Amit Raj Reddy Dharam

903-990-6812 | reddy.amitraj@gmail.com | linkedin.com/in/arrd19/ | ames-zero.github.io | Tempe, AZ

SUMMARY

Software engineer focused on AI-powered backend systems, cloud deployment, and agentic workflows. Experience shipping LLM-based products, optimizing latency at scale, and using AI coding assistants in daily development. Seeking new grad software developer (AI/ML), AI cloud engineer, data science, and data engineer positions. Authorized to work under OPT and STEM-OPT beginning May 2026.

EDUCATION

Masters in Computer Science, Arizona State University (Tempe, AZ) **May 2026**
Coursework: Artificial Intelligence, Statistical Machine Learning, Data Mining, Cloud Computing, Distributed Systems, IOT

Bachelors in Computer Science and MSc in Mathematics, BITS Pilani (Hyderabad, India) **May 2023**
Coursework: Data Structure and Algorithms (C/C++), OOP (Java), Deep Learning (Python), Databases(SQL), Computer Networks

TECHNICAL SKILLS

Languages and Technologies: Java, Python, JavaScript, HTML, Go, C/C++, Flask, FastAPI, PostgreSQL, MongoDB, Neo4j
ML frameworks and libraries: PyTorch, TensorFlow, Pandas, PySpark, Scikit-Learn, NumPy, Matplotlib, OpenCV, Open3D, XGboost
Natural Language Processing (NLP): Hugging Face Transformers, LangChain, CrewAI, Google ADK, FAISS, Pinecone, BERT, OpenAI APIs
Developer tools: Amazon Web Services, Google Cloud Platform (GCP), Docker, Kubernetes, Grafana, GitHub, Terraform, GitHub Copilot

PROFESSIONAL EXPERIENCE

AI Data Science Analyst (Part-time) **Sept 2024 – June 2025**
Enterprise technology at Arizona State University *Tempe, AZ*
- Implemented backend for a natural language conversational AI agent using Python and FastAPI, allowing users to query, filter, and display data from NoSQL and relational databases using conversational inputs, with query execution accuracy over 85%
- Integrated LangChain to orchestrate workflows and used GraphRAG with Neo4j for semantic query search. Deployed the system on AWS EKS with CI/CD pipelines, enabling scalable inference and achieving 30% faster response times for over 100 queries per session
- Designed a query optimization layer with caching and collaborated with other teams through code reviews and regular sprints to reduce database response times by 40% and improve system reliability.

Software Development Engineer - Machine Learning **July 2023 – June 2024**
TELUS Digital *India*
- Finetuned open-source LLMs using PyTorch and Hugging Face on domain specific scientific datasets using PEFT-based methods for creating synthetic datasets. Increased manual question generation throughput by 50%, reducing generation time by 40%
- Instrumented ML services with latency and error measurements using Grafana dashboards alongside AWS CloudWatch, resulting in 30-40% performance gains through rigorous data-driven iteration, Redis cache implementation, and query optimization tactics
- Mentored interns through targeted guidance to test the LLM against key NLP benchmarks (MATH, MMLU), achieving +6% score improvements and 50% faster evaluations through automated pipeline workflows and efficient parallel processing

Software Development Intern - Machine Learning **Aug 2022 – July 2023**
TELUS Digital *India*
- Prototyped advanced DeepSORT and ByteTrack algorithms for precise object bounding box tracking across video sequences, achieving 80% mAP object recognition and 75% IDF1 multi-object tracking accuracy, rigorously validated against MOT17 benchmarks
- Improved image segmentation precision by 23% across 15,000+ annotated frames by fine-tuning U-Net and Mask R-CNN architectures using the CityScapes dataset, reducing manual annotation time from 8 hours to 5 hours per dataset batch
- Designed pytest integration test strategies for FastAPI/TorchServe ML APIs serving 15+ projects, maintaining sub-200ms response times, 90%+ code coverage via GitHub Actions CI/CD automation, and JWT authentication security

PROJECTS

WardrobeAI (Recommendations and Image Generation) | *Winner Sunhacks, Sept 2025* **Project-Link**
- Built a robust Python FastAPI backend composed of 5+ microservices, integrating MongoDB, AWS S3, and CRM-style user profiles with Google Gemini APIs to provide virtual clothing try-ons and deliver personalized outfit suggestions across 3,000+ clothing items
- Coordinated a 4-person team to architect and deploy a sophisticated LinUCB recommendation engine integrated with a conversational chatbot, delivering highly personalized outfit suggestions tailored for 10+ occasions including birthdays, weddings, and formal events

Tom-Riddle's Diary (AI Chatbot Agent) | *Winner InnovationHacks, Apr 2025* **Project-Link**
- Engineered and integrated a Python FastAPI backend using ChromaDB (vector database) to enable long-term semantic memory and efficient information retrieval, reducing response time by 35% and supporting contextual AI reflections
- Integrated Google Gemini LLMs into a conversational backend pipeline, enabling natural language queries and persona-based tone generation, handling over 1,000 personalized interactions with >90% response relevance

ACTIVITIES AND LEADERSHIP

- Actively participated in hackathon teams to build end-to-end full-stack AI/ML applications, leveraging GitHub Copilot, debugging expertise, and problem-solving skills to deliver production-ready prototypes in 24-48 hours, securing multiple first-place wins
- Volunteered at OpportunityHack (ohack.org), proactively taking initiative to direct development teams with meticulous attention to detail while constructing robust websites and digital solutions for NPOs including the Heritage Square Foundation and NMTSA