MUSFIQUR RAHMAN

PhD Candidate in Software Engineering

Address: 2104-131 Torresdale Ave, North York, ON M2R 3T1 **Phone:** +1 (438) 870-0809

musfiqur.rahman@mail.concordia.ca | LinkedIn: musfiqur-rahman-37596444

GitHub: mrsumitbd | Twitter: @mrsumit2010

EDUCATION

PhD in Software Engineering *May 2022 – Expected Sep 2026*

Concordia University, Montréal, QC GPA: 4.30/4.30

Thesis: *Large Language Models in Coding: Generation, Detection, and Repair*

Advisor: Prof. Emad Shihab – Data-driven Analysis of Software (DAS) Lab

M.Sc. in Computer Science *Sep 2015 – Apr 2018*

Concordia University, Montréal, QC GPA: 3.90/4.30

Thesis: *Analyzing the Predictability of Source Code and its Application in Creating Parallel Corpora for English-to-Code Statistical Machine Translation*

Advisor: Prof. Peter C. Rigby – Concordia Empirical Software Engineering Lab (CESEL)

B.Sc. in Computer Science and Engineering *Jan 2011 – Dec 2014*

East West University, Dhaka, Bangladesh GPA: 3.62/4.00

RESEARCH INTERESTS

Large language models for software engineering (LLMs4SE) · Code generation, detection, and quality assurance · Code naturalness and predictability · Machine translation for software engineering · Empirical software engineering · AI/ML systems engineering

PUBLICATIONS

Conference Papers

1. **EASE 2025:** ***M. Rahman***, *S. H. Khatoonabadi, A. Abdellatif, et al. "The Impact of Environment Configurations on the Stability of AI-Enabled Systems." 29th International Conference on Evaluation and Assessment in Software Engineering, June 2025.*

2. **CAIN 2024:** *L. Barreto Simedo Pacheco,* ***M. Rahman****, F. Rabbi, et al. "DVC in Open Source ML-development: The Action and the Reaction." IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI, 2024.*

3. **HICSS 2022:** *Md. F. A. Bhuiyan,* ***M. Rahman****, Fairuza Laila, et al. "Tools and Techniques Adapted for Teaching Software Engineering Topics Remotely during the COVID-19 Pandemic." 55th Annual Hawaii International Conference on System Sciences, 2022.*

4. **ICSE 2019:** ***M. Rahman****, D. Palani, P. C. Rigby. "Natural Software Revisited." 41st IEEE/ACM International Conference on Software Engineering, 2019.*

5. **MSR 2019:** ***M. Rahman****, D. Palani, P. C. Rigby, et al. "Cleaning StackOverflow for Machine Translation." 16th IEEE/ACM International Conference on Mining Software Repositories (Data Showcase), 2019.*

6. **MIWAI 2014:** ***M. Rahman***, *S. Ripon. "Using Bayesian Networks to Model and Analyze Software Product Line Feature Model." International Workshop on Multi-disciplinary Trends in Artificial Intelligence, Springer, 2014.*

Preprints & Under Review

1. **arXiv 2025:** ***M. Rahman***, *S. H. Khatoonabadi, E. Shihab. "Beyond Synthetic Benchmarks: Evaluating LLM Performance on Real-World Class-Level Code Generation." arXiv:2510.26130, 2025. [Under Review - TOSEM]*

2. **arXiv 2025:** ***M. Rahman***, *S. H. Khatoonabadi, E. Shihab. "A Large-scale Class-level Benchmark Dataset for Code Generation with LLMs." arXiv:2504.15564, 2025. [Under Review - MSR 2026 Data track]*

3. **arXiv 2024:** ***M. Rahman***, *S. H. Khatoonabadi, A. Abdellatif, et al. "Automatic Detection of LLM-generated Code: A Case Study of Claude 3 Haiku." arXiv:2409.01382, 2024.*

Journal Papers

1. **IJST 2016:** *S. Ripon,* ***M. Rahman***, *J. Ferdous, et al. "Verification of SPL Feature Model by using Bayesian Network." Indian Journal of Science and Technology, Vol. 9, 2016.*

RESEARCH PROJECTS

CodeProbe: Universal Detection of AI-Generated Code *Sep 2025 – Present*
DAS Lab, Concordia University,

Primary contributions:

- Developed contamination-free benchmark dataset for training and evaluating AI-generated code detectors
- Built a robust detection system with high accuracy, stylometric and complexity features
- Designed a cross-language, cross-model detection pipeline

OpenClassGen: Large-Scale Class-Level Code Dataset *Apr 2025 – Nov 2025*
DAS Lab, Concordia University,

Primary contributions:

- Curated the largest open class-level code dataset with 400k+ real-world Python classes from GitHub
- Established a new baseline for class-level code generation research in the community

ConfigStability: Environment Configuration Impact on AI Systems *Sep 2022 – Dec 2024*
DAS Lab, Concordia University,

Primary contributions:

- Investigated variability and stability of AI-enabled software systems across different environment configurations
- Analyzed the impact of deployment environments on ML model performance and reliability
- Developed framework for assessing configuration-induced risks in production AI systems

NaturalCode: Code Predictability and English-to-Code Translation *Sep 2015 – Dec 2017*
CESEL Lab, Concordia University,

Primary contributions:

- Analyzed the predictability and naturalness of programming languages using NLP techniques
- Created parallel English-Code corpora from StackOverflow for statistical machine translation
- Developed data cleaning approaches and maximum likelihood models for code translation tasks

ACADEMIC SERVICE

Program Committees

Program Committee, Poster Track – CAIN 2026
Program Committee, ACM Student Research Competition – ICSE 2026
Junior Program Committee – TechDebt 2026
Junior Program Committee – MSR 2026
Junior Program Committee – MSR 2024

Reviewing

Referee – ACM Transactions on Software Engineering and Methodology (TOSEM), January 2025
Referee – ACM Transactions on Software Engineering and Methodology (TOSEM), September 2024
Reviewer – Montreal AI Symposium, 2022

PROFESSIONAL EXPERIENCE

Future Skills Board Member *Nov 2025 – Present*
General Assembly, Remote

Major responsibilities:

- Identify emerging skill gaps and workforce development opportunities across industries
- Guide curriculum development to ensure future-ready educational programs
- Explore innovative methodologies and technologies for skills acquisition and assessment

Subject Matter Expert (SME), Data and AI Domain *Feb 2024 – Present*
General Assembly, Remote

Major responsibilities:

- Design curriculum for data science and AI bootcamp programs
- Build lesson materials, labs, and theoretical frameworks for instructional delivery
- Perform quality assurance on developed educational content and hands-on exercises
- Train and mentor new instructors on pedagogical approaches and technical content

Lead Instructor, Data Science/Analytics Bootcamp *Feb 2020 – Dec 2025*
General Assembly, Remote

Major responsibilities:

- Delivered 50+ cohorts of the Data Science/Analytics Immersive program to diverse global audiences
- Mentored students in statistical machine learning, deep learning, and data engineering

Artificial Intelligence Engineer *Feb 2018 – Feb 2020*
Pentavere, Toronto, ON

Major responsibilities:

- Prototyped and productionized clinical NLP models processing millions of unstructured health records
- Developed end-to-end ML pipelines for healthcare data extraction and analysis

AWARDS & HONORS

- **Concordia University Doctoral Fellowship** (2022–2026) – Full PhD funding
- **Campaign for Concordia Graduate Award** (2022)
- **Graduate Student Support Program (GSSP) Fund** (2015)
- **Dean's List & Merit Scholarship** – East West University (2014)

TECHNICAL SKILLS

Languages: Python (expert), Java, C/C++, Scala, R, Bash
ML/DL Frameworks: PyTorch, Transformers, CatBoost, SHAP, Scikit-learn, Snorkel
Tools & Infrastructure: Git, Docker, GCP, AWS, PostgreSQL, BigQuery