

Arka Mukherjee

[Website] [GitHub]

Email : arka.mukherjee078@gmail.com

Mobile : +91-97490-60670

- **Research Focus:** LLM reasoning, multimodality, and AI agents with a HCI and agent-computer interaction focus

EDUCATION

- **Kalinga Institute of Industrial Technology (KIIT)** Bhubaneswar, India
Bachelor of Technology in Computer Science; Cumulative GPA: 9.73/10.0 *July 2023 - May 2027 (Expected)*
 - **Relevant Courses:** Agentic AI, Probability & Statistics, Data Analytics, Machine Learning, AI, Algorithms, Data Mining, Human-Computer Interaction (Coursera)

EXPERIENCE

- **Carnegie Mellon University - Dr. Min Xu** Remote
Research Intern *January - May 2026*
 - **Reinforcement Learning:** Applied GRPO, DPO, and PPO to study RL generalization to low-sampling decoding tasks (64-token action space). Improved VRAM efficiency and iterated GRPO on 800k+ datapoints.
- **IIT Bhubaneswar - Dr. Shreya Ghosh** Bhubaneswar, India
Undergraduate Research Fellow (Funded) *December 2024 - Present*
 - **Multimodal STEM reasoning:** Developed mmJEE-Eval, a multimodal reasoning benchmark of 1,460 STEM problems to combat saturation (Qwen 3 VL 235B scored 45% vs 85% on Math Vista). Evaluated 17 VLMs and discovered metacognitive barriers: VLMs detect 53% of errors but correct only 3.5%. (*IJCNLP-AAACL 2025*)
 - **Multimodal Cultural Competence:** Created the first systematic **evaluation framework for VLM cultural competence** via story generation on 5 VLMs with a novel evaluation metric. (*Oral @ ICCV 2025 ASI workshop*)
 - **Integrated Capability Evaluation (ICE):** Proposed **ICE**, a new evaluation paradigm unlike benchmarks and leaderboards that models real-world tasks. Tested 5 frontier LLMs across 7 scaffolds on a 5-tiered diagnostic setup, showing GPT-5.2 Codex, Claude Sonnet 4.6 collapse universally.
- **IIT Ropar - Dr. Sudarshan Iyengar** Ropar, India
Summer Research Fellow (IASc-INSA-NASI) *May - July 2025*
 - **VLM Benchmarks:** Developed **EduVLM-Bench** for STEM prerequisite detection and evaluated OSS LLMs. The top model, Gemma3 27B, achieved 38.5% accuracy.

PUBLICATIONS

1. Mukherjee, A., & Ghosh, S. (2025). [Toward Socially Aware Vision-Language Models: Evaluating Cultural Competence Through Multimodal Story Generation](#) (ICCV 2025 ASI workshop - Oral)
2. Mukherjee, A., & Ghosh, S. (2025). [mmJEE-Eval: A Bilingual Multimodal Benchmark for Evaluating Scientific Reasoning in Vision-Language Models](#) (IJCNLP-AAACL 2025 Findings, 12% acceptance rate)

PROJECTS

- **Automated VLM reasoning dataset generator:** Curated 10k samples for SFTing VLMs. Our method achieved the best VMCBench results, beating Walton Cold Start and MM MathInstruct. (*Ranked #6 in NeurIPS DCVLR challenge*)

ACHIEVEMENTS/EXTRACURRICULARS

- **Awards:** **KIIT Merit Scholar** (Dean's list) for 5 semesters, **IUSSTF-Viterbi Scholar** at the University of Southern California (one of 15 students selected from India), **Amgen Scholars** Summer Research Program (declined), **Aalto Science Institute (ASCI)** Summer Research Program in Finland (declined)
- **Outreach:** Published articles (10M+ reads) and hardware reviews in collaboration with Nvidia, AMD, and ASUS.
- **Nvidia GPU Sponsorship:** Received RTX 5090, 5080, and 5070 GPUs from Nvidia India for AI research.
- **Travel Awards & Grants:** D&I Subsidy award at IJCNLP-AAACL'25 (\$750), AI Grants India (API keys and credits)

TECHNOLOGIES AND SKILLS

- **Languages:** Python (advanced), Java, C, SQL
- **ML/NLP Frameworks:** PyTorch, HF Transformers, Datasets & TRL, Unsloth, LM Studio, Docker, Agentic AI (Langchain, Playwright, mini-SWE-agent)