Rajarshi Roy

Shyamnagar, India
royrajarshi0123@gmail.com Linkedin Google Scholar
Portfolio Github

Education

Kalyani Government Engineering College **August 2021 – July 2025**
B.Tech in Computer Science and Engineering *8.8 CGPA*
Final-year thesis: *ByDeWay: Boost Your multimodal LLM with DEpth prompting in a Training-Free Way*
Published at CVAM Workshop @ ICCV 2025 (This thesis was converted into a workshop paper)

Relevant Coursework

Artificial Intelligence, Machine Learning, Image Processing, Programming for Problem Solving, Object Oriented Programming, Data Structures and Algorithms, Design and Analysis of Algorithms.

Research Experience

Kalyani Government Engineering College, India **July 2024 - July 2025**
Undergraduate Thesis - Boost Your multimodal LLM with DEpth prompting in a training-free Way *Kalyani, India*

- Designed and led a **training-free, depth-aware prompting framework** to improve spatial reasoning in Multimodal Language Models (MLMs) without requiring expensive fine-tuning.
- Formulated core hypotheses and designed the experimental framework to systematically study how **Layered-Depth-Based Prompting (LDP)** influences depth understanding in MLLMs.
- Engineered a **zero-shot monocular depth estimation pipeline** using Depth Anything V2 combined with **percentile-based segmentation** (top 30%, middle 40%, bottom 30%) to partition images into foreground, mid-ground, and background layers.
- Developed a **region-specific captioning mechanism** using KOSMOS-2 to generate grounded captions for each masked depth layer, concatenating layer-wise descriptions into structured prompts.
- Implemented a **model-agnostic augmentation pipeline** that enhances input prompts with depth-enriched context, achieving compatibility across diverse MLM architectures (GPT-4o, Qwen2.5-VL, ViLT, BLIP).
- Achieved consistent accuracy improvements: GPT-4o's accuracy improved from 0.860 to 0.873, while Qwen2.5-VL's accuracy improved from 0.7267 to 0.9000 on POPE and 0.5007 to 0.6592 on GQA, demonstrating **hallucination mitigation and spatial reasoning enhancement**.
- Maintained a **fully reproducible codebase** and led our undergraduate team through experimental design, evaluation, and analysis.

Artificial Intelligence Institute of South Carolina **May 2024 - Present**
AI Research Intern *Kolkata, India (Remote)*

DPO-Kernels - Project (ACL Findings 2025) ***December 2024 - May 2025***

- Contributed to **DPO-Kernels**, a framework addressing semantic ambiguity in preference optimization by integrating kernel methods and diverse divergence measures into Direct Preference Optimization (DPO).
- Extended and modified the **original DPO codebase** to support the kernelized and divergence-generalized framework, enabling integration of Hierarchical Mixture of Kernels (HMK) and Hybrid Loss functions.
- Implemented some of the **reusable divergence and kernel utility functions** that became foundational for both DETONATE and DPO-Kernels projects.
- This framework achieved **state-of-the-art performances** over Factuality, Reasoning, Truthfulness, Safety, and Instruction Following, with HMK providing superior generalization and reducing overfitting by 15% compared to polynomial kernels.
- Reviewed manuscript sections and contributed stylistic refinements during later stages.

DETONATE - Project (AAAI Alignment Track 2026) ***November 2024 - November 2025***

- Contributed to **Project DETONATE**, a benchmark and alignment framework addressing hateful, toxic, or biased image generation by reframing alignment as a structural property of the model's latent space.
- Participated in implementing the **Diffusion-DPO (DDPO) method**, analyzing the original paper and proposing execution strategies for extending DDPO with geometry-aware kernel methods.
- Developed a **Kosmos-2-based image-processing utility** that performed region-aware extraction for image preprocessing, forming a key component of the project workflow.

- Implemented **image-to-embedding utility functions using JinaCLIP-v2**, enabling efficient generation of multimodal embeddings.
- Implemented some of the **reusable divergence and kernel utility functions** that became foundational for both DETONATE and DPO-Kernels projects.
- Provided coding assistance and debugging support for major components, ensuring smooth integration of the **Hybrid Loss function** and kernelized representations within the training pipeline.

Defactify-4 Workshop (AAAI 2025) — Web Chair ***November 2024 - April 2025***

- Served as the **Web Chair for DEFACTIFY-4 Workshop at AAAI 2025**, a workshop focused on multimodal fact-checking, AI-generated content detection, and hate-speech analysis, bringing together NLP and Computer Vision communities.
- Managed and continuously updated the **official workshop website**, timely leaderboard uploads, and completion of organizer details on the website.
- Oversaw **CodaLab competition platform operations** for two shared tasks (AI-Generated Text Detection and AI-Generated Image Detection).
- Provided **global participant support across time zones**, addressing queries on dataset usage, submission formats, evaluation procedures, and workshop paper guidelines throughout the competition cycle.
- Coordinated **internal administrative tasks** including dataset validation, **test-set redesign** to improve benchmark difficulty and representativeness, and EasyChair access management.
- Delivered **online presentations at AAAI 2025** covering competition results, participant paper summaries, and methodological approaches across both image and text detection tracks.
- Gained comprehensive experience in research-oriented competition management, workshop organization, cross-team coordination, and operational responsibilities required for delivering a workshop at a major AI conference.

Publications

Roy, R., Das, D., Banerjee, A., Bhattacharjee, A., Dasgupta, K. and Tripathi, S., 2025. ByDeWay: Boost Your multimodal LLM with DEpth prompting in a training-free Way. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6058-6064).

Das, A., Trivedy, S., Khanna, D., Narsupalli, Y., Ghosh, B., **Roy, R.**, Singh, G., Jain, V., Sharma, V., Reganti, A.N. and Chadha, A., 2025, July. Dpo kernels: A semantically-aware, kernel-enhanced, and divergence-rich paradigm for direct preference optimization. In Findings of the Association for Computational Linguistics: ACL 2025 (pp. 22174-22270).

Prasad, R., Borah, A., Abdullah, H.M., Shyalika, C., Singh, G., Garimella, R., **Roy, R.**, Surana, H., Imanpour, N., Trivedy, S. and Sheth, A., 2025. DETONATE: A Benchmark for Text-to-Image Alignment and Kernelized Direct Preference Optimization. arXiv preprint arXiv:2506.14903.

Certificates

- **Deep Learning Specialization - Coursera** **April 2023**
- **Machine Learning Specialization - Coursera** **March 2023**
- **Mathematics for Machine Learning Specialization - Coursera** **March 2023**

References

Dr. Amit Sheth
Professor, Computer Science and Engineering
University of South Carolina, USA
Email: amit@sc.edu

Dr. Kousik Dasgupta
Associate Professor and Head of the Department
Department of Computer Science and Engineering
Kalyani Government Engineering College, India
Email: kousik.dasgupta@kgec.edu.in

Dr. Subarna Tripathi
Principal Engineer, AI Research Science
Intel Corporation
Email: subarna.tripathi@intel.com