

CHARITH PURUSHOTHAM

charithcherry1100@gmail.com | +1 7206411756 | Boulder, CO, USA | [Github](#) | [Portfolio](#) | [LinkedIn](#)

EDUCATION

University of Colorado Boulder, USA, Master of Science, Computer Science | GPA: 3.93/4.00 August 2024 - May 2026
(Coursework: Deep Learning, Natural Language Processing, Computer Vision, Systems for ML, Distributed systems, Big Data Architecture)
Dayananda Sagar College of Engineering, India, Bachelor of Engineering, Computer Science | GPA: 9.68/10.00 August 2018 - June 2022

TECHNICAL SKILLS

Languages: Python, Java, SQL | **ML/DL:** PyTorch, TensorFlow, scikit-learn, XGBoost, Hugging Face, OpenCV, TensorRT | **LLM/RAG:** Embeddings, LlamaIndex | **Search/IR:** Azure Cognitive Search (hybrid + vector), re-ranking | **Data:** Databricks (Spark), Kafka, PostgreSQL, MySQL, MongoDB | **MLOps:** MLflow, Docker, CI/CD | **Web:** FastAPI, Flask, React.js, Node.js, REST | **Cloud/Tools:** Azure, AWS, GCP, Git/GitHub, Postman, Jira, Azure Boards | **OS/Viz:** Linux, Windows, Tableau, Matplotlib, Seaborn

PROFESSIONAL EXPERIENCE

Owl AI, Boulder, Machine Learning Engineer Intern

September 2025 - Present

- **Designed a distributed GPU data-curation pipeline for large-scale video analysis**, orchestrating multimodal models (Qwen-3VL, NVIDIA Cosmos, SAM) to automatically detect events, segment scenes, and extract structured signals from raw broadcast footage for downstream model training.
- Built and productionized a real-time snowboard video ML pipeline that ingests **live camera streams**, **extracts keypoints/pose**, **applies temporal modeling**, and **outputs trick-level metrics**, using a MediaMTX (SRT) streaming layer to support reliable broadcast feeds for downstream inference.
- Trained **keypoint detection + temporal models** on large-scale X Games video data; implemented dataset curation/automation and used **MLflow to track experiments, metrics, and artifacts** for reproducible model iteration and comparisons.
- **Engineered GPU inference pipelines** with TensorRT optimizations, quantization, and caching/prefetching, increasing throughput from **45 FPS to 58 FPS** for real-time analysis.
- Developed **Bayesian/MCMC models** to predict **event scores and podium probability** from trick metrics and athlete form; generated commentator-facing insights by conditioning on historical distributions and converting outputs into concise natural-language talking points.
- Leveraged **LLM-based agents** to **automate labeling, data QA, and dataset bookkeeping**, **reducing manual overhead** and **improving iteration speed**.

Tech-stack: PyTorch, CUDA, TensorRT, OpenCV, MLflow, MediaMTX/SRT, Python, GenAI (Qwen/Roo/Claude/Gemini workflows)

Idaho National Laboratory, Idaho Falls, Machine Learning Engineer Intern

June 2025 – August 2025

- Contributed to an **LDRD-funded project** developing and **fine-tuning deep learning models** (SAM/SAM2, YOLOv9, ResUNet, DenseNet, ResDense-UNet) for **semantic segmentation of pores and cracks in post-irradiated nuclear materials** using FIB, SEM, TEM, and X-ray CT data. Utilized **HPC infrastructure** for large-scale training and image characterization, enabling predictive modeling of radiation-induced material degradation.
- Analyzed how **data quality (resolution, contrast, noise, artifacts) impacts model accuracy**; explored optimization via **saliency maps** and attention mechanisms (**CBAM vs bottleneck self-attention**) across microscopy modalities.
- Co-authored: “**Bridging Multimodal Microscopy for Advanced Characterization on Nuclear Fuel Using Machine Learning**”, *Frontiers in Mechanical Engineering – Digital Manufacturing*, Vol. 11, 2025- [link](#)

Tech-stack: PyTorch, SAM/SAM2, YOLO, Scikit-learn, XGBoost, OpenCV, HPC (Alpine cluster), Matplotlib/Seaborn

Phamax, Bengaluru, Software Development Engineer

September 2022 - July 2024

- Co-architected **Ariya**, a **Microsoft-certified Teams SaaS assistant** that enables clients to query large internal document corpora with grounded, source-backed responses.
- Built Ariya’s document ingestion + embedding pipeline (**chunking, metadata normalization, embedding generation (OpenAI embeddings), indexing/storage**) and implemented **hybrid retrieval + re-ranking using Azure Cognitive Search** to improve relevance and **reduce hallucinations by grounding responses in retrieved documents**.
- Built and operated the backend service stack for **GPT-powered summarization/insights and feedback/logging services** to capture user signals; **leveraged feedback logs to iteratively improve answer quality and relevance** and **dockerized + productionized services** on Azure with monitoring, improving **resource efficiency 20–25%** and **cutting deployment time 50% (15 → 7 min)**.
- Led a data engineering project for Santen using **Databricks and Azure**, **halving processing time from 30 to 15 minutes** and improving data accuracy by implementing checkpoints for data evaluation. **Implemented a medallion architecture**, enhancing Tableau performance by offloading heavy processing to Databricks.

Tech-stack: Python, React, Azure Cognitive Search, OpenAI/Azure, Docker, Databricks (Spark), Tableau, AWS (Bedrock/SageMaker)

PROJECTS

- **Format Matters: An Empirical Study of Data Storage Format Impact on ML Training Pipelines (Systems for ML)** : Benchmarked data formats across two CPU-only workloads (CIFAR-10 image training; 1M-row tabular ML) and quantified impact on I/O throughput, load time, and storage efficiency; found performance is scale-dependent with large-scale loading/compression gains for optimized formats while training accuracy remained format-neutral, isolating effects to I/O- [link](#)
- **Reinforcement Learning for Abstractive Summarization (2025)**: Developed a **LSTM-based sequence-to-sequence model** with LSTMs for headline generation on the **SamSum** dataset; applied greedy decoding and fine-tuned using **Self-Critical Sequence Training (SCST)** to optimize BLEU scores, followed by beam search decoding with reinforcement learning to compare performance- [link](#)
- **Multimodal Zero-Shot Learning for Unseen Concepts (2025)**: Built a zero-shot image classifier using **CLIP image/text embeddings** and a **knowledge-graph R-GCN** to model semantic and contrast relations; added **edge-attention** and a **prototype refinement loss** before an MLP head, improving over the CLIP baseline by **+23% Top-1 accuracy (59% vs. 36%)** and **+13% F1 on AWA2**- [link](#)

ACHIEVEMENTS

- “**AWS Jam Hackathon**” – **University of Colorado Boulder**: Placed **7th (20 teams)**, solving **9/13 security challenges** using AWS services like Lambda, CloudTrail, DynamoDB, S3, VPC, IAM, and SageMaker, gaining hands-on experience in cloud security and threat mitigation.

EXTRACURRICULAR

- **Teaching Assistant, Distributed Systems, CU Boulder (Jan 2026 – Present)**: Support Prof. Mark Zhao in running the course by holding office hours, assisting with assignment/project design, grading, and helping students debug distributed systems concepts and implementations.