Soham Poddar

PMRF Research Scholar
Indian Institute of Technology, Kharagpur

PERSONAL DETAILS

Contact	sohampoddar26@gmail.com *OR* +91 9163642188
Website	https://sohampoddar26.github.io/
Google Scholar	https://scholar.google.com/citations?user=PIIZaaOAAAAJ&hl=en/
Research Interests	Deep Learning, Natural Language Processing, Agentic AI, Domain-specific applications.

EDUCATION

PhD in Computer Science and Engineering (Thesis submitted) 2020-2025
Indian Institute of Technology, Kharagpur.
Under the guidance of **Dr. Saptarshi Ghosh**. (Coursework CGPA: **9.3**)
Thesis title: *Effects of COVID-19 pandemic on Vaccine-Opinions through Social Media Analyses*

MTech Degree in Computer Science and Engineering 2018-2020
Indian Institute of Technology, Kharagpur
Graduated with CGPA: **9.6** (Department Rank: **2**)
Thesis Title: *Summarizing Legal Case Documents: Incorporating Domain Knowledge in Summarization Algorithms*

BTech Degree in Computer Science and Technology 2014-2018
Indian Institute of Engineering Science and Technology, Shibpur
Graduated with CGPA: **9.1**

EXPERIENCE

Research Intern Jun-Dec 2025
Hewlett Packard Enterprise Labs
Developing and benchmarking LLM Agentic frameworks for context engineering, long-term memory and data management.

SKILLS

Human Languages	English (fluent), Bengali (native), Hindi
Computer Languages	Python (extensive experience), C, C++, Java, HTML, CSS
Tools/Technologies	LangGraph, Transformers, Huggingface, PyTorch, Flask, Docker, GCP Github Copilot, Git, Linux, LaTeX
Hobbies	Bass Guitar, Motorcycling, Swimming, Gaming, watch Formula **1** & STEM documentaries

AWARDS

- Awarded **PMRF fellowship** by Ministry of Education, Govt. of India. December 2020 cycle.
- Received several **Travel Grants** to attend SIGIR 2022 (Madrid), ICWSM 2024 (Buffalo), NAACL 2025 (Albuquerque).
- **Best Student Paper** Award at ICAIL 2021
- Qualified for the onsite **regionals of ACM ICPC** in **Dec 2016** (IIT Kharagpur region) and **Dec 2017** (Kolkata region).
- Finalist and winner of several **coding and robotics challenges** in different tech fests (at IIT Kgp, IIEST Shibpur, etc.)
- Winner of **Music Cup** of the Inter IIT Cultural meet **2019** (at IIT Bombay) and **2023** (at IIT Kharagpur).

PUBLICATIONS

- Soham Poddar, Paramita Koley, Janardan Misra, Niloy Ganguly, Saptarshi Ghosh. **"Brevity is the soul of sustainability: Characterizing LLM response lengths"** *In Findings of the Association for Computational Linguistics: ACL 2025*

- Soham Poddar, Paramita Koley, Janardan Misra, Niloy Ganguly, Saptarshi Ghosh. **"Towards Sustainable NLP: Insights from Benchmarking Inference Energy in Large Language Models"** *In Proceedings of Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL), 2025.*

- Soham Poddar, Subhendu Khatuya, Rajdeep Mukherjee, Niloy Ganguly, Saptarshi Ghosh. **"How COVID-19 has Impacted the Anti-Vaccine Discourse: A Large-Scale Twitter Study Spanning Pre-COVID and Post-COVID Era"** *In Proceedings of the 18th International AAAI Conference of Web and Social Media (ICWSM) 2024.*

- Soham Poddar, Rajdeep Mukherjee, Azlaan Mustafa Samad, Niloy Ganguly, Saptarshi Ghosh. **"MuLX-QA: Classifying Multi-Labels and Extracting Rationale Spans in Social Media Posts"** *ACM Transactions on the Web (TWEB), 2024.*

• Soham Poddar, Azlaan Mustafa Samad, Rajdeep Mukherjee, Niloy Ganguly, Saptarshi Ghosh. **"CAVES: A dataset to facilitate explainable classification and summarization of concerns towards COVID vaccines"**. *In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) 2022.*

• Soham Poddar, Mainack Mondal, Janardan Misra, Niloy Ganguly, and Saptarshi Ghosh. **"Winds of Change: Impact of COVID-19 on Vaccine-related Opinions of Twitter users"** *In Proceedings of the 16th International AAAI Conference of Web and Social Media (ICWSM) 2022.*

• Abhay Shukla, Paheli Bhattacharya, Soham Poddar, Rajdeep Mukherjee, Kripabandhu Ghosh, Pawan Goyal and Saptarshi Ghosh. **"Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation"**. *The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (AACL-IJCNLP), 2022.*

• Paheli Bhattacharya, Soham Poddar, Koustav Rudra, Kripabandhu Ghosh, and Saptarshi Ghosh. **"Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents"** *In Proceedings of the 18th International Conference on Artificial Intelligence and Law (ICAIL) 2021.*

• Soham Poddar, Biswajit Paul, Moumita Basu, and Saptarshi Ghosh. **"ICPR 2024 Competition on Multilingual Claim-Span Identification"**. *In Proceedings of the 27th International Conference on Pattern Recognition (ICPR), 2024.*

• Rahul Pullanikkat, Soham Poddar, Anik Das, Tushar Jaiswal, Vivek Kumar Singh, Moumita Basu, and Saptarshi Ghosh. **"Utilizing the Twitter social media to identify transportation-related grievances in Indian cities"**. *Social Network Analysis and Mining (SNAM), 2024.*

• Soham Poddar, Mainack Mondal, and Saptarshi Ghosh. **"A Survey on Disaster: Understanding the After-effects of Super-cyclone Amphan and Helping Hand of Social Media."** *Advances in Urban Design and Engineering, Springer, 2022.*

• Ashish Kumar Layek, Soham Poddar, and Sekhar Mandal. **"Detection of Flood Images Posted on Online Social Media for Disaster Response."** *In Proceedings of the 2nd International Conference on Advanced Computational and Communication Paradigms (ICACCP) 2019.*

SYSTEMS DEVELOPED

• **MESSAGE CHECK** is a fact-checking dashboard, and a predictive learning platform to identify existing fact-checks from www.vishvasnews.com that match dis/misinformation claims going viral. It also enables fact-checkers to predict misinformation around events and identify seasonal or event-based trends that cause a surge in misinformation. Given a multilingual query, it was processed using BM25+ retriever, a supervised Fasttext classifier and a small BERTScore model. It was then merged using Reciprocal Rank Fusion method to efficiently and effectively match debunked claims from a database.
Link to website: https://mdp.vishvasnews.com/

SELECTED PROJECTS

Optimizing Energy Efficiency of LLMs 2024-2025
IIT Kharagpur

We first benchmarked the energy usage of different LLMs for various NLP tasks under different scenarios, and the effect of different optimizations methodologies [e.g. I/O compression, model compression, speculative decoding] in reducing energy consumption. We then showed that LLMs generate very long answers for factual queries, and formally categorize the information into different classes [e.g. minimal answer, additional information, reasoning, redundant information]. We highlight the trade-offs of such long answers that improve the user experience and utility but come at the expense of much higher energy consumption (which add up over time). We also explored some simple strategies (supervised fine-tuning, prompt-engineering with target length prediction, next token entropy mapping) to control the length and content composition of generated outputs, which can be used depending on the specific use case/user preferences.

Characterizing User Opinions towards Vaccines on Twitter (PhD Thesis Work) 2020-2025
IIT Kharagpur

We used automated NLP methods to systematically analyse and derive insights from the vaccine-opinions of various Twitter user-groups, and how these have evolved in course of the COVID-19 pandemic. Beyond high-level Anti- and Pro-vax categorization, we tried to understand specific reasons why people are hesitant to take vaccines and what causes them to change their opinions. We developed the large-scale annotated CAVES dataset for explainable multi-label classification of 12 vaccine concerns (e.g. side-effect, ineffectiveness, political, conspiracies). We also developed Transformer models (encoder-only, LLMs) and trained (SGD, SFT, LORA) them to identify concerns towards vaccines effectively, while extracting spans that explain the specific concerns. We also explored prompt-tuning and fine-tuning LLMs to provide personalised arguments to counter anti-vaccine content on social media to resolve misconceptions.

Summarizing Legal Case Documents (MTech Project and beyond) 2019-2022
IIT Kharagpur

Compared the performance of various summarization algorithms [generic and domain-specific; abstractive and extractive; classical ML vs Transformers based methods] on our curated set of UK and Indian Supreme courts' Legal Judgement Documents. We also created an unsupervised Linear Programming based method to systematically incorporate guidelines from legal experts into to create summaries with appropriate proportions of different rhetorical roles [e.g. fact, statutes, precedents, final judgement].