

Ajay Therala

+1 (623) 286-1552 | atherala@asu.edu | ajay-therala7 | Ajaytherala

SUMMARY

Software Developer & AI Engineer with **3** years of experience delivering scalable GenAI Solutions, deploying AI services on cloud infrastructure. Published researcher with proven ability to deliver end-to-end ML workflows, RAG systems, and domain-specific AI applications.

EDUCATION

Master of Science, Computer Science

Arizona State University, Tempe, AZ, United States.

August 2023 - May 2025

GPA: 3.93/4.0

Course Work: Topics in NLP, Data Mining, Data Visualization, Statistical ML, Data Processing at Scale, Knowledge Representation & Reasoning.

Bachelor of Technology, Information Technology

Jawaharlal Nehru Technological University, Hyderabad, TS, India

August 2017 - July 2021

GPA: 9.21/10.0

Course Work: DSA Using C, Java & Python, Operating Systems, Object Oriented Programming, Linear Algebra, Probability & Statistics, DBMS

TECHNICAL SKILLS

Programming Languages : C, Python, SQL, Java Script.

AI & Machine Learning : TensorFlow, PyTorch, Generative AI, scikit-learn, Lang Chain, NLP, Deep Learning, Pandas, NumPy.

Data Engineering & Big Data : PySpark, Spark SQL, Hadoop.

Frameworks & Databases : MongoDB, DynamoDB, PostgreSQL, Vector Databases (Open Search), Django, Streamlit, Tableau.

Cloud & Devops: AWS, CI/CD Pipelines, Git, Docker.

CERTIFICATIONS

- Microsoft Certified: Azure Data Scientist Associate [Verify Here.](#)
- Oracle Certified : Oracle Cloud Infrastructure 2024 Generative AI Certified Professional [Verify Here.](#)

PUBLICATIONS

- [1] Evaluating Multimodal Large Language Models across Distribution Shifts and Augmentations, Publisher : IEEE/CVF Conference on CVPR Workshops [Paper Link.](#)
- [2] NetMD- Network Traffic Analysis and Malware Detection, Publisher: IEEE [Paper Link.](#)

PROFESSIONAL EXPERIENCE

AI Data Engineer, Dexian LLC

Arizona State University (Client)

May 2025 - Present
Tempe, United States.

- Designed & implemented conversation truncation logic to efficiently manage context length, preventing context length errors and ensuring reliable LLM performance.

AI Full Stack Developer, Arizona State University

AI Acceleration Team

August 2024 - May 2025
Tempe, United States.

- **Optimized Data Ingestion Pipeline** - Spearheaded a **95% reduction** in file chunking times for large datasets (from 1500 s to 4.87s) by integrating an **advanced chunking approach**, enhancing data processing efficiency.
- Developed a **CI/CD pipeline** to streamline end-to-end deployment workflow: iteratively updated and tested code in SageMaker, managed version control, dockerized the application, published image to ECR & deployed to Lambda for serverless execution.
- Engineered a **RAG pipeline** by integrating the **Google Drive API** to extract & index into **Open Search**, empowering seamless **Retrieval-Augmented Generation** (RAG) for fast and precise content discovery.

Systems Engineer (ML Developer), Tata Consultancy Services Limited

TechBU, Digital Research & Innovation

August 2021 - August 2023
Hyderabad, India

- Engineered core components for a domain-specific search engine (Cognitive Product Support), **enhanced search accuracy by 30%**.
- Crafted Data Lens, a component for training custom NER models, achieving an **impressive 80% - 90% accuracy**.
- Demonstrated expertise in Generative AI, **Prompt Engineering** by developing advanced GPT-powered bots handling over **5,000+ interactions** daily. Delivered **impactful client demos**, earning high praise from esteemed clientele.
- **Led** the development of an automated document processing solution using **Texttract**, and built a POC for extracting data from handwritten forms.

ML Research Project Intern

Tata Consultancy Services Limited

Jan 2021 - Aug 2021
Hyderabad, India

- Improved detection accuracy by **6%** on NetML, CICIDS2017, and non-vpn2016 datasets through **data refinement techniques**, and **Bagging & Boosting** algorithms.
- **Achieved** a top 5 position in the NetML Network Traffic Analytics Challenge 2020 and presented research findings at **ICAIIC 2022** to over 300 peers and industry professionals.
- Orchestrated end-to-end machine learning workflow using Django, encompassing ingestion, EDA, training, and evaluation.

PROJECTS

AI Resume Analyzer.

- Developed an **AWS-powered AI Resume Analyzer** leveraging **Amazon Bedrock's LLM**, API Gateway, and Lambda to assess job fit, identify skill gaps, and provide **AI-driven** career insights through a Streamlit-based UI.

Business & User Level Analysis of YELP Dataset.

- **Optimized** Yelp dataset processing, leveraged **PySpark & Spark SQL** for scalable **ETL**, data transformations, and designed complex queries in a distributed environment. Built interactive visualizations in **Tableau** to analyze key business and user trends.

Data Analyst (Volunteer Research Assistant)

- **Analyzed** GB's of EEG signals, physiological metrics, and sensor data from 32 participants across 6 conditions, using **statistical analysis & visualization** to assess cognitive performance under varying insulation materials & temperatures.

Evaluating Multimodal Large Language Models across Distribution Shifts and Augmentations.

- **Benchmarked** state-of-the-art MLLMs across VQAv2 and CLEVR datasets under multimodal distribution shifts, revealing **key robustness gaps** in synthetic and complex reasoning settings.
- Generated **75K+ QA pairs** and conducted fine-grained analysis identifying performance degradation due to visual perturbations and logical connectives, especially in color and count-based tasks.