M SAIFUL BARI

+65 85052376 ⋄ bari0001@e.ntu.edu.sg

Ph.D. Student, Nanyang Technological University

sbmaruf.github.io

EDUCATION

Nanyang Technological University *2019 - Present*
Doctor of Philosophy (Ph.D.) CGPA: 4.63/5.0
Natural Language Processing
Computer Science and Engineering

Islamic University of Technology *2012-2016*
Bachelor of Science (BSc.) CGPA: 3.65/4.0
Computer Science and Engineering

SKILLS

Computer Languages	Python, C/C++
Tools & Library	Pytorch, DeepSpeed, fairseq, Huggingface
Problem Solving	Deep Learning, Algorithms & Data Structure

HONORS & AWARDS

Scholarship 2019
NTU Research 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 2014
Human Expedition on Mars Timeline 2018

Champion 2015
IUT Computer Programming Contest

2nd/100 in Inter University Programming Contest 2016
Daffodill International University ACM ICPC world finals warmup contest 2016

6th/100+ in Inter University Programming Contest 2016
NSU Cybernauts National Programming Contest

EXPERIENCE

Applied Scientist Intern July 2022 - Nov 2022
Amazon Development Service

Applied Scientist Intern (Part Time) July 2021 - Jan 2022
Amazon Web Service

· Working on prompt tuning for multi-lingual models.

Applied Scientist Intern Aug 2020 - Oct 2020
Amazon Web Service

· Collaborated with Amazon Lex team.
· Worked on Cross-lingual Few-shot Adaptation.

Research Assistant
Nanyang Technological University

Sep 2017 - Aug 2018

· Research on MT, NER and Adversarial Training.
· Maintain internal GPU servers
· Maintain group web-site

Software Engineering Intern
Aubichol IT Limited

Sep 2018 - Jan 2019

· Worked on an early startup where the key responsibility was to design the architecture of sports analytic and Translation System.

Lecturer
Southeast University

June 2017 - Sep 2017

· Teaching undergraduate CS student. Subject taught: * Introduction to C * Algorithms

Software Engineering Intern
XeonBD

Nov 2015 - Dec 2015

· The course of the internship goes through Cloud Computing and kernel virtualization.

Competitive Programmer
Islamic University of Technology

Nov 2012 - Nov 2016

· Solved different kinds of algorithmic and data structure based problem. Solve problem through codeforces, topcoder, lightoj, uva, codechef, hackerrank, projecteuler, coci, usaco. Solved around **2000 problems** in various online judge.

Competitive Programming Trainer

June 2014 - Nov 2016

· Taught Competitive Programming in different Universities in Bangladesh.

- Islamic University of Technology
- Khulna University of Engineering and Technology
- East West University
- Dhaka University of Engineering and Technology

PUBLICATION

1. **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, 2022

2. Mohammad Abdullah Matin Khan, **M Saiful Bari**, 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**, 2023

3. 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**, 2022

4. Zheng-Xin Yong, Hailey Schoelkopf, Niklas Muennighoff, Alham Fikri Aji, David Ifeoluwa Adelani, Khalid Almubarak, **M Saiful Bari**, Lintang Sutawika, Jungo Kasai, Ahmed Baruwa, Genta Indra Winata, Stella Biderman, Dragomir Radev, and Vassilina Nikoulina. Bloom+1: Adding language support to bloom for zero-shot prompting, 2022

5. 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

6. 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

7. 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 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

8. Teven et al, **BLOOM: A 176B-Parameter Open-Access Multilingual Language Model**, ArxiV preprint, 2022

9. **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

10. **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

11. 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

12. 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

13. **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

14. 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

15. **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)

PROJECTS

⋆ Prompt Engineering :: *Promptsource*
Details can be found here, https://github.com/bigscience-workshop/promptsource

⋆ LLM Pipelining :: *Megatron/DeepSpeed*
Details can be found here, https://github.com/bigscience-workshop/Megatron-DeepSpeed

⋆ Dataset :: *xCodeEval*
Details can be found here, https://github.com/ntunlp/xCodeEval

⋆ Deep learning :: *UXLA*

Details can be found here, https://github.com/sbmaruf/UXLA

⋆ Deep learning :: *Cross-lingual Few Shot Learning*

Details can be found here, https://github.com/amazon-science/nearest-neighbor-crosslingual-classification

⋆ Dataset :: *Promptsource*

Details can be found here, https://github.com/bigscience-workshop/promptsource

⋆ Deep learning :: *Zero-Resource Cross-lingual Named Entity Recognition*

Details of the project can be found here, https://github.com/ntunlp/Zero-Shot-Cross-Lingual-NER

⋆ Deep learning :: *Malay English Machine Translation System*

Details can be found here, https://sbmaruf.github.io/project/mt-system/

⋆ Machine learning system :: *A CBIR System*

Details of the project can be found here, http://103.82.172.44:8080/xmlui/handle/123456789/93

⋆ Algorithms :: *Algorithm-Code-Library*

Details of the project can be found here, https://github.com/sbmaruf/Algorithms-Code-Library