Arushi Sinha

arushisinha2509@gmail.com | 667-464-1582 | | | | Baltimore, Maryland

Work Experience

Oxbridge Health - *Business Intelligence & Data Engineer* Aug 2024 - Present

- Architected foundational **cloud data infrastructure** (85 tables, 8M+ rows) powering member portal and SaaS applications, by building PostgreSQL reference database on AWS RDS. Integrated data sources from external provider APIs, third-party vendors, and clinical sources with versioning, soft-delete tracking, and automated multi-environment migration protocols.
- Led end-to-end development of **Reporting Platform** and delivered 55 analytics dashboards across Finance, Operations, Claims, and Provider Network – driving initiatives from requirement gathering through production deployment. Designed a **star schema data warehouse** with robust staging architecture, built **scalable ETL pipelines** using AWS Glue and Step Functions, and developed **Looker's LookML semantic layer** to power dashboards.
- Reduced ETL deployment validation time from 5 days to 9 minutes across 87 Glue jobs by designing **automated CI/CD testing framework** using GitHub Actions, achieving 80% code coverage through parallel tests with integration testing, automated rollback, and **Jira/Xray** audit trail integration.
- Enabled proactive issue detection across production ETL workflows by building **HIPAA-compliant** Amazon CloudWatch dashboards with automated Amazon SNS alerts, improving visibility, reliability, and response times.
- Led evaluation, selection, and implementation of an enterprise analytics platform (Looker) across 5 business domains, comparing Looker, QuickSight, and Qlik on cost, security, CI/CD, embedded analytics, and AI/governance criteria—enabling **real-time stakeholder self-service** and eliminating ad hoc reporting requests.

Althea.ai - *Data Science Intern* Feb 2024 - Oct 2024

- Built **Social Determinants of Health** (SDoH) risk attribution algorithm achieving 95% validation accuracy against historical claims data by integrating US Census Geocoder with 58 socioeconomic variables to quantify population health risk factors.
- Predicted **HEDIS compliance** with 86% accuracy, directly informing client's 2024 member outreach strategy, by developing XGBoost model combining SDoH and claims data features.

Johns Hopkins University - *Machine Learning Research Assistant* Feb 2023 - May 2024
Artificial Intelligence for Engineering and Medicine Lab (AIEM)

- Improved downstream classification accuracy by 17.3% by leading research on synthetic-to-real domain adaptation using Stable Diffusion, ControlNet, and Vision-Language Models to generate domain-adapted training data. This work was **published** at the IEEE International Conference on Image Processing (**ICIP 2025**).
- Developed novel approach for **controlled synthetic-to-real image translation** by training textual embeddings to encode target domain styles, enabling realistic data generation for domains with limited real-world images.

Technical Projects

AI-Powered Conversational Analytics for Healthcare Reporting March 2025 - July 2025

- Eliminated SQL dependency for non-technical stakeholders across Finance, Operations, and Clinical teams by building a conversational analytics platform integrating MCP servers with the enterprise BI tool Looker.
- Configured Explore Assistant within Looker to enable natural language querying, and developed AI agent instruction libraries to standardize access to healthcare metrics.

Provider Utilization & Description Generation Pipeline Jan 2026 - Feb 2026

- Automated provider description generation for member-facing healthcare portal by building end-to-end pipeline integrating CMS public data, commercial claims, and multi-network provider rosters to compute utilization statistics with peer percentile rankings at scale.

Claims Reconciliation & Payment Reporting Pipeline June 2025 - Present

- Unified claims payment reporting across multiple third-party administrators by building automated reconciliation pipeline integrating TPA and internal claims data into consolidated financial views supporting payment decisions and audit compliance.

Cardiac Function Prediction using Deep Learning Jan 2023 - Apr 2023

- Achieved 93% AUC classifying at-risk cardiac patients by building video vision transformer analyzing 10,031 echocardiogram videos to predict Ejection Fraction using DeepLabv3 segmentation and ViViT architecture.

Education

Johns Hopkins University, Baltimore, MD — Aug 2022 - May 2024
Master of Science in Electrical and Computer Engineering — GPA: 3.86/4
Thesis: Diffuse2Adapt: Controlled Diffusion for Synthetic-to-Real Domain Adaptation | Advisor: Prof. Rama Chellappa

Manipal Institute of Technology, Manipal, India — Jul 2018 - Jun 2022
Bachelor of Technology in Electronics and Instrumentation Engineering — GPA: 3.83/4

Skills

Programming Languages: Python, SQL, PySpark, Bash

Cloud & Infrastructure: AWS - Glue, Lambda, Step Functions, RDS, S3, Redshift, SNS, EventBridge, SageMaker, Athena, CloudWatch, DynamoDB, CDK

Data Engineering & Modeling: PostgreSQL, Amazon Redshift, Star Schema, ETL/ELT

BI & Visualization: Looker, LookML, Tableau, AWS QuickSight, Plotly, Dash

Machine Learning & AI: PyTorch, Scikit-Learn, HuggingFace, TensorFlow, MCP Servers

Tools & Frameworks: GitHub Actions, Apache Airflow, Docker, Git, Jira, Xray, CI/CD, Confluence, Claude Code