using CNN - *Python, TensorFlow*

- Preprocessed 6000+ MRI scans using normalization and augmentation techniques.
- Implemented CNN model with 5-fold cross-validation.
- Reduced overfitting by 25% through hyperparameter tuning.
- Achieved 92% classification accuracy; research accepted for publication in Springer.

CERTIFICATIONS

- Google Cloud – Generative AI Virtual Internship
- AWS Academy Graduate – Cloud Foundations
- Languages: English (Fluent), Hindi (Fluent)