Sujith Reddy Bommareddy

(+91) 9133804687 bsujithreddy@gmail.com Linkdin github

Education

Blekinge Institute of Technology Jan 2024 – Jan 2026
Masters in Computer Science *Karlskrona, Sweden*

Jawaharlal Nehru Technological University Hyderabad Aug 2020 – Nov 2023
Bachelors Degree in Computer Science *Hyderabad, India*

Projects

AI Interview Simulator
(Python, LangChain, RAG, GPT4All)

- Built an AI-driven interview simulator using Python and LangChain to generate multi-round behavioral, technical, and coding questions for roles like Python Developer, Data Scientist, and ML Engineer.
- Designed an end-to-end LangChain workflow with Few-Shot and Chain-of-Thought prompts, sequential chains, structured output parsing, and a custom memory module to track performance and deliver contextual hints and targeted improvements.
- Added a RAG component with ChromaDB, HuggingFace embeddings, and a local GPT4All model to provide semantic tips, best-practice guidance, and real-time scoring (0–100), with future plans for a web UI, voice interviews, and expanded role-specific templates.

SmartSupport AI – Customer Service Chatbot
(Python, GPT4All)

- Built SmartSupport AI, an e-commerce customer service chatbot powered by Python and a local GPT4All LLM, handling 10 core support scenarios and achieving a 93/100 quality score.
- Created a rule-based evaluation system with 6 objective criteria and automated JSON reporting, boosting chatbot quality from 2.6 to 9.3 through iterative prompt engineering and testing.
- Defined a future roadmap with multi-turn conversations, a web chat UI, sentiment analysis, live chat integration, and multi-language features to move the prototype toward production readiness.

Smart Surveillance System with Natural Language Querying
(Python, YOLOv8, ByteTrack, LangChain, GPT4All, FastAPI)

- Built a smart surveillance system using YOLOv8, ByteTrack, and a LangChain RAG pipeline with ChromaDB and GPT4All to enable natural language querying of CCTV footage through a FastAPI backend.
- Processed 2,782 detections and 321 tracks with 5,553 behavior events and 3,103 indexed items, achieving 100% automated test success, around 30 FPS on GPU, and under 2-second query response time.
- Defined future improvements such as real-time streaming, web dashboard UI, cloud deployment with Docker, alerting, multi-camera support, and advanced analytics.

Microservices-Based Book Review Application
(Kubernetes, Docker, MongoDB, REST APIs)

- Architected a book review platform with 3 microservices (User, Review, Book) deployed on Kubernetes, Docker Hub, REST APIs, and persistent MongoDB storage.
- Achieved 99% uptime with horizontal pod autoscaling, health probes, and load balancing across replicas with sub-second API response times and independent scaling.
- Extensible to recommendation engines, GraphQL gateway, event-driven architecture with Kafka/RabbitMQ, distributed tracing, and multi-cloud deployment strategies.

Technical Skills

Programming: Python, C, C++, SQL
ML/DL: TensorFlow, Keras, PyTorch, scikit-learn, OpenCV, YOLOv8, ByteTrack
Cloud & DevOps: Kubernetes, Docker, FastAPI, MongoDB, SQLite, Git/GitHub
AI/ML Technologies: LangChain, GPT4All, Prompt Engineering, RAG (Retrieval-Augmented Generation),Few-Shot Learning, ChromaDB, Vector Embeddings
Domains: Computer Vision, NLP, CNNs, Object Detection/Tracking, Video Analytics, Semantic Search, AI Evaluation