

Vijay Fagaria

Software Development Engineer

✉ vijayfagaria97@gmail.com | ☎ +91-7014843462

in vijay-fagaria | 🌐 vijayfagaria-dev

PROFESSIONAL SUMMARY

Backend engineer with 3+ years in **Python**, progressing from multithreaded instrument-control systems to production **FastAPI** services on **SQL Server** and **MongoDB**. Owns **OMR** auto-grading and a multi-provider **LLM** pipeline that cuts inference cost by ~40% through batch-first routing, and built the centralized test architecture adopted across the backend team.

EDUCATION

Indian Institute of Technology, Kanpur — Bachelor of Science, Economics

2019 – 2023

Relevant Coursework: Data Structures & Algorithms, Database Management Systems, Object-Oriented Programming, Web Development

EXPERIENCE

SpeEdLabs (*EdTech SaaS*) | *Software Development Engineer II*

Nov 2025 – Present

- Designed **REST APIs** across assessment, practice, and reporting domains with **JWT** auth, **RBAC**, and **Pydantic** validation.
- Architected **OMR** evaluation pipelines for bulk sheet ingestion and automated grading using **OpenCV** and **NumPy**, applying perspective transformation and fiducial-based alignment to improve scan accuracy from **80% to 98%**.
- Designed the **pytest** architecture the team builds on, with transactional savepoint isolation, dependency overrides, and fixture-leak guards so integration tests run against live **SQL Server** and **MongoDB** without test-state leakage.
- Engineered an incremental metrics warehouse with **cron-scheduled**, state-tracked refreshes and precomputed aggregates, shifting dashboards from full-history scans to summary-table reads across **225K students** at **450+ schools**.
- Offloaded grading, reporting, and **LLM** workloads onto **Azure Service Bus** queues to keep the **Functions**-hosted API responsive under load, and routed file uploads directly to **Blob Storage** through scoped **SAS tokens**.
- Built a unified **transaction lifecycle** coordinating SQL Server and MongoDB writes within a single request boundary, with rollback and compensation handling that prevents **partial-state corruption** when either store fails.
- Optimized multi-provider LLM orchestration across **OpenAI**, **Anthropic**, and **Gemini**, cutting inference cost ~40% by routing work through batch APIs first, with inline fallback and tiered exponential-backoff retries.

Neuranics Pvt Ltd (*AI-Based Diagnostic Devices*) | *Software Development Engineer*

Jun 2023 – Oct 2025

- Architected a **multithreaded** Python instrument-control service, synchronizing camera, 3-axis stage motion, and image capture through a thread-safe command queue for stable unattended operation across long imaging runs.
- Designed a **config-driven** hardware abstraction layer that enabled new camera backends without application code changes.
- Engineered cell-detection and morphometric analysis using OpenCV and NumPy, validated by deploying the device on-site at AIIMS, running scans at the site and matching its output against AIIMS lab reports with **93% accuracy**.
- Developed real-time calibration and hardware-validation tooling that reduced operator setup effort by **roughly 70%**.
- Built a type-safe API bridge that generates **TypeScript** client interfaces directly from Python handler signatures, keeping frontend and backend contracts synchronized automatically as the API surface evolved.
- Automated full-range **autofocus** using edge-contrast sharpness scoring, eliminating manual focusing from every imaging run.

SmartTraveller | *Software Development Intern*

Jul 2022

- Built a **Node.js** and **Express** recommendation service aggregating travel inventory from multiple provider APIs into MongoDB, scoring options on weighted price, duration, and rating signals for per-user ranking.

PROJECTS

Hive (*shared-expense & settlement platform*) 🌐 hive-api · hive-web

Jun 2026 – Aug 2026

- Architected and shipped a full-stack expense and settlement platform on FastAPI, **SQLAlchemy**, and **Next.js**, covering bill splitting, rule enforcement, and debt settlement between flatmates; now in active daily use.
- Engineered a multi-strategy expense-splitting and settlement engine that partitions the balance graph into peer-to-peer obligations and settles in **n-1** transfers, reconciling every balance to an exact zero-sum result.
- Designed a derived-balance model computing every total from immutable source records rather than stored balances, with optimistic concurrency control using versioned aggregates and **HTTP 409** handling on conflicting writes.

TECHNICAL SKILLS

Languages

Python, TypeScript, JavaScript, SQL, C++

Backend & APIs

FastAPI, RESTful API design, asyncio, Pydantic, JSON-RPC, JWT, RBAC, Node.js, Express.js

Databases

Microsoft SQL Server, MongoDB, SQLAlchemy, Beanie, Alembic, SQLite, indexing, query optimization

Cloud & DevOps

Azure (Service Bus, Functions, Blob Storage), Docker, Git, Linux, Bash, cron

Testing

pytest, unit and integration testing, test automation, Postman

CV & Data

OpenCV, NumPy, pandas, image processing, Jinja2, WeasyPrint

AI & LLM

LLM pipelines, OpenAI, Anthropic, Gemini, batch inference, prompt engineering

Frontend

React, Next.js, TypeScript, Zod, Tailwind CSS

Concepts

OOP, multithreading, system design, event-driven architecture, message queues