Devansh Bhardwaj

github.com/fireshadow05 bhardwajdevansh398@gmail.com

Education

Indian Institute of technology Roorkee 2021 - Present
Bachelor of Technology in Electronics and Communication Engineering *Current CGPA: 8.144/10.0*
Senior Secondary School (Cambridge Court High School, Jaipur) 2019 - 2021
Specialization in Physics, Chemistry and Mathematics *Percentage: 94.8*

Skills and Interests

Languages and Tools: C/C++, Java, Python, Ros, LaTeX, Git/GitHub, Unix Shell
Libraries and Frameworks: Pytorch, Tensorflow, JAX, pandas, NumPy, Matplotlib, scikitlearn, RestAPI
Interests: Generative AI (specifically diffusion models), Reinforcement learning, Adversarial Machine Learning, Multimodal Learning,

Pre-Prints

Accelerated Smoothing: A Scalable Approach to Randomized Smoothing
Devansh Bhardwaj, *Kshitiz Kaushik, Sarthak Gupta* | *Arxiv* | *Code*

- Proposed a novel approach to address the compute-intensive nature of Monte Carlo sampling in randomized smoothing, replacing it with the training of a surrogate neural network.
- With various experiments on CIFAR-10 we showed the effectiveness of our approach in learning the robustness of the smoothened classifier.

Trust But Verify: A Survey of Randomized Smoothing Techniques
Anupriya Kumari,* ***Devansh Bhardwaj****, Sukrit Jindal*, Sarthak Gupta* | *Arxiv* |

- Reviewed the theoretical and fundamental foundations of **Randomized Smoothing**, highlighting both theoretical and application-based limitations and challenges of existing methodologies.
- This paper was a first of its kind in its attempt to provide a **concise summary** of randomized smoothing techniques.

Experience

Repello.ai | *ML Security Research Intern* March 2024 - Present

- Focused on Red Teaming applications that use Large Language Models (LLMs) to identify and exploit critical security vulnerabilities within them.

Indian Institute of Science, Bangalore | *Undergraduate Research Intern* | Hybrid Oct 2023 – Present

- At Professor Debasish Ghose's Lab, I am working on implementing computer vision and reinforcement learning-based algorithms for real-life robots.
- I am also working on setting up an environment in **NVIDIA's ISAAC Sim**, a realistic robotics simulator, followed by the validation of an algorithm that utilizes Kalman filters and Bayesian belief spaces for human-intent detection.

Projects

Behaviour and Content Simulation | *Inter-IIT Tech Meet 12.0* | *Code* Nov 2023 - Dec 2023

- Played key role in development of a multi modal LLM pipeline, which involved two tasks, first to simulate behaviour (likes) from the content of a tweet, second to simulate content (tweet text) from the tweet metadata.
- Tweet text generation utilised BLIP-2 for extracting visual information from images and videos, this information was then fed into a LLM along with other tweet meta-data, in the form of a prompt.
- Experimented with different techniques such as in-context learning and Wikipedia RAG.

AIKavach | *DSG. IITR* | *Code* March 2023 – May 2023

- Contributed significantly to the development of a flask based **web app** that certifies **robutsness of deep learning models** against adversarial attacks and returns a more robust model with a denoiser attached to the user provided model weights.
- Reviewed various techniques for robust radius certification and to decrease the time complexity of these techniques.
- Integrated Input specific sampling to Double Sampling Randomized smoothing for faster robust radius certification

Expert Answers in a Flash | *Inter-IIT Tech Meet 11.0* | *Code* Dec 2022 - Feb 2023

- Contributed to development of a domain-specific two stage retriever-reader based model for question-answering within 1000ms per query on a T4 GPU. Most notably our model took less than **2000ms** per query on a google colab CPU.

- Reviewed and experimented with various domain adaptation techniques for both retriever and reader models.
- Developed the final retriever model that combined results of BM25 and a Dense retriever using LEDR.

Reproducing Coin Flipping Neural Networks | *MLRC 2022* | *Code* Dec 2022 - Feb 2023

- Made a reproducibility report on the original research paper Coin Flipping Neural Networks that was published in ICML.
- Independently coded the entire novel CFNN architecture from scratch. Verified their claims, did further ablation studies and produced results beyond the original paper to get a better understanding of their work.

Achievements and Extra Co-curricular

Silver Medal, Inter-IIT Tech Meet 12.0
Bronze Medal, Inter-IIT Tech Meet 11.0
All India Rank 1410 in Jee Advanced 2021

Data Science Group | *Secretary* June 2022 – Present

Personal Side Projects

Graph Nets in JAX | *DSG. IITR* | *Code* March 2023

- Contributed to the Open Source blog series graph nets. Implemented a Graph Convolution Network using JAX.

Face Generator | **Code** March 2023

- Implemented a modified GAN Architecture from scratch for Anime Face Generation.