Rishabh Adiga

University of Illinois Urbana-Champaign

rishabhadiga.github.io 🔽 rishabh.adiga@gmail.com 间 linkedin.com/in/rishabhadiga

Education

University of Illinois Urbana-Champaign (UIUC); CGPA - $4.0/4.0$	Aug 2023 – May 2025
M.S. in Computer Science, Research Assistant and Teaching Assistant	Champaign, USA
Indian Institute of Technology Madras (IIT); CGPA - 9.48/10	2019 - 2023
B. Tech in Electrical Engineering (In top 5 out of 143 students), Minor in Artificial Intellige	e nce Chennai, India
Deeksha CFL PU College; 99.5/100	2017 - 2019
Department Of Pre-University Education, Karnataka	Bangalore, India

Publications

- Rishabh Adiga, Lakshminarayanan Subramanian, Varun Chandrasekaran; Designing Informative Metrics for Few-Shot Example Selection, Association for Computational Linguistics ACL 2024 [Paper]
- Rishabh Adiga, Besmira Nushi, Varun Chandrasekaran; Attention Speaks Volumes: Localizing and Mitigating Bias in Language Models, Association for Computational Linguistics ACL 2025 Paper under review with scores of 4, 3.5, 3.5, 3.5 at ARR]

Research & Professional Experience

ICL for reasoning in VLMs (Microsoft Research)

Dr. Vibhav Vineet (Microsoft), Prof. Varun Chandrasekaran (UIUC)

• Ongoing research which aims to explore in-context learning capabilities of VLM's for Spatial Understanding tasks (such as depth and height understanding).

Bias Mitigation in LLMs (Microsoft Research)

Dr. Besmira Nushi (Microsoft), Prof. Varun Chandrasekaran (UIUC)

- Researched the emergence of bias in Large Language Models (LLMs) using ambiguous comparative prompts, focusing on the role of attention mechanisms in bias formation.
- Introduced **ATLAS** (Attention-based Targeted Layer Analysis and Scaling), a two step novel technique involving 1) Localization of bias to specific layers of an LLM and 2) Mitigation of bias by scaling attention scores in these specific layers. (Paper submission to ACL 2025)
- Validated the effectiveness of ATLAS through extensive experiments across multiple models and datasets, demonstrating consistent improvements in bias mitigation with an average improvement of **0.28 points** in Exponential Bias Score.

LLMs for Privacy Policy Analysis

Prof. Lakshminarayanan Subramanian (NYU), Prof. Varun Chandrasekaran (UIUC)

- Contextual Integrity (CI) is a framework used for analysis of information flow in privacy policies. The classification task that is involved in this has currently been analyzed using improved semantic role labeling (**Paper**).
- My research has allowed the use of novel method for prompt selection through a few shot methodology to perform this classification task using LLMs (Paper accepted at ACL). Using models like LLaMA3, we created an auto tagging system for this task which is cost efficient.
- The bigger picture is to develop a **first order logic** for privacy policies using CI as its basis and then perform longitudinal analysis on these policies.

Evaluating Mathematical Reasoning Chains

Prof. Heng Ji (UIUC)

- Existing methods for CoT(Chain of thought) evaluation either perform poorly on math-based tasks or fail to measure the logical correctness of steps effectively since they only check math calculations.
- To bridge this gap, my team under the guidance of Professor Heng Ji trained a set of metrics using **Contrastive** Learning and Direct Preference Optimization on LLMs resulting in an 8% gain on an average in correlation scores.

Flipped.ai Research Internship

Prof. Lakshminarayanan Subramanian (NYU)

- Developing alternate methods for Question Answering called **Templatized Question Answering** and benchmarked it for various LLMs.
- Additionally, we created a fully functional document information extractor fine-tuned for CVs and Job descriptions using several natural language processing techniques. This work has been patented.

April 2024 - September 2024

November 2024 - Current

Sept 2023 - Current

August 2023 - Dec 2023

April 2023 – Aug 2023

Semantic segmentation in Histopathological images

Deep Learning Research Internship under Prof. T K Srikanth, Prof. Ramesh Kestur (IIITB)

- Research and development of a deep learning model implemented using a **novel version of UNet architecture** for early detection of necrosis through semantic segmentation at International Institute of Information Technology Bangalore.
- The new model surpassed the previously existing Conditional random field (CRF) model by 13.724% in the AUPRC metric (other metrics had a significant increase too).
- Creation of a high-quality ground truth dataset was a secondary task through data manipulation and augmentation techniques along with performance analysis.

Speech Technology and Handwriting Recognition

Prof. Hema Murthy (IITM)

- Developed a Speech-to-Text system for digits by using digit utterances. Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) was trained for this purpose and achieved accuracies of 93.33% and 98.3%, respectively.
- Also developed a handwritten Telugu character recognition system using DTW, HMM, and ANN. Achieved accuracies of 94%, 97%, and 98%, respectively.
- After experimenting and researching, the **number of symbols** used for the HMM model was set based on the **number** of phonemes (44) in the English language for the digit utterances and the number of different patterns across all the letters.

Wells Fargo Internship

Software Developer managed by Mr. Sunil Agarwal

- Implemented a production quality Trade Order Matching Engine modularized into UI, server, matching engine and database with 97% code coverage.
- Worked primarily on the back end of the engine and created a **novel data structure (based on Red-Black trees)** that enables matching of orders (and other operations such as insertion and amends) to be performed in O(1) time.

Google Research Week

Selected for discussions with esteemed researchers on seminal papers in AI (Primarily in CV and NLP)

Achievements

2021 Bajaj TORQ Engineering Quiz Awarded 2nd place in this engineering contest held between the top IITs 2019 Indian National Physics Olympiad (INPhO) Represented my state for this prestigious Olympiad

2019 **IIT-JEE 99.90034 percentile** out of 930k students from all over India (rank of 1315)

2019 COMEDK Exam (B.Tech entrance exam) 1st rank in state and 4th rank in the country (58k students)

2019 KCET (B.Tech entrance exam) 13th rank in this nation wide exam for engineering aspirants(194k students) 2018 National Standard Examination in Physics (NSEP) National Finalist and was awarded a certificate of merit for being placed in the top 1% in the country

Relevant Coursework

Computer Science: Advanced Topics in NLP, LLMs Post-pretraining, Pattern Recognition and Machine Learning, Artificial Intelligence, Natural Language Processing, User Centered Machine Learning, Reinforcement Learning, Non Linear Optimization, Convolutional Neural Networks for Visual Recognition, Advanced Topics in Security Privacy and Machine Learning, Introduction to Computer Networks, Introduction to Programming, Applied Programming Lab, Data Structures and Algorithms, Computer Organization, Database Systems

Skills

Programming Languages: Python, Matlab, Mathematica, HTML, CSS, C++ Frameworks/Libraries: PyTorch, NumPy, Pandas, Apache Kafka, Spring Boot, Git, MySQL, JUnit, React

Positions of Responsibility

Teaching Assistant

TA for CS225 - Data Structures course at UIUC. Conducting lab sessions and office hours.

Paper Reviewer for ICLR 2025

Responsible for reviewing papers submitted to ICLR

Centre For Innovation (CFI) Project Member in the accelerator for ray tracing group (Software team)

National Cultural Appreciation (NCA)

Dramatics community member

ICLR July 2021 - May 2022 Indian Institute of Technology Madras

October 2024 – December 2024

August 2023 – Current

UIUC

August 2019 - May 2020 Indian Institute of Technology Madras

June 2021 - Dec 2021

February 2022 – May 2022

May 2022 – July 2022

Jan 2023