Abhinav Rao

☎ +1-(412)-251-8197
✉ abhinavr@andrew.cmu.edu, abhinav.797c@gmail.com
GitHub Profile
LinkedIn Profile

`https://abhinavrao.netlify.app`
Master's student @ Carnegie Mellon University

EDUCATION

Carnegie Mellon University, School of Computer Science — Pittsburgh, PA
Master of Science (AI & NLP) in Intelligent Information Systems — Expected Dec '24, GPA: 4.25/4
Coursework: Multimodal Machine Learning, Advanced Natural Language Processing;
Ongoing: Computational Ethics, Quantitative Evaluation of language technologies

Birla Institute of Technology and Science (BITS), Pilani — Hyderabad, India
Bachelor of Engineering in Computer Science — Feb '22, GPA: 9.26 / 10.0
Coursework: Machine Learning, Data Structures and Algorithms, Software Engineering, Compilers

EXPERIENCE

Microsoft Turing — Bangalore, India
Research Fellow — Aug 2022 - Jul 2023

- Curated datasets for Content-harm and Jailbreaks on Bing Chat to improve classifier performance by 5% and 17% (F1-score), for jailbreak and content-harm detection respectively.
- Posited a moral-alignment framework for cross-cultural fairness in LLMs, and evaluated their ethical reasoning capabilities. Published at EMNLP Findings 2023.
- Formalized and studied the jailbreak phenomenon. Ideated a taxonomy of jailbreaks and evaluated their effectiveness against different GPT-based Large Language Models (LLMs). In review at LREC-CoLING 2024. Preprint on arxiv

Microsoft Research — Bangalore, India
Research Intern — Jan 2022 - Jul 2022

- Designed a multilingual data augmentation tool for query expansion as part of Project LITMUS. Sped up the pipeline for Bing's Defensive team by 10x using a multilingual topic model and Approximate Nearest Neighbors (ANNS).

Nanyang Technological University (NTU) — Singapore
Research Intern — Jun 2021 - Dec 2021

- Researched and developed a BERT-based model for multilingual automatic punctuation restoration of Automatic Speech Recognition (ASR) text as part of SpeechLab, NTU. Beat the SOTA Chinese Model by 4.2% F1-score. Published at APSIPA'22. Code available.

Oracle Corporation — Bangalore, India
Software Developer Intern — Jun 2021 - Jul 2021

- Engineered a system to map bugs to their associated features, incorporating text-mining from large databases for bug-feature-customer analytics. Achieved a 75% accuracy on bug analysis.

PROJECTS

Multilingual Sparse Federated Learning — Carnegie Mellon University
Course project for Advanced Natural Language Processing — September 2022 - Present

- Analyzed the impact of parameter efficient training methods for Machine translation in a federated learning setting.
- Maintained unseen language performance in a heterogenous setting with sparse subnet selection and LoRA. Accepted as a workshop paper at MOOMIN, EACL'24.

Compositionality of Vision Language Models — Carnegie Mellon University
Course project for Multimodal Machine Learning — September 2022 - Dec 2022

- Finetuned BLIP-2 using a multi-objective loss using synthetic counterfactual image generation with 5% improvement on Winoground.
- Pipelined the text and image data augmentation using Dependency parsing, Object detection, and inpainting. Report. Code.

Malware classification — BITS Pilani, Hyderabad
Advisor: Prof. Barsha Mitra — Jun 2021 - Dec 2021

- Treated malware detection as an image classification problem, facilitating the development of a lightweight model using 1D-CNNs and histogram-based classifications.
- Achieved an F1-Score of nearly 98% on the Microsoft BIG Dataset using 1D CNNs. Paper in review at IEEE TETC. Preprint on arxiv

Auto Code Commenting — BITS Pilani, Hyderabad
Advisor: Prof. N.L. Bhanu Murthy — Jan 2021 - Dec 2021

- Enhanced existing code-commenting models, using an encoder-decoder architecture with Pointer Generator networks and coverage attention. Achieved a BLEU-4 score of 40. Code available.

SKILLS

Prog. Languages & Libraries: Python; C++; SQL; NumPy, SciPy; Pandas; PyTorch; Tensorflow; Keras; Scikit-learn;
General skills: Machine Learning (ML); Natural Language Processing (NLP); Deep Learning (DL); Neural network