

M SAIFUL BARI (MARUF)

sbmaruf.github.io

sbmaruf@gmail.com

RESEARCH STATEMENT

My research focuses on **training Large Language Models (LLMs)** through (i) innovations in training dynamics (i.e., scaling laws, scalable truthfulness), (ii) evaluation methodologies (i.e., agentic eval via gamification, unit test based coding eval), and (iii) mid training (second modality acquisition, large scale post-training). I have contributed to two major LLM efforts, **BLOOM**, **ALLaM**. Recently I joined Amazon's **Nova Team**.

At the *National Center for Artificial Intelligence (Saudi Arabia)*, I served as the **Training Lead and Core Maintainer** of **ALLaM**, a sovereign bilingual LLM initiative. ALLaM has been integrated into IBM Watsonx and Microsoft Azure platforms and is part of a broader \$100B Sovereign AI mission. My contributions included:

- Orchestrating \sim \$60M [1] in compute infrastructure and optimizing the Megatron-LM and NeMo stacks.
- Achieving \sim 44% model flop utilization (MFU) during pretraining and \sim 38% MFU during alignment over 1024 GPUs.
- Leading and collaborating with a team of 4+ young engineers and a group of annotators in both research and deployment.

My research approach blends empirical rigor with system-level implementation. I integrate large-scale experimentation with methodologically sound modeling. I am also a contributor to open-source efforts such as [PromptSource](#), [xCodeEval](#), and [ExecEval](#), used widely by the NLP community.

Looking ahead, I aim to contribute to the development of safe, interpretable, and steerable foundation models through iterative de-risking cycles. My long-term goal is to contribute to human-aligned superintelligence, where alignment research and scaling laws **predictably** converge. I am particularly excited by research environments that value both engineering excellence and foundational research into *frontier AI systems and agents*.

EDUCATION

Nanyang Technological University

Jan 2019 - Aug 2023

Doctor of Philosophy (Ph.D.)

Natural Language Processing

Computer Science and Engineering

Islamic University of Technology

2012-2016

Bachelor of Science (BSc.)

Computer Science and Engineering

PROFESSIONAL & EMPLOYMENTS EXPERIENCE

Applied Scientist

SF Bay Area. May 2025 - Present

Amazon AGI

- Working on text pretraining with a focus on multilinguality.
- Working on measuring the value of a token during different stages of training.
- Working on improving function calling for complex multilingual scenarios.

Research Scientist Riyadh, Saudi-Arabia. Aug 2023 - Present
National Center for Artificial Intelligence

- One of the **Core Maintainer** and Research Lead of ALLaM- a nationwide LLM effort of Saudi Arabia.
- Wrote the basic orchestration and optimization of the LLM training framework, data, and training deployment.
- Optimized **Megatron-LM** and **NeMo** for Pretraining (~ 44% MFU) and Alignment (~ 38% MFU) for thousands of GPUs while training trillions of tokens.
- Performed **Second Language Acquisition** for ALLaM and produced the best performing Arabic LLM across all open and close sourced model. [\[paper link\]](#)
- For the work of ALLaM, [I received MIT Innovators Under 35 \(IU35\) for the year 2024 from the MENA Region.](#)

Applied Scientist Intern Santa Clara, California. July 2022 - Nov 2022
Amazon Development Service

- Proposed **SPT**, a memory-based prompt tuning method to improve multitask generalization for Parameter Efficient Multitask Inference (PEMI).
- Performed large scale soft prompt-tuning and evaluated zero-shot generalization on heldout tasks. Once of the early work that showed soft-prompt tuning can generalize at scale.[\[paper link\]](#)

BLOOM LLM training Remote. Sep 2021 - Mar 2022
International Research Effort

- Trained the first open sourced 176B parameter model with a focus on multilinguality. [\[paper link\]](#)
- Worked on scaling laws for training LLM under constrained resources. [\[paper link\]](#)
- Worked on the foundational work of Multitask Learning with Supervised finetuning. [\[paper link\]](#)[\[paper link\]](#)[\[paper link\]](#)

Applied Scientist Intern Santa Clara, California. July 2021 - Jan 2022
Amazon Web Service

- Worked on prompt tuning for multi-lingual models.
- Proposed efficient approach to finetune multilingual language model on specific tasks.[\[paper link\]](#)

Applied Scientist Intern Palo Alto, California. Aug 2020 - Oct 2020
Amazon Web Service

- Worked on Cross-lingual Few-shot Adaptation for multilingual models.
- Proposed transductive inference optimization for cross-lingual models outperforming all the multilingual benchmarks from the baselines.[\[paper link\]](#)

Research Assistant Singapore. Sep 2017 - Aug 2018
Nanyang Technological University

- Developed Malay-English Machine translation system. [\[Link\]](#)
- Worked on semi-supervised learning from cross-lingual generalization. [\[paper link\]](#)
- Apply adversarial training for task grounding and transfer knowledge from a high resource to a low resource language. [\[paper link\]](#)

HONORS & AWARDS

Award 2024
MIT Innovators Under 35 (IU35)

Scholarship 2019
NTU Graduate Scholarship, Fully funded Ph.D. scholarship for 4 years.

Scholarship 2012
OIC Scholarship for undergraduate study, Islamic University of Technology

Champion 2014
IUT Computer Programming Contest

Honorable Mention <i>Human Expedition on Mars Timeline 2018</i>	2014
Champion <i>IUT Computer Programming Contest</i>	2015
2nd/100 in Inter University Programming Contest <i>Daffodill International University ACM ICPC world finals warmup contest 2016</i>	2016
6th/100+ in Inter University Programming Contest <i>NSU Cybernauts National Programming Contest</i>	2016

TECHNICAL SKILLS

Computer Languages	Python (expert), C++ (intermediate), CUDA (basic), Bash
Model Architecture	Transformers, attention mechanisms, sparse models
Tools & Library	Pytorch, Megatron-LM, NeMo, Huggingface, DeepSpeed, vllm, Ray, SLURM
Infrastructure	Kubernetes, Docker, AWS, Azure, GPU optimization
Alignment Techniques	RLHF, Constitutional AI approaches, value alignment, function calling

PROJECTS

[ExecEval](#) (★ 35), [xCodeEval](#) (★ 69), [Algorithm-Code-Library](#) (★ 28), [Promptsources](#) (★ 2600)

PUBLICATION

1. **M Saiful Bari***, Yazeed Alnumay*, Norah A. Alzahrani, Nouf M. Alotaibi, Hisham A. Alyahya, Sultan AlRashed, Faisal A. Mirza, Shaykhah Z. Alsubaie, Hassan A. Alahmed, Ghadah Alabduljabbar, Raghad Alkhathran, Yousef Almushayqih, Raneem Alnajim, Salman Alsubaihi, Maryam Al Mansour, Majed Alrubaian, Ali Alammari, Zaki Alawami, Abdulmohsen Al-Thubaity, Ahmed Abdelali, Jeril Kuriakose, Abdalghani Abujabal*, Nora Al-Twairesh*, Areeb Alowisheq*, and Haidar Khan*. **ALLaM: Large Language Models for Arabic and English**, 2024
2. Long Phan et al. Humanity’s last exam, 2025
3. Hisham A. Alyahya, Haidar Khan, Yazeed Alnumay, Bülent Yener, and **M Saiful Bari**. Zerosumeval: An extensible framework for scaling llm evaluation with inter-model competition. In *Proceedings of the Systems and Demonstration of 63rd Annual Meeting of the Association for Computational Linguistics, ACL*, 2025
4. Md Tahmid Rahman Laskar, Sawsan Alqahtani, **M Saiful Bari**, Mizanur Rahman, Mohammad Abdullah Matin Khan, Haidar Khan, Israt Jahan, Amran Bhuiyan, Chee Wei Tan, Md Rizwan Parvez, Enamul Hoque, Shafiq Joty, and Jimmy Huang. A systematic survey and critical review on evaluating large language models: Challenges, limitations, and recommendations, 2024
5. **M Saiful Bari***, Mohammad Abdullah Matin* Khan, Xuan Long Do, Weishi Wang, Md Rizwan Parvez, and Shafiq Joty. **xCODEEVAL: A Large Scale Multilingual Multitask Benchmark for Code Understanding, Generation, Translation and Retrieval**. In *Proceedings of the 62th Annual Meeting of the Association for Computational Linguistics, ACL*, 2024
6. Norah Alzahrani, Hisham Abdullah Alyahya, Yazeed Alnumay, Sultan Alrashed, Shaykhah Alsubaie, Yusef Almushaykeh, Faisal Mirza, Nouf Alotaibi, Nora Altwairesh, Areeb Alowisheq, **M Saiful Bari**, and Haidar Khan. When benchmarks are targets: Revealing the sensitivity of large language model leaderboards. In *Proceedings of the 62th Annual Meeting of the Association for Computational Linguistics, ACL*, 2024
7. **Bari, M Saiful***, Laskar Tahmid*, Rahman Mizanur, Md Amran Hossen Bhuiyan, Shafiq Joty, and Jimmy Huang. **A Systematic Study of ChatGPT on Benchmark Datasets**. In *Findings of the 61th Annual Meeting of the Association for Computational Linguistics, ACL*, 2023

8. **M Saiful Bari**, Aston Zhang, Shuai Zheng, Xingjian Shi, Yi Zhu, Shafiq Joty, and Mu Li. **SPT: Semi-Parametric Prompt Tuning for Multitask Prompted Learning** (under review at the 2023 conference on empirical methods in natural language processing, emnlp'23), 2022
9. Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, **M Saiful Bari**, Sheng Shen, Zheng-Xin Yong, Hailey Schoelkopf, Xiangru Tang, Dragomir Radev, Alham Fikri Aji, Khalid Almubarak, Samuel Albanie, Zaid Alyafeai, Albert Webson, Edward Raff, and Colin Raffel. **Crosslingual Generalization through Multitask Finetuning**. In *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics, ACL*, 2023
10. Zheng-Xin Yong, Hailey Schoelkopf, Niklas Muennighoff, Alham Fikri Aji, David Ifeoluwa Adelani, Khalid Almubarak, **M Saiful Bari**, Lintang Sutawika, Jungo Kasai, Ahmed Baruwu, Genta Indra Winata, Stella Biderman, Dragomir Radev, and Vassilina Nikoulina. **BLOOM+1: Adding Language Support to BLOOM for Zero-Shot Prompting**. In *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics, ACL*, 2023
11. Teven Le Scao, Thomas Wang, Daniel Hesslow, Lucile Saulnier, Stas Bekman, **M Saiful Bari**, Stella Biderman, Hady Elsahar, Jason Phang, Ofir Press, Colin Raffel, Victor Sanh, Sheng Shen, Lintang Sutawika, Jaesung Tae, Zheng Xin Yong, Julien Launay, and Iz Beltagy. **What Language Model to Train if You Have One Million GPU Hours?** In *Findings in EMNLP, 2022*, 2022
12. Stephen H. Bach, Victor Sanh, Zheng-Xin Yong, Albert Webson, Colin Raffel, Nihal V. Nayak, Abheesht Sharma, Taewoon Kim, **M Saiful Bari**, Thibault Fevry, Zaid Alyafeai, Manan Dey, Andrea Santilli, Zhiqing Sun, Srulik Ben-David, Canwen Xu, Gunjan Chhablani, Han Wang, Jason Alan Fries, Maged S. Al-shaibani, Shanya Sharma, Urmish Thakker, Khalid Almubarak, Xiangru Tang, Dragomir Radev, Mike Tian-Jian Jiang, and Alexander M. Rush. **PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts**. In *Meeting of the Association for Computational Linguistics (ACL) Demonstration*, 2022
13. Victor Sanh, Albert Webson, Colin Raffel, Stephen Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, **M Saiful Bari**, Canwen Xu, Urmish Thakker, Shanya Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M Rush. **Multitask Prompted Training Enables Zero-Shot Task Generalization**. In *International Conference on Learning Representations, ICLR*, 2022
14. Teven et al, **BLOOM: A 176B-Parameter Open-Access Multilingual Language Model**, Arxiv preprint, (Under Review at Journal of Machine Learning Research, JMLR), 2022
15. **Bari, M Saiful**, Batool Haider, and Saab Mansour. **Nearest Neighbour Few-Shot Learning for Cross-lingual Classification**. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1745–1753, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics
16. Tasnim Mohiuddin, **M Saiful Bari**, and Shafiq Joty. **AugVic: Exploiting BiText Vicinity for Low-Resource NMT**. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Online, 2021. Association for Computational Linguistics
17. **Bari, M Saiful**, Tasnim Mohiuddin, and Shafiq Joty. **UXLA: A Robust Unsupervised Data Augmentation Framework for Zero-Resource Cross-Lingual NLP**. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL (Volume 1: Long Papers)*, pages 1978–1992, Online, August 2021. Association for Computational Linguistics
18. Tasnim Mohiuddin, **M. Saiful Bari**, and Shafiq R. Joty. **LNMap: Departures from Isomorphic Assumption in Bilingual Lexicon Induction Through Non-Linear Mapping in Latent Space**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Virtual, November 2020

19. **M Saiful Bari**, Shafiq Joty, and Prathyusha Jwalapuram. **Zero-Resource Cross-Lingual Named Entity Recognition**. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, AAAI '20, New York, USA, 2020. AAAI
20. Xiang Lin, Shafiq Joty, Prathyusha Jwalapuram, and **M Saiful Bari**. **A Unified Linear-Time Framework for Sentence-Level Discourse Parsing**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, ACL '19, Florence, Italy, 2019. ACL
21. **M Saiful Bari**. *Regression for Data Analytics*, chapter 2, pages 33–54. CRC Press, Boca Raton, September 2018. (in book: Data Analytics: Concepts, Techniques and Applications)