Anamika Lochab

(732) 895-5816 | anamikalochab@gmail.com | anamikalochab.github.io/AL/ | github.com/AnamikaLochab

EDUCATION

Purdue University, West Lafayette 09/2024 – 05/2029
PhD Computer Science *IN, USA*
- **Research Interests**: Alignment and stable post-training of large language models, with a focus on probabilistic methods and uncertainty quantification for trustworthy AI.
- **CGPA**: 4.00/4.00
- **Advisor**: Ruqi Zhang

Rutgers University, New Brunswick 09/2022 – 05/2024
MS Computer Science *NJ, USA*
- **CGPA**: 4.00/4.00, ***Outstanding Project Award***
- **Thesis**: Fine-Tuning Large Language Models, **Advisor**: Yongfeng Zhang

Vellore Institute of Technology, Bhopal 07/2018 – 07/2022
Bachelor of Technology in Computer Science and Engineering *MP, India*
- **CGPA**: 9.47/10, ***Gold Medalist(university topper)***

Publications

* denotes equal contribution

- **VERA: Variational Inference Framework for Jailbreaking Large Language Models** *NeurIPS 2025*
  Anamika Lochab*, Lu Yan*, Patrick Pynadath*, Xiangyu Zhang, Ruqi Zhang.
- **Energy-Based Reward Models for Robust Language Model Alignment** *COLM 2025*
  Anamika Lochab, Ruqi Zhang.
- **Cascade Reward Sampling for Efficient Decoding-Time Alignment** *COLM 2025*
  Bolian Li*, Yifan Wang*, **Anamika Lochab***, Ananth Grama, Ruqi Zhang.
- **VERA-V: Variational Inference Framework for Jailbreaking Vision-Language Models** *Under Review*
  Qilin Liao*, **Anamika Lochab***, Ruqi Zhang.

RESEARCH EXPERIENCE

Uncertainty Quantification in Automatic Vial Inspection 09/2025 – Present
Dr. Ruqi Zhang, Purdue University

Certified Robustness to AIP attacks in deep learning based Recommenders 04/2023 – 12/2023
Dr. Hao Wang, Machine Learning Group, Rutgers University
- Developed an approach to certify robustness against Adversarial Item Promotion (AIP) attacks on deep learning based recommender systems using auxiliary information i.e, textual item descriptions. Adapted continuous optimization techniques to introduce perturbation in embedding layers for gradient descent over discrete data.

Bias Mitigation in Large Language Models 08/2023 – 04/2024
Dr. Yongfeng Zhang, The Wise Lab, Rutgers University
- Engineered a multi-faceted adaptation strategy encompassing fine-tuning, prompt engineering, and instruction tuning to substantially elevate fairness and mitigate inherent biases in Large Language Models (LLMs).

WORK EXPERIENCE

SmartBridge Educational Services Private Limited 07/2021 – 08/2021
Data Analyst Internship
- Created a website to detect and predict e-payment phishing websites.

Excavate Research and Analysis 04/2019 – 06/2019
Summer Internship
- Worked on dynamic real-time web development in Angular and implemented statistical weighting and data analysis for various clients like Orange mobile company, VEON Bangladesh, and Banglalink.

TEACHING EXPERIENCE

Purdue University 2024 - 2025
CS 251: Data Structures and Algorithms

Rutgers University 2023 - 2024
CS 206:Intro to Discrete Structures II · CS 210:Data Management for Data Science · CS 439:Intro to Data Science

Awards and Honors

NSF ACCESS Discover Project Award, 2025-2026 · NeurIPS Scholar Award and Travel Award, 2025 · COLM Travel Award, 2025 · Outstanding Project Award, Rutgers University, 2024 · Gold Medalist, Vellore Institute of Technology, Bhopal, 2022 · Reviewer for ARR, February, 2025