Atharv Vani

(+1) 737-895-5379 | atharvv08@gmail.com | Austin, TX
Github | LinkedIn | Website

EDUCATION

The University of Texas at Austin, Austin, TX May 2027
Bachelor of Science in Statistics and Data Science, Minor in Computer Science
Relevant Coursework: Machine Learning, Data Structures, Algorithms, Probability & Statistics, Databases, Data Science, Linear Algebra, Calculus I-III

SKILLS

Tools: MySQL, MongoDB, Neo4j, GCP, CI/CD, Docker, VSCode, RStudio, Colab, MS365
Languages: Python, SQL, R, Scala, Java, C, C++
Frameworks: Pandas, NumPy, Seaborn, Streamlit, Scikit-learn, Keras, PyTorch, HuggingFace, LlamaIndex, tidyverse(R)
Certifications: IBM Data Science Professional, IT Certiport - Java Object Oriented Programming

WORK EXPERIENCE

GoTunda, Remote Sept 2025 - Present
Software Developer + Data Growth Intern

- Developed a **Django REST API backend** with MySQL + JWT authentication to present an agricultural marketplace platform, enabling secure multi-role user management and providing seamless frontend integration
- Providing analytic support in a cross-functional team by shipping BI dashboards presenting **valuable KPIs** for user, engagement, delivery routes, and product warehousing, driving strategy for ground operations
- Developing automated ingestion pipelines to collect user/merchant data, allowing for data valuation

Mercor Intelligence, Remote Nov 2025 - Nov 2025
Contract - Data Scientist

- Produced high-quality notebook samples for LLM training by developing complex Python workflows in scientific computation, statistical modeling, and ML to enhance **model reasoning**, debugging, and code-generation ability
- Authored detailed prompts, solutions, and pipelines with pytest and nbconvert, to serve large language models in understanding notebooks that use modern data frameworks for data science and software development tasks

Center for Autonomy, University of Texas Oden Institute, Austin, TX Jun 2025 - Aug 2025
Geoelements Group - Undergraduate Research Assistant

- Developed a LlamaIndex **AI agent** using retrieval augmented generation(RAG) to allow for reasoning steps and physics retrieval tasks, reducing manual work for a civil engineer by ~30% when designing reliable structures
- Fine-tuned an SLM using a reinforcement learning(RL) environment to evaluate structural systems and reward/penalize decision-making, helping reduce hallucinations and incorporate geotechnical domain expertise
- Built a POC data-driven physics simulator using Gaussian splatting to render an initial scene, merged with a graph neural network model for particle flow prediction as a 23% more **computationally efficient** alternative to MPM

PROJECTS

Near Earth Object Detection - Cockrell School of Engineering Aug 2025 - Dec 2025

- Co-authored a report improving state-of-the-art models in classifying hazardous objects with ~93% accuracy, and leveraged NASA's NEO API to develop a deep neural network highlighting strategy and global risk mitigation
- Optimized ML models by hyperparameter tuning with search algorithms and constructing an astrophysical feature set representing nonlinear interaction terms such as eccentricity, orbital radius, and diameter

IBM + SpaceX Simulation Capstone May 2025 - Oct 2025

- Developed an end-to-end pipeline to forecast Falcon 9 landing success with 84.6% classification accuracy, enabling optimized strategy initiatives for $60M+ cost-saving potential for reusable rockets and materials
- Built a front-end app to visualize launch outcomes, orbit success trends, and payload–performance relationships, presenting informative analytics for control system and stakeholder teams via a capstone presentation

University of Texas Medical Branch May 2023 - Aug 2023

- Presented a statistical method to analyze PLD1 expression post-inhibition using time-series data, by applying t-tests, causal inference, and A/B testing to evaluate treatment effects on AD models across cohorts
- Built an ingestion pipeline in R to prepare object recognition results by extracting data across experiments, such as Western blot analyses and electrophysiology recordings, for data governance, security, and future analysis