

Sai Yashwanth Dasari

☎ +91 9701149281 | ✉ saiyashwanthdasari@gmail.com | in LinkedIn | GitHub | </> Portfolio

ABOUT

2025 B.Tech graduate in AI/ML specialising in Python backend engineering and data infrastructure. Built production-grade async web crawlers, multi-agent LangGraph pipelines, and secure FastAPI backends with PostgreSQL at scale. Open source contributor to the FastAPI ecosystem (FastAPI Boilerplate) with **4 merged PRs**. Presented research at IHCI 2023 international conference on GAN-based Telugu audio classification.

EDUCATION

J.B. Institute of Engineering and Technology

Bachelor of Technology in Artificial Intelligence and Machine Learning | CGPA: 7.88/10.0

Hyderabad, Telangana

2021 – 2025

EXPERIENCE

Software Engineer Intern

Dec 2024 – April 2025

Workohol

Chennai, Tamil Nadu

- Built **15+** RESTful APIs using **Django REST Framework** for survey creation, response collection, and analytics in collaborative team of 4 developers
- Developed responsive **admin panel** using **HTML/CSS/JavaScript** with real-time dashboards, reducing manual report generation effort and improving workflow efficiency
- Optimized **MySQL** database queries achieving **40% faster** response times; designed schema supporting 10+ question types with complex relationships and proper indexing
- Implemented **JWT-based** authentication and role-based access control securing **3 user roles** (admin, manager, respondent) across the platform

Deep Learning Research Intern

Sep 2023 – Jan 2024

University of Hyderabad

Hyderabad, Telangana

- Authored research paper presented at **IHCI 2023 international conference** on Telugu song pattern recognition using Generative Adversarial Networks (GANs); collaborated with **research team of 3**
- Extracted audio features (MFCC, Chroma, Spectral Contrast, Zero-Crossing Rate) using **Librosa** and **Praat**; processed datasets with **Pandas** and **NumPy** achieving structured feature vectors for GAN training
- Implemented **SVM classifier** and **GAN** architecture for raga classification and emotion recognition across Telugu classical music corpus, annotated datasets using ASR transcription pipeline

PROJECTS

Mini Web Crawler | *FastAPI, Playwright, aiohttp, PostgreSQL, Docker* | [GitHub](#) | [Docker Hub](#)

May 2026

- Architected an **async BFS web crawler** with **dual fetch strategy**, i.e., **aiohttp** for static pages with automatic **Playwright** headless browser fallback for JS-rendered sites, detecting empty content via visible text extraction
- Implemented **robots.txt compliance** with async per-domain caching via **RobotsCache**; built **image extraction pipeline** with logo detection heuristics (alt text, class signals, homepage link traversal, size analysis)
- Designed **3-table PostgreSQL schema** (CrawlJob, CrawledPage, ExtractedImage) with **background job tracking** (pending → running → completed) and duplicate crawl detection with **force_recrawl** override; containerized and **published to Docker Hub**

Insurance Claim Document Processing Pipeline | *FastAPI, LangGraph, Gemini API, PyMuPDF* | [GitHub](#) April 2026

- Architected a **multi-agent LangGraph pipeline** to process insurance claim PDFs end-to-end — classifying pages into **9 document types** using Gemini vision, then routing to specialized extraction agents (identity, discharge summary, billing)
- Built 3 specialized **LLM agents** with exponential backoff retry logic for rate limit handling; implemented JSON parsing with graceful fallback for malformed LLM responses; built keyword-based **mock layer** to enable testing without API quota consumption
- Designed FastAPI endpoint accepting multipart PDF uploads; used PyMuPDF for **dual-mode extraction** (text + base64 image fallback for scanned documents), ensuring reliable parsing across digital and scanned PDFs

Finance Data Processing and Access Control Backend | *FastAPI, SQLAlchemy, JWT, pytest* | [GitHub](#) April 2026

- Built role-based finance dashboard backend with **3-tier RBAC (Admin, Analyst, Viewer)** enforcing access control across **15+** REST API endpoints using **JWT authentication and bcrypt password hashing**
- Implemented financial records management with soft delete, dynamic filtering, and pagination; designed analytics endpoints for income/expense summaries, category breakdowns, and monthly trends

- Achieved **33 passing pytest tests** organised by role with dependency injection overrides, shared fixtures, and isolated test database ensuring zero production data contamination

GhostDrop - Anonymous File Sharing | *FastAPI, NGINX, Docker* | [GitHub](#) July 2025

- Architected horizontally scalable system with **NGINX** load balancer across **3 FastAPI instances**; designed for zero-config horizontal scaling, supporting **2GB chunked file uploads**
- Implemented automated cleanup with **asyncio** background tasks; containerized with **Docker Compose** reducing setup time by **90%**

Hotel Booking Analytics & RAG Question Answering System | *FastAPI, LangChain, FAISS* | [GitHub](#) Mid 2025

- Built **RAG pipeline** over 119K+ hotel booking records — chunked and embedded domain documents using **sentence-transformers (MiniLM-L6-v2)**, indexed with **FAISS** (cosine similarity), integrated **Mistral-7B via HuggingFace Hub** for natural language Q&A
- Designed **LangChain PromptTemplate** with context injection; implemented hybrid retrieval combining vector similarity search with deterministic data queries for revenue, cancellation rate, and lead-time analytics
- Exposed 4 **FastAPI** endpoints (health, analytics, ask, query-history) with background embedding creation on startup; generated 5 analytical reports with Matplotlib/Seaborn charts returned as base64 for frontend rendering

TECHNICAL SKILLS

Backend & APIs: Python, FastAPI, Django, Django REST Framework, Flask, RESTful API Design, Pydantic, Microservices Architecture

Data & Crawling: Web Crawling, BFS Graph Traversal, aiohttp, BeautifulSoup4, Playwright, Data Pipelines, Pandas, NumPy

Databases: PostgreSQL, MySQL, SQLAlchemy ORM, Django ORM, Query Optimization, Schema Design, Indexing

Async & Concurrency: AsyncIO, aiohttp, Background Jobs, Concurrent Request Handling, Event-Driven Architecture

DevOps & Infrastructure: Docker, Docker Compose, NGINX Load Balancing, Git/GitHub, Linux, CI/CD concepts

Testing & Security: pytest (90%+ coverage), Unit/Integration Testing, JWT, OAuth2, RBAC, API Security

AI & LLMs: LangChain, LangGraph, RAG Pipelines, FAISS Vector Search, LLM Integration (Gemini, Claude 3.5 Sonnet, Mistral-7B), Prompt Engineering

Other: 300+ LeetCode, DSA, Open Source Contributor (4 merged PRs to fastapi-boilerplate), Agile, Code Reviews