Pranit Chawla

(412) 909-9852 | PranitChawla | linkedin.com/in/pranitchawla | ✉ pranitchawla98@gmail.com

Education

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Machine Learning | GPA: **4.24/4** — Dec 2023
Courses: Probability and Statistics, Adv. Intro. to Machine Learning, Convex Optimization
Probabilistic Graphical Models, Intermediate Deep Learning, Multi-modal Machine Learning

Indian Institute of Technology Kharagpur — Kharagpur, India
BTech & MTech in Electronics and Electrical Communication Engineering | GPA: **9.52/10** — Jun 2022
Minor in Computer Science and Engineering

Experience

Microsoft Corporation — Redmond, WA
Applied Scientist II — *Feb 2024 – Current*

- Collaborated closely with OpenAI's ChatGPT Agent team as an onsite-engineer to improve grounding, manipulation, and screenshot understand capabilities of o-series of models
- Performed large-scale online RL training with **600+ A100s**, and 2k+ Windows VMs and code execution containers to jointly train o3 on computer-use and power-point generation capabilities
- Curated a synthetic preference and SFT data generation pipeline using OpenAI CUA and o4-mini models to train Qwen2.5VL on GUI based web browsing tasks, surpassing UI-TARS 1.5 by **15%** on WebVoyager
- Designed a two-stage computer-using agent architecture with GPT-4o as the planner and fine-tuned Qwen2VL as the grounder to achieve **45%** success rate on **Windows Agent Arena**, improving baselines by 20%

Data and Applied Scientist Intern — *May 2023 – Aug 2023*

- Extended Bletchley, **Microsoft Turing's CLIP** style multi-modal **(vision and language)** model to enable seamless compatibility with variable-resolution images, without training on high-resolution images
- Boosted zero-shot performance on COCO retrieval by **5%** through the implementation of patch size randomization and leveraging sparsity techniques based on **mixture of experts** for vision transformer using **DeepSpeed**

Adobe Inc. — Noida, India
Research Intern — *May 2020 – Aug 2020*

- Investigated **text conditioned image retrieval** (TCIR) systems, demonstrating their ability to retrieve images closely resembling an original query image while adhering to applied constraints through natural language
- Enhanced existing TCIR models by **6% (R@10)** on three datasets by using a novel **cross-modal attention module** and a CNN + RNN **gated fusion technique** to generate fine-grained visio-linguistic features

Publications and Patents

[5] Scalable Data Synthesis for Computer Use Agents with Step-Level Filtering [Paper Link]
Preprint 2025 , Under Review

[4] Leveraging Diffusion Models for Test-time Adaptation via Pseudo-label Ensembling [Paper Link]
Accepted at Workshop on Distribution shifts at NeurIPS 2023

[3] HateProof: Are Hateful Meme Detection Systems really Robust? [Paper Link]
Accepted at TheWebConf'2023 (WWW'2023)

[2] Leveraging Style and Content Features for Text Conditioned Image Retrieval [Paper Link][Video Link]
Accepted at Proceedings of IEEE/CVF Conference on CVPR Workshops 2021 | Patent Filed at US-PTO

[1] SAC: Semantic Attention Composition for Text-Conditioned Image Retrieval [Paper Link]
Accepted at IEEE Winter Conference on Applications of Computer Vision WACV 2022 | Patent Filed at US-PTO

Research

University of Southern California — Los Angeles, CA (Remote)
*Advisor: Prof. Ram Nevatia, Department of Computer Science [**Paper Link**]* — *May 2021 – Aug 2021*

- Worked on **spatio-temporal video grounding** (STVG) to localize spatio-temporal tubes of queries in videos
- Revamped existing STVG methods and proposed a novel **video-text** transformer based architecture capable of performing STVG in a single pass by eliminating need for sliding windows or tube proposals

Technical Skills

ML/Data: PyTorch, Huggingface, DeepSpeed, LLamaFactory, vLLM **Software**: Python, Git, Linux, Flask