

AMIT RAJ REDDY DHARAM

+1 903-990-6812 | reddy.amitraj@gmail.com | [linkedin.com/in/arrd19/](https://www.linkedin.com/in/arrd19/) | ames-zero.github.io | Tempe, AZ

EDUCATION

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

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

TECHNICAL SKILLS

Languages and Technologies: Python, Java, JavaScript, C/C++, Flask, FastAPI, MySQL, MongoDB, Redis, Neo4j
ML frameworks and libraries: PyTorch, TensorFlow, Pandas, PySpark, Scikit-Learn, NumPy, Matplotlib, Open3D, XGboost
Natural Language Processing (NLP): Hugging Face Transformers, LangChain, FAISS, Pinecone, BERT, OpenAI APIs
Developer tools: Amazon Web Services (AWS), Google Cloud (GCP), Azure, Docker, Kubernetes, Github, Terraform
Concepts: Deep learning, AI cloud deployment, backend development and API design, generative AI, computer vision

PROFESSIONAL EXPERIENCE

AI Data Science Analyst (Part-time) | Arizona State University | Tempe, AZ **Sept 2024 – June 2025**

- Implemented the backend for a NL2SQL conversational agent using Python and FastAPI, allowing users to query, filter and display data from multi-table databases using conversational inputs, with query execution accuracy over 85%
- Optimized natural language to SQL module responses by incorporating team lead and engineer feedback into prompt engineering, fine-tuned LLMs, and GraphRAG implementation, leading to a 5% increase in accuracy
- Integrated LangChain for fast orchestration and Neo4j for efficient semantic search using embeddings; utilized AWS EKS for scalable inference; and reduced response time by 30% over 100+ questions per session

Software Development Engineer- ML | TELUS Digital | India **July 2023 – June 2024**

- Instruction-tuned open-source LLMs with PyTorch on scientific topics using PEFT-based methods for synthetic data creation using a RAG pipeline. Increased manual question generation throughput by 50%, reducing generation time by 40%
- Mentored an intern to test the LLM against conventional NLP benchmarks (MATH, MMLU) and internal domain-specific datasets, achieving a +6% improvement in benchmark scores and 50% faster evaluation timeframes
- Evaluated a custom-trained 3D object detection (voxel-based) and tracking models(DeepSORT) on the KITTI benchmark, achieving around 90% precision for object detection and 70% MOTA (Multi Object Tracking Accuracy) for tracking

Software Development Intern - ML | TELUS Digital | India **Aug 2022 – July 2023**

- Prototyped DeepSORT and ByteTrack algorithms for bounding box tracking over video sequences, resulting in 80% mAP object recognition and 75% IDF1 multi-object tracking accuracy, as confirmed against MOT17 benchmarks
- Improved image segmentation precision by 23% across 15,000+ annotated frames by uptraining U-Net and Mask R-CNN architectures using the CityScapes dataset, reducing manual annotation time from 8 hours to 5 hours per dataset batch
- Deployed ML APIs serving 15+ active projects with sub-200 ms response times by establishing FastAPI and TorchServe endpoints, containerizing the segmentation models with Docker, and hosting on AWS EC2 with JWT authentication

PROJECTS

WardrobeAI (Recommendations and Image Generation) | Winner Sunhacks, Sept 2025 **Project-Link**

- Built a Python backend that integrates with MongoDB and AWS S3, linking Gemini Chat API and Gemini Nano-Banana models across 5+ microservices to provide virtual try-ons and outfit recommendations for over 3000 clothing items
- Coordinated a 4-person team to architect and deploy a LinUCB-based recommendation engine with a conversational chatbot, generating personalized outfit suggestions for 10+ occasions like birthdays, weddings etc

Tom-Riddle's Diary (AI Chatbot Agent) | Winner InnovationHacks, Apr 2025 **Project-Link**

- Engineered and integrated a Python FastAPI backend using ChromaDB 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

- Participated in and won several hackathons by leading cross-functional teams to construct full-stack AI/ML solutions with Python/JavaScript backends, producing production-ready prototypes within 24-48 hours.
- Volunteered and supervised at OpportunityHack (ohack.org), directing development teams to construct and manage websites and digital solutions for Arizona organizations such as the Heritage Square Foundation and NMTSA.