Ansh Kumar Dev

520-599-5589 | anshkumardev@gmail.com | linkedin.com/in/ansh-kumar-dev/ | github.com/Anshkumardev

Education

University of Arizona — Tucson, AZ
Master of Science in Data Science — *Aug. 2023 – May 2025*

Gautam Buddha University — Greater Noida, India
B.Tech. in Artificial Intelligence — *Aug. 2019 – May 2023*

Experience

AI Engineer Intern — Jun 2025 – Present
Right Skale — *Pleasanton, CA*

- Designed and deployed **LLM evaluation systems** to parse SEC 10-Q filings, training Bedrock Llama 3 agents to detect hallucinations, misclassifications, and inconsistencies across **100+ filings/hour**.
- Built scalable, containerized microservices (**LlamaParse**, **LlamaExtract**, vector indexer, validators) with prompt chaining for **automated preference alignment and compliance checks**.
- Developed robust **RAG pipelines** for human-aligned content classification using **deep learning embeddings** and OpenSearch Serverless; processed **1M+ chunks** for semantic model QA tasks.
- Led **experimentation** and **benchmarking** of language models via AWS Agents and Step Functions; implemented automated ablation studies to compare inference quality and model hallucination rates.
- Implemented secure, compliant **AI infrastructure** with observability (CloudWatch/X-Ray), **CI/CD**, and **IAM-based access control**; ensured reproducible evaluation pipelines and model hosting.

Data Scientist (Graduate Research Assistant) — Aug 2024 – Jan 2025
University of Arizona — *Tucson, AZ*

- Led **time-series data modeling** and validation of ML workflows using CV–RNNs and **Bayesian hyperparameter tuning**; cut prediction RMSE **18%** on real-world aquifer datasets.
- Performed rigorous **data wrangling, EDA**, and statistical significance testing across hydrologic datasets; documented experiments in research journals and Jupyter notebooks.
- Mentored **15+ researchers** in **reproducible modeling**, model evaluation, and **statistical best practices**, contributing to the lab's open-source experimentation toolkit.

Research Data Scientist — Jun 2022 – Nov 2022
Gautam Buddha University — *Greater Noida, India*

- Conducted original research on anomaly detection and **QoS-aware streaming algorithms**; published in **e-Prime, Elsevier** and presented at **LICTIC 2023**.
- Designed **statistical model evaluations** and root-cause analyses across IoT streams; boosted model throughput **20%** and cut latency **25%**.
- Engineered real-time signal classifiers in **C++/Python**; improved anomaly detection accuracy **40%** using custom ML inference pipelines.

AI Engineer Intern — Apr 2022 – Jul 2022
Merkletree Technologies — *Delhi, India*

- Fine-tuned Faster R-CNN models for facial authentication with targeted augmentation; improved accuracy by **25%**.
- Productionized models via **Docker, Kubernetes**, and CI/CD, reducing latency **35%** and cloud cost **20%**.
- Integrated models into Android apps, enabling on-device inference for **50k+ users** with REST APIs and reproducible release cycles.

Projects

Healthcare Accessibility & Risk Mapping Platform | *Project Link* — Apr 2025

- Built a distributed **GeoPandas + OSMnx** pipeline unifying Geospatial, socioeconomic & clinical data across 3,000+ U.S. census tracts to identify underserved areas for policy action.
- Developed a **LangChain** multi-agent LLM workflow with **OpenAI API** and **FAISS** RAG and a **Streamlit** choropleth dashboard, enabling real-time Q&A and evidence-based insights.

Technical Skills

Languages/ML: Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, OpenCV), R, SQL, C++
Models: RoBERTa, SVM, Decision Trees, K-Means, PCA, Forecasting & Statistical Analysis
Cloud & MLOps: AWS (EKS, EC2, S3), Azure (ADF, Synapse, Databricks), GCP, Docker, Kubernetes, Spark, CI/CD, MLflow
Tools & APIs: Flask, REST APIs, TensorBoard, Git, Linux, Power BI, Streamlit