Koona Sai Vaishnavi

Bangalore, Karnataka, India

Phone: +91 7842841510 | Email: vaishnavikoona1@gmail.com | LinkedIn | GitHub

Education

National Institute of Technology Rourkela   Nov 2020 – May 2024
Bachelor of Technology in Ceramic Engineering   *CGPA: 7.78*

Certifications / Achievements

AWS Certified Data Engineer – Associate   Oct 2024
Data modeling, ETL design, Data lakes and warehouses, Schema evolution, AWS analytics (S3, Glue, Redshift, Kinesis).

AWS Certified Machine Learning Engineer – Associate   Jul 2025
ML lifecycle on AWS: Data preparation, Training and Evaluation, Model registries, Pipeline automation, MLOps.

Experience

Software Engineer – AI/ML   Jun 2024 – Present
FinTech Domain   *Bangalore, India*

- **Financial Risk Reporting Automation:** Designed and deployed an LLM-driven credit portfolio analytics platform automating nearly 70% of enterprise credit and financial risk reporting workflows across corporate lending teams.
- Implemented dual-agent architecture (SQL Query Agent + Commentary Agent), multi-period financial analytics pipelines, and PostgreSQL-backed APIs, integrating LLM-synthesized industry/news context with portfolio metrics to significantly reduce analyst turnaround time and manual reporting effort.
- **GCS Call Interpreter – AI Call Intelligence:** Built an end-to-end speech intelligence system using Faster-Whisper, Pyannote speaker diarization, and LLM-based summarization for large-scale enterprise customer support analytics.
- Improved call review efficiency by 60% and insight retrieval accuracy by 35% through semantic search, structured JSON summaries, sentiment analysis, and intent-level call understanding.
- **RegTech – Intelligent Legal Document Processing:** Developed scalable legal document intelligence pipelines for regulatory compliance analysis using PyMuPDF-based parsing, embedding-driven semantic comparison, and LLM-assisted clause interpretation and version drift detection.
- Reduced compliance false positives by 30% and enabled clause-level semantic review through embedding similarity, change attribution, and LLM-generated explanations surfaced via internal dashboards.
- **AI-driven Fraud Detection & Recommendation Systems:** Deployed real-time fraud detection and recommendation services using SageMaker-managed models, event-driven Lambda inference, and API Gateway across high-throughput banking environments.
- Improved recommendation precision by 40% while reducing inference latency by 45% through optimized feature pipelines, model caching, and asynchronous serving patterns.

Machine Learning Intern   Jun 2023 – Aug 2023
Tata Steel Ltd.   *Jamshedpur, India*

- **Predictive Maintenance – Torpedo Ladle Car Systems:** Developed a machine learning prototype to predict optimal repair intervals for torpedo ladle car column plans used in steel manufacturing operations.
- Integrated operational sensor data, thermal scan reports; performed exploratory analysis to identify key failure indicators.
- Implemented Random Forest regressor achieving **7.85%** mean squared error, enabling accurate maintenance planning.

Technical Skills

- **Languages:** Python, SQL, C++, C
- **Machine Learning & AI:** ML & DL Pipelines, NLP Systems, Semantic Similarity, Representation Learning
- **LLMs & Generative AI:** Transformer Architectures, Embeddings, RAG, Vector Search, Agentic Workflows
- **ML Systems Design:** End-to-End ML Pipelines, Data Validation, Skew Handling, Scalable Inference Design
- **MLOps:** Model Training & Evaluation, Feature Pipelines, Experiment Tracking, Model Versioning, CI/CD
- **Performance & Reliability:** Inference Optimization, Latency Reduction, Throughput Scaling, Drift Detection
- **Cloud ML Infrastructure:** AWS SageMaker, Lambda, API Gateway, S3, DynamoDB, ML Deployments

Relevant Coursework

Machine Learning ● Deep Learning ● Natural Language Processing ● Statistical Methods ● Computer Vision