SHADIKUR RAHMAN

+1 (416) 562-6670 ◇ North York, ON, Canada

shadikur@yorku.ca LinkedIn GitHub Γ Google Scholar Personal Website

OBJECTIVE

Machine Learning and NLP Researcher, and Adjunct Faculty with expertise in developing Code Generation LLM Agentic Frameworks for code quality improvement, information extraction, and multimodal reasoning. Skilled in fine-tuning, evaluating, and deploying scalable AI models that integrate language and vision for real-world decision support. Seeking opportunities in AI/ML, NLP, or Generative AI to advance reliable and interpretable model development.

EDUCATION

Master of Science in Computer Science Jan 2022 – Jan 2024
York University *Toronto, Canada*
GPA: 3.70/4.0
Thesis: Enhancing code review for improved code quality with language model-driven approaches.

Bachelor of Software Engineering Jan 2015 – Dec 2018
Daffodil International University (DIU) *Dhaka, Bangladesh*
GPA: 3.54/4.0
Thesis: Polynomial Topic Distribution with Topic Modeling for Generic Labeling.

EXPERIENCE

Adjunct Faculty of Information Technology May 2024 – Present
York University *Toronto, Canada*

- **ITEC 3020 - Web Technologies:** Web Technologies course covering HTML, CSS, Bootstrap, JavaScript, Node.js, and MySQL.
- **ITEC 3230 - Designing User Interfaces:** Focused on user-centered design, prototyping, and usability evaluation.
- Designing and delivering interactive lectures and hands-on coding sessions.
- Guiding students in developing real-world projects and improving their problem-solving skills.

Research Associate Nov 2024 – Present
Algoma University *Brampton, Canada*

- Developed **RefactorCoderQA**, a benchmark for multi-domain code reasoning with LLMs.
- Designed a **multi-agent LLM framework** (GuideLLM–SolverLLM–JudgeLLM) for structured reasoning.
- Fine-tuned **RefactorCoder-MoE** using QLoRA on DeepSeek-Coder-7B, achieving 76.84% accuracy.
- Built **JudgeLLM** using GPT-4 for automated accuracy and clarity evaluation.
- Performed **latency, ablation, and human–AI evaluation** studies across SE, DS, ML, and NLP domains.

Research Assistant Jan 2024 – Present
Intelligent Visualization Lab, York University *Toronto, Canada*

- Contributed to **ChartQAPro** and **DashboardQA** benchmarks for multimodal and GUI-based reasoning.
- Curated diverse real-world charts and interactive dashboards with expert-authored QA pairs.
- Developed evaluation pipelines for **VLMs and agentic models** using GPT-4o, Gemini, and Claude.
- Analyzed model grounding, planning, and reasoning gaps in multimodal LLM performance.
- Supported dataset design, experimentation, and benchmarking for vision–language understanding.

Graduate Research Assistant Jan 2022 – Jan 2024
York University *Toronto, Canada*

- Developed an application to identify and recommend similar code reviews to improve code quality.
- Fine-tuned a **BERT model** for code review classification, achieving a **96% F1-score**.
- Retrieved relevant Stack Overflow data using **chunking** and **NER** for contextual insights.
- Built a **chatbot with LLM reasoning** using the **Llama 2 model** to assist developers in issue resolution.

Teaching Assistant Jan 2022 – April 2024
York University *Toronto, Canada*

- Assisted in grading assignments, conducting lab sessions, and invigilating midterm and final examinations.
- Supported multiple undergraduate courses in Computer Science and Information Technology.

Research Assistant Jun 2020 – Jul 2021
Umm Al-Qura University *Saudi Arabia (Remote)*

- Implemented a CNN-based technique to classify COVID-19 from X-ray, comparing VGG and ResNet models.
- Proposed an optimized CNN model achieving **97% accuracy** in distinguishing COVID-19 cases from X-ray.

Software Engineer May 2019 – Nov 2019
Samsung R&D Institute Bangladesh Ltd. *Dhaka, Bangladesh*

- Analyzed code reviews from GitHub repositories to identify coding issues using **Machine Learning** and **NLP**.
- Applied **SVM with TF-IDF** for classifying ambiguous code reviews, achieving **90% accuracy** and an **F1-score of 92%**.
- Developed a developer-assistance tool using **Django** and **Django REST Framework** in Python.

AWARDS

Major CUPE 3903 Research Grant ($21,343) 2025–2026
York University
Funding for the project *"Optimizing Task-Specific Language Models for Multi-Domain Code Generation and Reasoning"*, supporting the next phase of the RefactorCoderQA benchmark and multi-agent LLM framework for reasoning-driven code generation.

CUPE 3903 Conference Travel Fund ($5,500) 2024–2025
York University
Awarded to present research at ACL, ICOA, and FLLM 2025 conferences in Vienna, Austria, showcasing work on Large Language Models, Code Generation, and Multimodal Reasoning.

DIU Research Award 2020 ($800) March 2020
Daffodil International University
Recognition for B.Sc. thesis published with Springer: *"Assessing the Effectiveness of Topic Modeling Algorithms in Discovering Generic Labels with Description" (FICC 2020, Vol. 2).*

PROJECTS

Optimizing Task-Specific Language Models for Multi-Domain Code Generation and Reasoning (RefactorCoderQA Pro) 2025 – 2026
Major CUPE 3903 Research Grant, York University

- Developed a large-scale benchmark and multi-agent framework (*Guide–Solver–Reviewer–Judge*) for reasoning-driven code generation.
- Evaluated LLMs across SE, DS, and NLP domains to study trade-offs in accuracy, clarity, and efficiency on cloud and edge environments.

Review2Code: Benchmarking LLM-Driven Code Generation from Code Review 2025 – 2026
Major CUPE 3903 Research Grant, York University

- Designed a pipeline to transform human review feedback into optimized code using instruction-tuned LLMs.
- Benchmarked models for actionable fix generation while maintaining logical and syntactic consistency.

UI2Code-Real: Bridging Visual Web Design and Front-End Code Generation 2025 – Present
York University

- Created a realistic design-to-code benchmark using 100+ student projects built with HTML, CSS, Bootstrap, and Node.js.
- Evaluated multimodal LLMs (GPT-4o, Claude 3.5, Gemini 1.5 Pro) for translating UI visuals into accurate front-end code.

SKILLS

Programming	Python, C, Java, JavaScript
ML / Data	scikit-learn, pandas, NumPy, Matplotlib, Weka, NLTK, Gensim
Deep Learning	TensorFlow, PyTorch, Keras, OpenCV
AI / NLP	Large Language Models (LLMs), LangChain, Ollama, Retrieval, NER, LLM reasoning
Backend	Django, Django REST Framework, Scrapy; SQL (PostgreSQL, MySQL)
Tools & Dev	Git, GitHub, Jira, Kanban; VLLM Framework, Quantization, Fine-tuning, Evaluation
Soft Skills	Teaching & Mentoring, Research, Communication, Collaboration

VOLUNTEER EXPERIENCE

Reviewer 2023 – Present
NAACL, ACL, ESEM, CSSE Conferences
Evaluated peer-reviewed submissions in NLP, Software Engineering, and Machine Learning, focusing on benchmark design, LLM reasoning, and applied research quality.

Supervision 2023 – Present
DIU Research Lab, Daffodil International University
Mentored student researchers on projects in code generation, benchmarking, and applied NLP; guided proposal review, experiment design, and academic writing.

Startup Collaboration 2024 – Present
TorontoRides Mobility Platform, Canada
Collaborating on an early-stage luxury ride service platform, contributing to service design, and web architecture

PUBLICATIONS

2025

1. **Rahman, S.**, Hameed, A., Srivastava, G., & Danish, S. M. (2025). *RefactorCoderQA: Benchmarking LLMs for Multi-Domain Coding Question Solutions in Cloud and Edge Deployment.* Submitted to Transactions on Services Computing, **Impact factor: 5.8**. *arXiv preprint* arXiv:2509.10436.[**Under Review**]

2. Masry, A., Islam, M. S., Ahmed, M., Bajaj, A., Kabir, F., Kartha, A., Laskar, M. T. R., Rahman, M., **Rahman, S.**, Shahmohammadi, M., Thakkar, M., Parvez, M. R., Hoque, E., & Joty, S. (2025). *ChartQAPro: A More Diverse and Challenging Benchmark for Chart Question Answering.* In *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 19123–19151). Vienna, Austria: Association for Computational Linguistics.

3. Ashraf, H., Danish, S. M., & Sattar, Z. (2025). *Toward Green Code: Prompting Small Language Models for Energy-Efficient Code Generation.*Accepted in Foundation and Large Language Models (FLLM2025). *arXiv preprint* arXiv:2509.09947.

4. Kartha, A., Masry, A., Islam, M. S., Lang, T., **Rahman, S.**, Mahbub, R., Rahman, M., et al. (2025). *DashboardQA: Benchmarking Multimodal Agents for Question Answering on Interactive Dashboards.*Submitted to **ACL Rolling Review**. *arXiv preprint* arXiv:2508.17398.[**Under Review**]

5. **Rahman, S.**, Shanto, H. K., Koana, U. A., & Danish, S. M. (2025). *Automated Research Article Classification and Recommendation Using NLP and ML.* Accepted in Foundation and Large Language Models (FLLM2025). *arXiv preprint* arXiv:2510.05495.

6. Shanto, H. K., Koana, U. A., & **Rahman, S.** (2025). *Multi-Armed Bandits-Based Optimization of Decision Trees. arXiv preprint* arXiv:2508.05957.

2024

1. Koana, U. A., Le, Q. H., **Rahman, S.**, et al. (2024). *Examining Ownership Models in Software Teams. Empirical Software Engineering, 29*(155). https://doi.org/10.1007/s10664-024-10538-5

2. **Rahman, S.** (2024). *Enhancing Code Review for Improved Code Quality with Language Model-Driven Approaches.* [**Thesis**]

3. **Rahman, S.**, Koana, U. A., & Hoque, E. (2024). *RefineCode: Enhancing Code Quality Through Actionable Code Review Recommendations and Intelligent Issue Resolution.* [**Submitted**]

2023

1. **Rahman, S.**, Ahmed, F., & Nayebi, M. (2023). *Mining Reddit Data to Elicit Students' Requirements During COVID-19 Pandemic.* In *IEEE 31st International Requirements Engineering Conference Workshops (REW).* IEEE.

2022

1. **Rahman, S.**, Koana, U. A., & Nayebi, M. (2022). *Example-Driven Code Review Explanation.* In *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM).*

2. Basalamah, A., & **Rahman, S.** (2022). *An Optimized CNN Model Architecture for Detecting Coronavirus (COVID-19) with X-Ray Images. Computer Systems Science & Engineering, 40*(1).

2021

1. **Rahman, S.**, et al. (2021). *Estimating the Effective Topics of Articles and Journals Abstract Using LDA and K-Means Clustering Algorithm.* In *Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020.* Springer International Publishing.

2020

1. Koana, U. A., **Rahman, S.**, Hasan, F., Ismael, A. M., & Hussein, K. (2020). *SW Release Challenges: A Case Study with Mitigation Plan Using Semi-Automated Process. Management, 29*(8), 4942–4949.

2. **Rahman, S.**, et al. (2020). *Measuring the Effectiveness of Code Review Comments in GitHub Repositories: A Machine Learning Approach. Machine Learning and Data Mining in Pattern Recognition, 16*, 35–48.

3. **Rahman, S.**, et al. (2020). *Assessing the Effectiveness of Topic Modeling Algorithms in Discovering Generic Labels with Description.* In *Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2.* Springer International Publishing.

2019

1. Hossain, S. S., Ul-Hassan, M. R., & **Rahman, S.** (2019). *Polynomial Topic Distribution with Topic Modeling for Generic Labeling.* In *Advances in Computing and Data Sciences: ICACDS 2019 (Part II).* Springer Singapore.

2. **Rahman, S.**, et al. (2019). *Context-Based News Headlines Analysis Using Machine Learning Approach.* In *Computational Collective Intelligence: ICCCI 2019 (Part II).* Springer International Publishing.

3. Hossain, S. S., et al. (2019). *Customer Feedback Prioritization Technique: A Case Study on Lean Startup.* In *Computational Science and Its Applications – ICCSA 2019 (Part V).* Springer International Publishing.