

# Swapnanil Mukherjee

🌐 [linkedin.com/in/swapnanil-mukherjee](https://www.linkedin.com/in/swapnanil-mukherjee)

🐙 [github.com/SwapnanilMukherjee](https://github.com/SwapnanilMukherjee)

✉ [swapnanil.mukherjee\[at\]ug25@ashoka.edu.in](mailto:swapnanil.mukherjee[at]ug25@ashoka.edu.in)

🌐 [swapnanilmukherjee.github.io](https://swapnanilmukherjee.github.io)

## EDUCATION

- **Ashoka University** Sonepat, Haryana  
*Bachelor of Science, Computer Science (3.66/4)* 2022 - 2026
- **Hem Sheela Model School** Durgapur, West Bengal  
*11th and 12th grades; Science (92%)* 2020 - 2022

## RESEARCH INTERESTS

Multimodal AI, Joint Vision and Language Tasks, Reasoning in AI, Neurosymbolic Methods, Commonsense Abilities

## PUBLICATIONS

- Aritra Dutta, **Swapnanil Mukherjee**, Deepanway Ghosal, and Somak Aditya. (2025). *NLKI: A lightweight Natural Language Knowledge Integration Framework for Improving Small VLMs in Commonsense VQA Tasks*. In EMNLP 2025 Findings.
- Debojyoti Chowdhury, Bhavesh Neekhara, Shreyansh Priyadarshi, **Swapnanil Mukherjee**, Debashruti Maity, Debayan Gupta, and Shubhasis Haldar. (2023). *Next-Gen Profiling of Tumor-resident Stem Cells using Machine Learning*. bioRxiv.
- **Swapnanil Mukherjee** and Sujit Das. (2021). *Application of Transformer-Based Language Models to Detect Hate Speech in Social Media*. In Journal of Computational and Cognitive Engineering, 2(4), 278-286.

## WORK EXPERIENCE

- **Precog Lab, IIIT Hyderabad** Hyderabad, India  
*Research Intern* June 2025 - Present
  - ◇ I am working with a group at the Precog Lab, in collaboration with MSR India, to explore the implications of multimodal representational alignment.
  - ◇ We aim to understand how alignment relates to performance across downstream vision-language tasks, investigating the mechanisms through which alignment emerges, and exploring how these insights generalize across different model families and downstream task.
- **SarvamAI** Bangalore, India (Remote)  
*ML Engineer Intern* June - August 2024
  - ◇ Independently developed the first prototype of Sarvam's document text recognition pipeline from scratch for 10+ Indic languages.
  - ◇ Set up the data curation, collection, and preparation pipelines. Experimented with different end-to-end multimodal architectures, including various encoder, decoder models, and came up with an appropriate modality fusion mechanism and fine-tuning recipe for this task. Trained the entire system end-to-end on large-scale data in a multi-node distributed training cluster framework.
  - ◇ Contributed to Sarvam's Parsing API by benchmarking, evaluating, and generating training data for the underlying models.
- **Koita Centre for Digital Health - Ashoka (KCDH-A)** Sonepat, India  
*Research Intern* March - October 2024
  - ◇ Working with Prof. Debayan Gupta in collaboration with Prof. Steven Clipman of Johns Hopkins Medicine Institute on machine learning-based methods for enhanced forecasting of critical antiretroviral therapy (ART) drugs for the National AIDS Control Organization, Ministry of Health.
  - ◇ Set up the data cleaning and processing pipeline and experimented with different models. Developed the first few iterations of the forecasting tool. Presented our initial results in a presentation to senior executives at NACO.
  - ◇ Our method is being adopted by NACO for country-wide forecasting of 17+ HIV drugs which will optimise drug availability and positively impact lakhs of patients living with HIV in India.

## RESEARCH EXPERIENCE

---

- **Undergraduate Thesis on Neurosymbolic Reasoning for VQA:** I am extending the framework proposed in NELLIE and TV-TREES family of work to enable grounded and verifiable reasoning for Visual Question Answering (VQA), supervised by Prof. Debayan Gupta and Prof. Somak Aditya. I am primarily working to (i) expand the scope of this framework to open-domain commonsense-based VQA, (ii) improve the mechanism of visual information integration into the inference engine, (iii) use small, local models instead of large closed-source models in the engine. (*In Progress*)
- **Commonsense Reasoning in Vision-Language Models:** Collaborated with Prof. Somak Aditya (IIT Kharagpur) on the NLKI framework, which enhances small Vision-Language Models (sVLMs) for commonsense VQA tasks. By integrating externally retrieved knowledge and LLM-generated explanations, our method improved the performance of sVLMs by up to 15%, enabling them to match or outperform larger generative models. [paper] [code] [webpage]
- **Precision Profiling of Tumor Stem Cells Using Machine Learning-Based Digital Cytometry:** Worked with Prof. Debayan Gupta and Prof. Subhashis Halder's lab to develop a novel ML-based method of classification and quantification of gene expression data corresponding to human stem cells for highly accurate profiling of cancerous tumors which will aid the diagnosis and treatment of the disease in patients. Responsible for collating and preparing the data used for training, running all ML-based experiments, and developing the primary codebase for a software package that integrates the aforementioned functionality into a user-friendly tool. (*Under review*) [pre-print]
- **Coreset Selection for Image Datasets:** Came up with a novel zero-shot method for coreset selection (on image datasets) using a simple graph-based approach for dataset distillation. The highlight of the method is that can identify important examples in a given dataset without requiring any training on the target dataset. It is comparable to current SOTA methods while being much more computationally simple and efficient. (*Sep '23*)
- **Miscellaneous:**
  - ◊ Presented our poster on machine learning-based enhanced forecasting of antiretroviral therapy demand in India at the Johns Hopkins GKII Meet (2024) **amongst 50+ groups sponsored by the GKII Breakthrough grant** in India.
  - ◊ Poster on Federated Learning with tree-based models on tabular medical data accepted to ICGA 2023. [Poster]
  - ◊ Led a research project for NeuroMatch Academy 2022 to explore whether the behaviours of the different visual cortices of a human brain upon exposure to a visual stimulus could be imitated by the layers of a convolutional network. [Code]
  - ◊ Developed the prototype of an emotion recognition module during my internship at BrainEnTech Neuroscience. [Code]

## SKILLS AND TOOLS

---

- **Languages:** Python, C, Java, SQL, HTML, CSS
- **Tools:** PyTorch, Transformers/HF, Pytorch Lightning, LlamaFactory, TransformerLens, vLLM, Keras, Numpy, SKLearn, SciPy, Pandas, Matplotlib, Gradio, Git, Docker, L<sup>A</sup>T<sub>E</sub>X
- **Special Skills:** Kyo-Kushin Karate

## HONORS AND AWARDS

---

- Science Academies' Summer Research Fellowship Programme 2025 Fellow
- Dean's List (for academic excellence in a semester; **thrice**).
- College Board India Scholar 2022.
- CCIR Summer TechCamp Distinguished Scholar Award for exemplary performance in the programme (Aug '21).