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 **multi-modal Generative AI applications** to extract structured insights from SEC 10-Q filings, incorporating **financial analytics**, **customer behavior tagging**, and **dashboard-ready formats** for downstream reporting.
- Built and monitored **self-healing ML pipelines** using AWS Lambda, Step Functions, and Bedrock Agents, automating 90% of anomaly detection and ingestion failures.
- Developed KPI dashboards and **data quality monitors** using **CloudWatch, Excel**, and GitHub CI/CD; integrated **DevOps practices** to maintain production-grade reporting workflows.
- Engineered production-grade **RAG vector search systems** using OpenSearch, Titan embeddings, and custom retrievers across **1M+ financial and behavioral data chunks**.
- Documented pipeline logic, maintained clear **data lineage**, and collaborated with finance SMEs and developers to influence analytics strategy.
- Tracked and resolved ingestion issues using **JIRA-based support tickets** and contributed to agile sprint cycles.

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

- Built an end-to-end **Python (Pandas, SQL)** ETL pipeline to ingest and preprocess sensor data and satellite imagery; developed time-series CV–RNN model that reduced 12-month aquifer RMSE by **18%**.
- Led data onboarding, cleansing, and **exploratory reporting** on geospatial and temporal patterns using **Jupyter**, Excel, and SQL.
- Created executive summaries and analytics reports to inform stakeholders on usage trends and anomalies, with traceable **data lineage** documentation.
- Mentored **15+ students** in statistical modeling, feature engineering, and communicating ML insights to **non-technical stakeholders**.

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

- Fine-tuned **TensorFlow Faster R-CNN** for biometric verification, improving face-auth accuracy **25%** via synthetic data augmentation.
- Deployed models via **Docker + Kubernetes** with Flask APIs and integrated GitHub Actions for CI/CD in a real-time customer-facing environment.
- Collaborated with Android developers to support **real-time model inference for 50k+ users**, tracking performance via metrics dashboards.
- Supported QA and ops teams by documenting model behaviors and triaging bug reports to improve resilience.

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.

E-commerce Analytics Lakehouse | *Project Link* — Mar 2025

- Engineered an **Azure Data Factory to Databricks** (**PySpark**) ETL ingesting 1.2 TB/day of order & clickstream data to power near-real-time KPI dashboards.
- Modeled star-schema views in **Synapse SQL** and built live **Power BI** reports, Reduce manual reporting by 90%

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