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

- Lead **multi-modal AI** pipeline for SEC 10-Q & balance-sheet ingestion, standardizing formats via containerized microservices (LlamaParse/Extract, vector-DB, etc.), processing **100+ filings/hour** with enhanced **operational automation**.
- Developed **real-time anomaly detection** and **self-healing** workflows leveraging logs, metrics, and events to predict and resolve system failures in cloud infrastructure.
- Implemented **incident automation** and **controlled restarts** using AWS Step Functions, IAM, CloudWatch, and GitHub Actions, ensuring **proactive monitoring** and SOX-compliant deployments.
- Built production-grade RAG vector search **AI system** integrating **deep learning**–based Titan embeddings across 1M+ chunks; enhanced performance of **operational analytics dashboards**.
- Collaborated with **DevOps/SRE teams** to design **automation rollouts**, performance alerts, and **infrastructure observability** using OpenSearch Serverless and X-Ray.

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

- Designed and deployed **ETL pipelines** in Python (Pandas, SQL) to preprocess satellite & sensor data for forecasting aquifer levels, enabling **operational decision support**.
- Built time-series models using CV–RNN; conducted **Bayesian hyperparameter tuning**, model validation, and **statistical significance testing** to reduce 12-month RMSE by **18%**.
- Led **data wrangling**, exploratory analysis, and visualization; ensured **reliable data onboarding** and reproducibility of model outputs with strict adherence to **data quality protocols**.

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

- Built **anomaly detection models** on IoT network traffic, improving pipeline reliability by **20%** and reducing latency **25%** through adaptive streaming analysis.
- Designed **QoS-aware data pipelines** and automated root-cause analysis protocols using **statistical time-series modeling** and Python.
- Established **data control measures** to preserve **confidential information** in high-throughput systems and ensured compliance with academic research integrity standards.

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

- Improved model reliability by **25%** through fine-tuning TensorFlow Faster R-CNN with extbfautomated error detection and augmentation for biometric face-auth systems.
- Containerized inference services using **Docker & Kubernetes**, implemented model tracking with **MLflow CI/CD**, and reduced inference latency by **35%**.
- Participated in **data onboarding**, anomaly reconciliation, and deployment monitoring to ensure **system uptime** and **incident response readiness**.

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