Mohammed Saidul Islam

saidulislam143.si@gmail.com LinkedIn Github Google Scholar saidul-islam98.github.io

SUMMARY

Applied ML Specialist with **3+ years** delivering end-to-end ML systems across research and production settings, from **data preprocessing and large-scale pipeline development** to **model training**, **optimized inference**, and **deployment-ready** implementations. Strong **Python** engineer with experience building **robust, modular** ML software, profiling and optimizing performance (e.g., **vLLM** batched decoding on A100/H100), and maintaining reproducible workflows on Linux/HPC. Proven track record partnering with researchers and technical stakeholders, producing thorough **documentation** and clear design decisions, and iterating through experimentation and error analysis to improve model and system quality. Publications at **ACL, EMNLP, LREC-COLING, EACL**.

CORE SKILLS

ML Software Engineering: Modular Python systems, scalable ML pipelines, data preprocessing/QA, performance profiling and optimization, reproducibility on Linux/HPC
End-to-End ML Lifecycle: Problem framing, data exploration, experimentation, model evaluation, iterative improvements, production-oriented inference pipelines
Generative AI Systems: LLM/VLM workflows, agentic planning/reflection, tool-augmented generation, fine-tuning (SFT) and post-training (DPO, GRPO), human-in-the-loop evaluation
Evaluation & Reliability: Robustness testing, error analysis, regression-style comparisons, quality standards and failure-mode tracking
Data & Pipelines: Large-scale dataset construction, preprocessing, batch processing, automated quality checks, versioned artifacts
Collaboration & Documentation: Partnering with research/product teams, stakeholder updates and technical advisory, documenting workflows and design decisions

TECHNICAL SKILLS

Languages: Python, JavaScript, SQL, C++, C, Java, HTML, CSS

ML/GenAI Stack: PyTorch, Hugging Face Transformers, Scikit-Learn, NumPy, Pandas, DeepSpeed, vLLM, LLaMA-cpp

Agent Frameworks: LangChain, LangGraph (agent workflows, tool-augmented generation)

Post-training & Evaluation: GRPO, DPO, reward modeling (LLM/VLM-as-judge), post-execution evaluation, robustness/error analysis

Engineering/Infra: Git (source control), Linux, Slurm, GCP Vertex AI

EXPERIENCE

Vector Institute – Toronto Sep 2025 – Present
Associate Applied Machine Learning Specialist

- Built an automated benchmark-generation tool that leverages **external knowledge sources** such as textbooks and technical references to create grounded evaluation tasks for foundation models.
- Developed an automated **OCR-based text extraction pipeline** to convert reference PDFs into structured inputs for downstream generation.
- Designed a **multi-agent LLM pipeline** with separate **designer** and **verifier** agents for task generation, iterative repair, trace-aware validation, and final quality control.
- Implemented **prompt-level anti-duplication** and **embedding-based deduplication** using semantic similarity filtering to remove near-duplicate samples.
- Owned an end-to-end **data extraction and preprocessing pipeline** for large-scale medical imaging figures (compound figures, subfigures, captions, embedded subcaptions), producing training-ready artifacts for downstream ML.
- Implemented **subcaption extraction** with **Qwen2.5-VL-32B-Instruct** and optimized GPU inference using **batched decoding in vLLM**, reducing latency from **4.0s/sample** to **1.2s/sample** on A100s (**70% faster**).
- Developed a **summary generation pipeline** using image context, full captions, and extracted subcaptions with **Qwen2.5-14B**; improved generation speed from **4.0s/sample** to **0.6s/sample** via **vLLM batch generation** (**85% speedup**).
- Established **quality checks** (spot checks, error categorization, iterative prompt/workflow refinement) and documented pipeline assumptions and design decisions to support troubleshooting and maintenance.
- Leveraged extracted pairs to support training of **OpenCLIP** on **18M image–text pairs**, strengthening multimodal alignment for medical imaging research.

Intelligent Visualization Lab, York University – Toronto Sep 2023 – Aug 2025
Graduate Research Assistant

- Built **robust, reusable** research-to-production style pipelines for multimodal ML tasks, including dataset construction, preprocessing, experimentation, and standardized evaluation with thorough documentation.
- Designed and shipped **agentic LLM workflows** for complex generation tasks, including an **Actor–Critic** framework that operationalizes planning, reflection, narration, and revision using modern LLM APIs (e.g., **GPT-4o**, **Gemini**, **LLaMA**); accepted at **EMNLP 2024**.
- Built **DataNarrative**, a benchmark of **1,449** real-world data stories, enabling evaluation of **content generation** and multi-step reasoning; implemented preprocessing and quality control to ensure consistent data artifacts.
- Proposed **RL-Text2Vis**, a GRPO-based post-training framework for **text-to-visualization** that uses **post-execution multimodal feedback** (text correctness, code executability, visual clarity/correctness) via a **multi-objective reward**, improving code executability and chart quality over strong prompting baselines (**EACL 2026 Main**).
- Created **DashboardQA**, a benchmark for **interactive dashboard question answering** requiring **multi-step GUI interactions** (filters, tabs, view switches) and coordinated multi-view reasoning; evaluated agent behavior and documented failure modes in grounding, planning, and visual reasoning (**EACL 2026 Findings**).
- Created **ChartQAPro** (**1,300+** charts, **1,900+** QA pairs) and evaluated **21** open/closed-source VLMs; demonstrated robustness gaps that motivate better **reliability** and **regression testing** (**ACL 2025 Findings**).
- Trained a **2B** judge model that matched the performance of a **7B** baseline while significantly **reducing inference latency**; additionally benchmarked compact **VLMs as judge/evaluator models** in both point-wise and pair-wise protocols (bias, length/position effects), reinforcing practical, deployment-ready **LLM evaluation** workflows (**ACL & EMNLP 2025 Industry Track**).

Islamic University of Technology – Bangladesh Jul 2021 – Aug 2023
Lecturer

- Designed and delivered curricula for theory and lab courses including **Structured Programming (C)**, **Object-Oriented Programming (C++)**, and **System Analysis and Design**.
- Mentored students on building reliable software artifacts and debugging complex programming assignments, strengthening communication and technical leadership skills.

EDUCATION

York University Sep 2023 – Aug 2025
MSc in Computer Science, CGPA: 8.6/9.0

Islamic University of Technology Jan 2016 – Dec 2019
BSc in Computer Science and Engineering, CGPA: 3.77/4.0

SELECTED PROJECTS

Paper Reproducibility Challenge
- Reproduced results for contrastive learning research; strengthened skills in experiment design, debugging training pipelines, and reporting discrepancies with clear written documentation.

Co-Author Network Analysis
- Analyzed academic/industry co-authorship trends and research topics; applied data processing and statistical analysis to derive interpretable insights.

SELECTED PUBLICATIONS

- *Mizanur Rahman,* ***Mohammed Saidul Islam*** *et al.* "RL-Text2Vis: Aligning Text, Code, and Visualizations with GRPO and Post-Execution Multimodal Feedback." **EACL 2026 (Main)**.
- *Aaryaman Kartha, Ahmed Masry,* ***Mohammed Saidul Islam*** *et al.* "DashboardQA: Benchmarking Multimodal Agents for Question Answering on Interactive Dashboards." **EACL 2026 (Findings)**.
- ***Mohammed Saidul Islam*** *et al.* "DataNarrative: Automated Data-Driven Storytelling with Visualizations and Texts." **EMNLP 2024 (Main)**.
- *Ahmed Masry*,* ***Mohammed Saidul Islam***** et al.* "ChartQAPro: A More Diverse and Challenging Benchmark for Chart Question Answering." **ACL 2025 (Findings)**.
- *Md Tahmid Rahman Laskar,* ***Mohammed Saidul Islam****, et al.* "Deploying Tiny LVLM Judges for Real-World Evaluation of Chart Models: Lessons Learned and Best Practices." **EMNLP 2025 (Industry Track)**
- *Md Tahmid Rahman Laskar,* ***Mohammed Saidul Islam****, et al.* "Judging the Judges: Can Large Vision-Language Models Fairly Evaluate Chart Comprehension and Reasoning?." **ACL 2025 (Industry Track)**

LEADERSHIP & SERVICE

Bangladeshi Graduate Student Association at York Aug 2024 – Aug 2025
President

Peer Reviewer Jan 2023 – Present
Neurips, ACL, EMNLP, COLM, NAACL, ACL ARR