Krishna vamsi Dhulipalla

(+1) 540-558-3528 | dhulipallakrishnavamsi@gmail.com | GitHub: Krishna-dhulipalla | LinkedIn: krishnavamsidhulipalla
Portfolio | Personal Chatbot

Summary

ML Engineer with 3+ years building scalable AI systems and hybrid data pipelines. Skilled in Python, Java, and SQL with experience in multi-agent workflows (LangChain, LangGraph, AutoGen, MCP) and deep learning (PyTorch, TensorFlow). Delivered production-ready LLMs, semantic search, and containerized microservices (Docker, Kubernetes, Jenkins) across cloud platforms (AWS, GCP, Snowflake). Proven ability to reduce latency, optimize pipelines, and collaborate with cross-functional teams.

Skills

Programming: Python, R, SQL, JavaScript, TypeScript, Node.js, MongoDB, PostgreSQL, MySQL, FastAPI, Go

ML & AI Frameworks: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, OpenCV

Multi-Agent & LLM: LangChain, LangGraph, AutoGen, CrewAI, MCP, FastMCP, LLM Fine-Tuning, Retrieval-Augmented Generation (RAG)

Data Engineering&Cloud: Apache Spark, Kafka, Airflow, dbt, ETL Pipelines, AWS (S3, Redshift, ECS, SageMaker, CloudWatch), GCP (GCS, BigQuery, Dataflow), Snowflake, Real-time Pipelines

MLOps & Infrastructure: Docker, Kubernetes, Jenkins, MLflow, CI/CD, Weights & Biases, Git/GitHub, Linux

Other: Semantic Search, LangSmith, PowerBI, Plotly, Synthetic Data Generation, Cross-Domain Adaptation, Hyperparameter Optimization, A/B Testing, Model Evaluation, Pandas, EDA

Experience

AI Engineer
Jul. 2025 - Present

Cloud Systems LLC — *Remote*

- Delivered internal knowledge automation agents (LLM-powered retrieval + workflow bots) that reduced manual query handling by 40% and supported early adoption across product teams
- Built and optimized data analytics pipelines for the company's reporting dashboards, improving SQL/ETL performance by 25% and cutting daily pipeline costs
- Deployed containerized agent services on Kubernetes with Docker + Helm, ensuring scalable uptime for internal AI products
- Collaborated in Agile sprints with cross-functional teams, ensuring timely delivery of ML features

ML Research Engineer
Aug. 2024 - Jul. 2025

Virginia Tech — *Blacksburg, VA*

- Developed genomic sequence classification, fine-tuning LLMs with LoRA and soft prompting, as part of a bioinformatics platform and accelerating research iterations by 30%
- Automated sequence preprocessing pipelines (1M+ samples) with Airflow, reducing data prep time by 40% and enabling faster experimentation on the lab's genomics platform
- Built a LangChain-based semantic search tool for genomics literature, transforming manual literature review into an on-demand search product for lab researchers
- Deployed fine-tuned LLMs via Docker + MLflow, enabling reproducible experimentation and cutting deployment friction for researchers by 25%

Data Engineer
Jul. 2021 - Dec. 2022

UJR Technologies Pvt Ltd — *Hyderabad, India*

- Migrated client ETL pipelines into a real-time financial reporting platform using Kafka and Spark, lowering latency by 30% and improving client decision-making
- Deployed Dockerized financial microservices to AWS EKS (Kubernetes), strengthening the company's SaaS product reliability and cutting release rollback time by 40%
- Optimized Snowflake-based analytics platform by redesigning schemas and materialized views, boosting query speed by 40% for customer dashboards
- Built proactive monitoring dashboards in AWS CloudWatch, maintaining 99.9% uptime for client-facing reporting systems
- Collaborated with frontend and BI teams to deliver interactive dashboards (Snowflake + Power BI) for clients, improving visibility into KPIs and accelerating reporting cycles

Projects

Multi-Agent AI System: Community & Hazard Intelligence Map

- Built an agentic system combining official feeds (USGS, NWS, EONET, FIRMS) with community reports on a live map, enabling real-time disaster and safety intelligence
- Designed LangGraph workflows (reporting, classification, verification) and a React + FastAPI stack with GeoJSON overlays, reducing duplicate entries and improving engagement by 35%
- Solved concurrency and overlay issues with request guards and optimistic UI, delivering a smooth, reliable mapping experience
- Integrated MCP protocol for tool orchestration across agents, aligning with emerging industry standards for interoperability

Proxy TuNER: Advancing Cross-Domain Named Entity Recognition through Proxy Tuning

- Implemented a proxy-tuning approach for BERT using logit ensembling with domain-specific expert models, improving F1-score by 8% across diverse datasets
- Reduced computational overhead by 70% and accelerated inference by 30% through distributed training optimizations
- Applied gradient reversal for domain-invariant feature learning, boosting cross-domain accuracy by up to 15%

IntelliMeet: AI-Enabled Decentralized Video Conferencing App

- Built a secure video conferencing platform with federated learning and encryption to protect user data
- Integrated on-device RetinaFace ML models for real-time attention tracking, reducing call dropouts and increasing engagement by 25%
- Deployed a Transformer CNN-RNN speech-to-text pipeline with summarization, enabling automated meeting notes and enhancing productivity
- Ensured GDPR-aligned data practices with federated learning and encrypted pipeline
- Maintained 99.9% uptime with CI/CD pipelines and modular microservices architecture

Education

Virginia Tech Jan. 2023 - Dec. 2024
M.S. in Computer Science CGPA - 3.95/4 *Blacksburg, Virginia*

Vel Tech University Chennai Jul. 2018 - May. 2022
Bachelor's in computer science CGPA - 8.24/10 *Chennai, India*

Publications

Predicting Circadian Transcription in mRNAs and lncRNAs — IEEE BIBM 2024
Applied ML models to genomic data; improved transcription prediction accuracy, enabling deeper insights into biological rhythms. [DOI: 10.1109/BIBM62325.2024.10822684]

DNA Foundation Models for Cross-Species TF Binding Prediction — NeurIPS ML in CompBio 2025
Leveraged DNABERT-style architectures for plant genomics; advanced cross-species binding prediction with improved generalization. [bioRxiv: 10.1101/2025.07.14.664780v1]

Certifications

- Building RAG Agents with LLMs - NVIDIA (2025)
- Unleashing the Power of AI Agents - IBM (2025)
- Deploying RAG Pipelines for Production at Scale -NVIDIA (2025)
- Delivering Data- Driven Decisions with AWS – AWS (2024)
- End-to-End Data Engineering with Snowflake -Snowflake (2024)
- Google Cloud Data Engineering Foundations -GCP (2024)
- AWS Cloud Foundations - AICTE-EduSkills (2022)
- Machine Learning Specialization - Coursera (2021)