Srinikhil Durisetti

USA | (+1) 689-275-5670 | srinikhildurisetti@gmail.com | linkedin | github

Professional Summary

AI Software Engineer with end-to-end experience designing, training, and deploying machine learning and generative AI systems across healthcare, enterprise automation, and intelligent agents. Adept at building explainable models, scalable data pipelines, and autonomous agent workflows using modern AI stacks. Passionate about bridging data, design, and human context to craft AI solutions that are impactful, interpretable, and production-ready. Skilled in building and deploying multi-agent systems using LangGraph, CrewAI, and LangFlow, integrating adaptive RAG and MLOps workflows for scalable, production-ready agentic systems. Currently applying agentic AI principles to automate operational workflows and enhance system resilience in enterprise environments.

Education

University of Central Florida — Orlando, FL, USA
[Master of Science in Computer Science], CGPA - 3.85/4.0 — *Jan. 2024 – May 2025*

VNR Vignana Jyothi Institute Of Engineering and Technology — Hyderabad, India
[Bachelor of Technology in Computer Science Engineering], CGPA - 8.93/10 — *Jul. 2017 – Apr 2021*

Skills

Programming Languages: Python, SQL, Go (Golang)
AI & ML: Deep Learning (CNNs, Transformers), NLP, LLMs, RAG, Reinforcement Learning, Generative AI, Prompt Engineering, Agentic AI Frameworks (LangChain, LangGraph, CrewAI, LangFlow), Adaptive/Agentic RAG, Multi-Agent Coordination
Frameworks: TensorFlow, PyTorch, Keras, Hugging Face, OpenCV
Cloud & Infrastructure: AWS (EC2, Lambda, S3, Bedrock), Docker, Kubernetes, Terraform, CI/CD for AI Agents (GitHub Actions, BentoML), Monitoring & Optimization (LangSmith, Opik)
Databases: PostgreSQL, MongoDB, Vector DBs (FAISS, Chroma)
Data Processing: Pandas, NumPy, Matplotlib, feature engineering, analytics pipelines
Tools & Platforms: Git, Jira, Postman, VS Code, Jupyter Notebook, Linux, Power BI, Tableau, n8n for AI workflow, Model Context Protocol (MCP) integration

Experience

AI Developer — Jun 2025 – Present
Impacter AI — *San Francisco, CA, USA*

- Led the design and implementation of an AI-driven personalization engine using **LangChain** based agent workflows that analyzed behavioral signals and psychographic data to tailor communication tone and content, improving enterprise outreach response rates by 45%.
- Built **multi-agent orchestration pipelines** with **LangGraph** to coordinate lead-profiling, message-generation, and follow-up agents, enabling **autonomous** execution of multi-step outreach tasks and reducing manual supervision by 60%.
- Integrated **LangSmith** for experiment tracking, evaluation, and real-time debugging of LLM outputs establishing feedback loops that improved response relevance and reduced hallucination rates.
- Developed **adaptive RAG pipelines** combining **FAISS**, **LangChain** retrievers, and domain-aware prompts to retrieve context from structured and behavioral datasets, cutting irrelevant responses by 38%.
- Implemented **reinforcement learning** based reward models using user interaction data to **fine-tune LLM agents** dynamically and personalize communication strategies.
- Deployed and monitored **agentic workloads** on **AWS Lambda** and **EC2** with automated CI/CD pipelines via **GitHub Actions** and **BentoML**, reducing compute cost by 20% while maintaining sub-second inference latency.
- Collaborated with UX and product teams to translate psychological insights into user-centric flows that enhanced trust, engagement, and retention across the outreach platform.

AI Software Engineer — Jul 2021 – Dec 2023
LTIMindtree (Johnson & Johnson) — *Hyderabad, India*

- Developed and deployed a **BioBERT, GPT hybrid chatbot** to classify and resolve patient inquiries, reducing response time by 30% and automating 60% of recurring queries.
- Designed a **ResNet-50 CNN pipeline** for pneumonia detection from **X-ray images**, trained on internal **DICOM** data and **ChestX-ray14**, achieving 85% accuracy and high AUC-ROC.

- Applied **Grad-CAM interpretability** visualizations for radiologist validation, improving clinical trust and explainability.
- Created **ML pipelines** (**XGBoost**, **TensorFlow**) to predict ER workflow delays and triage outcomes, improving throughput efficiency by 15%.
- Built robust **ETL systems** using **Spark** and **AWS** data services to handle large medical imaging datasets securely under HIPAA compliance.
- Containerized and deployed ML microservices using **Docker** and **Kubernetes**, enabling modular, version-controlled model updates.
- Established automated **CI/CD pipelines** with **GitHub Actions** and coverage gates to ensure reliable model deployments.
- Collaborated with cross-disciplinary teams, data engineers, clinicians, and architects to align AI solutions with patient-centric KPIs.

Software Engineer Apr 2020 – Jul 2021
PGK Technologies *Hyderabad, India*

- Contributed to the design and development of **C2Hire**, a recruitment and assessment platform connecting students, universities, and employers to streamline campus hiring and training workflows.
- Engineered dynamic dashboards and workflow modules using **React.js**, **TypeScript**, and **REST APIs** to integrate real-time data and improve decision visibility across stakeholders.
- Collaborated with UI/UX designers to translate wireframes from **Adobe XD** into interactive web components, enhancing usability and engagement by 25%.
- Built **data analytics dashboards** using **Python**, **Power BI**, and custom data pipelines to track candidate performance and training outcomes.
- Analyzed user engagement and behavioral trends to inform UX improvements, leading to a 30% rise in feature adoption during beta testing.
- Implemented data filtering and reporting logic on **Node.js** and **MongoDB**, enabling universities and companies to access insights on candidate readiness and hiring metrics.
- Collaborated with backend teams to design and test **RESTful APIs**, improving data delivery reliability and reducing latency by 35%.
- Worked in **agile sprints** with designers, QA, and product managers to deliver end-to-end features from concept to release.
- Mentored interns on **React.js**, **API integration**, and **data visualization** best practices, improving team efficiency and code consistency.
- Supported data processing pipelines on **AWS** to aggregate multi-stakeholder statistics, experience that built the foundation for later AI and data-engineering roles.

Projects

Notion TaskSense | *[Python, LLM (OpenAI), MCP, Notion API, Gradio, LangChain]*

- Developed an AI workflow that translated unstructured user goals into structured Notion tasks using LangChain reasoning and MCP integration.
- Integrated an automated feedback mechanism to refine intent parsing and keep tasks synchronized across connected tools.

Financial Agentic AI Application | *[Python, Phi SDK, LLaMA 3, OpenAI API, RAG, Yahoo Finance]*

- Designed a multi-agent architecture in LangGraph where agents collaborated for live data retrieval, market summarization, and risk analysis.
- Implemented adaptive RAG for context retrieval from both live and historical data, increasing analytical accuracy by 50%.

RAG Question & Answering System | *[Python, LangChain, LLaMA 8B, FAISS, Hugging Face]*

- Built a retrieval-augmented Q&A engine that used LangChain retrievers and FAISS indexing to generate fact-grounded answers from document knowledge bases.
- Applied LangSmith for tracing and evaluation, improving factual accuracy and reducing hallucinations by 35%

Certifications

Agentic AI And Generative AI Application With Cloud Bootcamp (Tutor: Krish Naik), **Oracle Cloud Infrastructure 2024 Generative AI Certified Professional**, **Oracle AI Vector Search Certified Professional**, **Machine Learning** (by Stanford University Tutor: Andrew Ng), **Convolutional Neural Networks** (Tutor: Andrew Ng), **Sequence Models** (Tutor: Andrew Ng), **Python for DS, AI & Dev** (by IBM)