Atharv Vani

737-895-5379 | atharvv08@gmail.com | LinkedIn | Github | Website

Education

University of Texas at Austin Austin, TX

Bachelor of Science in Statistics and Data Science, Minor in Computer Science *Expected: May 2027*

- **Coursework:** Machine Learning, Software Design, Databases, Statistics/Probability, Linear Algebra

Experience

Data Scientist - Contract Nov. 2025 - Dec. 2025

Mercor Intelligence *Remote*

- Enhanced **large language model** inference for software engineering and data science tasks by authoring detailed prompts and tested python notebooks to serve as **high-quality data samples**
- Developed Python workflows in computation, statistics, and machine learning topics to improve **model reasoning**

Undergraduate Research Assistant Jul. 2025 – Aug. 2025

Oden Institute for Computational Engineering and Sciences, Center of Autonomy *Austin, TX*

- Worked in the Geoelements Group focusing on SciML model development for data-driven simulation
- Developed AI agents using RAG to reference **multi-step** physics problems and civil design documents, reducing hallucinations in queries by 33%, providing support tools for geotechnical engineers designing reliable solutions
- Fine-tuned a Gemma-2B SLM within a **RL environment** by incorporating structural domain expertise to analyze load-bearing systems and design quality, enabling data-driven model improvements
- Prototyped a **graph network simulator** with PAC-NeRF, serving as a surrogate model for particle flow

Research Assistant May 2023 – Aug. 2023

UT Medical Branch, Mitchell Center for Neurodegenerative Diseases *Galveston, TX*

- Applied **A/B testing** and analysis methods to lab results and observational memory deficits, including novel **object recognition data** and protein expression records, under the mentorship of Dr. Balaji Krishnan
- Presented analytics report with emphasis on the research workflow and **key findings** from PLD1 inhibition trials

Projects

Translative Transformer | *PyTorch, HuggingFace, NLTK, FastAPI, Docker*| Github Sep. 2025 – Present

- Researched self-attention in transformers to leverage the mechanism in the development of a multilingual translation model, presented to users as an inference service, reducing request speeds and latency by 40%
- Improved multi-step dynamic model training with partial-epoch checkpointing, statistic evaluation, and tuned parameters, resulting in faster debugging cycles and stable convergence during large-scale Seq2Seq training
- Deployed a production-grade, translation inference service by developing a microservice drawing on the model, creating a scalable foundation for containerization, CI/CD integration, and future model versioning

NEO Hazard Classification | *Pandas, Seaborn, Scikit-learn, PyTorch* | Github Aug. 2025 – Dec. 2025

- Led development efforts in a 3 person team to author a paper within Cockrell's Engineering Mechanics department, presenting SOTA models trained on NASA's API to classify hazardous near earth objects with 93% accuracy
- Improved classification by training on astrophysical features representing nonlinear interaction terms (eccentricity, orbital radius, magnitude, diameter) and tuning model hyperparameters with efficient search algorithms
- Demonstrated the viability of a computationally light deep learning approach for risk analysis and driving strategic planetary safety efforts by establishing a baseline multilayer perceptron neural network with standardized features

GoTunda | *Django, MySQL, pytest, PowerBI* | Website Jan. 2025 – Sep. 2025

- Worked in a team to develop software for a local agricultural e-commerce service in Nairobi, incorporating complete support for user orders by drawing on product databases and analytics to improve farm-to-table workflows
- Formulated strategic insights based on raw data and automated business logic to support data-centric components such as product analytics, delivery systems, inventory tracking, and financial payout pipelines
- Implemented JWT-based role validation and achieved 98% API test coverage for production-grade operations

Technical Skills

Languages: Python, SQL, R, Java, Cypher, Scala
Frameworks: BeautifulSoup, FastAPI, Django, Pandas, NumPy, Matplotlib, Seaborn, Plotly/Dash, Streamlit, Scikit-learn, XGBoost, PyTorch, JAX, LangChain, LlamaIndex, tidyverse (R)
Tools: PostgreSQL, MongoDB, Neo4j, GCP (BigQuery, Firestore), CI/CD, VSCode, RStudio, Colab, MS 365