Srinath Tummala

AI Engineer | Data Scientist

linkedin.com/in/srinath-tummala | github.com/SrinathTummala | srinathtummala.github.io
Email: srinathtummala7@gmail.com | Mobile: +91 75699 74666

Professional Summary

Highly accomplished **AI Engineer and Data Scientist** with **4 years of experience** specialising in production-ready **GenAI, LLMs, and RAG pipelines**. Proven expertise in performance engineering, achieving an **80% reduction in training time** and a **60% reduction in response latency**. Strong background in deploying scalable AI microservices within enterprise infrastructures.

Technical Skills

Generative AI: GPT-4, Llama, RAG Pipelines, Prompt Engineering, LangChain, Vector Databases (FAISS), LoRA/PEFT
Machine Learning: NLP, Multilingual AI, Computer Vision, PyTorch, TensorFlow, Scikit-learn, MLOps
Architecture: Python, SQL, NoSQL (MongoDB), FastAPI, PostgreSQL, EDA, Data Wrangling, ETL Pipelines
Cloud & DevOps: AWS (S3, SageMaker), Docker, Kubernetes, CI/CD (Jenkins), Grafana, Git

Experience

Data Scientist   December 2023 – Present
Apporchid   *Hyderabad, India*

- Reduced model training time by **80%** and response latency by **60%** by implementing **Stochastic Variational Inference (SVI)** for high-dimensional models.
- Developed enterprise-grade **RAG pipelines** using FAISS and LangChain; optimised retrieval via **semantic chunking** and hybrid search.
- Architected and containerised **FastAPI microservices** deployed via Jenkins CI/CD on **AWS infrastructure**.

Junior Data Scientist   July 2022 – December 2023
Onpassive   *Hyderabad, India*

- Engineered multilingual NLP systems using **Transformer-based architectures** to support global product scaling.
- Developed deep learning **eKYC modules** using CNNs for face detection and identity verification.
- Collaborated with technical leadership to integrate modular AI microservices into production ecosystems.

Data Science Intern   January 2022 – April 2022
Onpassive   *Hyderabad, India*

- Developed an NLP-based email classification system categorising communications with high precision.

Key AI Projects

RAG-Powered Document Search Assistant
GPT-4, LangChain, FAISS, FastAPI

- Built a personal RAG QA assistant using **GPT-4** and **text-embedding-ada-002** for high-precision semantic search.
- Optimised retrieval relevance by fine-tuning chunking granularity and prompt heuristics for enterprise performance.

Master's Thesis: Malicious URL Detection
Python, AdaBoost, Random Forest

- Engineered a Voting Ensemble classifier using 17 lexical features achieving **99.8% accuracy** on 450K+ entries.
- Optimised model performance by architecting a hybrid system combining **AdaBoost** and **Random Forest** architectures.

Official College Communication App
Java, Android Studio, XML

- Engineered and deployed an official Android application for BVC to streamline academic updates and communications.
- Maintained production impact with **1,000+ active downloads** and a 4.4/5 rating on the Play Store.

Education

Master of Science in Data Science   October 2020 – June 2022
GITAM University   *Hyderabad, India*

- **CGPA: 8.89** | Core focus on Machine Learning Algorithms, SQL and Data Structures.

Bachelor of Science in Computer Science   June 2017 – September 2020
Bhavan's Vivekananda College   *Hyderabad, India*

- **CGPA: 8.81** | Specialised in Mathematics, Statistics and Computer Science.

Additional Information

Certifications: Advanced Certification in ML and DL – IIIT Bangalore.
Languages: English (Professional), Telugu (Native), Hindi, German (Beginner).
Notice Period: Immediate / Negotiable.