FARDIN AHSAN SAKIB

Fairfax, VA 571-668-1676 fsakib@gmu.edu Website LinkedIn

EXPERIENCE:

Agentic AI Research Intern **November 2025-January 2026**
Fujitsu Research of America **Santa Clara, CA**

- Trained a verifier model for coding agents that evaluates and selects the best solution from multiple agent attempts on real-world software engineering tasks (SWE-Bench), improving resolve rate by 15%.
- Designed and implemented an end-to-end evaluation pipeline for scoring and ranking agent-generated solutions across multiple runs, supporting the team's research on inference-time scaling for coding agents.

Applied Research Scientist Intern **June 2025-August 2025**
Amazon (AWS Support) **Seattle, WA**

- Developed comprehensive evaluation datasets and frameworks for multi-step tool calling workflows in distributed cloud systems, contributing methodologies for agentic AI assessment
- Engineered simulation-based testing infrastructure that eliminated benchmarking costs while achieving 100% reproducibility across complex tool calling and API integration scenarios
- Conducted systematic analysis of leading large language models (LLMs) on multi-step tool calling orchestration, identifying key failure modes and establishing quantitative performance baselines for production deployment

Applied Research Intern **May 2024 - August 2024**
Brillient Corporation **Reston, VA**

- Developed key components of BRAG (Brillient Retrieval-Augmented Generation) using PyTorch, implementing query distillation and synonym generation modules to enhance contextual understanding across multiple domains.
- Optimized information retrieval by designing efficient chunking algorithms and re-ranking methods, improving the relevance of extracted data by 22% as measured by Mean Average Precision (MAP), and reducing processing time by 35%.
- Finetuned domain-specific Large Language Models (LLMs) and performed advanced prompt engineering (e.g., chain-of-thought, tree-of-thought), successfully deploying the system across medical, HR, and IRS domains with an 85% user satisfaction rate.

Graduate Teaching and Research Assistant **August 2021 - Present**
George Mason University **Fairfax, VA**

- Designed evaluation frameworks that expose reliability failures in large language models, quantifying how models exploit spurious correlations to produce biased predictions with up to 37-percentage-point error gaps across demographic groups.
- Built inference-time correction methods leveraging interpretability signals to selectively fix biased model predictions, reducing errors by up to 78% without retraining, counterfactual data, or knowledge of the bias source.
- Conducting labs and facilitating discussions for over 100 students, providing individualised feedback on assignments and exams.

Machine Learning and AI Intern **May 2023 - August 2023**
Brillient Corporation **Reston, VA**

- Engineered and fine-tuned advanced Natural Language Processing models for tasks including question answering, summarisation, emotion detection, and readability analysis, improving model performance by 12-18% as measured by F1 score, ROUGE, and accuracy metrics, thus enhancing capabilities for critical IRS and USCIS projects.
- Architected and implemented an end-to-end machine learning pipeline from model development to AWS deployment and API creation using AWS SageMaker, Lambda, and API Gateway, reducing deployment time by 40%.

EDUCATION:

Ph.D. in Computer Science (5th Year) — August 2021 - May 2027
George Mason University — Fairfax, VA
Bachelor's in Computer Science and Engineering — January 2017-March 2021
Islamic University of Technology — Dhaka, Bangladesh

TECHNICAL EXPERTISE:

- **Research Areas:** LLMs, Foundation Models, Generative AI, Mechanistic Interpretability, Inference-time Steering, Post-training, Bias Mitigation, AI Safety, Model Alignment, Clinical NLP, Healthcare AI
- **Methods:** Fine-tuning, LoRA, RAG, Verifier Models, Reward Modeling, Model Evaluation, Benchmarking, Agentic AI, Coding Agents, Tool Use
- **Frameworks & Tools:** PyTorch, Hugging Face, Transformers, Multi-GPU Training, HPC Clusters
- **Programming:** Python, C++, SQL
- **Cloud & MLOps:** AWS (SageMaker, Lambda, API Gateway), Azure, Docker, Kubernetes
- **Data Engineering:** Vector Databases, Pandas, NumPy
- **Development Tools:** Git, API Development

RECENT RESEARCH PUBLICATIONS

- **Sakib, Fardin Ahsan**, Ziwei Zhu, and Özlem Uzuner. "Trajectory-Guided Steering: Mitigating Spurious Feature Bias in Clinical Classification." *Under review at COLM 2026*.
- **Sakib, Fardin Ahsan**, Ziwei Zhu, Karen Trister Grace, Meliha Yetişgen, and Özlem Uzuner. "Spurious Correlations and Beyond: Understanding and Mitigating Shortcut Learning in SDOH Extraction with Large Language Models." Accepted at The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL 2025**).(Link)
- Rahman, Md Mushfiqur, **Fardin Ahsan Sakib**, Fahim Faisal, and Antonios Anastasopoulos. "To token or not to token: A Comparative Study of Text Representations for Cross-Lingual Transfer." In Proceedings of the 3rd Workshop on Multi-lingual Representation Learning @ **EMNLP**, pp. 67-84. 2023. (Link)
- **Sakib, Fardin Ahsan**, AHM Rezaul Karim, Saadat Hasan Khan, and Md Mushfiqur Rahman. "Intent Detection and Slot Filling for Home Assistants: Dataset and Analysis for Bangla and Sylheti." In The First Workshop on Bangla Language Processing @ **EMNLP,** p. 48. 2023. (Link)