Shivank Garg

shivankgarg21@gmail.com | Shivank Garg | github.com/shivank21 | Google Scholar | Webpage

Education

Indian Institute of Technology Roorkee **Current CGPA- 8.73/10**
Bachelor of Technology in Data Science and Artificial Intelligence *June 2022 – Present*

Experience

Adobe **May 2025 – Present**
Research Intern, Media and Data Science Research (MDSR) Lab

- Working on the task of Video Virtual Try-On using Diffusion Transformers as a part of Adobe Firefly Services.
- Fine-tuned diffusion models with image conditioning and garment warping, achieving 39% improvement over SoTA.

Sony AI **July 2024 – April 2025**
Data Science Intern

- Designed and implemented a multimodal graph-based framework to improve video recommendations for SonyLIV.
- Conducted research at the intersection of Large Language Models and Diffusion Models for user-centric recommendation systems, including behavioral intent recognition and collaborative knowledge infusion in LLMs.

Carnegie Mellon University **February 2024 – May 2025**
Research Intern

- Worked on benchmark creation using model confidence and VQA reasoning to jailbreak vision-language models.

Selected Publications

- "Text2Arch: A Dataset for Generating Scientific Architecture Diagrams from Natural Language Descriptions" Under review at Association for the Advancement of Artificial Intelligence (AAAI) , 2026
- "Evaluating Reflective Behavioral Self-Awareness, Latent Policy Generalization and Faithful Reasoning in Post-Trained LLMs" Under review at Neural Information Processing Systems (NeurIPS), 2025. [Paper]
- "CCR: Complete Category Removal from Text-to-Image Diffusion Models" Under review at Neural Information Processing Systems (NeurIPS), 2025. [Paper]
- "Do Biased Models have Biased Thoughts?" Conference on Language Modelling (COLM), 2025. [Paper]
- "IPO: Your Language Model is Secretely a Preference Classifier" Association for Computer Linguistics (ACL) main track, 2025. [Paper]
- "Unmasking the Veil: An Investigation into Concept Ablation for Privacy and Copyright Protection in Images." Transactions on Machine Learning Research (TMLR). [Paper]

Projects

LLM Efficiency Challenge | *NeurIPS 2023 Challenge* | Problem Statement | Code

- Fine tuned LLMs on a single V100 GPU for Question Answering and Document, achieving a global rank of Top 20.
- Used huggingface transformers and PEFT for fine-tuning LLMs with 4-bit quantization utilising QLoRA.

Self Improving LLMs | Code

- Worked on LLMs that self-generate data and preferences for finetuning, enabling autonomous self-improvement.

Machine Unlearning | Code

- Applied parameter pruning and LoRA-based finetuning for machine unlearning on LLMs and classification models.

References

Prof. Durga Toshniwal
Head of Department of Data Science and AI, IITR | *Webpage*

Manish Gupta
Principal Applied Scientist at Microsoft | *Webpage*