Mithun Venkatesh Gowda

New York, NY | +1 (607) 381-0177 | gowdamithunv@gmail.com | Github: mgowda23 | LinkedIn:mgowdav

SUMMARY

Software Engineer with 3+ years building Python backend services on AWS, plus hands-on RAG and multi-agent LLM work, AI integrations, and production-oriented backend systems

WORK EXPERIENCE

Binghamton University - SUNY | AI Research Assistant *Jan 2026 – Present*

- Improved collaborative reasoning accuracy by 30% in a multi-agent LLM system by engineering shared vector memory with Python and Hugging Face Transformers.
- Cut memory retrieval latency by 25% across 10K+ reasoning vectors by optimizing the embedding-based retrieval layer.

Extern | AI Engineer *Sep 2026 – Present*

- Designed a RAG document Q&A system over pharmaceutical PDFs, by extracting and structuring content with PyMuPDF and pdfplumber and indexing it with LlamaIndex using FAISS and Chroma for retrieval.
- Implemented OCR for scanned documents and product labels using Tesseract, PaddleOCR, and EasyOCR.
- Compared LLMs including Gemini, Mistral, and Phi-2 for document Q&A, and deployed a Gradio chatbot for uploading documents and querying them in real time.

ITC Infotech Pvt Ltd | Software Engineer *Jan 2022 – June 2024*

- Developed Python/Flask REST APIs supporting an interactive forecasting platform across 200+ countries, processing database queries and serving structured JSON responses.
- Optimized PostgreSQL data access with Redis server-side caching, reducing database load by 40% and improving API response performance.
- Built React dashboards with interactive filters and geospatial visualizations, improving page-load performance by 28%.
- Containerized and deployed backend services using Docker, AWS ECS, and Amazon RDS, reducing API response time by 62% and supporting 2× the previous traffic volume.

PROJECTS

Zomato AI Data Platform | *Python, S3, Snowflake, Airflow, Docker, OpenAI API*

- Delivered an end-to-end ELT pipeline handling 10M orders, 23M order items, and 300K reviews, by loading S3 data into Snowflake through a keyless IAM integration and modeling it with dbt into Bronze/Silver/Gold layers with incremental `MERGE` facts that process only new rows instead of full rebuilds.
- Automated the full workflow into one daily run across 4 dependent tasks (load → dbt build → LLM enrichment → AI mart), by orchestrating an Airflow DAG in Docker with credentials injected through environment variables.
- Enabled plain-English analytics through 3 AI capabilities (review enrichment, RAG chat, text-to-SQL), by using `gpt-4o-mini` to structure sentiment/topic from free-text reviews, embedding-based retrieval for grounded answers, and a SELECT-only guard on generated SQL, all served in Streamlit.

Change Request System | *Python, Flask, SQLAlchemy, SQLite, Flask-Login, Docker*

- Developed a full-stack change-management platform with requester, approver, and admin roles, improving UI interaction efficiency by 40%, by implementing hashed-password session auth, CSRF-protected validated forms, and role-scoped dashboards, request lists, and access checks.
- Modeled a Draft → Submitted → Under Review → Approved/Rejected lifecycle, by designing a relational schema (users, requests, approvals, comments) with per-level approval tracking, timestamped reviewer decisions, and cascading relationships, containerized with Docker.

EDUCATION

Binghamton University - SUNY, New York

Master's in Computer Science | *3.67 / 4.00 GPA* *Aug 2024 – May 2026*

SKILLS

GenAI & Agentic: RAG, GraphRAG, Agentic Architectures, MCP, OpenAI API, FAISS, BM25, Prompt Engineering
Cloud & DevOps: AWS (EC2, S3, DynamoDB), Docker, Kubernetes, GitHub Actions, CI/CD Pipelines
Languages: Python, SQL, JavaScript, TypeScript, Shell Scripting, Java
Backend Frameworks: FastAPI, Django, Flask, Django ORM, RESTful APIs, Microservices, PyTest
Databases, BI & Frontend: PostgreSQL, MongoDB, MySQL, Redis, Power BI, Grafana, React, Next.js, HTML5, CSS3
Developer Tools: Git, Postman, OpenAPI, GitHub Copilot, AI-Assisted Coding Workflows

CERTIFICATIONS

- AWS Certified AI Practitioner and AWS Certified Cloud Practitioner