# FAHIM FAISAL

Fairfax, Virginia, USA
Linkedin: linkedin.com/in/faisalfahim
Website: fahimfaisal.info

Email : ffaisal@gmu.edu
Mobile : +001-5715361901
Github: ffaisal93
Google Scholar

## Education

- **George Mason University** — Virginia
  Ph.D in Computer Science (ongoing) — January. 2020 – Spring 2025 [expected]
  · GPA: **4.00\*** [current]

- **George Mason University** — Virginia
  MS in Computer Science — January. 2020 – May 2023
  · CGPA: 4.00 (out of 4.00)

- **Islamic University of Technology** — Gazipur
  B.Sc. in Computer Science and Engineering — Dec. 2012 – Nov. 2016
  · CGPA: 3.78 (out of 4.00)
  · Class Rank: 12 (out of 48)

## Research Interests

- Multilingual adaptation of language models, Safety and policy alignment of LLMs, Evaluation of language models

## Research & Professional Experience

- **Research Assistant** — Virginia, USA
  George Mason NLP Lab — June. 2020 - Present

- **PhD Research Intern** — California, USA
  eBay Inc. — Summer 2023 & 2024
  · Aspect-based product recommendation system (2023)
  · Safety and policy alignment of e-commerce specific LLM (2024)

- **Teaching Assistant** — -
  George Mason University — January. 2020 - May, 2021

- **NRT trainee** — -
  NSF Research Traineeship (NRT) Program, CASBBI — June. 2020 - May. 2021
  · Community-engaged project designing
  · Project: Students with disabilities in COVID-19: Parents' perspective

- **Teaching & Research Assistant** — Dhaka, Bangladesh
  BRAC University — June. 2019 - Present

- **Teaching Assistant** — Gazipur, Bangladesh
  Islamic University of Technology — Jan. 2018 - Oct. 2018

- **Software Engineer** — Dhaka, Bangladesh
  Samsung R&D Institute — Mar. 2017 - Sept. 2017

## Volunteer experience and recognition

- Reviewer: ACL, EMNLP, NAACL, EACL

- Best social impact paper award (ACL 2024)

- Outstanding PhD Student Award (George Mason CS, 2024)

- Student Volunteer: ACL 2022, NAACL 2022, ACL 2024

# Publications

## 2024

○ Faisal, F., Ahia, O., Srivastava, A., Ahuja, K., Chiang, D., Tsvetkov, Y., & Anastasopoulos, A. (2024, August). DIALECTBENCH: An NLP benchmark for dialects, varieties, and closely-related languages. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 14412–14454). Bangkok, Thailand: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2024.acl-long.777 doi: 10.18653/v1/2024.acl-long.777

○ Faisal, F., & Anastasopoulos, A. (2024b, November). An efficient approach for studying cross-lingual transfer in multilingual language models. In J. Sälevä & A. Owodunni (Eds.), *Proceedings of the fourth workshop on multilingual representation learning (mrl 2024)* (pp. 45–92). Miami, Florida, USA: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2024.mrl-1.4

○ Faisal, F., & Anastasopoulos, A. (2024a, June). Data-augmentation-based dialectal adaptation for LLMs. In Y. Scherrer, T. Jauhiainen, N. Ljubešić, M. Zampieri, P. Nakov, & J. Tiedemann (Eds.), *Proceedings of the eleventh workshop on nlp for similar languages, varieties, and dialects (vardial 2024)* (pp. 197–208). Mexico City, Mexico: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2024.vardial-1.17 doi: 10.18653/v1/2024.vardial-1.17

## 2023

○ Faisal, F., & Anastasopoulos, A. (2023). Geographic and geopolitical biases of language models. *Accepted at Multilingual Representation Learning (MRL) Workshop 2023, Co-located with EMNLP 2023*. Retrieved from https://arxiv.org/abs/2212.10408

○ Song, Y., Cui, C., Khanuja, S., Liu, P., Faisal, F., Ostapenko, A., ... Neubig, G. (2023). Globalbench: A benchmark for global progress in natural language processing. *Accepted at EMNLP 2023 main conference*. Retrieved from https://arxiv.org/abs/2305.14716

○ Rahman, M. M., Sakib, F. A., Faisal, F., & Anastasopoulos, A. (2023). To token or not to token: A comparative study of text representations for cross-lingual transfer. *Accepted at Multilingual Representation Learning (MRL) Workshop 2023, Co-located with EMNLP 2023*.

○ ibn Alam, M. M., Xie, R., Faisal, F., & Anastasopoulos, A. (2023, July). Gmnlp at semeval-2023 tasks 12: Sentiment analysis with phylogeny-based adapters. In *Proceedings of the 17th international workshop on semantic evaluation (semeval-2023)*. Toronto, Canada: Association for Computational Linguistics. Retrieved from https://arxiv.org/abs/2304.12979 [code]

## 2022

○ Faisal, F., Wang, Y., & Anastasopoulos, A. (2022, May). Dataset geography: Mapping language data to language users. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 3381–3411). Dublin, Ireland: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.acl-long.239 doi: 10.18653/v1/2022.acl-long.239 [code]

○ Faisal, F., & Anastasopoulos, A. (2022, November). Phylogeny-inspired adaptation of multilingual models to new languages. In *Proceedings of the 2nd conference of the asia-pacific chapter of the association for computational linguistics and the 12th international joint conference on natural language processing (volume 1: Long papers)* (pp. 434–452). Online only: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.aacl-main.34 [code]

## 2021

○ Faisal, F., Keshava, S., Alam, M. M. I., & Anastasopoulos, A. (2021, November). SD-QA: Spoken dialectal question answering for the real world. In *Findings of the association for computational linguistics: Emnlp 2021* (pp. 3296–3315). Punta Cana, Dominican Republic: Association for Computational Linguistics. Retrieved from https://aclanthology .org/2021.findings-emnlp.281 [code]

○ Faisal, F., & Anastasopoulos, A. (2021, November). Investigating post-pretraining representation alignment for cross-lingual question answering. In *Proceedings of the 3rd workshop on machine reading for question answering* (pp. 133–148). Punta Cana, Dominican Republic: Association for Computational Linguistics. Retrieved from https://aclanthology .org/2021.mrqa-1.14 [code]

○ Mahmud, J., Faisal, F., Arnob, R. I., Anastasopoulos, A., & Moran, K. (2021, August). Code to comment translation: A comparative study on model effectiveness & errors. In *Proceedings of the 1st workshop on natural language processing for programming (nlp4prog 2021)* (pp. 1–16). Online: Association for Computational Linguistics. Retrieved from https:// aclanthology.org/2021.nlp4prog-1.1 doi: 10.18653/v1/2021.nlp4prog-1.1

## 2020

○ Choudhury, N., Faisal, F., & Khushi, M. (2020). Mining temporal evolution of knowledge graphs and genealogical features for literature-based discovery prediction. *Journal of Informetrics*, *14*(3), 101057. Retrieved from https://www .sciencedirect.com/science/article/pii/S1751157719304468 doi: https://doi.org/10.1016/j.joi.2020.101057 [code]

## References Available to Contact

○ **Dr. Antonios Anastasopoulos**
Assistant Professor
Dept. of Computer Science (CS)                                                    antonis@gmu.edu
George Mason University
4400 University Dr, Fairfax, VA 22030

○ **Dr. Yulia Tsvetkov**
Associate Professor
Paul G. Allen School of Computer Science & Engineering                  yuliats@cs.washington.edu
University of Washington