Aryan Ballani

Aspiring Data Engineering Intern |Full Stack, ML, Data pipelines, and Backend focus | UBC CS + Stats

aryanballani@gmail.com | aryanballani.github.io | linkedin.com/aryanballani | +1 (236) 997-3533

Technical Skills

Languages	: Python, TypeScript , C, C++, Java, SQL
ML & Data Libraries	: PyTorch, Scikit-learn, NumPy, Pandas, NLTK, XGBoost, Matplotlib
Frameworks & APIs	: Flask, FastAPI, React, Next.js, REST, GraphQL, WebSockets
Cloud & Infrastructure	: AWS (S3, Lambda, Bedrock), Azure, Docker, Kubernetes, Terraform, Docker
Databases & Data Systems	: PostgreSQL, MongoDB, DynamoDB, Redis, Apache Spark
Automation & CI/CD	: GitHub Actions, Jenkins, Azure Pipelines, CI/CD Pipelines
Concepts	: OOPs, Microservices, TDD, BDD, Async Programming, Distributed Systems

Work Experience

Machine Learning Engineer | *Railtown AI Inc., Vancouver*, BC *Jan 2025 – Present*

- **Co-Architected and open-sourced Railtracks:** a framework for building complex **tool-calling agentic AI systems and pipelines** in a few lines of code with visualization; **30+ stars** and adopted by **20+ developers company-wide**.
- **Testing Leadership:** Independently refactored and wrote **600+ unit tests**; mocked various objects and improving **testing efficiency and latency by 300%** (5 mins → 1.6 mins).
- **Scalable APIs:** Engineered **20+ RESTful API endpoints** with async execution and **multi-model backend support**, following **DRY principles** and modular design patterns.
- **System Migration:** Led complete backend migration from **LlamaIndex to LiteLLM**, achieving **loose coupling** and completing the transition with **zero downtime**.
- **CI/CD & Deployment:** Established CI/CD workflows with **92% code coverage**, reducing deployment cycles by **35%**; **Dockerized benchmarking pipeline** for standardized evaluation using SWE-bench.
- **Library Distribution:** Packaged and deployed the Railtracks library to **Azure Artifacts** with GitHub Actions, enabling seamless internal distribution and versioning across **5+ microservices**.

Software Engineer in Test | *BC Liquor Distribution Branch*, Burnaby, BC *May 2024 – Dec 2024*

- Automated **80%** of manual test cases using Selenium and **Oracle SQL,** raising QA throughput by **25%**.
- Enhanced application performance by **12%** through profiling and optimizing **30+ database queries** and integrating **100+ unit tests**.
- Spearheaded CI/CD integration with Azure DevOps, reducing release cycle time by **30%** across **5 QA modules**.

Relevant Projects

UBC Lens | *python, JavaScript, BERT, NLTK, Vue.js*| link *Nov 2024*

- **End-to-End NLP Pipeline**: Built an ETL workflow to extract and clean 5,000+ Canvas threads via REST APIs, enabling sentiment and topic modeling using BERT and keyword extraction.
- **Data Cleaning & Preprocessing:** Cleaned raw thread data using regex, NLTK, and custom filters to remove noise, null entries, and sensitive information.
- **Smart Course Search Engine:** Built a Vue.js frontend with real-time search, dynamic keyword filtering, and interactive sentiment charts.
- **Secure Offline Deployment:** Deployed a privacy-preserving chatbot using LLaMA 3.2 via Ollama, anonymizing inputs to support course ranking and ensure secure NLP interaction.

WasteNet | *React Native, python, flask, AWS bedrock, MongoDB*| link *Oct 2024*

- **GenAI Recipe Engine:** Integrated AWS Bedrock with RAG to generate 100+ personalized recipes, powering LLM responses via secure Flask endpoints and MongoDB user context.
- **Scalable API & Data Layer:** Designed 10+ RESTful APIs, modeled NoSQL schemas for efficient AI-driven filtering.

- **Cloud-Backed Architecture:** Integrated MongoDB for structured ingredient storage and implemented scalable backend services for recipe generation and recommendation.

Dep2Phrase | *NLP, python, NLTK trees, Penn Tree Bank, LLMs, RoBERTa, MLP classifier* *Sept 2024*

- **Syntactic Chunking & Mapping:** Developed robust methods for both constituency and dependency chunking to assess syntactic correlations, engineering rule-based logic to map trees.
- **Performance Optimization:** Achieved 88% precision and 89% recall by refining rule-extraction algorithms and reverse-engineering binarization strategies to improve constituency tree accuracy.
- **Automation & Scalability:** Integrated RoBERTa embeddings with MLP classifiers to automate preprocessing pipelines and significantly scale up chunking operations across datasets.
- **Research & Collaboration:** Collaborated with a professor to align implementation with linguistic theory, performing comparative evaluations to validate model performance against theoretical benchmarks.

Campus Capture | *TypeScript, JavaScript, HTML, CSS* | link *Jan 2024*

- **Crash-Resilient:** Developed a robust backend for querying and managing 20K+ datasets using async patterns, delivering <500ms response times and rejecting 100% of malformed inputs.
- **System Reliability & Testing:** Designed comprehensive fail-safes and implemented test-driven logic across 50+ edge cases, ensuring continuous uptime and backend stability.
- **Test-Driven Development:** Verified 100+ test cases across dataset uploads, query structures, and edge case errors following software engineering best practices.
- **RESTful API Design:** Developed modular RESTful endpoints for dataset upload, query execution, and error reporting, enabling scalable interaction across 100+ API calls in testing.

NYC Airbnb Regression Analysis | *R, leaps, regressio3n, Model Selection, ggplot2* | link *Jan 2024*

- **Exploratory Data Analysis (EDA):** Analyzed 5,000+ NYC Airbnb listings using ggpairs, histograms, and transformations to normalize distributions and reveal behavioral patterns.
- **Feature Engineering:** Created *minimum_price* from *price* × *minimum_nights* and removed multicollinearity and outliers to enhance model reliability and reduce variance.
- **Model Building & Selection:** Trained multiple regression models with Mallow's Cp, adjusted R^2, and AIC; selected final model with $R^2 = 0.676$ and RMSE = 2.64 using key interaction terms.
- **Model Validation:** Applied 2-fold cross-validation and residual diagnostics to validate model assumptions and minimize overfitting for better generalization.

LEADERSHIP

Undergraduate Teaching Assistant | UBC Computer Science Dept., Vancouver, BC *Jan 2024 – Present*

- Led labs for **50+ students weekly** and answered **200+ questions** on Piazza over the term, earning praise as a `Piazza beast` for daily engagement and exceptional clarity in technical communication.

Hackathon Mentor | *BCS Hacks 2025, UBC, Vancouver, BC* *May 2025*

- **Mentored 100+ student participants** by providing guidance on project ideation, implementation, and debugging across a wide range of technologies and domains.

EDUCATION

BSc. (Computer Science and Statistics), University of British Columbia *Sep 2021 – Apr 2026*

- Dean's Honor List sessional standing for 3 consecutive years: 4.0/4.33 GPA
- Awarded **the International Student Scholarship** for outstanding community presence and academic excellence.