

Akhil Vallala

Delray Beach, FL | <https://www.linkedin.com/in/akhil-fau> | 561-451-6518 | akhilvallalajs@gmail.com

EDUCATION

Florida Atlantic University

MS in Data Science (GPA: 3.95)

Boca Raton, FL

Graduation Date: May 2025

- Organizations/Awards: Graduate Leadership Award, Responsible AI Hackathon (3rd Place), CMBB Biotech Hackathon (3rd Place), PACE Club Founder and President.
- Relevant Coursework: Machine Learning, Deep Learning, NLP, and Big Data Analytics.

ACADEMIC PROJECTS

Tiny-Physics – Master’s Thesis Research Project (ProQuest: [Tiny-Physics](#))

May 2025

- Fine-tuned SmolLM2-360M and SmolLM2-135M using LoRA and full instruction-tuning strategies on domain-specific datasets for solving physics word problems on mobile devices.
- Built a multi-agent synthetic data generation pipeline using GPT-4o to create and validate 6,000+ high-quality question-answer pairs for training.
- Designed and implemented instruction tuning datasets from camel-ai/physics and custom LaTeX-processed physics textbook data, optimizing domain adaptation.
- Benchmarked model performance on MMLU College Physics using the Lighteval framework, achieving 0.2451 accuracy with instruction-tuned SmolLM2-360M, outperforming the base model by over 30%.
- Deployed quantized GGUF versions (Q4_K_M, Q8_0) of the fine-tuned model for real-time inference on mobile devices, with end-to-end support for GGUF conversion and deployment.
- Evaluated model trade-offs across LoRA, instruction tuning, and full fine-tuning, and collaborated on carbon footprint estimation for sustainable AI practices.

NSF Research Project – Generative AI in Education ([Paper Link](#))

Boca Raton, FL

Computational Modeling of Teacher and Student Perspectives in Social Media Discourse

Jan 2024-Aug 2024

- Conducted large-scale data collection from Reddit and news APIs to capture evolving educational discourse on generative AI using PySpark.
- Conducted large-scale data collection and curation of ~15,000 Reddit posts and comments using numpy and pandas, and news APIs to analyze stakeholder discourse on generative AI in education.
- Applied NLP techniques (NLTK, VADER, TextBlob) to benchmark LLM and classical sentiment models for capturing diverging attitudes of teachers and students on AI-assisted learning.
- Designed pipeline for theme extraction and discourse classification to identify key issues like false positives in AI detection and the emotional impact on students falsely accused of cheating.
- Contributed to manuscript submitted to *IEEE Transactions on Computational Social Systems*; presented research at 4 poster events and STELAR’s “Accelerating Research on AI in K–12 Education – RAPID Project Briefings.”
- Contributed to manuscript submitted to *IEEE Transactions on Computational Social Systems*; presented research at 4 poster events and STELAR’s “Accelerating Research on AI in K–12 Education – RAPID Project Briefings.”

Florida Atlantic University

Responsible AI Hackathon

Boca Raton, FL

Feb 2024

- Won 3rd Place and Innovative Solution Award out of 20+ teams for building ARA (Academic Responsible Assistant), an AI-powered tool designed to proactively detect academic dishonesty using Retrieval-Augmented Generation (RAG) architecture.
- Integrated Hugging Face models, LangChain pipelines, and OpenAI APIs to implement an explainable and responsible academic assistant.

WORK EXPERIENCE

Aventusoft LLC

AI Research Intern

Boca Raton, FL

2025 - Present

- Developed AIMEE, a voice-based AI assistant for post-discharge heart failure patient monitoring, integrating OpenAI’s LLM with voice APIs like ElevenLabs and telephony APIs such as Twilio.
- Owned full lifecycle of a RAG-based AI assistant (AIMEE) for heart failure patient monitoring: from data pipeline design to real-time deployment and evaluation, optimizing relevance, safety, and clinical utility.
- Optimized system prompts for medical Q&A using iterative prompt engineering and implemented a prompt evaluation pipeline to assess accuracy, coherence, and patient safety.
- Built an automated escalation feature that triggers email alerts to doctors when patient symptoms indicate serious conditions, reducing response time in critical scenarios.

Accenture

Data Analyst

Hyderabad, India

Aug 2021 – Jul 2023

- Improved data processing efficiency by 40% by building and optimizing ETL pipelines in Python and SQL, resulting in faster execution and streamlined reporting workflows for enterprise clients.
- Expanded data platform functionality by writing modular Python scripts for ingestion and validation, reducing manual intervention by 70% and accelerating deployment cycles.
- Resolved over 50+ data quality issues by writing advanced SQL queries and conducting root cause analysis, leading to 99% data accuracy in executive dashboards.
- Designed and deployed a recommendation engine using collaborative filtering and TensorFlow, leading to a 24% increase in conversion rates for an e-commerce client.
- Built and operationalized a dynamic pricing model using XGBoost and scikit-learn, handling over 50K products per hour and boosting annual revenue by 16%.
- Processed multi-terabyte datasets using Apache Spark on Databricks, improving batch data transformation speeds by 3x and supporting critical business reporting.

Akhil Vallala

Delray Beach, FL | <https://www.linkedin.com/in/akhil-fau> | 561-451-6518 | akhilvallalajs@gmail.com

Infosys

Data Engineer

Mysore, India

Mar 2019 - Jul 2021

- Enhanced reporting accuracy by 25% by designing ETL pipelines with Python and SQL, resulting in improved trust in BI metrics across the enterprise.
- Reduced validation time by 90% by developing a Python-based data quality check library, enabling daily data audits with zero manual effort.
- Cut deployment time in half by implementing configuration-driven ETL logic using YAML/JSON, which allowed rapid onboarding of new data sources.
- Collaborated with the DevOps team to implement CI/CD pipelines in Azure DevOps for deploying ETL workflows and data processing scripts.
- Automated infrastructure and release processes using PowerShell and Azure DevOps Services, ensuring smooth deployments and reducing manual errors.
- Resolved recurring data anomalies by running ad hoc Spark SQL queries, leading to a 95% reduction in error escalation incidents.

LEADERSHIP EXPERIENCE

PACE Club ([PACE Club](#))

Founder & President

Boca Raton, FL

April 2024 – Dec 2024

- Founded and led the PACE (Professional and Academic Collaboration for Excellence) Club, impacting 200+ students.
- Organized and hosted 30+ sessions featuring IEEE Distinguished Speakers, university researchers, and Industry professionals.
- Promoted student engagement in research, leadership, and career growth through tech talks, mentoring events, and networking opportunities.

SKILLS

AI & ML: OpenAI API, Hugging Face, Langchain, Jupyter notebook, Scikit-learn, RAG, and Instruction tuning, Pandas

Data Engineering: Python, SQL, Apache Spark, Azure Data Factory, Databricks, ETL pipelines, and Data Modeling

Data Analysis & Viz: Power BI, Tableau, Statistical Testing

Tools and Practices: Git, Agile, Jira, and Confluence