Jahnavi Ashok

jahnaviashok99@gmail.com | (817) 883-9958 | http://www.linkedin.com/in/jahnavi-ashok | https://github.com/JahnaviAshok

SUMMARY

Data Scientist & Software Engineer with 4+ years of experience in Python, C++ and deep learning optimization. Skilled in LLMs, RAG pipelines, graph-based learning, and time-series modeling with experience deploying production-grade models. Analyze complex datasets and collaborate with cross-functional teams to drive impactful, data-driven decisions.

EDUCATION

MS. in Data Science, University of Texas at Arlington — August 2024 – Expected May 2026 | Arlington, United States

TECHNICAL SKILLS

- **Languages & Libraries:** Python, SQL, C++, Pandas, NumPy, Matplotlib, Scikit-learn, TensorFlow, Keras, PyG
- **Deep Learning and ML:** Model optimization, GNNs, LSTMs, recommendation systems, hyperparameter tuning
- **GenAI & NLP:** LangChain, Hugging Face Transformers, ChromaDB, embeddings (Sentence-Transformers, OpenAI)
- **Cloud-based ML Services:** AWS SageMaker, EC2, S3
- **Tools & Platforms:** Linux, Git, Docker, Kubernetes, VS Code, Power BI, Streamlit, Jupyter notebook
- **Frameworks & Databases:** Spark, PySpark, MySQL

WORK EXPERIENCE

University of Texas at Arlington, *Machine Learning Research Assistant* — May 2025 – Present | Arlington, United States

- Formulating a physics-informed ML framework to model soil water retention curves from soil texture data (~660 samples).
- Implementing hybrid models that combine physics-based baselines with Gaussian Processes and Neural Networks, reducing prediction error by ~20% compared to empirical fits on SWRC (660+ samples).
- Optimized deep learning training efficiency on GPUs by 25% through feature engineering and workload tuning.
- Applied parallelized data preprocessing pipelines to handle 660+ soil samples efficiently.
- Benchmarked model performance using RMSE, MAE, and R^2, achieving >0.85 R^2 across validation folds.

Western Digital, *Sr. Software Engineer* — January 2021 – July 2024 | Bangalore, India

- Progressed from Intern to Senior Software Engineer in 3.5 years, leading validation and automation initiatives for NVMe protocol based SSDs across 4+ product lines
- Contributed to the prototyping and validation of NVMe protocol-based storage devices, accelerating firmware testing cycles by ~15% and improving test coverage in Agile/Scrum teams.
- Engineered a Python-based UI integrated with low-level C++ device drivers via Boost.Python, cutting manual intervention by 80% and enabling end-to-end test automation for daily regression suites.
- Designed and implemented enhancements for NVMe command support based on evolving product requirements, increasing validation efficiency by 15%.
- Replaced direct device queries with a **database-backed framework** improving performance by 20%.
- Developed and optimized large-scale validation frameworks in C++ and Python, reducing runtime by 40% through lazy-loading and parallelization.

RELEVANT PROJECTS

Data Science Tutor Bot (LLM + RAG + Agent) — August 2025

- Developed an LLM tutor using GPT-4o-mini and LangChain to explain ML algorithms step-by-step with math and code, supporting 20+ core ML topics.
- Implemented a RAG pipeline with Sentence-Transformers embeddings, and ChromaDB for context-grounded answers.
- Extended to a Streamlit-based RAG Agent with safe Python execution, enabling live demos and plots, reducing manual explanation effort by ~40% during peer study sessions.

Well Log Prediction using LSTM + GCN — June 2025

- Developed a hybrid LSTM + GCN model to predict resistivity from gamma ray and porosity across 142 wells.
- Generated multiplex graphs using Spearman correlations to capture dependencies and improve predictive accuracy.
- Fine-tuned the model with Optuna hyperparameter tuning and regularization, achieving ~30% error reduction and $R^2 > 0.8$ on test wells.
- Built Python pipeline for preprocessing, training, and evaluation, supporting real-time inference and visualization.

ACHIEVEMENTS

- Awarded a $9,500 research grant from Intera Incorporated to support graduate-level research.
- Winner of Datathon, recognized for delivering innovative, data-driven solutions in a competitive setting.