

# Ajay Therala

(623) 286-1552 | [atherala@asu.edu](mailto:atherala@asu.edu) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

### Master of Science, Computer Science

Arizona State University, Tempe

May 2025

AZ, United States

*Coursework: Topics in Natural Language Processing, Data Mining, Data Visualization, Statistical Machine Learning, Data Processing at Scale, Foundations of Algorithms, Statistical Learning Theory, Software Verification, Validation, and Testing, Knowledge Representation and Reasoning.*

### Bachelors of Technology, Information Technology

Jawaharlal Nehru Technological University, Hyderabad

July 2021

TS, India

*Coursework: Data Structures & Algorithms Using C, Java & Python, Operating Systems, Computer Networks, Software Engineering, Business Intelligence, Object Oriented Programming, Linear Algebra, Probability & Statistics, DBMS, Design & Analysis of Algorithms.*

## TECHNICAL SKILLS

Programming Languages	C, Python, SQL, PySpark, Java Script, D3.js, Spark SQL.
Technologies & Frameworks	Generative AI, Machine Learning, Deep Learning, Natural Language Processing, Big Data, Hadoop, Hive, Django, AWS (S3, Lambda, Athena, EMR, EC2, GLUE), Oracle Cloud Infrastructure, Fast API, Elastic Search, TensorFlow, Pytorch.
Skills	Data Structures & Algorithms, Git, Databases (MongoDB, DynamoDB, MySQL, Oracle, PostgreSQL), Object Oriented Design, OpenSearch.

## CERTIFICATIONS

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

## PUBLICATIONS

- [Evaluating Multimodal Large Language Models across Distribution Shifts and Augmentations.](#)
- [NetMD- Network Traffic Analysis and Malware Detection.](#)

## PROFESSIONAL EXPERIENCE

### AI Full Stack Developer, AI Acceleration Team

Arizona State University

August 2024 - Present

Tempe, U.S.A

- 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.
- AWS Lambda Pipeline Deployment** - Containerized & Deployed docker-image to **AWS ECR**, and integrated it with **AWS Lambda** for serverless execution. Enhanced pipeline efficiency by instantiating **multiprocessing** in the chunking code to fully leverage AWS Lambda's vCPUs.
- Crafted a script that leverages the **Google Drive API** to extract and index course content from a Google Drive into **OpenSearch**, empowering seamless **Retrieval-Augmented Generation (RAG)** for fast and precise content discovery.

### Systems Engineer (ML Developer), Digital Research & Innovation

Tata Consultancy Services Limited

August 2021 - August 2023

Hyderabad, India

- Engineered core components for TCS Cognitive Product Support, an intelligent domain-specific search engine, **improving search accuracy by 30%**.
- Crafted Data Lens, a component for training custom NER models, achieving an **impressive 80% - 90% accuracy** by leveraging Ontology.
- Demonstrated expertise in Generative AI, and Prompt Engineering by developing advanced GPT-powered bots handling over **5,000+ interactions** daily. Delivered **impactful client demos**, earning high praise from esteemed clientele.
- Mastered AWS services to optimize data storage and automate document processing. Strategically planned and developed a proof of concept (POC) for extracting key-value pairs from handwritten forms, resulting in streamlined data management.

### Research Project Intern

Tata Consultancy Services Limited

January 2021 - August 2021

Hyderabad, India

- Investigated data refinement and balancing techniques while evaluating Machine Learning and Deep Learning algorithms on NetML, CICIDS2017, and non-vpn2016 datasets, **achieving a 6% improvement** in detection accuracy through Bagging & Boosting Algorithms.
- Accomplished **top 5 position** in the **NetML - Network Traffic Analytics Challenge 2020**, surpassing baseline metrics.
- Presented research findings at **ICAHC 2022**, sharing key insights with over 300 peers and industry professionals.

## PROJECTS

### [Text Similarity Model for Question Answer Validation for Online Learning Platforms](#)

Tech stack: Python, NLP, Siamese Networks, LLM

January 2024 - May 2024

Tempe, U.S.A

- Engineered cost-effective automated answer validation systems for e-learning platforms, eliminating manual labor.
- Developed a sophisticated "Text Similarity Model for Answer Validation" utilizing deep learning technologies, **attaining 87% accuracy** in evaluating student responses.
- Implemented a dual validation approach by operationalizing **Siamese Networks** and **Large Language Models**, attaining a **6% accuracy improvement over baseline** approach.

### [Analyzing and Mitigating Hallucinations in Multi modal LLM's](#)

Tech stack: Python, Instruct BLIP, LLM's

September 2023 - December 2023

Tempe U.S.A

- Analyzed behavior of **multi modal LLM's**, including Instruct BLIP, across diverse question types, **covering over 15,000 test cases**.
- Utilized Mistral LLM to curate over **75,000 QA pairs** from COCO Captions for identifying hallucinated responses.
- Conducted in-depth analysis on **5000+ count** and color-based questions, **revealing a 47% of fabricated instances**.