PRITAM SARKAR

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

Final-year PhD candidate with prior industry experience (internships at Google and Borealis AI, along with 3 years as a Software Engineer), solid research background (NeurIPS Spotlight, AAAI Oral), and a perfect academic record (GPA 4.3/4.3), looking for a Research Scientist role.

RESEARCH INTERESTS

multimodal learning (i.e., vision-language, audio-visual), video understanding, representation learning, unsupervised learning, self-supervised learning, multimodal LLMs, large vision language models, foundation models, computer vision, human-centric AI

EXPERIENCE

Google **Fall 2023**
Student Researcher/Research Intern *Sunnyvale, USA*
Hosts: Sayna Ebrahimi and Sercan Ö. Arık.
Introduced phrase-level alignment, to mitigate vision-language hallucinations in multimodal LLMs while preserving their ability on general tasks, 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 developing AugESeq, a conditional diffusion model that generates realistic augmented event sequences for improved representation learning.

Queen's University **2020 - Present**
Research Assistant (PhD) *Kingston, Canada*
Supervisor: Ali Etemad
During this time I am also affiliated with the **Vector Institute** (2021 - Present).
• Currently, I am working on analysing and mitigating the limitations of multimodal LLMs in video understanding.
• For the first time studied the behaviour of popular video self-supervised methods in response to various forms of natural distribution shift, uncovering a series of intriguing findings and interesting behaviors.
• Introduced XKD to improve alignment between audio and visual modalities in video representation learning.
• Introduced CrissCross to learn generalized representations leveraging the asynchronous relationships between audio and visual modalities.
• Introduced AVCAffe, the largest recorded affective video dataset for human behavioural understanding.

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

Queen's University **2018 - Present**
Teaching Assistant *Kingston, Canada*
I worked as a Guest Lecturer/Head TA/TA 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) **May 2020 - Apr 2025 (expected)**
Dept. of Electrical and Computer Engineering, Queen's University *Kingston, Canada*
Thesis topic: Multimodal Visual Understanding
Advisor: Ali Etemad
GPA: 4.3/4.3

Master Applied Science (MASc) **Sept 2018 - Apr 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) **Aug 2011 - Jul 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 approx. 700. Please find the full list here.

10. **P. Sarkar**, S. Ebrahimi, A. Etemad, A. Beirami, S. Arik, T. Pfister, "Data-Augmented Phrase-Level Alignment for Mitigating Object Hallucination", *Under review*. [Paper]
9. **P. Sarkar**, A. Beirami, A. Etemad, "Uncovering the Hidden Dynamics of Video Self-supervised Learning under Distribution Shifts", *NeurIPS 2023*. Spotlight [Paper]
8. **P. Sarkar**, A. Etemad, "XKD: Cross-modal Knowledge Distillation with Domain Alignment for Video Representation Learning", *AAAI 2024*. [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*. Virtual [Paper]
2. **P. Sarkar**, A. Etemad, "Self-supervised ECG Representation Learning for Emotion Recognition", *IEEE Transactions on Affective Computing, 2020*. [Paper] 300+ citations
1. **P. Sarkar**, A. Etemad, "Self-supervised Learning for ECG-based Emotion Recognition", *ICASSP, 2020*. Oral [Paper] 100+ citations

TECHNICAL SKILLS

Deep Learning Frameworks: PyTorch, TensorFlow, Keras
Programming Languages: Python, MATLAB, C, SQL

ACADEMIC ACHIEVEMENTS/AWARDS

- First prize in IEEE Research Excellence Award (PhD), 2023, IEEE Kingston Section.
- Best Poster Award at Robotics and AI Symposium at Ingenuity Labs, 2023. Title: *Cardiac Insights On-the-Go: Inexpensive Continuous ECG Monitoring from PPG Using Diffusion Models*
- Honourable Mention at FEAS Research Symposium at Queen's University, 2022. Title: *The First Large-Scale Audio-Visual Dataset of Cognitive Load and Affect for Remote Work*
- Honourable Mention at Robotics and AI Symposium at Ingenuity Labs, 2022. Title: *The First Large-Scale Audio-Visual Dataset of Cognitive Load and Affect for Remote Work*
- Best Poster Award at Robotics and AI Symposium at Ingenuity Labs, 2021. Title: *Toward Wearables of the Future: Affordable Acquisition of Continuous ECG with Deep Learning*
- Postgraduate Affiliate Award, Vector Institute, 2021 - 2023.

- Graduate Research Fellowship, Queen's University, 2020 - 2025.
- Graduate Research Scholarship, Queen's University, 2019 - 2020.

ACADEMIC SERVICE

Mentorship

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

Reviewing/PC Member

I regularly review for the following venues:

- NeurIPS, ICLR, AAAI, CVPR, ICCV, ECCV, ICML, ICASSP
- 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.

REFERENCES

- Dr. Ali Etemad, Associate Professor at Queen's University, email: `ali.etemad@queensu.ca`
- Dr. Ahmad Beirami, Research Scientist at Google DeepMind, email: `beirami@google.com`