

# Md Mohaiminul Islam

📍 Edmonton, AB, Canada | 🌟 Permanent Resident | 📞 +1 825-888-4849 | ✉️ mohaiminul2810@gmail.com | in linkedin.com/in/md-mohaiminul-islam

## PROFILE

Senior Data Scientist with 8+ years of applied ML/NLP experience building production AI systems for regulated claims and healthcare-adjacent workflows. Recent work includes 70M-document PII detection, employee-facing RAG application, LLM/RAG evaluation, Databricks-based release controls, model monitoring, and AI roadmap contribution. Combines hands-on ML engineering with stakeholder communication, mentorship, and hiring support.

## SELECTED IMPACT

|                                               |                                                      |                                                   |                                                                 |
|-----------------------------------------------|------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------|
| <b>70M+</b><br>historical documents processed | <b>~5K/day</b><br>new documents scored in production | <b>~91% F1</b><br>PII detection across 14 classes | <b>~40% to ~10%</b><br>form-return drop indicated in evaluation |
|-----------------------------------------------|------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------|

## PROFESSIONAL EXPERIENCE

**Senior Data Scientist** Aug 2023 - Present

Workers' Compensation Board (WCB) | Edmonton, AB

- Designed and operationalized end-to-end PII detection for a roughly 70M-document historical claims corpus and ~5K new documents/day, combining prompt-routed LLM inference, deterministic validation, post-processing, and reviewable evidence; achieved ~91% F1 across 14 PII/entity classes on human-reviewed holdout labels.
- Established governed release controls on Azure Databricks Jobs with MLflow/Unity Catalog model registry, audit logging, release gates, token/cost tracking, and developer alerts.
- Built an agentic maintenance harness using Databricks Model Serving LLM endpoints (Claude) that monitors daily production jobs, diagnoses failures, validates fixes on sandboxed data samples, and stages deploy-ready changes for human approval; extended to automated code review.
- Built a Helpdesk Virtual Agent as a Databricks App using Databricks Vector Search and Model Serving LLM endpoints for employee support knowledge retrieval; a two-week helpdesk-team trial found ~70% of answers helpful, and it is approved for company-wide rollout to ~3,500 employees.
- Improved RAG quality with HyDE, multi-query retrieval, reranking, validated-answer capture, human-verified eval sets, LLM-as-judge scoring, and prompt/retrieval error analysis.
- Prototyped internal medical report-generation agents supporting extraction, validation, pre-fill recommendations, routing, human review, and audit tracking; evaluation on historical data indicates return-for-modification can drop from ~40% to ~10%, now under departmental review.
- Put a 3-class XGBoost claim-duration model into production after benchmarking NLP/Transformer alternatives for accuracy, speed, and cost; trained on ~30K/tested on ~10K records, now scoring ~1K claims/day with weekly monitoring/retraining alerts.
- Led applied-AI R&D and POC initiatives for regulated claim workflows, spanning use-case scoping, hands-on prototyping, evaluation, and coordination of assigned contributors; presented recommendations to CTO and director-level stakeholders.
- Acted as the ML point person for enterprise AI adoption; contributed to AI roadmap and release-gate decisions, managed 2 direct-report interns, mentored 5 junior data scientists during project delivery, and supported 5 hiring/interview processes.

**Senior Data Scientist** May 2023 - Jul 2023

Quantolio | Montreal, QC

- Prototyped a client-oriented Vision Transformer chest X-ray classifier using public medical imaging datasets; achieved ~96% accuracy and presented internally for handoff.
- Refactored a Python ML codebase and developed a Streamlit-based portfolio analytics prototype, including time-series analysis and reinforcement-learning experiments for portfolio optimization.

**Lead Data Scientist, ML/NLP** Feb 2023 - May 2023

Blue Guardian Canada Inc. | Toronto, ON

- Managed 4 direct-report data science interns to deliver a 29-class mental-health sentiment classifier using Transformers and LLM-generated synthetic text; achieved ~92% F1.
- Set up an AWS Lambda-based workflow for LLM data generation, validation, and annotation, avoiding an estimated \$25K in manual labeling cost.

CORE EXPERTISE

|                                                                                                                                       |                                                                                                                                                                                       |                                                                                                                                                         |                                                                                                                                                                        |
|---------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| APPLIED AI / GENAI                                                                                                                    | ML & MODEL QUALITY                                                                                                                                                                    | DATA / MLOPS / CLOUD                                                                                                                                    | AI ENGINEERING & DELIVERY                                                                                                                                              |
| LLMs, RAG, AI agents<br>LlamaIndex, LangChain<br>Databricks Apps, Vector Search<br>HyDE, multi-query, reranking<br>Hugging Face, vLLM | XGBoost, PyTorch,<br>Transformers (BERT)<br>clustering, unsupervised ML<br>reinforcement learning<br>LLM-as-judge, holdout evals<br>human-verified validation<br>error analysis, SHAP | Python, SQL, PySpark<br>Databricks Jobs, MLflow<br>Unity Catalog registry<br>Docker, model monitoring<br>retraining alerts<br>Azure Databricks, AWS/GCP | FastAPI, model serving<br>release gates, audit logs<br>token/cost tracking<br>AI roadmap planning<br>executive technical<br>communication<br>mentoring, hiring support |

EARLIER APPLIED ML EXPERIENCE

|                                                                                                                                                                                                                                                                                                                                                                                                    |                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| Applied ML Research Assistant                                                                                                                                                                                                                                                                                                                                                                      | Sep 2015 - Dec 2023 |
| University of Manitoba   Winnipeg, MB                                                                                                                                                                                                                                                                                                                                                              |                     |
| <ul style="list-style-type: none"><li>Modeled high-dimensional prediction with a privacy-preserving deep learning approach; achieved AUROC 0.94 on ~11K samples.</li><li>Created a PyTorch GNN model for protein-ligand binding scoring; achieved RMSE 1.24.</li><li>Applied XGBoost and SHAP to high-dimensional proteomics data for interpretable classification; achieved AUROC 0.83.</li></ul> |                     |
| Data Scientist Intern, Mitacs                                                                                                                                                                                                                                                                                                                                                                      | Jul 2021 - Jun 2022 |
| Servier Canada Inc.   Laval, QC                                                                                                                                                                                                                                                                                                                                                                    |                     |
| <ul style="list-style-type: none"><li>Automated config-driven preprocessing and embedding pipelines for ~2M SMILES molecular strings using PySpark, Pandas, and NumPy, reducing model-preparation time by ~30%.</li><li>Trained a Reformer-based generative sequence model, producing ~98% valid outputs and ~85% desired-property rate.</li></ul>                                                 |                     |
| Data Scientist Intern                                                                                                                                                                                                                                                                                                                                                                              | May 2019 - Sep 2019 |
| Sightline Innovation Inc.   Winnipeg, MB                                                                                                                                                                                                                                                                                                                                                           |                     |
| <ul style="list-style-type: none"><li>Cleaned and preprocessed ~14K graph relationships and risk-gene records in R, then implemented a maximum-flow candidate-prioritization workflow for downstream analysis.</li></ul>                                                                                                                                                                           |                     |

ADDITIONAL EXPERIENCE

|                                                                                                                                                                                                     |                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| Lecturer, Computer Science                                                                                                                                                                          | May 2014 - Jul 2015 |
| Green University of Bangladesh   Dhaka, Bangladesh                                                                                                                                                  |                     |
| <ul style="list-style-type: none"><li>Taught 5 undergraduate computer science courses to 500+ students and coordinated 2 departmental semesters across 10+ courses and faculty schedules.</li></ul> |                     |

EDUCATION

|                                                                                       |             |
|---------------------------------------------------------------------------------------|-------------|
| PhD Candidate, Computer Science (candidacy completed; left for industry)              | 2018 - 2023 |
| University of Manitoba, Canada   University of Manitoba Graduate Fellowship           |             |
| M.Sc., Computer Science                                                               | 2017        |
| University of Manitoba, Canada   Manitoba Graduate Scholarship                        |             |
| B.Sc., Computer Science and Engineering                                               | 2014        |
| University of Chittagong, Bangladesh   Ranked 1st in class; Prime Minister Gold Medal |             |

SELECTED RECOGNITION

|              |                                                                                                                                                                               |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Awards       | Prime Minister Gold Medal; VADA Big Data Challenge Winner; Manitoba Graduate Scholarship; University of Manitoba Graduate Fellowship; Research Completion Award.              |
| Publications | 12 applied AI/ML publications with 200+ citations across deep learning for omics, privacy-preserving ML, and drug discovery, including a first-author paper cited 100+ times. |