

Giripranay Kona

AI / ML Engineer

☎ (469) 583-9601

✉ giripranay.kona@gmail.com  linkedin.com/in/giripranaykona  github.com/giripranay

Professional Summary

- AI / Applied ML / GenAI Engineer with 6.5+ years of experience designing and delivering production-grade AI systems across enterprise environments.
- Proven technical leader driving innovation, rapid experimentation, and scalable AI solutions from ambiguous business problems.
- Expertise in LLMs, NLP, Deep Learning, prompt engineering, and autonomous agent systems using LangChain, LangGraph, Google SDK, CrewAI and MCP.
- Specialized in multi-agent architectures, tool chaining, A2A communication, and enterprise workflow automation.
- Engineered AI-driven knowledge graph and semantic data intelligence solutions by integrating structured and unstructured enterprise data sources, breaking down organizational data silos and enabling unified, context-aware information discovery across business systems.
- Strong MLOps and cloud deployment experience across AWS and GCP using Docker, Kubernetes, and CI/CD pipelines.
- End-to-end ownership mindset with focus on scalability, governance, observability, and performance.

Technical Skills

Languages: Python, JavaScript, TypeScript, Go, SQL, Java, C++

AI/ML: OpenAI API, Hugging Face, LangChain, LangGraph, LlamaIndex, MCP, RAG, Prompt Engineering

Deep Learning: PyTorch, TensorFlow, Keras, scikit learn

Cloud: AWS (Lambda, S3, EC2, DynamoDB), GCP (Vertex AI, BigQuery, Cloud Run)

MLOps: Docker, Kubernetes, MLflow, CI/CD, GitHub Actions, Jenkins

Databases: MySQL, PostgreSQL, MongoDB, Redis, Vector DBs (FAISS, Pinecone)

Backend: Node.js, Django, Spring Boot, REST APIs, GraphQL

Professional Experience

SWBC Finance

Applied AI Engineer

San Antonio, TX

Aug 2024 – Present

- Resolved recurring ETL production failures caused by inconsistent and malformed incoming files by developing an AI-driven file validation agent that automatically detected file types, validated schema and format integrity, and intelligently routed files to downstream pipelines or quarantine workflows, resulting in 0% production failures and significantly improving data pipeline reliability and operational efficiency.

- Addressed a highly manual OCR validation process where QA teams spent months verifying extracted data across thousands of documents by designing and implementing an AI-powered validation agent using Google OCR, Azure OCR, and LLM-driven token mismatch analysis, ultimately reducing processing time from 2 months to overnight while improving scalability, validation accuracy, and QA efficiency.
- Identified friction in client-facing AP portal where users navigated multi-step workflows for reports, claims, and transactions, leading to poor user experience and delayed task completion.
- Led design and development of **QuickLinks**, an AI-powered feature leveraging multi-agent workflows (Strands framework) to interpret user intent and autonomously execute end-to-end actions via a simplified interface.
- Engineered backend agent orchestration to gather contextual data, trigger downstream services, and complete multi-step financial workflows (report generation, claims, payments) with minimal user interaction.
- Delivered a seamless embedded AI experience that reduced navigation overhead and significantly improved usability, driving **20% increase in revenue** and **30% growth in sales conversions**.
- Architected enterprise MLOps platform using SageMaker, MLflow, and Azure DevOps.
- Implemented Explainable AI, governance, and compliance for financial regulatory requirements.

Erik Johnson School of Engineering

AI Engineer

Dallas, TX

May 2023 – May 2024

- Automated large-scale SDS document acquisition by developing a Python and Puppeteer-based web scraping framework, enabling continuous extraction, validation, and ingestion of chemical safety documents from public websites into the AI knowledge base pipeline.
- AI-Powered Document Processing: Built a document extraction pipeline using NLP and LLM, reducing manual data processing time by 50%.
- Designed and optimized a Retrieval-Augmented Generation (RAG) architecture for SDS document analysis by integrating web-scraped datasets, embedding-based indexing, semantic chunking, and vector similarity search, enhancing knowledge discovery and reducing information lookup time for safety and compliance workflows.

The Live Green Company

Senior Software Developer

Bengaluru, India

Oct 2020 – Aug 2022

- Led end-to-end development of the *Know Your Food* mobile platform, owning the complete SDLC from requirement analysis, wireframing, system architecture, backend integration, testing, and production deployment, delivering a cross-platform solution for food science research workflows.
- Architected and developed a scalable cross-platform mobile application using Flutter and Dart for iOS and Android, integrating Firebase Authentication, Firestore, and a Django REST API backend deployed on Google Cloud, successfully launching the application on the Google Play Store with real-time data synchronization capabilities.
- Engineered a centralized food science data management portal that replaced fragmented Excel-based workflows with a concurrent, cloud-backed data platform, improving data consistency, enabling multi-user collaboration, and accelerating research data processing efficiency by 50%.
- Developed an AI-powered text summarization pipeline leveraging LLMs, BERT, Transformer

architectures, and ChatGPT-based NLP workflows, integrating SpaCy for tokenization, preprocessing, and entity parsing, improving summarization accuracy and research document analysis efficiency by 30%.

- Designed and trained a Named Entity Recognition (NER) model using NLP and deep learning techniques to automatically extract molecule names and scientific entities from research papers and SDS documents, reducing manual annotation effort by 80% and significantly improving research productivity for food scientists.

Dvara Solutions

Software Developer

Bengaluru, India

May 2019 – Sep 2020

- Engineered and deployed scalable RESTful APIs and microservices using Spring Boot within a containerized AWS ECS environment, optimizing backend performance, implementing secure IAM-based access controls, and improving application scalability and reliability for high-volume enterprise workloads.
- Optimized end-to-end system performance by restructuring complex SQL queries in MySQL and PostgreSQL, implementing advanced indexing strategies, and resolving AngularJS frontend bottlenecks using browser developer tools, significantly reducing data retrieval latency and improving UI responsiveness and user experience.
- Collaborated with cross-functional engineering and DevOps teams to streamline CI/CD workflows using Jenkins scripted pipelines, automating build, testing, and deployment processes for faster feature releases, seamless AWS ECS deployments, and improved development-to-production transition efficiency.

Education

M.S. in Computer Science (Data Science)

B.S. in Electronics and Communication Engineering

Projects

ResGen – AI-Powered Resume Generator

- Built LLM-powered resume generator using OpenAI GPT, Streamlit, and FastAPI.
- Generates ATS-optimized resumes with real-time customization.
- github.com/giripranay/ResGen