VIBHHU SHARMA

vibhhus@cs.cmu.edu GScholar vibhhusharma.github.io vibhhu-sharma

EDUCATION

Carnegie Mellon University, *Pittsburgh, PA* *Aug 2023 - Dec 2024*
Master of Science in Machine Learning; ***GPA: 4.0/4.0***

Indian Institute of Technology (IIT), Madras, *Chennai, India* *Aug 2019 - Jun 2023*
Bachelor of Technology in Electrical Engineering; ***CGPA: 9.35/10***

SCHOLASTIC ACHIEVEMENTS

- Ranked **6 out of 121** students in the Electrical Engineering Department. *2023*
- Secured **All India Rank 539** in **JEE (Advanced)** out of **200,000+** candidates *2019*
- Secured **All India Rank 421** in **JEE (Mains)** out of **1.5 million+** candidates *2019*
- Recipient of the prestigious **KVPY (Kishore Vaigyanik Protasahan Yojana)** scholarship in the SX stream with an **All India Rank 421** out of **50,000+** students *2019*
- Placed among the **top 300** students in the country in the National Standard Examinations in Physics and Chemistry, as a part of the **International Olympiads** selection procedure *2019*

PUBLICATIONS

- **Vibhhu Sharma**, Neham Jain, and Gaurav Sinha: **Counterfactual Explanations for Visual Recommender Systems**, The Web Conference 2024 (**WWW 2024**) [Paper] [Video]
- Khurram Yamin, **Vibhhu Sharma**, Edward Kennedy, Bryan Wilder: **Accounting for Missing Covariates in Heterogeneous Treatment Estimation**, (in submission) [Paper]
- **Vibhhu Sharma**, Shantanu Gupta, NJ Akpinar, Zachary Lipton, Liu Leqi: **A Unified Causal Framework for Auditing Recommender Systems for Ethical Concerns**, (presented at FAccTRec Workshop RecSys 2024, in submission) [Paper]
- **Vibhhu Sharma**, Bryan Wilder, **Comparing Targeting Strategies for Maximizing Social Welfare with Limited Resources** (in submission) [Paper]

RESEARCH EXPERIENCE

Comparing Targeting Strategies for Maximizing Social Welfare with Limited Resources | CMU Feb 2024 - Oct 2024
Guide: Prof. Bryan Wilder *Pittsburgh, PA*

- Analyzed data from real world RCTs in varied settings to compare the efficacy of targeting interventions based on **baseline risk** vs **biased** estimates of **treatment effect** after artificially introducing different levels of confounding.

Accounting for Missing Covariates in Heterogeneous Treatment Estimation | CMU Apr 2024 - Sept 2024
Guide: Prof. Bryan Wilder *Pittsburgh, PA*

- Developed novel statistical methodology to estimate heterogeneous treatment effects when **generalizing** from study populations to target populations with **previously unobserved covariates**.
- Derived provably **tight bounds** on conditional treatment effects using **ecological inference** techniques.
- Created bias-corrected estimator achieving $O(1/\sqrt{n})$ convergence rates and asymptotic normality.

A Unified Causal Framework for Auditing Recommender Systems | CMU Sep 2023- May 2024
Guide: Prof. Zachary Lipton *Pittsburgh, PA*

- Developed a general causal framework for defining and categorizing recommender system auditing metrics.
- Proposed future and past **reachability** & **stability** as metrics to audit user agency in dynamic recommendation processes.
- Provided gradient-based and black-box approaches for computing proposed metrics under different access levels.

Natural Language Counterfactual Generation for Indic Languages | Bachelor Thesis, IIT Madras Jan 2023- May 2023
Guide: Prof. Mitesh Khapra *Chennai, India*

- Created a flexible counterfactual generator for Indic Languages with **customizable perturbations**.
- Proved **counterfactual augmentation**'s value in NLP tasks like sentiment analysis and paraphrase identification.

Deep Learning for Extreme Multilabel Classification (XMC) | Aalto University Jun 2021-Nov 2021
Guide: Prof. Rohit Babbar *Espoo, Finland*

- Devised a model that made use of a deep Probabilistic Label Tree for label clustering and a Graph Convolutional Network based on document-document similarity for label ranking to assign correct labels to short text documents.

Work Experience

Machine Learning PhD Engineer Intern | **Instacart** May 2024-Aug 2024
Manager: Shishir Kumar Prasad *San Francisco, CA*

- Reduced sequence recommendation latency by **29.6%** using approximate nearest neighbor search for candidate retrieval.
- Improved recall for tail end retailers by **3%** via retailer-specific candidate retrieval using exact nearest neighbor search.
- Boosted overall Recall@200 by **1.5%** after testing/implementing multiple approaches for pretraining item embeddings.

Research Intern | **Adobe Research** May 2022-Oct 2022
Guide: Dr. Gaurav Sinha *Bangalore, India*

- Proposed a method to generate counterfactual explanations for a multimodal recommender system's recommendations.
- Developed an algorithm to identify the minimal change in an item's image to remove it from a user's recommended list and used CLIP to connect the perturbed image features to textual features in order to lend meaning to the perturbations.
- **Outperformed the existing state of the art** by 4% on **Explanation Fidelity** and 26.5% on **Explanation Number**.

Key Courses

- **Machine Learning:** Advanced Introduction to Machine Learning (10715) | Deep Learning for Imaging | Deep RL and Control (10703) | Multi-Armed Bandits | Probabilistic Graphical Models (10708)
- **Mathematics:** Probability and Mathematical Statistics (36700) | Linear Algebra | Convex Optimization (10725)
- **Programming:** Numerical Methods | Design and Analysis of Algorithms | Applied Programming Lab
- **Miscellaneous:** Introduction to Game Theory | French | Principles of Economics

Key Technical Projects

Using LLMs to enhance Graph Learning on Text Attributed Graphs Feb 2024-May 2024
Guide: Prof. Andrej Risteski [Link] *Pittsburgh, PA*

- Used LLMs to enhance node information in text attributed graphs, by using them for text augmentation and encoding.
- Benchmarked the method on 4 popular TAG datasets, beating standard TAG methods in both low and high label settings.

Multi-Armed Bandit in a game of Cricket Mar 2022- Apr 2022
Course Project under Prof. Chandrashekar Lakshmi Narayanan *Chennai, India*

- Used the Upper-Confidence Bound(UCB) algorithm to decide effective batting and bowling strategies in a game of cricket.

Software Engineer, Team Anveshak Apr 2020- Aug 2021
Mars Rover Team, IIT Madras *Chennai, India*

- Implemented algorithms for autonomous navigation, path planning and object detection on a ROS Based Framework for a rover capable of withstanding Mars-like conditions and carrying out scientific tasks effectively.
- Tested approaches to the above tasks extensively using Gazebo and RViz.

Analysis of Recommendation Systems May 2020- Jul 2020
vRhythms Software Pvt Ltd *Chennai, India*

- Worked in a team of four to analyze recommendation algorithms' performance on ranking metrics.
- Optimized the performance of traditional collaborative filtering & matrix factorization on ranking metrics by 22%.
- Analyzed models' susceptibility to popularity bias & cold start issue using novelty/coverage metrics.

Skills

- **Languages**: Python, Java, Bash, C++, MATLAB, C, Octave
- **Web Development**: HTML5, CSS3, Javascript
- **Data Analysis**: MATLAB, Octave, NumPy, Pandas, Matplotlib, Keras, TensorFlow, PyTorch
- **Other Libraries and Tools**: ROS, Eagle, Arduino, LaTeX

Extra Curricular Activities

- Served as a **reviewer** for **ICLR 2025.** *2024*
- Organized a department-wide quiz night for the Machine Learning Department at Carnegie Mellon University. *2024*
- Led a team of 50 students as the **Executive Editor** for The Fifth Estate, the official student news body of IIT Madras. *2022-23*
- Regularly **participated in and conducted** quizzes all over India as a part of the IIT Madras quiz contingent. *2020-23*
- Conducted a public workshop on "**Python Algorithms for Robotics**" as a part of Shaastra 2020. *2021*
- Provided quality **mentorship** as a part of Avanti Fellows to **underprivileged students** in JNV Puducherry with regard to their academics and entrance exam preparation. Both students cleared JEE Main-2020 with >**99 percentile**. *2019-20*