SOUMYA SHARMA

soumyasharma20@gmail.com | Google Scholar: LBNLcm8AAAAJ | +91-98761699406 |
https://soummyaah.github.io/

RELEVANT PUBLICATIONS

- **Sharma, Soumya**, Preksha Nema, Abhirut Gupta, Aravindan Raghuveer. "CoPSTr: Compositional Planning for Unsupervised Style Transfer". Under review in ACL Rolling Review.

- **Sharma, Soumya**, Subhendu Khatuya, Manjunath Hegde, Afreen Shaikh, Koustuv Dasgupta, Pawan Goyal, and Niloy Ganguly. 2023. "Financial Numeric Extreme Labelling: A dataset and benchmarking." *In Findings of the Association for Computational Linguistics: ACL 2023*, pages 3550–3561, Toronto, Canada. Association for Computational Linguistics.

- Kumar, Dibyakanti, Vivek Gupta, **Soumya Sharma**, and Shuo Zhang. "Realistic Data Augmentation Framework for Enhancing Tabular Reasoning." In Findings of *the Association for Computational Linguistics: EMNLP 2022*

- **Sharma, Soumya**, Tapas Nayak, Arusarka Bose, Ajay Kumar Meena, Koustuv Dasgupta, Niloy Ganguly, and Pawan Goyal. 2022. "FinRED: A Dataset for Relation Extraction in Financial Domain." *In Companion Proceedings of the Web Conference 2022 (WWW '22).* Association for Computing Machinery, New York, NY, USA, 595–597. https://doi.org/10.1145/3487553.3524637

- Nandy, Abhilash, **Soumya Sharma**, Shubham Maddhashiya, Kapil Sachdeva, Pawan Goyal, and Niloy Ganguly. "Question Answering over Electronic Devices: A New Benchmark Dataset and a Multi-Task Learning based QA Framework." In Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 4600-4609. 2021.

- **Sharma, Soumya***, Bishal Santra*, Abhik Jana, Santosh Tokala, Niloy Ganguly, and Pawan Goyal. "Incorporating Domain Knowledge into Medical NLI using Knowledge Graphs." *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 6092-6097. 2019.

Please find the complete list on Google Scholar

RESEARCH EXPERIENCE

Google DeepMind, India   July 2023 -
PreDoctoral Researcher in Agent Quality and Understanding (AQUA) team

- Worked on a planning based Unsupervised writing style and tone transfer pipeline. Currently under review in ACL Rolling Review.

- Working on evaluation benchmark and auto-raters to measure creativity through implicit rubrics. Working on improving creativity in LLM response.

Indian Institute of Technology, Kharagpur   September 2020 - June 2023
Graduate Student, Complex Networks Research Lab under Prof. Pawan Goyal and Prof. Niloy Ganguly

- Worked in collaboration with **Goldman Sachs, India** on creating datasets from financial documents including large-scale labeling of numericals, table understanding, summarization, and relation extraction.

- Published multiple works in top conferences including ACL'23, EMNLP'22 and WWW'21.

Indian Institute of Technology, Kharagpur   February 2019 - June 2023
Language Analyst Research, Complex Networks Research Lab under Prof. Pawan Goyal and Prof. Niloy Ganguly

- Worked in collaboration with **Samsung R&D Institute, India** on different aspects of knowledge graphs including *knowledge graph completion* on dynamic graphs, and *hyperbolic space embeddings*.

- Published a paper in EMNLP'19 on incorporating medical domain knowledge using knowledge graphs in deep learning models.

Microsoft Research (MSR), India   January 2018 - July 2018
Research Intern, with Dr. Aseem Rastogi

- Worked on EzPC: Easy Secure Multi-Party Computation - a tool to write secure programs without worrying about underlying cryptographic computations.
- Built a transpiler in OCaml that converts EzPC code to C++, formalized the EzPC language, and created benchmarks by writing a decision tree program in EzPC.

EDUCATION

Indian Institute of Technology, Kharagpur   Sep, 2020 - Present
M.S. (Research) in **Computer Science & Engineering**   CGPA: 9.3

Indraprastha Institute of Information Technology, Delhi, India   Aug, 2013 - April, 2017
Bachelor of Technology in **Computer Science & Engineering**   CGPA: 7.1

INDUSTRY EXPERIENCE

Myntra Ltd, Bengaluru   August 2017 - January 2018
Software Engineer

- Worked on detecting viral trends in the Indian fashion industry using multi-modal sources. Created a distributed, scalable data collection app using Actor models and Redis. Data collected was used to create a nuanced mix-bag model using ARIMA and LSTM models.
- Created a Django based web app used by internal sales team to obtain insights into the latest trends. The recommendations based on our pipeline were tested on 30% of the users and led to an increase in 2% revenue of total sales.

SCHOLARSHIPS AND AWARDS

Qualcomm Innovation Fellowship 2021: Finalist in QIF, submitted a proposal for "Question Answering in E-Manuals"; related paper published in EMNLP 2021

WomenTech Makers Scholarship 2016: Was awarded the scholarship by Google Inc. for excellence in computer science, leadership, and passion for technology.

LEADERSHIP & SERVICE

Teaching Assistant (IIT-Kharagpur): AI Ethics'22, NLP'22, Algorithms'21, Programming & Data Structures'20

Teaching Assistant (IIIT-Delhi): DSA'16, Applied CS with Android in collaboration with Google Inc.'17

Lead, Mentor Program: Mentored 8 BTP, MTP, and Internship students on their project/thesis.

Reviewer: Reviewer for NAACL 2024, ACL 2024, EMNLP 2023

Rails Girls Summer of Code: Engineering Intern. Mentored by Keith-Magee (ex-president of Django Foundation) | VOC, Pybee Project

Google Code Jam for Women 2016: Ranked 111th globally, 6th in India.

Build The Shield Capture-The-Flag 2016 (Microsoft): Placed 26th all over India in the finals.

INTERESTS AND HOBBIES

Sports & Fitness: Mixed Martial Arts, Swimming, and Tennis.
Reading: Interdisciplinary research papers, Historical accounts, and Fiction.
Others: Music, Baking, Travel, and Nature walks.