Debrup Das

⚲ Amherst, MA   ✉ debrupdas@umass.edu   in DebrupDas   ⌗ Debrup-61

Education

Ph.D. *Computer Science, University of Massachusetts, Amherst, MA, USA*   Sept 2024 – Ongoing
- **Advisor:** Prof. Negin Rahimi ↗
- **Member of CIIR Lab** (Center of Intelligent Information Retrieval ↗)
- **Anticipated Graduation Date:** September 2030

M.Sc. *Mathematics and Computing, Indian Institute of Technology Kharagpur, India*   Aug 2019 - Apr 2024
[*Integrated Master of Science 5 yr*]
- **Thesis Advisor:** Prof. Somak Aditya ↗
- **CGPA:** 9.00/10

Research Interests

My research focuses on building retrievers capable of complex reasoning to identify relevant documents, with a particular interest in retrieval for solving challenging mathematical reasoning problems. I am also interested in developing multi-agent and tool-augmented systems using LLMs, enabling effective coordination between components to tackle hard reasoning tasks.

Experience

Research Assistant, University of Massachusetts, Amherst, CIIR   Research Assistant April 2025 - Current
- Research on developing strong retrievers through RL training in an agentic setting, guided by rewards specifically designed for retrieval.
- Mentored by Prof Negin Rahimi ↗.

Research Assistant, University of Massachusetts, Amherst, CIIR   Research Assistant Sep 2024 – April 2025
- Research on developing reasoning-intensive retrievers from mathematical data (without any notion of document relevance), which shows strong generalizablity to diverse domains.
- Mentored by Prof Negin Rahimi ↗.
- Accepted as Main Conference Paper at **EMNLP 2025**, China

Rakuten India, Language and Speech Processing Team   Research Intern May 2023 – Dec 2023
- Research on *tool-augmented LLMs* for complex math reasoning.
- Integrating LLMs with symbolic tools such as *Python*, *Wolfram Alpha* and web search with *Bing API*.
- Accepted as a Main Conference Paper at **NAACL 2024**, Mexico City.

Indian Institute of Technology, Kharagpur (in collaboration with **MBZUAI**),

Research Intern
Dec 2023 – May 2024

- Research on *sensitivity* of LLMs to different segments of the input, in classification tasks such as hate speech, sentiment analysis, etc.
- *Adversarial example generation*, *attacks* using the notion of sensitivity.
- Mentored by Prof Monojit Choudhury ⬀ and Prof Somak Aditya ⬀ and funded by Microsoft AFMR ⬀
- Accepted as a Main Conference Paper at **NAACL 2025**, New Mexico.

McGill University, Montreal, Genome Quebec Innovation Center

Research Intern
May 2022 – Dec 2022

- Research on genome-wide association studies (*GWAS*) to detect rare disease causing variants.
- Mentored by Prof Simon Gravel ⬀ and funded by the MITACS Globalink Research Scholarship ⬀.

Indian Institute of Technology, Kharagpur

Research Intern
May 2021 – Dec 2021

- Research on deep learning methods, using tree-based models with nature-inspired optimization using genetic algorithms and evolutionary algorithms.
- Mentored by Prof Nirupam Chakraborti, Czech Technical University ⬀.

Publications

RaDeR: Reasoning-aware Dense Retrieval Models

May 2025

Debrup Das, Sam O' Nuallain, Negin Rahimi

ArXiv Paper Link ⬀, EMNLP 2025 Main Conference
Website ⬀

MATHSENSEI: A Tool-Augmented Large Language Model for Mathematical Reasoning

June 2024

Debrup Das, Debopriyo Banerjee, Somak Aditya, Ashish Kulkarni

NAACL 2024, Main Conference ⬀, MATHSENSEI featured on *Rakuten News* ⬀

SMAB: MAB based word Sensitivity Estimation Framework and its Applications in Adversarial Text Generation

April 2025

Saurabh Kumar Pandey, Sachin Vashistha, **Debrup Das**, Somak Aditya, Monojit Choudhury
NAACL 2025, Main Conference ⬀

Languages and Technologies

Languages: Python, Bash, Matlab, R, SQL, C, C++

Libraries: Pytorch, Transformers, Numpy, Scikit-Learn, Scipy, Sympy

Tools: Wolfram Alpha, Git, Latex, HuggingFace

Responsibilities

- Teaching Assistant, UMass Amherst, Fall 2024, **Search Engines** (CS 446) by Prof James Allan ↗

- Teaching Assistant, UMass Amherst, Spring 2025, **Search Engines** (CS 446) by Prof Negin Rahimi ↗

- Teaching Assistant, UMass Amherst, Summer 2025 Online, **Practice and Applications of Data Management** (CS 345) by Prof Gordon Anderson ↗

- Former member of National Social Service Scheme, India (NSS India)

English Proficiency

TOEFL: Total: 112/120, Speaking: 30/30, Reading: 29/30, Listening: 27/30, Writing: 26/30

Relevant Coursework/Certifications

Coursework: *Natural Language Processing, Information Retrieval, Neural Networks, Machine Learning, Graphical and Generative AI models, Probability and Statistics, Stochastic Processes.*

Certifications: *Machine Learning Stanford University by Andrew Ng, Deep Learning by deeplearning.ai, Python for Everybody by UMichigan*