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 **Generative AI applications** using **multi-modal LLMs** (Llama 3) to process SEC 10-Q filings, integrating with **vector databases** and achieving ingestion speeds over **100+ filings/hour**.
- Developed and monitored **self-healing ML pipelines** on **AWS** using Step Functions, Lambda, and Bedrock, automating error handling and reducing manual recovery by **90%**.
- Implemented robust **AI infrastructure** with GitHub Actions CI/CD, **CloudWatch, X-Ray**, and IAM roles to ensure secure and scalable model deployments.
- Engineered semantic **RAG pipelines** leveraging **Bedrock Titan** and **ONNX-compatible embeddings**, optimizing for low-latency vector search across **1M+ documents**.
- Tracked **operational KPIs** and anomaly events, presenting performance improvements to technical leads.

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

- Built an **IoT sensor-driven** ETL pipeline using Python (Pandas, SQL) to process satellite data and forecast aquifer levels using CV-RNNs, improving RMSE by **18%**.
- Led data onboarding, cleaning, exploratory analysis, and **anomaly detection** using time-series analytics in **Jupyter** and **SQL**.
- Mentored **15+ students** on data wrangling, **Bayesian hyperparameter tuning**, and reproducible ML evaluation techniques.
- Communicated model findings and technical risks to stakeholders and provided actionable AI recommendations.

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

- Analyzed **IoT network data** to build **real-time anomaly detection systems**, improving throughput by **20%** and reducing latency **25%**.
- Implemented statistical models in **C++ and Python**, conducted root cause analysis and boosted prediction accuracy by **40%**.
- Published **QoS-aware streaming protocols** in e-Prime, optimizing message delivery by **25%**.
- Presented ML advancements and pipeline optimizations to research committees and advisors.

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

- Fine-tuned **TensorFlow Faster R-CNN** for biometric verification with targeted data augmentation, improving accuracy by **25%**.
- Containerized ML models using **Docker, Kubernetes**, and deployed via Flask REST APIs with GitHub CI/CD.
- Integrated inference into Android application workflows, enabling real-time edge inference for **50k+ users**.
- Performed model monitoring and KPI tracking to ensure stable, scalable deployment across cloud platforms.

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