Subhajit Chaudhury

CONTACT INFORMATION

Email: subhajit.utokyo@gmail.com **Phone:** +1-914-498-3047
Address: 2 Canfield Avenue, White Plains, NY 10601, USA

EDUCATION

The University of Tokyo, Japan — Apr 2018 - Mar 2021
PhD, Graduate School of Information Science and Technology

Indian Institute of Technology (IIT), Bombay, India — Jul 2012 - Aug 2014
M.Tech, Department of Electrical Engineering

Jadavpur University, India — Jul 2008 - Jun 2012
B.E.(Hons.) Department of Electrical Engineering

WORK EXPERIENCE

Senior Research Scientist, IBM Research, New York — Apr 2017- present
My work focuses on improving the safety, reliability and trustworthiness of LLMs.

- **Safe & Trustworthy LLMs**: Developed **Granite Guardian**, a suite of safeguard models designed to detect risks such as social bias, jailbreaking, and hallucination in LLMs. Granite Guardian ranks at the top of **GuardBench**, the leading benchmark for evaluating guardrail models. Our model is an industry-leading solution for hallucination detection in RAG settings.
- **Memory-Augmented LLMs**: Designed **Larimar**, a brain-inspired architecture for enhancing LLMs with distributed episodic memory, enabling dynamic, one-shot knowledge updates without costly re-training. Introduced **EpMAN**, a method for processing long contexts using episodic memory, improving recall and question-answering robustness across 16k–256k tokens.

Researcher, NEC Central Research Labs, Japan — Oct 2014 - Mar 2017
Developed advanced computer vision algorithms for real-time anomaly detection in video surveillance systems. My work was integrated into NEC's flagship products for **road maintenance** and **airport runway monitoring**, contributing to enhanced infrastructure safety and efficiency. The project gained recognition and was featured in **Japanese media**, highlighting its industry impact.

SELECTED PUBLICATIONS

- **Subhajit Chaudhury***, Payel Das*, Sarathkrishna Swaminathan, Georgios Kollias, Elliot Nelson, Khushbu Pahwa, Tejaswini Pedapati, Igor Melnyk, Matthew Riemer, **EpMAN: Episodic Memory AttentioN for Generalizing to Longer Contexts**, **ACL** 2025

- Inkit Padhi*, Manish Nagireddy*, Giandomenico Cornacchia*, **Subhajit Chaudhury***, Tejaswini Pedapati* *et al.* **Granite Gaurdian**, **NAACL** 2025 (Industry Track)

- Payel Das*, **Subhajit Chaudhury***, Elliot Nelson, Igor Melnyk, Sarathkrishna Swaminathan, Sihui Dai, Aurelie Lozano, *et al.* **Larimar: Large Language Models with External Episodic Memory Control**, **ICML** 2024.

- Kinjal Basu, Ibrahim Abdelaziz, **Subhajit Chaudhury**, Soham Dan *et al.* **API-BLEND: A Comprehensive Corpora for Training and Benchmarking API LLMs**, **ACL** 2024

- Zhang, Shuai, Hongkang Li, Meng Wang, Miao Liu, Pin-Yu Chen, Songtao Lu, Sijia Liu, Keerthiram Murugesan, and **Subhajit Chaudhury**, **On the Convergence and Sample Complexity Analysis of Deep Q-Networks with -Greedy Exploration**, **NeurIPS** 2023

- **Subhajit Chaudhury**, Sarathkrishna Swaminathan, Daiki Kimura, Prithviraj Sen, Keerthiram Murugesan *et al.*, **Learning Symbolic Rules over Abstract Meaning Representations for**

Textual Reinforcement Learning, **ACL** 2023

- Maxwell Crouse, Pavan Kapanipathi, **Subhajit Chaudhury**, Tahira Naseem *et al.* **Laziness Is a Virtue When It Comes to Compositionality in Neural Semantic Parsing**, **ACL** 2023

- Heshan Devaka Fernando, Han Shen, Miao Liu, **Subhajit Chaudhury**, Keerthiram Murugesan, and Tianyi Chen, **Mitigating Gradient Bias in Multi-objective Learning: A Provably Convergent Approach**, **ICLR** 2023

- Subhajit Chaudhury, Sarathkrishna Swaminathan, Chulaka Gunasekara *et al.* **X-FACTOR: A Cross-metric Evaluation of Factual Correctness in Abstractive Summarization**, **EMNLP**, 2022.

- Keerthiram Murugesan, **Subhajit Chaudhury**, and Kartik Talamadupula, **Eye of the Beholder: Improved Relation Generalization for Text-based Reinforcement Learning Agents**, **AAAI**, 2022.

- Subhajit Chaudhury, Prithviraj Sen, Masaki Ono *et al.* **Neuro-symbolic Approaches for Text-based Policy Learning**, **EMNLP** 2021.

- Subhajit Chaudhury, Daiki Kimura, Kartik Talamadupula, Michiaki Tatsubori, Asim Munawar, and Ryuki Tachibana, **Bootstrapped Q-learning with Context Relevant Observation Pruning to Generalize in Text-based Games**, **EMNLP** 2020.

Technical Skills

- **Technical Skills**: *Natural Language Processing, Reinforcement Learning, Machine Learning, Computer Vision, Human-Computer Interaction*
- **Programming Languages**: *Python, C/C++, Java*
- **Frameworks and Tools**: *Pytorch, Tensorflow, scikit-learn, Matlab, ROS, Gazebo, OpenGL*

Mentorship

- **Akshat Gupta**, Ph.D. Student, UC Berkeley, USA
- **Sihui (Sophie) Dai**, Ph.D. Student, Princeton University, USA
- **Shib Sankar Dasgupta**, Ph.D. Student, University of Massachusetts Amherst, USA
- **Heshan Fernando**, Ph.D. Student, Rensselaer Polytechnic Institute, USA
- **Maurício Gruppi**, Ph.D. Student, Rensselaer Polytechnic Institute, USA
- **Thomas Carta**, Ph.D. Student, INRIA, France
- **Tristan Matthieu Stampfler**, MS Student, École Polytechnique, France

Blogs

1) **IBM's safety checkers top a new AI benchmark**, Apr 2025

2) **How memory augmentation can improve large language model efficiency and flexibility**, Sep 2024

3) **Coverage of the crack-detection work at NEC in Nikkei newspaper**, Nov 2021

4) **Coverage of sports activity detection work for Olympics 2020 in Nikkei**, Mar 2020

Professional Activities

- Senior Program committee member for AAAI'23, AAAI'24.
- Program Committee member: IJCAI'23, AAAI'22, AAAI'20, IJCAI'20.
- Reviewer: ACL'23, TMLR'23, CVPR'21, CVPR'22, IJCAI'19, ICRA'20, ICRA'18, IROS'18.

Website: https://subhajitchaudhury.github.io
See full list of publications here: https://scholar.google.com/citations?user=EBTpFrQAAAAJ&hl=en