Harshvardhan Srivastava

New York City, NY | hvars.github.io
+1 (646)-407-2847 hs3447@columbia.edu linkedin.com/in/harshvardhansrivastava github.com/hvarS

EDUCATION

Columbia University **New York, NY**
Master of Science in Computer Science Expected Dec 2024
Awards: Google CS Research Mentorship Fellow 2023B, Outstanding Paper Award ACL 2024
Relevant Courses: Machine Learning, Databases, NLP Gen. Summ., AML Cloud, Reinforcement Learning, Adv. SLP

Indian Institute of Technology Kharagpur **Kharagpur, IN**
B.Tech in Electrical Engineering, Minor in Computer Science | GPA 9.03/10.00 May 2021
Relevant Courses: NLP, Artificial Intelligence, Algorithms, OS, Statistics and Stochastic Processes, Linear Algebra

WORK EXPERIENCE

Articulate AI **New York, NY**
Machine Learning Engineer Jun 2024 - Aug 2024
- Spearheaded patent description generation project, developed long-context methods for generation, improving model accuracy and coherence using ***RoPE scaling*** and transformer optimization with ***LoRA***
- Fine-tuned state-of-the-art models such as ***Llama3.1***, ***Mixtral 8x7b***, ***Gemma 2***, and ***Phi-3***, achieving a significant increase in model performance metrics, leading to more precise and legally sound patent drafts.
- Deployed scalable machine learning pipelines for training and inference reducing processing time by **30%**

Oracle Corporation **Bangalore, IN**
Member Technical Staff Jul 2021 - Aug 2022
- Created robust framework using ***Oracle PL/SQL*** for loading of transactional data in the Oracle Data Warehouses
- Formulated validation services for ***CX***, ***HCM*** and ***ERP*** module in Fusion Analytics Warehouse deploying Spring Boot
- Proposed and built an end-to-end automation framework in Python and Terraform deploying Oracle Cloud Infrastructure (OCI) Docker Images resulting in accelerated deployments by **8x** and collaborated with 2 cross-functional teams

Samsung R&D Insitute **Bangalore, IN**
Visual Intelligence Intern - Low Light Imaging May 2020 - Jul 2020
- Devised algorithm for deep burst denoising of **1 TB** image dataset extending the CNN with parallel recurrent networks to integrate information of all frames in burst-set; developed on Attention-enhanced Kernel Prediction networks (KPNs)
- Analysed images in ***4 channel bayer domain*** to enhance quality of output video for low light video denoising
- Attained SOTA results with **29.8 PSNR**(Peak Signal to noise Ratio) and **0.87 SSIM** (Structural Similarity Index)

PROJECTS & RESEARCH

RL Assisted Portfolio Management Bot | *Python, Flask, PyTorch, Tensorflow* **Jan 2024 - April 2024**
- Developed a trading bot using ***DDPG***, ***A2C***, ***PPO***, ***SARL*** algorithms on DJ30 and FX datasets to maximize cumulative rewards; created an environment incorporating relative price values and moving averages
- Achieved a Sharpe Ratio of **1.380**, reduced volatility to **11.5%**, and attained an overall return of **18%**

Roofline Analysis | *Python, AWS EC2, NVIDIA Nsight Compute, PyTorch Profiler* **Feb 2024 - Apr 2024**
- Analyzed the impact of model complexity and batch size on arithmetic intensity and achievable FLOPs across ***T4***, ***L4***, and ***V100*** GPUs. Identified bottlenecks affecting GPU efficiency in the forward pass, backward pass, and data loading
- Quantified the speedup in training time across different GPUs and achieved max performance of **2.7 TFLOPs**

AI Assisted Clinical Detection of Celiac Disease | *Python, Qt, PyTorch, Tensorflow, Flask* **Aug 2022 - Aug 2023**
- Constructed an end to end clinical computer vision based product with ***6 product features*** based on collaboration between AIIMS Delhi and IIT Delhi; developed multiple vision algorithms for clinical detection of celiac disease
- Integrated features in app for Tissue-Type Detection, Villi Tip Identification, Cell counting; achieved weighted average dice score for tissue type detection close to **0.72**, with a mAP score increase of about **12%**

Dynamic and Robust Server Allocation | *Java, Apache Kafka* **May 2019 - Jul 2019**
- Managed a multi-node multi partition Kafka Cluster to store the data from live stream data on Twitter accounts
- Collected, analysed native Kafka performance metrics with custom metrics devised to verify the robustness of cluster

Selected Publications

- Crafting Humor Datasets with Unfunny LLMs [**Outstanding Paper Award**] **ACL 2024**
- MMER: Multimodal Multi task learning for Speech Emotion Recognition **InterSpeech 2023**

TECHNICAL SKILLS

Languages: Python, C/C++, Java, LaTeX, Javascript **Developer Tools**: Git, Docker, Kubernetes, AWS, Terraform
Technologies/Frameworks: Oracle DB, MySQL, Django, Flask, Spring Boot, OpenCV, TensorFlow, Keras, PyTorch