

# Shraman Pramanick

## Curriculum Vitae

108 W 39th St, Apt 28  
Baltimore, MD 21210  
☎ +1 (443) 722 4993  
✉ [spraman3@jhu.edu](mailto:spraman3@jhu.edu)  
🌐 [Personal Website](#)

### Research Interests

- **Multimodal Learning** (Vision + Language, Vision + Other Modalities)
- **Multimodal LLMs, Egocentric Vision, Video-Language Pre-training**

### Education

- Jan 2021 - Present **Johns Hopkins University, Baltimore, MD, USA.**  
Ph.D. (with M.S.) in Electrical and Computer Engineering  
**Advisor:** [Rama Chellappa](#), [AIEM Lab](#), ECE (**GPA:** 4.0/4.0)  
**Research:** Self-Supervised Learning, Multimodal Learning, Image Geo-localization
- 2016 - 2020 **Jadavpur University, Kolkata, WB, India.**  
Bachelor of Engineering (B.E.) in Electronics & Telecommunication Engineering  
**Advisor:** [Amit Konar](#), AI Lab, ETCE (**GPA:** 9.41/10.0)

### Research Experience

- October 2023 - Present **Student Researcher, Google Research.**  
**Collaborators:** [Subhashini Venugopalan](#), [Dan Liebling](#)  
• **AI for Science:** Advanced question answering and grounding on scientific research papers.
- June 2023 - October 2023 **Research Scientist Intern, Meta AI.**  
**Collaborators:** [Nicolas Ballas](#), [Amjad Almahairi](#), [Guangxing Han](#), [Rui Hou](#), & [Qifan Wang](#)  
• **Multimodal LLMs:** Working on LLM-based framework for open-ended, customizable and unified coarse- and fine-grained vision-centric tasks over single and multiple input images.  
• **Ego-Exo4D:** Pre-training EgoVLPv2 on Ego-Exo4D dataset for developing strong baselines.
- May 2022 - Mar 2023 **Research Scientist Intern, Meta AI.**  
**Collaborators:** [Pengchuan Zhang](#), [Li Jing](#), [Yale Song](#), [Hardik Shah](#), & [Yann LeCun](#)  
• **Egocentric Video-Language Pre-training:** Proposed [EgoVLPv2](#), the second generation of egocentric video-language foundational model using cross-modal *fusion* in backbones.  
• **Multimodal Dimension-Contrastive Pre-training:** Proposed [VoLTA](#), a dimension-contrastive pre-training for image-caption pairs with explicit region-level understanding.
- Feb 2021 - Present **Graduate Research Assistant, Johns Hopkins University.**  
**Advisor:** [Rama Chellappa](#), [AIEM Lab](#), ECE  
• **Planet-scale Single Image Geo-localization:** Proposed [TransLocator](#), a dual-branch transformer network for planet-scale image geo-location as a part of [WRIVA](#) program.  
• **Real-time Detection of Activities in Untrimmed Videos:** Working on proposal-based solution to spatio-temporal action detection in untrimmed videos as a part of [DIVA](#) program.
- May 2020 - Jan 2021 **Research Associate, QCRI (Doha) & IIIT-Delhi Collaboration.**  
**Advisor:** [Preslav Nakov](#) & [Tanmoy Chakraborty](#)  
• **Data Efficient and Scalable Multimodal Learning (Vision, Language & Speech)**  
- Multimodal Abstractive Summarization, Detecting Harmful Internet Memes and Their Targets.
- May 2019 - Aug 2019 **Mitacs Globalink Research Intern, University of Montreal, Canada.**  
**Advisor:** [Antoine Saucier](#), Mathematical and Industrial Engineering  
• Worked on classical NR algorithms that preserve details, edges and fine patterns in images.

### Teaching Experience

- Spring 2023 **Course Assistant: Machine Intelligence (EN.520.650), Johns Hopkins University.**
- Spring 2022 **Course Assistant: Machine Intelligence (EN.520.650), Johns Hopkins University.**

## Selected Publications

Please see [Google Scholar](#), [Semantic Scholar](#) for the complete list of publications

### Pre-prints

- **Pramanick S.**, Han G., Hou R., Nag S., Lim S., Ballas N., Wang Q., Chellappa R., Almahairi A., “Jack of All Tasks, Master of Many: Designing General-purpose Coarse-to-Fine Vision-Language Model”. [[Paper](#) | [Project](#)]
- Grauman K. et al., “Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives”. [[Paper](#) | [Project](#) | [Blog](#) | [Video](#)]

### Conference Proceedings

- **Pramanick S.**, Song Y., Nag S., Lin K., Shah H., Shou M., Chellappa R., Zhang P., “EgoVLPv2: Egocentric Video-Language Pre-training with Fusion in the Backbone”. *International Conference on Computer Vision (ICCV)*, 2023 [[Paper](#) | [Project](#) | [Code](#) | [Poster](#) | [Slides](#)]
- **Pramanick S.**, Nowara E.M., Gleason J., Castillo C.D., Chellappa R., "Where in the World is this Image? Transformer-based Geo-localization in the Wild", *European Conference on Computer Vision (ECCV)*, 2022 [[Paper](#) | [Code+Data](#) | [Poster](#) | [Slides](#) | [Video](#)]
- **Pramanick S.\***, Roy A.\*, Patel V., “Multimodal Learning using Optimal Transport for Sarcasm and Humor Detection”, *Winter Conference on Applications of Computer Vision (WACV)*, 2022. [[Paper](#)]
- **Pramanick S.\***, Sharma S\*, Dimitrov D., Aktar S., Nakov P., Chakraborty T., “MOMENTA: A Multimodal Framework for Detecting Harmful Memes and Their Targets”, *Findings of Empirical Methods in Natural Language Processing (EMNLP)*, 2021. [[Paper](#) | [Code+Data](#) | [Poster](#) | [Slides](#)]
- **Pramanick S.**, Dimitrov D., Mukherjee R., Sharma S., Aktar S., Nakov P., Chakraborty T., “Detecting Harmful Memes and Their Targets”, *Findings of Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021. [[Paper](#) | [Code+Data](#) | [Slides](#) | [Video](#)]

### Journals

- **Pramanick S.\***, Jing L.\*, Nag S.\*, Zhu J., Shah H., LeCun Y., Chellappa R., “VoLTA: Vision-Language Transformer with Weakly-Supervised Local-Feature Alignment”, *Transactions on Machine Learning Research (TMLR)*, 2023. [[Paper](#) | [Project](#) | [Code](#)]
- Atri Y.\*, **Pramanick S.\***, Goyal V., Chakraborty T., “See, Hear, Read: Leveraging Multimodality with Guided Attention for Abstractive Text Summarization”, *Knowledge-Based Systems* [IF - 8.139], Elsevier, Sept. 2021. [[Paper](#) | [Code+Data](#)]

## Selected Honors & Awards

- Jan 2021 [JHU ECE Departmental Fellowship](#), awarded to outstanding incoming PhD students.
- May 2019 [Mitacs Globalink Research Internship](#), awarded to top-ranked applicants from 15 different countries to participate in a 12-week research internship in Canadian universities.
- Oct 2016 [JBNSTS Senior Scholarship](#), 4-year scholarship for academic excellence during B.E.
- Jan 2015 [Regional Mathematical Olympiad \(RMO\)](#), ranked among top-10 students in the state.

## Technical Skills

**Programming Languages:** Python, MATLAB, C/C++, Mathematica, L<sup>A</sup>T<sub>E</sub>X

**Libraries & Tools:** Pytorch, Keras, Tensorflow

## Voluntary Services

Reviewer for CVPR, ECCV, ICCV, WACV, ARR, EMNLP, ACL, TPAMI, TNNLS, TAI, TAFFC.

## References

[Rama Chellappa](#), Bloomberg Distinguished Professor, Johns Hopkins University  
[Yale Song](#), Research Scientist, FAIR, Meta AI  
[Subhashini Venugopalan](#), Research Scientist, Google Research