RAKSHIT NAIDU NEMAKALLU

Email ⋄ LinkedIn ⋄ Website ⋄ Google Scholar

OBJECTIVE

My research interests hover around topics in Ethical Machine Learning (ML), Trustworthy/Responsible Artificial Intelligence (AI), and AI for societal good. I'm interested in creating applications that have a direct impact on society through my research. Hence, this has led me to pursue understanding ML models through the following important questions : (1) How do we quantify/explain disparities that arise in decisions made by ML models? (2) How do we create novel algorithms which mitigate the above effects?

EDUCATION

Doctor of Philosophy (Ph.D.) in Computer Science and Engineering, Georgia Institute of Technology
2023 - (Expected) 2028

Master of Science (M.Sc.) in Information Technology (Privacy Engineering), Carnegie Mellon University
2021 - 2022
Selected Courses: Ethics in Machine Learning, Foundations of Privacy, Privacy Policy, Law and Technology (PPLT) and ML with Large Datasets

Bachelor of Technology (B.Tech.) in Computer Science and Engineering, Manipal Institute of Technology
2017 - 2021
Minor in Computational Mathematics
Selected Courses: Computational Linear Algebra, Distributed and Cloud Computing, Graph Theory and Matrices.

EXPERIENCE

Graduate Research Intern Apr 2023 - July 2023
Carnegie Mellon University *Pittsburgh, PA, USA*

- Working with Prof. Hoda Heidari as a Research Assistant in the Machine Learning Department (MLD) at CMU.
- My responsibilities entail of collecting, assimilating, and analyzing both qualitative and quantitative data from prior academic publications, with the goal of creating a tool that offers a pipeline-aware view of Fairness for Machine Learning to researchers and practitioners.

Visiting Research Scholar Jun 2022 - Aug 2022
Syracuse University *Syracuse, NY, USA*

- Working with Prof. Ferdinando Fioretto on topics related to Differential Privacy and Fairness in AI.

Application Engineering Intern Jan 2021 - Jul 2021
BlackRock *Gurugram, India (Remote)*

- Part of the Client-End Fund Reporting Team. Improved test coverage on FRED (Factsheet Reporting Engine and Distribution) and fixed code issues, blockers and bugs.
- Received an honourable mention for our internal hackathon project on "BlackRock's Cultural Heatmap" which provides a forum for both employees (to assess their mental and cultural well-being) and managers (to maintain a cultural pulse throughout the organization).

PUBLICATIONS & PROJECTS

An Empirical Study of Input Regurgitation: Probing Leakage, Bias and Hallucinations in Chatbot Responses for Privacy-Sensitive Applications Preprint

In this article, we aim to answer three key questions: (1) How much private information gets copied from in-context examples to the output?, (2) Are there any biases with regards to the sensitive attributes (such as gender) in the data that is copied and regurgitated? and (3) Does the model hallucinate in-context examples that do not really exist? (*Under Review*)

Toward Operationalizing Pipeline-aware ML Fairness: A Research Agenda for Developing Practical Guidelines and Tools Preprint

We consult the fair-ML literature to understand the progress to date toward operationalizing the pipeline-aware approach: we systematically collect and organize the prior work that attempts to detect, measure, and mitigate various sources of unfairness through the ML pipeline. We utilize this extensive categorization of previous contributions to sketch a research agenda for the community. We hope this work serves as the stepping stone toward a more comprehensive set of resources for ML researchers, practitioners, and students interested in exploring, designing, and testing pipeline-oriented approaches and guidelines to algorithmic fairness. (*Under Review*)

Can Causal (or Counterfactual) Representations benefit from Quantum Computing? Preprint

We discuss on how latent causal representations can possibly be modelled through quantum computing. (*Accepted as an extended abstract at Algorithmic Fairness through the Lens of Causality and Privacy (AFCI) workshop at NeurIPS'22*)

Pruning has a disparate impact on model accuracy Preprint

We show that accuracy disparities in pruned models arise due to the presence of two key factors: (1) disparity in gradient norms across groups, and (2) disparity in Hessian matrices associated with the loss function computed using a group's data. (*Accepted at NeurIPS'22*. 🏆 Spotlight Lightning Talk)

Fair Context-Aware Privacy Threat Modelling Preprint

We examine notions of fairness in privacy threat modelling due to different causes of privacy threats within a particular situation/context and that across contexts. (*Presented at PTM workshop at USENIX-SOUPS'22*)

Can Causal (and Counterfactual) Reasoning improve Privacy Threat Modelling? Preprint

We discuss what causal and counterfactual reasoning is and how this can be applied in the field of privacy threat modelling (PTM). (*Presented at PTM workshop at USENIX-SOUPS'22*)

Efficient Hyperparameter Optimization for Differentially Private Deep Learning Preprint

We study three different hyperparameter optimization approaches for DP-SGD to achieve the best privacy-utility tradeoffs. (Accepted at *PPML workshop at ACM CCS'21*)

Privacy Enabled Financial Text Classification using Differential Privacy and Federated Learning

We apply DP and FL on Financial Text data and provide results and intuition for the same. (*Accepted at ECONLP workshop at EMNLP'21*)

Benchmarking Differential Privacy and Federated Learning for BERT models Preprint

We benchmark BERT-based models with DP and FL on two Twitter datasets (Depression and Sexual harassments). We provide open source implementations for the same to accelerate Private NLP research. (Accepted at ML4Data workshop at ICML'21)

Towards Quantifying Carbon Emissions of Differentially Private Machine Learning Preprint

We quantify carbon emissions for DP-SGD in three different environments : NLP (News Classification), CV (MNIST Digit Classification) and RL (Cartpole problem). (*Accepted at SRML workshop at ICML'21*)

DP-SGD vs PATE: Which Has Less Disparate Impact on Model Accuracy? Preprint

We compare DP-SGD with PATE, another DP-based approach in terms of Fairness. We infer that as PATE uses a teacher-student setup where disjoint data is distributed among the teachers, it suffers less disparity than DP-SGD (due to diversity in training). (*Accepted at ML4Data workshop at ICML'21 and PPML workshop at ACM CCS'21*)

FedPerf: A Practitioners' Guide to Performance of Federated Learning Algorithms Publication

I worked with this team on extending their NeurIPS short paper with the Stragglers and Robustness experiments. We propose an empirical investigation on four prominent FL algorithms to discover the relation between the FL System Parameters (FLSPs) and their performance. (*Accepted for publication at PMLR*)

When Differential Privacy Meets Interpretability: A Case Study Preprint — Poster

We investigate the tradeoffs between Differentially-Private SGD (DP-SGD) and Interpretability specifically through CAMs on the APTOS dataset. (*Accepted as extended abstract at RCV workshop at CVPR'21; full paper accepted at PPML workshop at ACM CCS'21*)

Improved variants of Score-CAM via Smoothing and Integrating Poster

We improve Score-CAM by adding smoothing and integration functions as suggested in the SmoothGrad and IntegratedGrad papers respectively. (*Accepted as extended abstract at RCV workshop at CVPR'21*)

FedPandemic: A Cross-Device Federated Learning Approach Towards Elementary Prognosis of Diseases During a Pandemic Preprint

We come up with a simple noise algorithm (inspired by Randomized Response and integrated with Federated Learning) to retrieve prominent COVID-19 symptoms in a privacy-preserving fashion.

(*Accepted at DPML and MLPCP workshops at ICLR'21*)

SS-CAM: Smoothed Score-CAM for sharper visual feature localization Preprint

We introduce Smoothing to the Score-CAM algorithm, which is a state-of-the-art CAM algorithm. Smoothing allows us to capture more features of the focused object in the image, which leads to better visually attributed results.

IS-CAM: Integrated Score-CAM for axiomatic-based explanations Preprint

We borrow the idea of integration from "IntegratedGrad" and combine it with Score-CAM to conduct faithfulness evaluations. IS-CAM performs better than SS-CAM and Score-CAM in terms of faithfulness evaluations, considering the VGG-16 as our baseline model.

TeleVital: Enhancing the quality of contactless health assessment Paper — News

Our team came 2nd in a pan-Indian hackathon called #CODE19 and won $5000 for this solution to detect vitals from the webcam itself, thereby promoting remote diagnosis during COVID-19. I worked on the Respiratory rate calculations via webcam and was responsible for documenting the entire project for presenting at the hackathon.

PROFESSIONAL SERVICES

- Teaching Assistant Quantum Computing Theory and Lab (11-860), Programming Quantum Computers (17617-A1), Quantum Circuit Mappings (17-620)
- Served on the Program Committee at the GenLaw workshop @ ICML'23
- Reviewer and Ethics Reviewer at NeurIPS'23
- Reviewer at the Algorithmic Fairness through the Lens of Causality and Privacy workshop at NeurIPS'22
- Talk at Comcast Cybersecurity team (headquartered in Philadelphia, PA) on "Context-Aware Privacy Threat Modeling".
- Served on the Program Committee as a reviewer at PPAI-AAAI'22, PPAI-AAAI'23.
- TEDxMAHE Countdown 2020 Speaker on *Federated Learning for Climate Change.* Event Link — Talk
- Manipal Conclave 2020 Student Speaker on *Privacy for ML.* Memento
- Poster Presented at PyCon India 2019 on *Secure and Private AI with PySyft.* Poster
  Volunteered at PyCon India 2020.

EXTRA-CURRICULAR ACTIVITIES

- Played for the CMU Badminton team in the Fall 2022 Eastern Collegiate MidAtlantic Conf (Badminton Tournament Regionals) held at University of Maryland, College Park in October 2022.
- Finished a full marathon (42 km) at Manipal Marathon 2020 with a timing of 6 hours and 33 minutes. Certificate