

Ansh Kumar Dev

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

EDUCATION

University of Arizona

Master of Science in Data Science

Tucson, AZ

Aug. 2023 – May 2025

Gautam Buddha University

B.Tech. in Artificial Intelligence

Greater Noida, India

Aug. 2019 – May 2023

EXPERIENCE

AI Engineer Intern

Right Skale

Jun 2025 – Present

Pleasanton, CA

- Lead **multi-modal AI** LLM-driven pipeline for SEC 10-Q balance-sheet ingestion, standardizing formats via containerized microservices (LlamaParse/Extract, vector-DB, etc.), processing **100+ filings/hour**.
- Drive orchestration of **machine learning pipeline** using AWS AI Agents and Step Functions, crafting dynamic prompt templates for Bedrock Llama 3 agent with **90% automation**.
- Develop a production-grade RAG vector search **AI system** by designing OpenSearch Serverless schemas and integrating **deep learning**-based Bedrock Titan embeddings across 1M+ chunks for semantic data retrieval.
- Implement **AI infrastructure** with observability, security, and CI/CD; CloudWatch, X-Ray, IAM, GitHub Actions, ensuring SOX-compliant deployments and robust **natural language processing** pipelines.

Data Scientist (Graduate Research Assistant)

University of Arizona

Aug 2024 – Jan 2025

Tucson, AZ

- Built an end-to-end **Python (Pandas, SQL)** ETL pipeline to ingest and preprocess satellite imagery & sensor data, engineered time-series features for a CV-RNN model, and cut 12-month aquifer RMSE by **18%**.
- Led data collection, cleaning, exploratory analysis, and visualization in **Jupyter** and **SQL**; mentored **15+ students** on data wrangling and statistical modeling.
- Validated **ML workflows**, conducted model evaluation, Bayesian hyperparameter tuning, and statistical significance testing to ensure reproducible, high-accuracy predictions.

AI Engineer Intern

Merkletree Technologies

Apr 2022 – Jul 2022

Delhi, India

- Improved face-auth accuracy **25%** by fine-tuning a TensorFlow Faster R-CNN with targeted data augmentation.
- Productionised TensorFlow model with Docker & Kubernetes, cutting inference latency **35%** and cloud cost **20%**.
- Partnered with Android team to embed on-device inference for biometric login across **50 k+ users**.
- Exposed models via Flask REST API and MLflow-tracked CI/CD for rapid, reproducible deployments.

PROJECTS

Healthcare Accessibility and Risk Mapping Platform — [Project Link](#)

Apr 2025

- Built a distributed **GeoPandas** and **OSMnx** pipeline unifying geospatial, socioeconomic, and clinical data across 3000US census tracts to identify underserved areas for policy action.
- Developed a **LangChain** multi-agent LLM workflow using **OpenAI API** and **FAISS** for RAG and a **Streamlit** dashboard enabling real-time QandA and evidence-based insights.

Ecommerce Analytics Lakehouse — [Project Link](#)

Mar 2025

- Engineered an **Azure Data Factory** to **Databricks PySpark** ETL pipeline ingesting 1.2TB per day of orders and clickstream data to power near real-time KPI dashboards.
- Modeled star-schema views in **Synapse SQL** and built live **Power BI** reports reducing manual reporting by 90 percent.

TECHNICAL SKILLS & COMPETENCIES

Languages & ML: Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), SQL, C++

Models & Algorithms: RoBERTa, SVM, Decision Trees, K-Means, NLP (prompt engineering, embeddings), Deep Learning architectures

Domain Expertise: Machine Learning Model Architecture Design, Deep Learning Systems Engineering, Predictive Analytics & Statistical Modeling, Generative AI Application Development

Cloud, MLOps & Infra: MLOps Pipeline Optimization, Kubernetes, AWS, Azure (ADF, Synapse, Databricks), Docker, CI/CD, MLflow

Linguistic AI: Large Language Model Fine-Tuning, AI Strategy Development & Implementation