


Ayush Kakkar

Delhi, India

☎ +91 8920815655 • ✉ kakkarayush0@gmail.com •  [LinkedIn](#) •  [GitHub](#) • [Leetcode](#)

EDUCATION

B.Tech. - Computer Science
Vellore Institute of Technology

2022 -2026
CGPA 8.7/10

EXPERIENCE

Data Engineering and Automation Trainee – LabelNest India Pvt. Ltd.

Dec, 2025 – Present

- Built and maintained scalable **ETL data pipelines** for structured datasets, ensuring reliable **ingestion, transformation, and storage**.
- Implemented data validation, schema enforcement, and integrity checks in **relational databases** to improve data quality and consistency.
- Use **Linux commands** to navigate systems, inspect logs, and assist in troubleshooting issues.
- Supported **QA and production issue resolution**, identifying edge cases and improving overall **pipeline reliability by ~25%**.
- Collaborated with **cross-functional teams** to enhance pipeline reliability and operational stability.


SKILLS

Languages & Frameworks: Java, Python(scripting, debugging), fastAPI

Tools & Technologies: Git, GitHub, Linux, VS Code, REST APIs, MySQL, PySpark


Core Concepts: Data Structures & Algorithms (DSA), OOPS, DBMS

PROJECTS

Video-Gen: AI video generation and scoring system |  [Python](#), [Gemini API](#), [Kling](#), [Docker](#)

Nov, 2024 – Feb, 2025

- Architected a high-concurrency backend using **FastAPI** to manage AI video synthesis, reducing server response time by **40%** via asynchronous task execution.
- Engineered **RESTful API** endpoints to orchestrate data flow between Gemini Flash 2.0 and **Kling models**, improving metadata accuracy by 25%.
- Optimized media processing pipelines by integrating **Cloudinary**, decreasing video rendering latency by 50% for end-users.
- Operationalized a server-side scoring algorithm with **6 evaluation metrics**, ensuring 95% consistency in output quality across varied prompts.

NYC Taxi Data Engineering Pipeline |  [PySpark](#), [Azure Data Factory](#), [Azure Data Lake](#), [Databricks](#)

Nov, 2025 – Jan, 2025

- Developed an **end-to-end batch data pipeline** processing **10+ million NYC Taxi trip records**, following **Bronze–Silver–Gold architecture**.
- Orchestrated ingestion workflows using **Azure Data Factory**, enabling **scheduled, fault-tolerant data ingestion** from raw sources.
- Applied **PySpark transformations** to clean, standardize, and enrich datasets, improving downstream analytical reliability by **40%**.
- Curated **analytics-ready Gold tables** optimized for reporting, enabling faster query performance for trend and revenue analysis.

ACHIEVEMENTS & EXTRA CURRICULAR ACTIVITIES

- **DSA Proficiency** — Completed 250+ coding challenges on LeetCode and GFG, demonstrating consistent logic building and algorithmic understanding.
- **Led a team to 2nd Runners-up at National Hackathon by OnlineSale.ai** — Secured 3rd rank among 200+ teams by building an AI-powered video ad platform using React and FastAPI that automates professional video creation from product details.
- **Smart India Hackathon 2025 Internal Round Finalist** — Achieved Top 75 in Smart India Hackathon 2025 for an innovative civic-tech application.
- **Core Team Member at VIT's Software Development Club** — Volunteered in hackathons and webinars that upskilled 300+ students.

CERTIFICATIONS

- **Oracle** : Oracle Cloud Infrastructure Developer Professional
- **IBM** : Devops, Agile & Design Thinking