Laiba Mehnaz

+1 929 584 4743| lm4428@nyu.edu | Website | linkedin | github | Earliest start date: October 2023

Education

New York University, Tandon School of Engineering May 2023
Master of Science in Computer Science
Relevant coursework: Design and Analysis of Algorithms, Machine Learning, Deep Learning

Delhi Technological University, New Delhi, India May 2020
Bachelor of Technology in Software Engineering
Relevant coursework: Data Structures in Java, Database Management Systems, Object Oriented Programming

Technical Skills

Languages: Python, Java, C/C++, R, SQL
Technologies: PyTorch, TensorFlow, AWS EC2, AWS SageMaker, HPC, SLURM, Singularity, Keras, Numpy, Pandas, Scitkit-Learn, Git, Shell Scripting, Linux, SpaCy, Data Visualization, Tableau, XGBoost

Work Experience

Applied Scientist Intern, Natural Language Processing May 2022 – Aug 2022
Amazon *Seattle, USA*

- Built a pipeline to experiment with different model architectures and tokenizers for domain adaptation of LLMs to tasks/domains/datasets, using PyTorch and Huggingface transformers on AWS.

Assistant Research Scientist July 2023 – Present
New York University *New York, USA*

- Developing scalable R packages for Generative Adversarial Networks, Monte Carlo simulation, and statistical analysis(regression) on NYU HPC(High-performance computing cluster) for use cases in Econometrics. [**blog**]

Research Graduate Assistant(Open-Source Software Development) Mar 2023 – May 2023
New York University, VIDA lab *New York, USA*

- Developed Huggingface support for the scikit-learn based alpha-automl library to use any LLM for feature extraction and model training, on NYU high-performance computing cluster(HPC). [**code**]
- Provided unit tests for the same.

Research Assistant, Natural Language Processing June 2020 – June 2021
MIDAS lab, IIIT-Delhi *New Delhi, India*

Conversation summarization and translation, conversational AI *[**paper**]*

- Led a team of 8 students in data annotation to release a large-scale dataset for code-switched Hindi-English conversations.
- Built a Python framework using object-oriented design to compute and visualize statistical code-mixed metrics for analyzing Hindi-English code-mixed conversations. [**code**]
- Trained and fine-tuned several deep learning language models (LLMs) to summarize and translate conversations written in code-mixed Hindi-English to English.
- Created an evaluation framework for summaries generated by these LLMs, to show the limitations of the models for conversational data and code-switched data.

Probing the embeddings of large language models for domain robustness *[**paper**]*

- Created a pipeline for data cleaning, data preprocessing, and training a classifier, at each layer of BERT for the task of probing.
- Conducted an analysis using the probing results to show the quality of the pretrained word embeddings when used in a different context/domain/dataset compared to the the training dataset.

Leadership Experience

Research and Development Group Lead Jan 2023 – Present
AI for Scientific Research, New York University(part-time) *New York, USA*

- Leading the AIfSR group at NYU, by establishing collaborations with professionals in natural sciences to help with AI/ML, interviewing students, training and supervising students in product/paper delivery in a cross-disciplinary environment.
- Created the tutorial for using LLMs using Huggingface on NYU HPC cluster for feature extraction. [**tutorial code**]
- Presented and showcased group's work at NYU's annual Research Expo 2023.

Publications

"GupShup: An Annotated Corpus for Abstractive Summarization of Open-Domain Code-Switched Conversations."
EMNLP 2021, Dominican Republic.
***Laiba Mehnaz**, Debanjan Mahata, Rakesh Gosangi, Uma Sushmitha Gunturi, Riya Jain, Gauri Gupta, Amardeep Kumar, Isabelle Lee, Anish Acharya, Rajiv Ratn Shah.* [**paper**] [**code**]

"Analyzing the Domain Robustness of Pretrained Language Models, Layer by Layer."
AdaptNLP at EACL 2021.
*Abhinav Ramesh Kashyap, **Laiba Mehnaz**, Bhavitvya Malik, Abdul Waheed, Devamanyu Hazarika, Min-Yen Kan, Rajiv Ratn Shah.* [**paper**]

"Automatic classification of tweets mentioning a medication using pre-trained sentence encoders."
SMM4H at COLING 2020, Italy, Spain.
Laiba Mehnaz. [**paper**] [**poster**]

"Using Transfer Learning for detecting drug mentions in tweets."
ICT4SD 2020, Goa, India.
***Laiba Mehnaz** and Rajni Jindal.* [**paper**]

"MIDAS@SMM4H-2019: Identifying Adverse Drug Reactions and Personal Health Experience Mentions from Twitter."
SMM4H at ACL 2019, Italy.
*Sarthak Anand, Debanjan Mahata, Haimin Zhang, Simra Shahid, **Laiba Mehnaz**, Yaman Kumar, Rajiv Ratn Shah.* [**paper**] [**poster**] [**code**]

"MIDAS at SemEval 2019 Task 6: Identifying Offensive Posts and Targeted Offense from Twitter."
SemEval at NAACL HLT 2019, Minneapolis, USA.
*Haimin Zhang, Debanjan Mahata, Simra Shahid, **Laiba Mehnaz**, Sarthak Anand, Yaman Kumar, Rajiv Ratn Shah, Karan Uppal.* [**paper**]

"MIDAS at SemEval-2019 Task 9: Suggestion Mining from Online Reviews using ULMFiT."
SemEval at NAACL HLT 2019, Minneapolis, USA.
*Sarthak Anand, Debanjan Mahata, Kartik Aggarwal, **Laiba Mehnaz**, Simra Shahid, Haimin Zhang, Yaman Kumar, Rajiv Ratn Shah, Karan Uppal.* [**paper**]

"Identification of Emergency Blood Donation Request on Twitter."
SMM4H at EMNLP 2018, Brussels, Belgium.
*Puneet Mathur, Meghna Ayyar, Sahil Chopra, Simra Shahid, **Laiba Mehnaz**, Rajiv Ratn* Shah.[**paper**]