

ANMOL TRIPATHI

Data Scientist | Machine Learning Engineer | Applied AI

Fort Collins, Colorado | tripathianmol74@gmail.com

linkedin.com/in/anmol-tripathi-60311917a | github.com/unit-mole

Professional Summary

Data scientist building applied machine learning, NLP, and retrieval-augmented generation systems for quality intelligence and operational analytics. Combines professional experience in hybrid classification, deep-learning research, and Python/SQL automation with evaluated Agentic RAG projects spanning investigation, code intelligence, and financial research.

Professional Experience

Quality Data Scientist (Permanent)

Sep 2024 – Present

Hach Company — United States

- Designed and deployed an internal RAG/LLM quality-intelligence agent integrating records dating to 2015 and thousands of nonconformance PDFs, enabling source-grounded investigation for R&D and quality teams.
- Implemented ingestion, chunking, embeddings, semantic retrieval, and contextual prompting; delivered typical responses in approximately 10–15 seconds with traceable supporting evidence.
- Developed a multi-stage NLP framework for Failure Mode, Defect Symptom, and Root Cause Function using Transformer representations, word/character sparse features, LightGBM, calibrated ensembles, and chronological validation.
- Improved external Defect Symptom accuracy from 69.3% to 70.1%, macro F1 from 56.1% to 57.7%, and high-confidence auto-accept coverage from 38.4% to 46.0% in internal full-row evaluation.
- Automated recurring quality analytics with Python, SQL Server, Power BI, and Excel, integrating data validation, product KPI analysis, Chemkey testing, and executive reporting.

Machine Learning Engineer (Part-time)

May 2023 – Aug 2024

University of Arizona College of Nursing — Tucson, AZ

- Built an end-to-end research workflow for wearable-ring temperature data from approximately 135 participants, covering SQL extraction, normalization, temporal features, sequence generation, and validation.
- Trained and compared SVM, LSTM, Bidirectional LSTM, CNN, and Autoencoder models; used ARIMA/SARIMA trend analysis and prediction-error analysis to investigate longitudinal and circadian patterns.
- Refined the estimated labor-prediction window from approximately 14 days to one day in internal research evaluations; communicated findings and limitations to interdisciplinary collaborators.

Student Assistant Manager (Part-time)

Dec 2022 – Aug 2023

University of Arizona BookStores — Tucson, AZ

Previously Student Assistant, Sep 2022 – Dec 2022

- Analyzed approximately 80,000–100,000 monthly sales records across 30+ categories using Excel; developed Tableau dashboards to support demand forecasting and inventory decisions.
- Contributed to an approximately 15% reduction in overstock and stock-shortage conditions; trained 8–10 team members and earned promotion within four months.

Network Operations Center Engineer (Permanent)

Apr 2022 – Aug 2022

Orange Business Services — India

Associate NOC Engineer, Dec 2019 – Mar 2022; Graduate Engineering Trainee, Jun 2019 – Dec 2019

- Analyzed 10,000+ daily network-performance metrics; used Python, SQL, and Excel to investigate anomalies, correlate incidents, and automate extraction, validation, and reporting.
- Developed a network-failure prediction proof of concept achieving approximately 85% accuracy in internal evaluation; enhanced Tableau dashboards tracking 5+ critical network KPIs.

Selected AI and Machine Learning Projects

ReliabilityOps — Agentic Root-Cause Investigation GitHub

- Engineered a LangGraph investigation system combining BGE-M3/BM25 hybrid retrieval, reciprocal rank fusion, Qdrant, structured SQL/statistical tools, local Qwen3 reasoning, and independent claim/citation verification.
- Indexed 14,030 section-aware chunks derived from synthetic incidents and procedures; achieved 93% root-cause Top-1 accuracy on 100 frozen synthetic test cases with 12.88-second mean end-to-end latency.
- Implemented bounded evidence-gap retries, human approval, FastAPI/Gradio interfaces, and OpenTelemetry/Phoenix tracing; selected hybrid retrieval over reranking through development ablations.

RepoAtlas — Repository Intelligence and Safe Coding GitHub

- Built syntax-aware code ingestion, symbol provenance, dependency graphs, and test discovery around BM25/BGE-M3 hybrid retrieval and a bounded Qwen3 investigation workflow.
- Achieved 84.3% File Recall@20 on five frozen historical HTTPX issue-to-fix cases; isolated pre-fix repository evidence from evaluator-only gold fixes to prevent future-fix leakage.
- Gated candidate edits within isolated workspaces and network-disabled Docker execution, verifying changes with pytest, Ruff, and Bandit while preserving the original repository.

FilingsGraph — Temporal Financial Research and GraphRAG GitHub

- Integrated 20 SEC filings and 121,745 XBRL facts across five companies using DuckDB, Qdrant, and NetworkX; produced 2,521 hierarchical text chunks and a provenance-preserving risk graph.
- Routed questions to deterministic financial tools, hybrid retrieval with BGE reranking, temporal comparison, and GraphRAG; grounded local Qwen3 synthesis with claim-level citation checks.
- Validated 100% fact-selection accuracy on 12 financial test questions; evaluated textual retrieval, temporal classification, and graph reasoning separately to expose component-level limitations.

Instruction-Tuned Domain LLM — FLAN-T5, LoRA, and PEFT GitHub

- Fine-tuned FLAN-T5-base through two controlled LoRA experiments in PyTorch; compared adapters with the untouched model on an independent 80-prompt benchmark using ROUGE-L, BERTScore, and human review.
- Improved ROUGE-L F1 from 0.0951 to 0.1461 for the selected adapter; applied release-quality gates and retained the weaker second experiment without promotion.

Technical Skills

Programming and Data: Python, SQL, pandas, NumPy, SciPy, SQL Server, PostgreSQL, DuckDB
Machine Learning: scikit-learn, LightGBM, PyTorch, TensorFlow/Keras, Transformers, NLP, LSTM, CNN, feature engineering, ensemble learning, calibration, chronological validation
Generative AI and Retrieval: RAG, Agentic AI, LangGraph, Hugging Face Transformers, LoRA/PEFT, embeddings, BM25, reciprocal rank fusion, reranking, Qdrant, GraphRAG, NetworkX
Engineering and Analytics: FastAPI, Gradio, Docker, Git, GitHub Actions, pytest, OpenTelemetry, Phoenix, Power BI, Tableau, Excel, statistical analysis, automated data pipelines

Education

University of Arizona — Master's Degree, Data Science Aug 2022 – Dec 2023
GPA: 3.889/4.0
Texas McCombs School of Business — Post Graduation Program, Data Science and Business Analytics 2021 – 2022
Amity University — Bachelor's degree, Electronics and Telecommunications 2015 – 2019

Certification

Google Data Analytics Professional Certificate