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, C | Backend: Node.js, Express.js, FastAPI, Flask, REST APIs | Frontend: React.js, Redux | Databases & Caching: PostgreSQL, MySQL, MongoDB, Redis | Data/Streaming: Kafka, Databricks (Spark) | Cloud/DevOps: Azure, AWS, GCP, Docker, CI/CD | Search & Retrieval: Azure Cognitive Search (hybrid + vector), embeddings, re-ranking, LlamaIndex | Applied ML: PyTorch, TensorFlow, OpenCV, TensorRT, MLflow | Tools/OS: Git/GitHub, Postman, Azure Boards, Linux, Windows

PROFESSIONAL EXPERIENCE

Owl AI, Boulder, Machine Learning Engineer Intern **September 2025 - Present**

- Built and productionized a real-time snowboard analytics system that ingests live camera streams, runs pose/keypoint + temporal inference, and emits trick-level metrics, using a MediaMTX (SRT) streaming layer to deliver reliable broadcast feeds to downstream services.
- Engineered a GPU inference pipeline with TensorRT, quantization, and caching/prefetching, increasing processing throughput from 45 FPS → 58 FPS for real-time workloads.
- Built a reproducible training/evaluation workflow for keypoint + temporal models on large-scale X Games video by automating dataset curation/labeling pipelines and tracking experiments, metrics, and artifacts with MLflow for fast iteration and regression-safe comparisons.
- Built probabilistic scoring services using Bayesian/MCMC modeling to estimate event scores and podium probability from trick metrics; generated commentator-facing insights by conditioning on historical distributions and converting outputs into concise 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 microscopy ML pipeline by training and fine-tuning segmentation models (SAM/SAM2, YOLOv9, ResUNet variants) to detect pores/cracks in post-irradiated nuclear materials across FIB/SEM/TEM/X-ray CT modalities on HPC infrastructure.
- Built and ran systematic experiments to quantify how resolution, contrast, noise, and imaging artifacts affect segmentation performance; evaluated modality-specific improvements using saliency-guided approaches and attention modules (CBAM vs bottleneck self-attention).
- Co-authored: "Bridging Multimodal Microscopy for Advanced Characterization on Nuclear Fuel Using Machine Learning," Frontiers in Mechanical Engineering – Digital Manufacturing (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 Microsoft Teams SaaS assistant enabling enterprise users to query large internal document corpora with grounded responses.
- Built the document ingestion pipeline (SFTP/Blob ingestion, validation, metadata normalization, chunking) and automated guardrails (document type/expiry/naming checks) to prevent bad inputs from corrupting downstream workflows.
- Implemented hybrid search + vector retrieval + re-ranking using Azure Cognitive Search, integrating Azure OpenAI embeddings for indexing and improving relevance and increasing answer grounding with source-backed responses.
- Built and operated backend microservices for retrieval + generation + summarization (Node), adding Redis caching to reduce latency and increase throughput for repeated queries and shared document segments.
- Developed feedback/logging and monitoring services (Azure log management) to capture user signals and conversation traces; used feedback to iteratively improve ranking/answer quality, and productionized deployments with Docker + CI/CD, cutting release time 50% (15 → 7 min) and improving resource efficiency 20–25%.
- Led Databricks + Azure data engineering work (medallion architecture, Spark/Delta pipelines) reducing ETL runtime 30 → 15 min and improving downstream Tableau performance.

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

PROJECTS

- **Full-Stack E-Commerce application (2024):** Developed a scalable, responsive web application using **React.js, Node.js, Express.js, and MongoDB, JWT authentication, RESTful APIs, and state management with Redux**. Integrated secure payment gateways (**Stripe**), real-time order tracking, **Redis** for caching, and a **dynamic admin dashboard** for managing orders, inventory, and user interactions- link
- **Human pose estimation in fitness tracking and guidance (2023):** Developed an AI-based human pose detection tool using BlazePose for real-time exercise monitoring. Built a website with HTML, CSS, Node.js, and JavaScript, integrating exercise metrics and deep learning yoga pose classification. Automated data reporting via email for user engagement and tracking. **Research paper**: "**Spatial Structure-oriented and Angle-based Human Pose Estimation for Pose Classification**" in International Journal for Multidisciplinary Research (IJFMR) - link
- **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

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.
- **Co-Founder, Data Science and AI-ML Student Club, Layers (Dec 2021 - Present)**: Lead Layers Club to promote knowledge sharing, networking, and skill development in Data Science and AI-ML.