Mohna Chakraborty

716-748-5386 | mohna0310@gmail.com | https://mohna0310.github.io

RESEARCH INTERESTS

My research lies at the intersection of Generative AI, natural language processing, and machine learning, with a strong emphasis on real-world applicability. As a postdoctoral researcher, I focused on enhancing the practical use of large language models (LLMs), particularly their ability to demonstrate human-like social behaviors in controlled environments. My work aims to improve the personalization, explainability, interpretability, and robustness of LLMs in socially grounded contexts.

During my doctoral studies, I concentrated on the domain of data mining, natural language processing, and machine learning, with a specific focus on low-cost information extraction using minimal supervision. My research tackled the scarcity of labeled textual data by developing methods that significantly streamline the annotation process, minimizing human input while ensuring cost-effective, accessible solutions compatible with everyday systems. This focus on efficient, budget-friendly approaches has been central to my research philosophy.

I have authored seventeen papers published in leading conferences such as ACL, UAI, SIGKDD, PAKDD, SIAM, and ESEC/FSE, journals such as TKDD, and notable workshops, including ICML, ICLR, WWW, PAKDD, RANLP, and the ML Reproducibility Challenge. With additional papers under review, I remain committed to pushing the boundaries of knowledge in these domains.

EDUCATION

University of Michigan — Ann Arbor, MI, USA
PostDoctorate Fellow — *Sep 2024 - July 2025*
- **Research Area:** Social Science, Data mining, Machine Learning, Natural Language Processing
- **Mentor:** Dr. David Jurgens and Dr. Lu Wang

Iowa State University — Ames, IA, USA
Ph.D in Computer Science — *Aug 2020 - Aug 2024*
- **Research Area:** Data mining, Machine Learning, Natural Language Processing
- **Advisor:** Dr. Qi Li

State University of New York at Buffalo — Buffalo, NY, USA
M.S in Computer Science — *Aug 2018 - July 2020*
- **Research Area:** Data Mining, Machine Learning, Natural Language Processing
- **Advisor:** Dr. David Doermann
- **Thesis Title:** Using Machine Learning for Predicting Aspect-Wise Satisfaction Ratings by Semantic Analysis of Text

West Bengal University of Technology, Kolkata — Kolkata, WB, India
B.Tech in Electronics and Communication Engineering — *June 2011 - May 2015*

CONFERENCE PUBLICATIONS

Conferences

ACL, EMNLP, UAI, AAAI, SIGKDD, PAKDD, SIAM, TKDD, ESEC/FSE

- Adit Desai, Sudipta Roy, **Mohna Chakraborty**, "How Reasoning Influences Intersectional Biases in Vision Language Models." **Association for the Advancement of Artificial Intelligence, (AAAI) 2026**.

- **Mohna Chakraborty**, Lu Wang, David Jurgens, "Structured Moral Reasoning in Language Models: A Value Grounded Evaluation Framework." **Conference on Empirical Methods in Natural Language Processing, (EMNLP) 2025**.

- **Mohna Chakraborty***, Adithya Kulkarni*, Qi Li, "Weakly Supervised Open-Domain Aspect-based Sentiment Analysis." **Transactions on Knowledge Discovery in Data, (TKDD) 2024**.

- Seok Hwan Song, **Mohna Chakraborty**, Qi Li, Wallapak Tavanapong "Is Large Language Model Performance on Reasoning Tasks Impacted by Different Ways Questions Are Asked?.",**63rd Annual Meeting of the Association for Computational Linguistics (ACL Findings), 2025**.

- Adithya Kulkarni, **Mohna Chakraborty**, Sihong Xie, Qi Li, "Budget Allocation Exploiting Label Correlation between Instances.",**41st Conference on Uncertainty in Artificial Intelligence, (UAI) 2025**.

- **Mohna Chakraborty***, Adithya Kulkarni*, Qi Li, "Modeling Data Diversity for Joint Instance and Verbalizer Selection in Cold-Start Scenarios.", **Pacific-Asia Conference on Knowledge Discovery and Data Mining, (PAKDD) 2025**.

- Adithya Kulkarni*, **Mohna Chakraborty***, "Reducing Performance Gap between Commercial and Open-Source LLMs.", **SIAM International Conference on Data Mining, (SDM) 2025**.

- **Mohna Chakraborty***, Adithya Kulkarni*, Qi Li, "Zero-shot Approach to Overcome Perturbation Sensitivity of Prompts.", **61st Annual Meeting of the Association for Computational Linguistics (ACL), 2023**, url: https://aclanthology.org/2023.acl-long.313. (Acceptance Rate: 20.8%).

- Adithya Kulkarni, **Mohna Chakraborty**, Sihong Xie, Qi Li, "Optimal Budget Allocation for Crowdsourcing Labels for Graphs.", **39th Conference on Uncertainty in Artificial Intelligence (UAI), 2023**, url: https://proceedings.mlr.press/v216/kulkarni23a.html. (Acceptance Rate: 29.3%).

- **Mohna Chakraborty***, Adithya Kulkarni*, Qi Li, "Open-Domain Aspect-Opinion Co-Mining with Double-Layer Span Extraction.", **28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD), 2022**, url: https://doi.org/10.1145/3534678.3539386. (Acceptance Rate: 14.9%).

- **Mohna Chakraborty**, "Does reusing pre-trained NLP model propagate bugs?", **29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), 2021**, url: https://dl.acm.org/doi/10.1145/3468264.3473494.

WORKSHOP PUBLICATIONS

Conferences

ICML, PAKDD, ICLR, WWW, RANLP, ML Reproducibility Challenge

- Xiaofeng Liu, Xiaobo Xia, Xuechen Zhang, **Mohna Chakraborty**, Xiyuan Chang, Kuai Fang, William S. Currie, Samet Oymak, "Self-Imputation and Cross-Variable Learning Improve Water Quality Prediction with Sparse Data.", **Foundation Models for Structured Data Workshop, ICML 2025**.

- **Mohna Chakraborty***, Adithya Kulkarni*, Qi Li, "Empirical Evaluation of Prompting Strategies for Fact Verification Tasks.", **PromptEng Workshop at the ACM WebConf, WWW 25**.

- Adithya Kulkarni*, **Mohna Chakraborty***, Yonas Sium*, Sai Charihsma Valluri, Wei Le, Qi Li, "From Pseudo-Code to Source Code: A Self-Supervised Search Approach", **Deep Learning for Code (DL4C) Workshop, ICLR 25**.
- Adithya Kulkarni, **Mohna Chakraborty**, Oliver Eulenstein, Qi Li, "Beyond Single Parsers: An Empirical Analysis of Dependency Parse Tree Aggregation.", **Research and Applications of Foundation Models for Data Mining and Affective Computing (RAFDA), PAKDD 2025**.
- Abhishek Kumar Mishra*, **Mohna Chakraborty***, "Does local pruning offer task-specific models to learn effectively?.", **Student Research Workshop Associated with 13th International Conference on Recent Advances in Natural Language Processing RANLP, 2021**, url: https://aclanthology.org/2021.ranlp-srw.17.
- Richard D Jiles, **Mohna Chakraborty**, "[Re] Domain Generalization using Causal Matching.", **ML Reproducibility Challenge, 2021**, url: https://openreview.net/forum?id=r43elaGmhCY

HONORS AND AWARDS

- **Best Vision Paper:** Our paper on "Reducing Performance Gap between Commercial and Open-Source LLMs" has been awarded the third-place Best Vision Paper at the Blue Sky Track, SDM Conference 2025. (May, 2025)

- **DAIR3 Program (Mississippi):** Selected for the Data and AI Intensive Research with Rigor and Reproducibility (DAIR3) program with a full scholarship, to be held at Jackson State University in Jackson, Mississippi. (March, 2025)

- **DAAD AInet fellowship (Germany):** Selected as a DAAD AInet fellow for the Postdoc-NeT-AI, where I will be participating in the virtual networking week, along with organizational support for a one-week networking visit to Germany and the opportunity to apply for short-term research funding to conduct research at a German institution. (October, 2024)

- **MINK WIC Conference Poster Competition:** Best Research Poster Award for my paper "Zero-shot Approach to Overcome Perturbation Sensitivity of Prompts" (accepted at ACL, 2023) at MINK WIC Conference. (October, 2023)

- **MINK WIC (Missouri, Iowa, Nebraska, Kansas Women in Computing):** Selected to represent Iowa State University at MINK WIC (An ACM celebration of Women in Computing). (October, 2023)

- **Guest Lecturer Invitation:** I have taught a graduate-level course at Iowa State University as a guest lecturer for COM S 571X (Responsible AI: Risk Management in Data Driven Discovery.) to teach the students about representation learning, transformer models, and future research directions to develop trustworthy machine learning methods for natural language processing.

- **Research Presentation Competition:** Second place at the 7th Annual Research Day Competition in the Computer Science Department at Iowa State University. (May, 2023)

- **Grace Hopper Celebration:** Selected to represent Iowa State University for the prestigious and competitive Grace Hopper Celebration. (September, 2022)

- **Travel Award:** One among 46 students selected by SIGKDD for the student travel award among all the applicants. (August, 2022)

- **Research Presentation Competition:** First place at the 6th Annual Research Day Competition in the Computer Science Department at Iowa State University. (May, 2022)

TEACHING

Teaching Assistant

Iowa State University   *Ames, IA, USA*

COM S 309 - Software Engineering Practises   *Fall 2020, Spring 2021, Fall 2021 and Spring 2022*

- This is a core, mandatory undergraduate course that introduces methods for managing software development and developing a large software from scratch to the end product with Android client, Spring back-end, concurrent features, etc.
- I have graded, and closely supervised 10 project teams (40+ students) from a class of 300+ student.
- In three out of four semesters, my supervised teams won the best project award in the whole class.

COM S 363 - Introduction to Database Management Systems   *Spring 2022 and Fall 2023*

- This course is an undergraduate course covering basic topics of database systems.
- I created, mentored, and graded the Assignments/Quizzes/Exams of the students in this course and others.

COM S 227 - Object Oreinted Programming   *Spring 2023*

- This course is an undergraduate course covering basic topics of Object Oriented Programming. I worked as a head teaching assistant for this course.
- I created, mentored, and graded the Assignments/Quizzes/Exams of the students in this course and others.

WORK EXPERIENCE

Data Science Intern   May 2023 - July 2023

Home Depot   *Atlanta, GA, USA*

- ∗ I worked on building a **sentence-based transformer personalized search ranker** using personalized signals to drive better engagements.
- ∗ The current search ranker in production uses a **tree-based model** which does not catch semantic similarity between the search query and product and does not consider personalized information about the customers. Therefore, the need was to create a generalizable modeling framework that can handle multi-modality and understand the correlation between different customers to mitigate the cold start problem.
- ∗ The proposed model improved MAP@8 by 20.38% MRR@8 by 21.27% NDCG@8 by 16.31% compared with the PROD baseline. Thus, improved relevancy for top-ranked products (i.e., higher precision).

Data Science Intern   May 2022 - Aug 2022

Epsilon Data Management, LLC   *Chicago, IL, USA*

- ∗ The current production system uses **Spark SQL**, which does not support atomic operations like **upsert, delete**, etc., so the system overwrites the entire table for every update, resulting in higher resource consumption and time.
- ∗ **Apache Hudi** is proposed as an alternative tool to solve issues like upsert and delete operations. We tested multiple workflows using Apache HUDI, including parameter tuning to validate its effectiveness by testing real-life scenarios with 1 billion data for up to 10% upsert operations, and the result shows a 37% increase in run time compared with Spark SQL.

Data Science Intern
Epsilon Data Management, LLC

May 2021 - Aug 2021

Chicago, IL, USA

* The ability to accurately classify individual names plays a crucial role in the quality of the final product. Yet this ability is hampered due to heterogeneity in data collection and validation.
* Current production methods validate the name data using **rule-based approaches**, limiting its ability to update or scale. Therefore, to alleviate this problem, we propose using **machine learning algorithms** on top of rule-based features and encoded features with 191 million data.

* Based on the classifiers' performance, **Random Forest** achieved a 91% F1 score with **oversampling** and **customized features**, explaining the need to incorporate better features to help the learning model better and faster.

Data Analyst Intern
Delaware North

May 2019 - March 2020

Buffalo, NY, USA

* Worked on training a **Logistic Regression model** to predict passenger occupancy across the US Airport and used its prediction to train another Logistic Regression model that predicts the number of transaction counts and labor force needed during various days in various kiosks or restaurants across airport terminals.
* Worked with **Beautiful Soup** to web scrape data like attendance, duration of the game, home and opponent team details, and other details for the games like NBA, NHL, NFL, and MLB from their official website, and used the data to train a **predictive model** to understand the trend of the crowd for all these games.

Engineer
Ericsson India Global Pvt. Ltd

Sep 2015 - Dec 2017

Mumbai, MH, India

* Two years of work experience as an Engineer in an **agile environment** with hands-on experience in designing and implementing Machine Learning Algorithms.
* Utilized alarms to perform **Anomaly detection** that captures unusual site behavior of base stations via **supervised and unsupervised machine learning models**. The developed models significantly reduced the alarms by 30%. The models helped prevent the faults from happening through early predictions and proactive decisions.
* Developed a **regression model** to identify the cause of weak network connection by performing a descriptive analysis to gain insights into the dataset, summary statistics, and analyzing features impacting the target correlation among variables. The developed regression model achieved a 10% more precise prediction than the previous year.