

# 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 - **Johns Hopkins University, Baltimore, MD, USA.**

Present Ph.D. (with M.S.) in Electrical and Computer Engineering  
**Advisor:** [Rama Chellappa](#), [AIEM Lab](#), ECE (**GPA:** 4.0/4.0)

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)

### Selected Publications

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

#### Pre-prints

- **Pramanick S.**, Mavroudi E., Song Y., Chellappa R., Torresani L., Afouras T., “Enrich and Detect: Video Temporal Grounding with Multimodal LLMs”.

#### Conference Proceedings

- **Pramanick S.\***, Chellappa R., Venugopalan S.\*, “SPIQA: A Dataset for Multimodal Question Answering on Scientific Papers”. **NeurIPS D&B**, 2024. [[Paper](#) | [Dataset](#) | [Code](#) | [Poster](#)]
- **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”. **CVPR**, 2024. (Highlight, Top 2.8%) [[Paper](#) | [Project](#)]
- Grauman K. et al., “Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives”. **CVPR**, 2024. (Oral, Top 0.8%) [[Paper](#) | [Project](#) | [Blog](#) | [Video](#)]
- **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”. **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”. **ECCV**, 2022. [[Paper](#) | [Code+Data](#) | [Slides](#)]
- **Pramanick S.\***, Roy A.\*, Patel V., “Multimodal Learning using Optimal Transport for Sarcasm and Humor Detection”. **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 **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 **ACL**, 2021. [[Paper](#) | [Code+Data](#) | [Slides](#)]

#### 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”. **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**, Elsevier, 2021. [[Paper](#) | [Code+Data](#)]

---

## Research Experience

- June 2024 - Present **Research Scientist Intern**, FAIR, Meta.  
**Collaborator:** Triantafyllos Afouras, Yale Song, Effrosyni Mavroudi, & Lorenzo Torresani.  
• Fine-grained video temporal grounding.
- October 2023 - June 2024 **Student Researcher**, Google Research.  
**Collaborator:** Subhashini Venugopalan.  
• Proposed SPIQA, a dataset for multimodal QA and grounding on scientific papers.
- June 2023 - October 2023 **Research Scientist Intern**, GenAI, Meta.  
**Collaborators:** Nicolas Ballas, Amjad Almahairi, Guangxing Han, Rui Hou, & Qifan Wang.  
• **Multimodal LLMs:** Proposed VistaLLM, a LLM-based framework for open-ended, customizable and unified coarse-to-fine 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**, FAIR, Meta.  
**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  
• Multimodal LLMs, vision-language pre-training, planet-scale single image geo-localization.
- May 2020 - Jan 2021 **Research Associate**, QCRI (Doha) & IIIT-Delhi Collaboration.  
**Advisor:** Preslav Nakov & Tanmoy Chakraborty.  
• Multimodal abstractive summarization, detecting harmful internet memes & 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 Honors & Awards

- June 2024 **EgoVis 2022/2023 Distinguished Paper Award**, for EgoVLPv2.  
June 2024 **Spot Bonus from Google**, for exceptional contributions while being student researcher.  
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.

---

## Voluntary Services

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

---

## References

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