PRITAM SARKAR

Toronto, Canada | google scholar | github | linkedin | pritam.sarkar@queensu.ca | www.pritamsarkar.com

I am a recent PhD graduate from Queen's University & Vector Institute, with solid research background (1K+ citations and 10+ first-authored papers at top-tier venues e.g., NeurIPS, ICLR, AAAI) and industry experience (Research Intern at Google and Borealis AI, along with 3 years as a Software Engineer), looking for a research position.

RESEARCH INTERESTS

I am interested in advancing safe multimodal intelligence and in designing algorithms that require minimal human supervision.
Broadly interested in: generative models (LLMs, multimodal LLMs, diffusion models), foundation models (video, image, vision-language, audio-visual), representation learning (self-supervised, unsupervised), alignment, reasoning, AI agents, world models, computer vision.

EXPERIENCE

Queen's University **2020 - 2025**
Research Assistant (PhD) *Kingston, Canada*
Supervisor: Ali Etemad
During this time, I have also been affiliated with the **Vector Institute** (2021 - 2025).
• Introduced VCRBench, the first benchmark to study video-based long-form causal reasoning capabilities of multimodal LLMs.
• Introduced Recognition Reasoning Decomposition, a simple and modular approach to address multi-step reasoning directly from visual observations.
• Proposed Refined Regularized Preference Optimization, enabling fine-grained alignment of multimodal LLMs during post-training.
• Designed a self-alignment framework enabling multimodal LLMs to learn from their own successes and failures.
• The first to systematically study video self-supervised methods under real-world distribution shifts, uncovering novel insights and failure modes that guide future self-supervised algorithm design.
• Introduced XKD, a multimodal video foundation model, with improved audio-visual representation alignment through feature refinement and cross-modal knowledge distillation.
• Introduced CrissCross, a multimodal video foundation model, to learn more generalized video representations by leveraging temporally asynchronous audio-visual relationships.
• Introduced AVCAffe, the largest recorded multimodal video dataset capturing affective states and cognitive load to facilitate human behavior understanding from videos.

Google **Fall 2023**
Student Researcher/Research Intern *Sunnyvale, USA*
Host: Sercan Ö. Arık.
Proposed phrase-level alignment to reduce vision-language hallucinations in multimodal LLMs without hurting general capabilities, unlike existing finetuning methods.

Borealis AI **Fall 2022**
Machine Learning Research Intern *Toronto, Canada*
Host: Fredrick Tung
Addressed the limitation of standard augmentation techniques in event sequence time-series data by introducing AugESeq, a conditional diffusion model that generates realistic augmented event sequences for improved representation learning.

Queen's University **2018 - 2020**
Research Assistant (MASc) *Kingston, Canada*
Supervisor: Ali Etemad
• I was the first to introduce self-supervised learning with ECG (Electrocardiograms).
• I was the first to introduce PPG (Photoplethysmogram) to ECG translation using deep learning for continuous cardiac activity monitoring.

Queen's University **2018 - 2025**
Guest Lecturer/Lead Teaching Assistant/Teaching Assistant *Kingston, Canada*
I worked for the following courses: Artificial Intelligence & Interactive Systems (ELEC 872 in F'23), Artificial Intelligence (ELEC 472 in W'25, W'23, W'22, W'21, S'19), Electronics I (ELEC-252 in F'20), Introduction to Computer Programming for Engineers (APSC-143 in F'19)

Infosys Ltd. **2017 – 2018**
Senior System Engineer *Bangalore, India*
I worked on Oracle Cloud and SQL.

Tech Mahindra Ltd. **2015 – 2017**
Software Engineer *Hyderabad, India*
I worked on Oracle Cloud and SQL.

EDUCATION

Doctor of Philosophy (PhD) **2020 - 2025**
Dept. of Electrical and Computer Engineering, Queen's University *Kingston, Canada*
Thesis title: Multimodal Learning from Videos: Self-supervised Pre-training, Post-training Alignment, and Benchmarks
Advisor: Ali Etemad
GPA: 4.3/4.3

Master Applied Science (MASc) **2018 - 2020**
Dept. of Electrical and Computer Engineering, Queen's University *Kingston, Canada*
Thesis title: Self-Supervised ECG Representation Learning for Affective Computing. [Link to thesis]
Advisor: Ali Etemad
GPA: 3.8/4.3

Bachelor of Technology (B.Tech) **2011 - 2015**
Dept. of Electrical Engineering, West Bengal University of Technology *Kolkata, India*
Rank 4 of 150 in graduating class of Electrical Engineering.
GPA: 8.84/10

SELECTED PUBLICATIONS

As per google scholar, my publications have total citations of approximately 1000. Please find the full list here.

12. **P. Sarkar**, A. Etemad, "VCRBench: Exploring Long-form Causal Reasoning Capabilities of Large Video Language Models", *Preprint. Under Review.* [Paper]
11. **P. Sarkar**, A. Etemad, "Self-alignment of Large Video Language Models with Refined Regularized Preference Optimization", *NeurIPS 2025*. [Paper]
10. **P. Sarkar**, S. Ebrahimi, A. Etemad, A. Beirami, S. Arik, T. Pfister, "Mitigating Object Hallucination in MLLMs via Data-augmented Phrase-level Alignment", *ICLR 2025*. [Paper]
9. **P. Sarkar**, A. Etemad, "XKD: Cross-modal Knowledge Distillation with Domain Alignment for Video Representation Learning", *AAAI 2024*. [Paper]
8. **P. Sarkar**, A. Beirami, A. Etemad, "Uncovering the Hidden Dynamics of Video Self-supervised Learning under Distribution Shifts", *NeurIPS 2023*. Spotlight [Paper]
7. **P. Sarkar**, A. Etemad, "Self-supervised Audio-Visual Representation Learning with Relaxed Cross-Modal Synchronicity", *AAAI, 2023*. Oral [Paper]
6. **P. Sarkar**, A. Posen, A. Etemad, "AVCAffe: A Large Scale Audio-Visual Dataset of Cognitive Load and Affect for Remote Work", *AAAI, 2023*. [Paper]
5. D. Shome, **P. Sarkar**, A. Etemad, "Region-Disentangled Diffusion Model for High-Fidelity PPG-to-ECG Translation", *AAAI 2024*. [Paper]
4. **P. Sarkar**, A. Etemad, "Method and Apparatus for Generating an Electrocardiogram from a Photoplethysmogram", *US20230363655A1. Patent* [Link]
3. **P. Sarkar**, A. Etemad, "CardioGAN: Attentive Generative Adversarial Network with Dual Discriminators for Synthesis of ECG from PPG", *AAAI, 2021*. [Paper] 100+ citations
2. **P. Sarkar**, A. Etemad, "Self-supervised ECG Representation Learning for Emotion Recognition", *IEEE Transactions on Affective Computing, 2020*. [Paper] 400+ citations
1. **P. Sarkar**, A. Etemad, "Self-supervised Learning for ECG-based Emotion Recognition", *ICASSP, 2020*. Oral [Paper] 150+ citations

TECHNICAL & INTERPERSONAL SKILLS

Deep Learning Frameworks: PyTorch, TensorFlow, Keras

Programming Languages: Python, MATLAB, C, SQL
Soft skills: Low-ego, High integrity, Focused

ACADEMIC ACHIEVEMENTS/AWARDS

- Queen's University Travel Award, 2025.
- Queen's University Travel Award, 2024.
- Smith Engineering Graduate Student Conference Travel Award, 2023.
- First prize in IEEE Research Excellence Award (PhD), 2023, IEEE Kingston Section.
- Best Poster Award at Robotics and AI Symposium at Ingenuity Labs, 2023.
- Queen's University Travel Award, 2023.
- Honourable Mention at Faculty of Engg. and Applied Science Research Symposium at Queen's University, 2022.
- Honourable Mention at Robotics and AI Symposium at Ingenuity Labs, 2022.
- Best Poster Award at Robotics and AI Symposium at Ingenuity Labs, 2021.
- Postgraduate Affiliate Award, Vector Institute, 2021 - 2023.
- Graduate Research Fellowship, Queen's University, 2020 - 2025.
- Queen's University Travel Award, 2019.
- Graduate Research Scholarship, Queen's University, 2019 - 2020.

INVITED TALKS

- June 2024 at Google: Data-Augmented Phrase-Level Alignment for Mitigating Object Hallucination
- July 2023 at Ingenuity Labs: Learning without Human Supervision
- January 2023 at Borealis AI: Augmentation Improves Event Sequence Prediction

ACADEMIC SERVICE

Mentorship

- Shayan Samakosh, Master's at York University, 2025 - Present.
- Seth Grief-Albert, Bachelor's at Queen's University, Summer 2024.
- Vishal Narnaware, Visiting Student at University of Cambridge, co-mentored with Nikhil Churamani, 2023 - 2024.
- Debaditya Shome, Master's at Queen's University, 2022 - 2023.
- Aaron Posen, Bachelor's at ECE, Queen's University, 2021 - 2022.
- Rachel Phinnemore, Bachelor's at Queen's University, 2020 - 2021.

Area Chair

I am serving as an AC for the following venues:

- WACV

Reviewing/PC Member

I regularly review for the following venues:

- NeurIPS, ICLR, AAAI, CVPR, ICCV, ECCV, ICML, ICASSP, ACII
- IEEE Transactions on - PAMI, Affective Computing, Artificial Intelligence

Organizing workshops and conferences

- Session chair for computer vision tracks at AAAI 2023.
- Co-organizer of AAAI 2023 Workshop on Representation Learning for Responsible Human-centric AI (R2HCAI).
- Co-organizer of AAAI 2022 Workshop on Human-centric Self-supervised Learning (HCSSL).
- Volunteer at AI/GI/CRV Conference, 2019.

Others

- Student Rep. in Graduate Studies Academic Advisory Committee, Dept. of ECE, Queen's University, 2020 - 2021.
- PhD Rep. at Graduate Electrical and Computer Engineering student council, Queen's University, 2020 - 2021.