AYUSH MALL

Seattle, WA | amall@uw.edu
LinkedIn | ayushmall.com

Data Scientist with 4+ years of experience designing data systems and driving actionable insights. Currently contributing to an agentic data platform start-up while pursuing an MS in Data Science at the University of Washington.

Skills & Tech-stack

Core Competencies: Data Engineering, ETL, Distributed Computing, Machine Learning, NLP, Agentic Systems, Computer Vision, Statistical Modeling, Data Warehousing, Cloud Data Infrastructure, Data Visualization
Languages and Frameworks: Python, SQL, BAML, Polars, PySpark, mQuery, Scikit-learn, TensorFlow, PyTorch
Platforms & Tools: Temporal.io, Databricks, DuckDB, Docker, Apache Iceberg, Huggingface Transformers, Microsoft Azure (ADF, ADLS), AWS, GCP, Tableau, Git, CICD, RESTful Web Services

Professional Experience

Machine Learning Engineer Intern — Corvic AI, Mountain View ***Jun 2025 – Present***

GenAI-native enterprise data-platform turning complex, multimodal data into explainable, actionable intelligence.

- **Automated end-to-end evaluation frameworks** for LLM-based information composition systems, implementing automated benchmarking pipelines with LLM-as-a-judge validation across multiple domain-specific datasets, enabling systematic quality measurement and model performance tracking.
- **Designed and delivered an end-to-end ingestion framework** integrating multi-cloud object storage (S3, Azure Blob, GCS) with incremental loads and delta management, **capable of processing 50M+** files in production.
- **Built a scalable LLM-augmentation feature** enabling million-row scale data enrichment via user prompts (e.g., identifying sustainability keywords and assigning ESG scores), unlocking diverse downstream use cases.

Data Engineer — Shell India Markets Pvt. Ltd, Bangalore ***Oct 2021 – Sept 2024***

Shipping and Maritime: Key contributor to project centralizing emission data management and analytics using the Azure tech stack, automating reporting for compliance, and enabling sustainable decision-making through PowerBI.

- **Engineered and deployed data pipelines** using ADF, Databricks, ADLS, and SQL to integrate 5 diverse sources, collaborating with 8 cross-functional teams to automate emission reporting, achieving **97.85% reduction** in turnaround and **$80,000** in annual operational savings.
- **Designed a DBSCAN-based clustering framework** utilizing metrics like AER and EEOI to group voyages into efficiency tiers, enabling ranked emission-performance profiling.
- **Developed an ADF solution template** in an individual initiative submitted to the organization's Continuous Improvement (CI) program, saving ∼**$65,000 annually** in FTE costs from a single data solution use case.
- **Articulated portfolio objectives and strategic goals** to a C-suite executive by presenting a business value review, securing resource allocation that **supported 4 key initiatives** throughout the fiscal year.
- **Led Knowledge Sharing Team** of the Logistics Portfolio, disseminating domain knowledge across 18 natural teams.

Co-founder — Aegis Soft Solutions Pvt. Ltd, Ahmedabad ***May 2020 – Aug 2021***

Co-founded a technology consulting firm delivering digitization and analytics-driven business solutions for small and mid-sized enterprises across multiple industries.

- **Architected and deployed a unified business analytics platform** integrating automation, data processing, and visualization layers to streamline operational workflows.
- **Digitized key supply-chain processes** in the textile domain, **improving raw-to-product turnaround by 27%** and enabling faster, data-driven decision-making through real-time KPI tracking.

Education

Master's in Data Science (MSDS) — University of Washington, Seattle ***Sept 2024 – Mar 2026***

Coursework focused on statistical modeling, experiment design, and ML | GPA: 3.85/4.0

Bachelor of Technology — Amity University, Mumbai ***Mar 2017 – Jun 2021***

Major: Computer Science & Engineering | Minor: Photography | GPA: 9.26/10.0

Research & Projects

PET Reconstruction from MRI Using CNN and Diffusion Frameworks* ***Jan 2025 – Apr 2025***

Conducting research to synthesize PET images from MRI scans using CNNs and diffusion-based stochastic models. Aims to offer non-invasive, cost-effective alternatives to PET imaging for use in neurology, oncology, and precision medicine.

Deepfake Face Swapping Detection Using Generative Adversarial Networks* ***Published Feb 2024***

Investigated GAN-based detection methods for identifying deepfake manipulations in visual media. Implemented adversarial training framework using CelebA dataset. Published in IEEE Conference Proceedings (DOI: 10.1109).

Underwater Acoustic Monitoring for Marine Conservation (Orcasound Project)* ***Apr 2025 – Jun 2025***

Collaborated with the eScience Institute to build scalable data pipelines for marine noise monitoring in the Puget Sound. Project supports marine conservation efforts and long-term ecological research through ambient sound analysis.