

Sourabrata Mukherjee

Prague, Czechia
☎ +420 778046278
✉ soura1990@gmail.com
📁 souro.github.io



Education

- 2019–Present **Ph.D. in Computational Linguistics**, *Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague, Czechia.*
Thesis: *Text Style Transfer using Neural Models*
- 2011–2013 **Master in Computer Science and Engineering**, *National Institute of Technology (NIT), Durgapur, India, Grade - 9.06 on a scale of 10.*
Thesis: *Solving 'GFRP Composites Machining' and 'Data-Paths Design During High-Level Synthesis' using Evolutionary Algorithms*
- 2007–2011 **Bachelor in Computer Science and Engineering**, *West Bengal University of Technology (WBUT), Kolkata, India, Grade - 8.79 on a scale of 10.*
Thesis: *Pseudo Random Number Generator*

Selected Publications

- Mukherjee, S., Ojha, A. K., Bansal, A., Alok, D., McCrae, J. P., & Dušek, O. (2024). Multilingual Text Style Transfer: Datasets & Models for Indian Languages. [To appear]. Proceedings of the 17th International Natural Language Generation Conference (INLG 2024).
- Mukherjee, S., Ojha, A. K., & Dušek, O. (2024). Are Large Language Models Actually Good at Text Style Transfer? [To appear]. Proceedings of the 17th International Natural Language Generation Conference (INLG 2024).
- Mukherjee, S., Lango, M., Kasner, Z., & Dušek, O. (2024). A Survey of Text Style Transfer: Applications and Ethical Implications. [Communicated] Northern European Journal of Language Technology (NEJLT).
- Mukherjee, S., & Dušek, O. (2024). Text Style Transfer: An Introductory Overview. [Accepted] 4EU+ International Workshop on Recent Advancements in Artificial Intelligence.
- Mukherjee, S., Bansal, A., Ojha, A. K., McCrae, J. P., & Dušek, O. (2023). Text Detoxification as Style Transfer in English and Hindi. Proceedings of the 20th International Conference on Natural Language Processing, ICON 2023, 133–144.
- Mukherjee, S., Bansal, A., Majumdar, P., Ojha, A. K., & Dušek, O. (2023). [Best Paper Award] Low-Resource Text Style Transfer for Bangla: Data & Models. Proceedings of the First Workshop on Bangla Language Processing (BLP-2023), EMNLP 2023, 34–47.
- Mukherjee, S., & Dušek, O. (2023). Leveraging Low-resource Parallel Data for Text Style Transfer. Proceedings of the 16th International Natural Language Generation Conference, INLG 2023, 388–395.

Mukherjee, S., Hudeček, V., & Dušek, O. (2023). Polite Chatbot: A Text Style Transfer Application. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, EACL 2023, 87–93.

Mukherjee, S., Kasner, Z., & Dušek, O. (2022). Balancing the Style-Content Trade-off in Sentiment Transfer using Polarity-Aware Denoising. International Conference on Text, Speech, and Dialogue, TSD 2022, 172–186.

Internships

- PhD Internship Culture+LLM, from Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE.
- PhD Internship Low resource Machine Translation, from Panlingua, India.
- Research Internship Conducted data analytics and time series analysis on river flood and rain data project at the Indian Institute of Science, Bangalore.
- Development Internship Completed summer internship utilizing DB2 from IBM.

Work Experience

- Dec, 2018 – **Research Assistant**, *UKP Lab, TU Darmstadt*, Germany.
- May, 2019
 - Prepared a data-to-text generation corpus using automatic detection of evidence related to structured data like tables from scientific articles.
- May, 2016 – **Technical Lead**, *Tricon Infotech*, Bangalore, India.
- Sept, 2019
 - Analytical recommendation engine leveraging usage statistics.
- April, 2016 – **Senior Software Engineer**, *Avaya Research Lab*, Bangalore, India.
- May, 2016
 - Full-stack log monitoring with optimized persistence log data mechanism.
- May, 2015 – **Software Engineer**, *O9 Solutions*, Bangalore, India.
- March, 2016
 - Batch job management framework with notification capability.
- June, 2013 – **Senior Software Engineer**, *AMDOCS*, Pune, India.
- April, 2015
 - Worked on Recommendation Engine of a web application, Search engine.

Research Interest

Natural Language Generation (*Controllable NLG, Evaluation*), Miscellaneous (*Low-Resource NLP, Model interpretability, Reasoning, LLMs, Culture*)

Certificates

LLM, ML, DL, and, Statistics from Coursera

Miscellaneous

- Received the Charles University Grant Agency (GAUK) research grant as PI and earned an honorable mention upon project completion.
- Awarded a government scholarship for the Graduate Aptitude Test in Engineering (GATE) during my Master's degree.
- Won IBM national level Technical Competition.
- Taught introductory courses on Machine Learning (ML), Deep Learning (DL), Data Structures, Programming, and Search Engine Optimization