

LINGESH S

Email: lingeshsrinivasan1@gmail.com

Phone No: [7299307308](tel:7299307308)

Portfolio: <https://lingesh-dev.netlify.app>

GitHub: github.com/Lingesh-S

ABOUT ME

Data Scientist with 1.8 years of experience building production-grade ML and GenAI systems at a product-based company. Strong background in XBRL auto-tagging using LSTM & NER, taxonomy-specific ML pipelines, and RAG-based systems for real-time applications. Experienced in deploying scalable ML services, working across multiple regulatory domains, model development, data preprocessing, data analysis, ML pipelines and building LLM-powered features for enterprise products.

WORK EXPERIENCE

Data Scientist

DataTracks | July 2025 – Present

- Built and maintained LSTM + NER-based XBRL auto-tagging models for multiple region-specific taxonomies across core compliance products.
- Designed independent ML pipelines per taxonomy, and trained & deployed two new taxonomy pipelines end-to-end (data prep → training → production integration).
- Developed a RAG-based next-word prediction system for a Word-like editor in a flagship product using Haystack, FastAPI and RabbitMQ.
- Implemented production monitoring and observability for ML services using Grafana and Prometheus.
- Deployed and managed ML services in production (on-premise infrastructure) using Docker, FastAPI, and REST APIs, ensuring scalable model serving, low-latency inference, and reliable system performance.

LLM Intern

Dstyle.AI (IIT Madras Incubated) | Jan 2025 – June 2025

- Appointed as a virtual intern to work on Large Language Models and Generative AI.
- Collaborating with the tech team under the CTO's mentorship on NLP-centric tasks.
- Contributing to prompt engineering, fine-tuning, and evaluation of LLM-based systems.
- Engaging in research, experimentation, and practical application of transformer models.
- Ensuring data privacy and following responsible AI practices in project workflows.
- Developing RAG-based architectures using Flan-T5 and LangChain for modular, context-aware Q&A applications.

ML Intern

Cognibot Labs | June 2024 – Dec 2024

- Built end-to-end classical ML pipelines for text analytics using Scikit-learn, including preprocessing, feature engineering, model training, and evaluation.
- Developed an LDA-based topic modeling system using Gensim & TF-IDF to extract latent topics from large text datasets and improve document clustering.
- Applied NLP techniques with NLTK, spaCy (tokenization, lemmatization, stopword removal) and trained models like Logistic Regression, Naive Bayes, SVM.
- Performed model tuning & evaluation using GridSearchCV, Cross-Validation, and metrics like Accuracy, Precision, Recall, F1-score, with data handling via Pandas, NumPy.

SKILLS SUMMARY

- **Languages:** Python, SQL
 - **Databases:** MySQL, Vector Databases (ChromaDB / FAISS)
 - **ML & DL:** Scikit-learn, TensorFlow, PyTorch, LSTM, NER, Transformers, Fine-tuning(LoRA, PEFT)
 - **GenAI / RAG:** LangChain, Haystack, Retrieval-Augmented Generation
 - **Cloud Platforms:** AWS (S3, SageMaker, Lambda, EC2, Glue, Athena)
 - **Big Data & Streaming:** Apache Spark (PySpark), Apache Kafka
 - **Backend, Infra & Monitoring:** FastAPI, RabbitMQ, Grafana, Prometheus
 - **Data & Tools:** Pandas, NumPy, Git, Streamlit, Flask , Fast API
-

PROJECTS

1. AWS Bedrock RAG-Based Knowledge Assistant

- Designed a Retrieval-Augmented Generation (RAG) system using AWS Bedrock (Titan LLM) and LangChain, enabling context-aware question answering over custom knowledge sources.
- Built a data pipeline to ingest, chunk, and embed documents using Amazon Titan Embeddings, storing vectors in ChromaDB for efficient semantic search and retrieval.
- Integrated retrieval + generation workflow to provide accurate, grounded responses, significantly reducing hallucinations in LLM-based applications.
- Developed a real-time interface using Streamlit and currently extending the system by integrating AWS S3 for scalable data storage and EC2-based deployment to enhance production readiness of the Generative AI system.
- GitHub Link: (<http://bit.ly/4ql10rY>)

2. Codebase RAG Assistant

- Designed a Retrieval-Augmented Generation (RAG) system that answers questions about code using Mistral-7B built on PyTorch, combining transformer-based generation with semantic search from a vector store.
- Built a pipeline to load and chunk .py, .md, and .json files, then embed them with SentenceTransformers and store in a Chroma/FAISS vector database for high-precision code retrieval.
- Integrated LangChain, Mistral-7B, ChromaDB, and Streamlit to deliver real-time code understanding, explanation, and developer support via a modular assistant.
- GitHub Link: (bit.ly/4kxpzFu)

3. LDA-Powered Resume Screening Dashboard

Presented at ICCCAI-2024 (Taylor's University, Malaysia)

- Developed an interactive Streamlit dashboard using LDA (Latent Dirichlet Allocation) implemented with Scikit-learn for extracting dominant topics and keywords from resumes.
- Designed a custom NLP-based job matching algorithm with token indexing, achieving >80% accuracy in simulated resume screenings.
- Integrated OpenAI API to auto-generate personalized interview questions, enhancing recruiter productivity through AI-powered automation.
- GitHub Link: (bit.ly/438nhXw)

EDUCATION

B.E. Computer Science Engineering (Fulltime)

2020 - 2024

Sathyabama University, Chennai

CGPA: 8.41

CERTIFICATES

- Machine Learning Internship – Cognibot Labs
 - Cloud Data Management Associate - Oracle
 - Data Analytics Virtual Experience - Accenture
-

ACHIEVEMENTS

- Published and presented a research paper titled "LDA-Powered Resume Screening Dashboard with Token Indexing & Streamlit" at ICCCAI-2024, in association with Taylor's University, Malaysia.
- Developed and deployed a personalized learning chatbot for students using LangChain and Flan-T5, enabling real-time, context-aware Q&A from academic notes.
- Engineered a RAG-powered Codebase Assistant using LLMs, LangChain, and vector databases to help programmers understand, query, and explore codebases through natural language.