

Dhrumil Ankola

(951) 213-8774 | ankoladhrumil@gmail.com | [linkedin.com/in/dhrumil-ankola](https://www.linkedin.com/in/dhrumil-ankola) | github.com/dhrumilankola | dhrumilankola.netlify.app

Summary

AI/ML Engineer with an M.S. in Computer Science, delivering production-scale systems that boost throughput (15%), cut costs (\$50K+/yr), and scale to 10M+ node knowledge graphs. Skilled in multi-agent RAG workflows on AWS/GCP and large-scale pipelines with Spark. Recent projects include deploying AI agents with Bedrock, MCP tools, and Google's A2A for explainable and interoperable AI

Experience

Shifa Precision Inc <i>Software Engineer - AI/ML</i>	06/2025 - Present <i>Boston, MA (Remote)</i>
- Built a predictive engine on a weighted knowledge graph, delivering fully traceable clinical insights for healthcare providers - Unified terabytes of biomedical data into a knowledge graph of 10M+ nodes and 250M+ relationships, enabling large-scale AI-driven analysis - Designed and deployed a GNN-based risk model that generated patient embeddings, improving classification accuracy and powering explainable AI outputs	
Medlaunch Concepts <i>AI/ML Engineer - Internship (Part-time)</i>	08/2025 - Present <i>Clearwater, FL (Remote)</i>
- Deployed Clara AI (AWS Bedrock, RAG) to automate compliance and policy inquiries, reducing staff research time by 40% - Optimized retrieval pipelines with OpenSearch hybrid search and intelligent chunking, improving precision by 25% - Built a multi-agent Bedrock workflow where a supervisor orchestrated tools via OpenAPI and AWS Lambda; also prototyped agent-to-agent handoffs (Google A2A) for escalation and verification across workflows	
Center for Environmental Research and Technology <i>Assistant Software Developer</i>	07/2024 - 04/2025 <i>Riverside, CA</i>
- Built Laravel (PHP) backend modules and integrated third-party APIs for California Air Resources Board's vehicle inspection system, supporting 740K+ trucks statewide - Refactored database queries and implemented caching, reducing API response times by 40% during peak scheduling - Strengthened CI/CD pipeline by adding comprehensive PHPUnit tests in Jenkins, cutting deployment rollbacks by 30%	
Edicon Paper <i>Systems Engineer</i>	01/2022 - 07/2023 <i>Gujarat, India</i>
- Built a real-time data pipeline (Python, Kafka, Spark) to predict KPIs like Sheet Break Probability, boosting throughput by 15% - Deployed a QA monitoring system (Python, InfluxDB) with proactive alerts, cutting material waste by 5% and saving \$50K+ annually - Developed a B2B client portal (Node.js, PostgreSQL) with atomic transactions, reducing client inquiries by 30%	

Projects

LLM-Powered Anomaly Detection for Microservices Reliability
- Designed a serverless, event-driven pipeline on GCP to process real-time errors from microservices deployed on GKE with Docker/ Kubernetes - Implemented a zero-shot Vertex AI Gemini agent for log classification, reducing incident resolution time by 40% using OpenTelemetry insights
Study Buddy - UC Berkeley CalHacks
- Built a full-stack RAG app (LangChain, ChromaDB) with persistent chat sessions and a hybrid retriever for contextual Q&A - Integrated ElevenLabs AI for a voice-driven interface, enabling dynamic knowledge-base attachment to a conversational agent
PyAsterix - Database Driver for AsterixDB
- Authored the open-source Python/JavaScript driver for Apache AsterixDB with a PEP 249-compliant API and Pandas-style interface, reducing query latency by 15% - Published a lightweight MCP tool server for AsterixDB queries, enabling AI agents to call SQL++ endpoints directly with authentication and built-in observability (OpenTelemetry, Prometheus, structured logging)

Education

University of California, Riverside <i>Master of Science in Computer Science (GPA: 3.7)</i>	03/2025
BVM Engineering College, Anand, India <i>B.Tech in Information Technology (GPA: 3.8)</i>	07/2023

Skills

Languages: Python (expert), Go, JavaScript, SQL (PostgreSQL, MySQL), C++, PHP, NoSQL (MongoDB, Neo4j, InfluxDB)
AI/ Machine Learning: PyTorch, TensorFlow, LangChain, RAG, LLMs (Gemma, Llama3, BERT), GNNs, Fine-Tuning (PEFT, LoRA)
Agents & Protocols: Agents & Protocols: Google ADK (A2A), AWS Bedrock Agents, MCP tools/servers, LangGraph
Data Engineering & Databases: Spark, Airflow, ChromaDB, AsterixDB, PostgreSQL, MySQL, MongoDB, Neo4j, InfluxDB
Cloud & DevOps: AWS (S3, Glue, Lambda), GCP (GKE, Vertex AI), Docker, Kubernetes, Jenkins, OpenTelemetry, Git
Frameworks & APIs: React, Node.js, Django, Express, Laravel, REST, GraphQL, Microservices, Distributed Systems
Testing & Methodologies: pytest, PHPUnit, Selenium, Agile/Scrum, Jira