Shivank Garg

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

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 aimed at integration into Adobe Express, and publication at CVPR 2025.

Sony Research **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.

Selected Publications

- "Evaluating Reflective Behavioral Self-Awareness, Latent Policy Generalization and Faithful Reasoning in Post-Trained LLMs" Under review at Neural Information Processing Systems (NeurIPS), 2025.
- "CCR: Complete Category Removal from Text-to-Image Diffusion Models" Under review at Neural Information Processing Systems (NeurIPS), 2025.
- "Do Biased Models have Biased Thoughts?" Under review at Conference on Language Modelling (COLM), 2025.
- "IPO: Your Language Model is Secretely a Preference Classifier" Association for Computer Linguistics (ACL) main track, 2025. [Paper]
- "Attention Shift: Steering AI Away from Unsafe Content." The Conference on Neural Information Processing Systems (NeurIPS) Workshop on Responsibly Building the Next Generation of Multimodal Foundational Models, 2024. [Paper]
- "Give me a hint: Can LLMs take a hint to solve math problems?" The Conference on Neural Information Processing Systems (NeurIPS) MATH AI Workshop, 2024. [Paper]
- "Unmasking the Veil: An Investigation into Concept Ablation for Privacy and Copyright Protection in Images." Transactions on Machine Learning Research. [Paper]
- "Snowy Scenes, Clear Detections: A Robust Model for Traffic Light Detection in Adverse Weather Conditions." ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2024. [Paper]

Projects

Machine Unlearning | *NeurIPS 2023 Challenge* | *Code*

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

References

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

Manish Gupta
Principal Applied Scientist at Microsoft | *Webpage*