Gautama Shastry Bulusu Venkata

Fairfax, VA — +1 571-653-0056 — satya.2k02@gmail.com
linkedin.com/in/satya2603 — github.com/GautamaShastry — gautamportfolio.com

Professional Summary

Software Engineer (Backend/AI) building **RAG + tool-using agents** with **FastAPI** and **LangGraph/LangChain** and vector retrieval (**FAISS/ChromaDB/Qdrant**). Delivered measurable retrieval speedups (e.g., **150ms → 15ms**) and shipped production-minded APIs (rate limiting, retries, structured logs, tests). Published **2 NLP papers (arXiv)**.

Technical Skills

Languages: Python, Java, TypeScript/JavaScript, SQL, C/C++
Backend: FastAPI, Spring Boot, REST, async, validation, JWT/OAuth basics, Flask, Node.js/Express
AI/RAG: LangGraph, LangChain, RAG, embeddings, Sentence-Transformers, spaCy, FAISS, ChromaDB, Qdrant, Hugging Face, Ollama, ChatGPT, Claude
Cloud/DevOps: AWS (EC2, S3, RDS, IAM, CloudWatch, Lambda), Docker, Kubernetes, Jenkins, Rancher, Git/GitHub
DB/Tools: PostgreSQL, MySQL, MongoDB, JPA/Hibernate, HikariCP, Postman, Maven, Confluence

Professional Experience

Backflipt Hyderabad, India
Associate Software Engineer *Jan 2023 – Dec 2023*

- **UI modernization:** Re-architected a monolithic frontend into modular React/Redux feature slices and reusable components, accelerating feature delivery and supporting a measured **20%** increase in engagement.
- **Performance engineering:** Diagnosed render hotspots and optimized memoization/selectors/component boundaries, reducing re-render overhead by **40%** and improving page load time by **10%**.
- **Responsiveness + DevEx:** Improved client-server responsiveness using normalized state + caching (latency **15%**); authored **10+** Confluence docs that standardized patterns and reduced onboarding time **25%**.

Technical Projects

Support Sage — AI Customer Support Agent — *Python, LangGraph, ChromaDB, FastAPI, React* *Oct 2025 – Dec 2025*

- **Retrieval speed:** Built semantic search over **8 policy docs** and tuned chunking/overlap to cut retrieval latency **150ms → 15ms** while preserving answer quality.
- **Tool-using agents:** Implemented LangGraph **ReAct** with **9 tools** (order/profile/policy/escalation) to automate **7** cancellation scenarios with policy guardrails and escalation paths.
- **Production API:** Shipped FastAPI service with TTL session memory, IP rate limiting, Jira creation (retry/backoff), structured JSON logs, health checks, and a focused unit test suite (**32 tests**).

Document Assistant — *FastAPI, LangChain, FAISS, Ollama, Hugging Face* *Aug 2025 – Sep 2025*

- **High-throughput ingestion:** Built a FastAPI RAG backend and achieved **5.57 chunks/sec** ingestion via optimized chunk sizing (avg **555 chars**) and clean indexing flow.
- **Fast retrieval:** Implemented FAISS Similarity + **MMR** for diversity-aware retrieval; delivered **sub-50ms** average query time (**0.038s**, k=4).
- **Operationalization:** Added vector-store lifecycle endpoints (create/load/stats), citation-grounded prompting, and automated tests for ingestion/indexing/retrieval.

Student Survey Application — *Spring Boot, Docker, Kubernetes, Jenkins, AWS* *Mar 2025 – May 2025*

- **HA deployment:** Dockerized Spring Boot and deployed **3 replicas** on Kubernetes (Rancher) on AWS EC2 to support pilot high availability.
- **CI/CD automation:** Built Jenkins pipeline to build Maven artifacts, publish versioned images, and execute rolling updates; reduced deploy time to **less than 5 minutes**.
- **Reliable data layer:** Provisioned RDS (MySQL), tuned HikariCP pooling for concurrency, and verified strict validation + API behavior with **10+** Postman tests.

Resume Analyzer — *React, Spring Boot, Flask, PostgreSQL, GPT-4, LangGraph* *Feb 2025 – Nov 2025*

- **Multi-agent workflow:** Orchestrated **6 agents** (parse/score/match/feedback) using LangGraph to produce consistent, structured recommendations vs single-prompt baselines.
- **Scoring engine:** Built hybrid fit score using Sentence-Transformers (**60%**) + spaCy NER rules (**40%**) to improve matching precision.
- **Service stability:** Designed Spring Boot ↔ Flask boundary with retries/safeguards; generated analytics (Matplotlib) and PDF reporting outputs.

TRUST Agents — Multi-Agent Fact-Checking — *Python, GPT-4, LangGraph, RAG* *Sep 2025 – Dec 2025*

- **End-to-end pipeline:** Built a **4-agent** system (claim → evidence → verification → aggregation) producing explainable verdicts with evidence trails.
- **Structured reasoning:** Implemented claim decomposition via **NER + dependency parsing** and computed compound verdicts with **AND/OR/IMPLIES** logic.
- **Evaluation:** Benchmarked on **LIAR (12.8K)** and analyzed uncertainty behavior to drive human-in-the-loop escalation for ambiguous cases.

Education

George Mason University Fairfax, VA
M.S. Computer Science — GPA: 3.87/4.00 *Jan 2024 – Dec 2025 (Graduated)*

- **Relevant:** Algorithms, Advanced NLP, Data Mining, Systems Programming, Deep Learning, DevOps

Andhra University Visakhapatnam, India
B.S. Computer Science — GPA: 8.15/10 *Aug 2019 – May 2023*

Certifications & Publications

Certifications: AWS Cloud Practitioner (CLF-C02) — AWS AI Practitioner (AIF-C01)
Publications: arXiv:2502.10632 — arXiv:2502.04269