

AJAY KALLEPALLI

✉ ajay.r.kallepalli@gmail.com

🌐 [linkedin.com/in/ajaykallepalli/](https://www.linkedin.com/in/ajaykallepalli/)

☎ (424) 354-7363

📍 San Francisco, CA

EXPERIENCE

Senior Software Engineer, AI

San Francisco, CA

PwC

Aug 2025 - Present

- Architected and shipped a **Python-based microservices platform** powering an internal AI-enabled opportunity tracking system used by firm leadership, integrating **Salesforce and enterprise data sources** to unify \$50M+ in active commercial pipeline visibility
- Designed and implemented a **document ingestion, indexing, and semantic search pipeline** to support an internal AI assistant, enabling leadership to retrieve context-aware insights across proposals, client records, and knowledge assets
- Proposed and prototyped a **Deep Research Agent** that automatically synthesizes client intelligence from internal and external sources, accelerating partner preparation workflows and expanding opportunity scope
- Owned database migrations, CI/CD automation, and containerized deployments across multiple services, improving release velocity and reliability in a pre-launch, multi-team environment
- Delivered full-stack features using **TypeScript frontend components** aligned with Figma-driven design systems, ensuring consistent UX across analytics dashboards and AI workflows

Software Engineer, Agentic AI

Atlanta, GA

Honeywell (Master's Intern)

Jun 2025 - Sep 2025

- Delivered a **LangGraph-orchestrated** chatbot that enabled **Databricks** users to query data and launch SQL, GraphRAG, and visualization agents with natural language
- Containerized services with Docker and deployed to **Azure Kubernetes Service (AKS)**; used GitHub Actions and Octopus Deploy for end-to-end CI/CD pipelines to enable **zero-downtime** blue-green releases.

Software Engineer, AI/ML

San Francisco, CA

ArangoDB (Part-Time)

Oct 2024 - Jun 2025

- Architected an open-source **GraphRAG** solution (FastAPI, Docker/K8s, LangChain, vLLM, ArangoDB) that boosted answer accuracy from **50% to 90%** through knowledge-graph semantic search and graph traversals
- Reinforcement **finetuned** an LLM Copilot using PyTorch, **improving** Arango Query Language generation accuracy by **50%** and enabling natural-language graph querying for ArangoDB customers
- Expanded and productionized the **langchain-arangodb** Python package—added **hybrid search**, grew PyTest coverage **10% to 90%**, automated CI/CD with GitHub Actions, and published docs that drove adoption by **NVIDIA's** Video Search & Summarization team

Machine Learning Engineer, Support

San Diego, CA

Qualcomm

Dec 2021 - Jun 2024

- Built from scratch end-to-end computer vision data pipelines leveraging Python, AWS (S3, IAM), and Bash, orchestrating secure and scalable handling of **petabyte-scale** data for facial modeling and synthetic avatar generation
- Achieved a **16x speedup** in CV data processing workflows by parallelizing processing tasks (Color Correction, Normal Map Generation, Mesh Generation etc.) using multiprocessing and distributed computing strategies
- Designed and implemented a custom image processing pipeline for facial avatar reconstruction using RealityCapture, OpenCV, NumPy, and RawTherapee across an 87-camera photogrammetry rig, eliminating **\$35,000** in third-party consulting costs
- Trained and led a remote annotation team, developing modular video tutorials and automating Quality Control protocols using Python scripts, ultimately resulting in **4000 hours saved** for engineering staff

EDUCATION

University of San Francisco (USF)

San Francisco, CA

M.S. in Data Science, Data Engineering Concentration GPA: 3.95/4.0

Expected Dec 2025

University of California, Los Angeles (UCLA)

Los Angeles, CA

B.S. Statistics; B.A. in Economics

Sep 2017 - Jun 2021

SKILLS

Languages: Python, R, SQL, JavaScript, PowerShell, Bash, Matlab, TypeScript

Frameworks & Tools: PyTorch/TensorFlow, Git, Docker, Kubernetes, AWS, Apache (Hadoop, Spark, Kafka), Selenium

Libraries: NumPy, Pandas, scikit-learn, Hugging Face Transformers, LangChain, LlamaIndex, MLFlow, OpenCV