Hanumantha Rao Vulli

Data Engineering | Artificial Intelligence | Applied Machine Learning
Denton, Texas, USA | vullihanumantharao@gmail.com
Portfolio: https://sunnyvulli.github.io/ | GitHub: https://github.com/Sunnyvulli
LinkedIn: https://www.linkedin.com/in/hanumantharaovulli/

RESEARCH PROFILE

Data engineer with an MS in Data Engineering and experience in cloud pipelines, SQL, backend development and data validation. Independent work spans evidence support in language-model outputs, phishing detection, scientific time-series classification and reliable analytics systems. Strengths include reproducible experiments, baseline comparisons, data provenance and failure analysis, supported by public code and documented results.

RESEARCH INTERESTS

Reliable and responsible AI; natural language processing and information retrieval; cybersecurity and model robustness; data quality and governance; scientific machine learning and time-series analysis.

EDUCATION

Master of Science in Data Engineering | University of North Texas, USA | August 2023–December 2024

Coursework degree. Relevant study: natural language processing, machine learning, data visualisation, big data and empirical analysis.

Bachelor of Technology in Computer Science and Engineering | KKR and KSR Institute of Technology and Sciences, India | 2019–2022

Diploma in Mechanical Engineering | Kallam Haranadhareddy Institute of Technology, India | 2016–2019

INDEPENDENT RESEARCH AND SELECTED PROJECTS

Applied ML Research Suite | 2026

https://github.com/Sunnyvulli/Data-Engineer_AI

• Citation-support evaluation: studied 2,604 human-labelled ALCE sentences and a separate 900-answer RAGTruth set. Compared lexical and entailment approaches, examined length-related confounding and documented exploratory results and uncertainty.

• Phishing robustness: compared classifiers on 11,055 historical UCI records using grouped splits and feature-outage tests. Examined how missing page and URL signals affect recall; historical performance is not presented as current-campaign validation.

• Scientific time series: ranked early astronomical transients from simulated PLAsTiCC light curves. Compared a forest model with a RAPID-inspired GRU adaptation at a shared three-day cutoff. Observed-data validation remains incomplete.

Evidence Ledger RAG | 2026

https://github.com/Sunnyvulli/evidence-ledger-rag

• Built incremental document indexing, hybrid lexical retrieval and source-hashed extracts, with 18 tests. On 14 authored queries, expected-file hit at 3 was 11/14 for hybrid retrieval, versus 10/14 for BM25. This small development evaluation concerns extractive answers, not free-form generation.

Lakehouse Guardian | 2026

https://github.com/Sunnyvulli/lakehouse-guardian

• Built a local Spark/Iceberg workflow with 25 dbt data checks, Airflow orchestration and LangGraph approval gates. Across 21 controlled fault batches, row precision was 1.00 and recall 0.80; count inflation could go undetected. Terraform was validated; AWS deployment was not performed.

SELECTED PROJECTS CONTINUED

Catalog Sentinel | 2026
https://github.com/Sunnyvulli/catalog-sentinel

• Implemented a read-only MCP catalog server with host-assigned roles, bounded inputs and a hash-chained audit log. Across 16 requests in real stdio sessions, all 13 expected denials were enforced. This is a local capability-control experiment, not a remote authentication service or model-security audit.

Mobility Quality Lab | 2026
https://github.com/Sunnyvulli/mobility-quality

• Built repeat-safe Python/SQLite ingestion and a read-only analytics API over 17,379 UCI Bike Sharing observations. Recorded 13 passing tests covering system and temporal protocols. Added transactional recovery checks, data contracts and a rolling-origin evaluation covering 84 test days. Retained source hashes, fold boundaries and baseline errors.

DocForge AI | Engineering prototype
https://github.com/Sunnyvulli/docforge-ai

• Developed a local document-generation application with streaming responses, SQLite credit accounting and optional Stripe test-mode checkout. Repository tests cover concurrent reservations, failure refunds, duplicate payment events and stream parsing. No commercial deployment or revenue is claimed.

Research status: the studies and systems above are independent portfolio work. Manuscripts are unpublished drafts; experiments and local prototypes are not claimed as accepted publications or production validation.

PROFESSIONAL EXPERIENCE

Data Engineer (Contract) | PRC Cloud Tech, USA | January 2026–present

• Develop and maintain AWS ingestion and ETL pipelines and reporting SQL. Investigate integration issues across Lambda, API Gateway and RDS.

Data Engineer (Contract) | Tech Force, USA | September 2025–January 2026

• Built AWS data pipelines, optimised PostgreSQL and Oracle queries, and implemented data-quality checks. Investigated anomalies and supported CI/CD workflows.

Data Engineer (Contract) | Infosys Now, USA | April 2025–September 2025

• Developed Spring Boot backend services using AWS EC2, S3 and RDS. Added validation checks and maintained workflow documentation.

Java Full Stack Developer (Full-Time) | DBS Technologies, India | July 2022–May 2023

• Built Angular interfaces and Spring Boot REST APIs. Supported AWS deployments and Jenkins/Docker automation.

TECHNICAL SKILLS

Programming: Python, SQL, Java; JavaScript and TypeScript

Data systems: Spark, Iceberg, dbt, Airflow, PostgreSQL, Oracle, SQLite

Research methods: grouped and temporal evaluation, baseline comparisons, error analysis, uncertainty estimation, data provenance and reproducibility

AI and applications: information retrieval, citation-support evaluation, machine learning, LangGraph, MCP, Spring Boot and Angular

Cloud and DevOps: AWS EC2, S3, RDS, Lambda and API Gateway; Terraform; Docker; Jenkins; GitHub Actions

UNPUBLISHED WRITING AND TEACHING

Length Baselines and Evidence-Support Triage: An ALCE-to-RAGTruth Audit — manuscript draft.

Data contracts and evaluation leakage in historical bike-rental analytics — technical manuscript draft.

Reliable analytics from real observations — SQL/Python teaching material with exercises and assessment rubric.

CERTIFICATIONS

AWS Certified Data Engineer – Associate

Oracle Certified Professional Java SE 17 Developer