

Ishika Johari

ishijohari@gmail.com | (510)-284-9015 | Seattle, WA, 98105 | [linkedin.com/in/ijohari](https://www.linkedin.com/in/ijohari) | [ishijo.github.io](https://github.com/ishijohari)

EDUCATION

MS in Information Management | University of Washington | GPA: 3.9/4.0

Seattle, WA | Sep 2024 - Jun 2026

Specializations: Data Science, Artificial Intelligence, Business Intelligence

B. Tech in Computer Science Engineering | Manipal University Jaipur | CGPA: 8.5/10

Jaipur, India | Jul 2018 - Jun 2022

PROFESSIONAL EXPERIENCE

AI Research Scholar | eScience Institute, UW

Seattle, WA | Sep 2025 - Jun 2026

- Engineered an LLM-powered ADA accessibility remediation pipeline for UW research PDFs combining IBM Docling extraction, pikepdf structural rewriting, and a swappable multi-model backend (Claude, Hugging Face-hosted Gemma and Olmo) via unified API gateway, automating WCAG 2.1 AA fixes and cutting manual remediation effort **~70%** with human-in-the-loop approval.
- Built an open-source Python library providing scalable lazy-loading access to Caltech's **96TB** radio astronomy archive by implementing a unified API over S3 and local backends with Dask-backed chunked computation, enabling **~10x** faster exploratory analysis for Caltech and Rice researchers vs. full-download workflows.
- Shipped two open-source Claude Code agent plugins for UW's research software engineering toolchain, automating cloud-native data workflow decisions: one benchmarks Zarr chunk-size tradeoffs across read/write/compute patterns; the other standardizes Zarr/Xarray compression and S3 upload.

Data Science Intern | UW IT

Seattle, WA | Jun 2025 - Sep 2025

- Built a multi-year revenue forecasting system for UW's financial planning team by engineering hierarchical temporal features over SQL Server data and training LSTM models in PyTorch, achieving **~12%** MAPE on held-out periods and informing FY26 budget decisions.

Data Scientist 2 | Dell Technologies, Experimentation Team

Bengaluru, India | Feb 2024 - Aug 2024

- Lifted conversion rates **15%** by designing a behavioral and transactional K-means customer segmentation pipeline in PySpark, then validating segment-targeted campaigns against generic baselines via controlled A/B tests with two-proportion z-tests.
- Built an implicit-feedback ALS collaborative filtering recommender on sparse behavioral signals (clicks, cart adds, purchases) and ran a controlled A/B test on **10-15%** of live traffic, delivering **20%** CTR lift ($p < 0.10$) over the production rule-based baseline.
- Cut manual reporting overhead **30%** by automating ETL pipelines across multiple datasets via GitLab CI/CD, with monitoring and alerting that maintained **97%** extraction reliability across daily runs.

Data Scientist 1 | Dell Technologies, Experimentation Team

Bengaluru, India | Aug 2022 - Jan 2024

- Engineered a production-deployed end-to-end self-serve NLP chatbot with custom NER, **96%** intent classification accuracy, and dynamic response generation over Dell's **200+** annual A/B test repository, cutting experimentation-insight retrieval time **45%** and replacing manual documentation search; **patented Aug 2024** and adopted across the experimentation org.
- Trained one-vs-all LSTM classifiers in PyTorch on sequential hit-level clickstream data for per-category purchase propensity scoring across Dell's 5 laptop lines, achieving **88%** per-class accuracy and generating A/B test hypotheses for targeted marketing campaigns.
- Simulated Bayesian inference and Thompson Sampling Multi-Armed Bandit strategies on historical A/B test data, demonstrating **30%** test-duration reduction at equivalent statistical power, and authored a proposal driving org-wide adaptive experimentation adoption.
- Built an automated balanced-assignment clustering tool for marketplace A/B test groups, saving **300+** engineering hours annually and standardizing randomization across the experimentation team.

Data Analyst Intern | Dell Technologies, Data Governance Team

Bengaluru, India | Jun 2021 - May 2022

- Resolved **50%** null inconsistencies and drove **25%** data completeness for downstream ETL via a Power BI data-quality dashboard; eliminated **15%** data inconsistencies via data catalog mapping and lineage gap analysis in Manta, Collibra for GDPR-aligned governance.

SKILLS

Programming Languages: Python, SQL, R

Statistical Methods & Experimentation: A/B Testing, Multi-Armed Bandits, Bayesian Inference, Causal Inference, Hypothesis Testing, Regression Analysis, Power Analysis, Statistical Modeling, Probability Theory

Machine Learning & AI: PyTorch, TensorFlow, Scikit-learn, Hugging Face, LangChain, LangGraph, LSTMs, Transformers, BERT, SBERT Embeddings, spaCy, LLM Fine-tuning & Evaluation, RAG, ChromaDB, Multi-Agent Systems, MCP, Prompt Engineering, K-means Clustering, ALS Collaborative Filtering, Thompson Sampling, Model Validation

Data Engineering & Cloud: AWS (S3), Azure, Dask, Xarray, Zarr, Spark, Hadoop, FastAPI, GitLab CI/CD, Docker, Git, SQL Server, Teradata

BI & Analytics: Power BI, Tableau, Adobe Analytics, Jupyter, Large-Scale Telemetry Analysis, Clickstream Analytics, User Engagement Metrics

PROJECTS

- Experimentation Intelligence Agent** - Agentic RAG over historical A/B test corpora using multi-node LangGraph with ChromaDB embeddings, bias detection for underpowered tests and segment imbalance, served via FastAPI on Hugging Face Spaces with MCP tools.
- Google Vanir (Capstone)** - Extended Google's security patch detection tool to Python via Tree-Sitter CST integration; validated on CVE/fix pairs from OSV (**26/26** tests), expanding Android OEM patch-compliance coverage. **workshop paper** in progress.
- Cross-content Recommendation System** - Cross-domain recommender matching **18K+** books to movies via contrastive learning (InfoNCE), Hugging Face SBERT embeddings, and LLM genre mapping; **99.8%** Precision@10.

ACHIEVEMENTS

Game Changer and President's Award, Dell Technologies - Recognized by SVP Greg Bowen for self-serve chatbot impact

Apr 2024

Kaggle Datasets Expert - Top **2.6%** globally (391/15,111); Notebooks & Datasets Expert

May 2023

Guest Speaker & Mentor: Mentored **100+** students at Manipal Kagglethon; keynoted Kaggle Days 2023 on A/B testing

Jan 2026