Anindya Mondal

Researcher and Engineer
a.mondal@surrey.ac.uk / mondalanindya.github.io / Google Scholar / Guildford, UK

Career Highlights

- 7 peer-reviewed publications with first-author papers in AAAI and ICCV Workshops, plus current submissions under review on unified count-aware VLMs and agent-guided generation.
- 4+ years of research experience across multimodal learning, vision–language systems, diffusion models, and object counting.
- Built reproducible research pipelines spanning model training, ablation studies, benchmarking, and evaluation tooling for counting and generation tasks.
- Teaching and mentoring experience across MSc AI cohorts and advanced computer vision modules at the University of Surrey.

Area of Expertise

Large-scale model training and post-training; multimodal foundation models (VLM); diffusion-based image generation; controllable generation; open-vocabulary object counting; benchmark and metric design; Python/PyTorch research engineering.

Professional Experience

Research Intern, (Adobe Research) Jun 2026 – Present
Bengaluru, India

- Building rubric-aware VLM judges for generated and edited visual content with stronger specificity and calibration.
- Designing criterion-level scoring for prompt fidelity, edit correctness, design quality, layout preservation, and artifact severity.

Doctoral Researcher, (University of Surrey) Oct 2022 – Present
Guildford, UK

- Developing multimodal systems for count-aware perception and generation, including ABACUS and CountLoop.
- Conducting research on VLM post-training, diffusion generation, open-vocabulary counting, and actor-agnostic action recognition.
- Building end-to-end reproducible codebases with strong evaluation and ablation support.

Research Intern, (Indian Institute of Science) May 2022 – Aug 2022
Bengaluru, India

- Developed a source-free domain adaptation framework for image classification under target-domain shift.

Undergraduate Research Assistant, (Jadavpur University) Oct 2020 – May 2022
Kolkata, India

- Worked on graph learning and event-based vision for signal recovery, semi-supervised segmentation, and moving-object detection.

Education

PhD Candidate in Artificial Intelligence Oct 2022 – Present
University of Surrey, Guildford, UK

Research focus: multimodal learning, vision–language integration, generative AI, object counting, reproducible evaluation.

B.E. (Hons.) in Electronics and Telecommunication Engineering; GPA: 8.79/10 Aug 2018 – Jun 2022
Jadavpur University, Kolkata, India

Selected Publications*

- **ABACUS: Adapting Unified Foundation Model for Bridging Image Count Understanding and Generation**.
  A. Mondal, S. Nag, A. Dutta. Under review.
- **CountLoop: Iterative Agent Guided High Instance Image Generation**.
  A. Mondal, A. Banerjee, S. Nag, J. Lladós, X. Zhu, A. Dutta. Preprint / under review. [Project]
- **OmniCount: Multi-label Object Counting with Semantic–Geometric Priors**.
  A. Mondal, S. Nag, X. Zhu, A. Dutta. AAAI 2025. DOI: 10.1609/aaai.v39i18.34151
- **Actor-Agnostic Multi-label Action Recognition with Multi-modal Query**.
  A. Mondal, S. Nag, J.M. Prada, X. Zhu, A. Dutta. ICCV Workshops 2023.
  DOI: 10.1109/ICCVW60793.2023.00086
- **Time-Varying Signals Recovery via Graph Neural Networks**.
  J.A.C. Correa, J.H. Giraldo, **A. Mondal**, et al. ICASSP 2023. DOI: 10.1109/ICASSP49357.2023.10096168

Honors and Awards

- AAAI 2025 Conference and Travel Grant.
- ICCV 2023 Conference Grant.
- University of Surrey Postgraduate Studentship (full PhD funding).
- ACM SIGKDD India Chapter Uplink Research Internship Award.

Teaching and Service

- Teaching Assistant / Demonstrator, University of Surrey (2023 – 2025): Applied Machine Learning (EEEM068), Advanced Topics in Computer Vision and Deep Learning (EEEM071).
- Reviewer: CVPR, ICCV, ECCV, NeurIPS, ICASSP, ICPR, IEEE Transactions on Signal Processing, IEEE TSIPN.

Technical Skills

- Programming: Python, MATLAB, C; Linux, Git, Docker, Jupyter, LaTeX.
- ML stack: PyTorch, scikit-learn, NumPy, SciPy, Pandas.
- Core areas: vision–language models, diffusion models, transformers, multimodal learning, object counting, graph neural networks.