SUBHAJIT CHAUDHURY

✉ subho.chaudhury@outlook.com ◇ https://subhajitchaudhury.github.io/ ◇
2 Canfield Avenue, White Plains, New York, USA 10601 ◇ ☎(+1) 914-498-3047

RESEARCH INTERESTS

Large-Scale Generative AI for Enterprise, Safe and Reliable LLMs, Neuro-Symbolic Reasoning,

EDUCATION

- **The University of Tokyo, Japan** 2018 – 2021
  Ph.D., Graduate School of Information Science and Technology
  Thesis: *Generalization in Neural Networks for Robustness against Adversarial Vulnerabilities.*
- **Indian Institute of Technology (IIT), Bombay, India** 2012 – 2014
  M.Tech, Department of Electrical Engineering
- **Jadavpur University, India** 2008 – 2012
  B.E.(Hons.) Electrical Engineering

RESEARCH AND INDUSTRY EXPERIENCE

- **IBM Thomas J. Watson Research Center**, NY, USA April 2017 – Present
  Senior Research Scientist
  I specialize in developing large language models for enterprise use-cases with expertise in synthetic data generation and training using multi-GPU architectures. My current focus is on enhancing ***reasoning capabilities*** and ***ensuring robust safety measures*** for IBM's Granite language models. I led the code-assisted synthetic data generation effort for **IBM Granite models** to improve algorithmic reasoning and was the lead researcher for developing **retrieval-augmented generation (RAG) hallucination detection** and **agentic safety** for **IBM Granite Guardian 3.1** models, ensuring safer and reliable outputs. My work has directly contributed to **IBM's WatsonX Governance** platform, reinforcing enterprise-grade AI safety.

  Beyond model training and deployment, I actively explore cutting-edge ML topics such as **memory-augmented LLMs** and **function-calling**, contributing regularly to top-tier ML conferences.
- **NEC Central Research Labs**, Kawasaki, Japan Oct 2014 – Mar 2017
  Researcher
  Lead researcher for the development of an AI-based **crack detection system** for infrastructure surveillance by finding discontinues in dense 2D motion fields using energy minimization on a Conditional Random Fields (CRF).

SELECTED RECENT PUBLICATIONS

- Inkit Padhi*, Manish Nagireddy*, Giandomenico Cornacchia*, **Subhajit Chaudhury***, Tejaswini Pedapati* and *other authors*, **Granite Gaurdian**, **NAACL Industry Track** 2025 (* denotes equal contribution)
- Ching-Yun Ko, Pin-Yu Chen, Payel Das, Youssef Mroueh, Soham Dan, Georgios Kollias, **Subhajit Chaudhury**, Tejaswini Pedapati, Luca Daniel, **Large Language Models can be Strong Self-Detoxifiers**, **ICLR** 2025

- Payel Das*, **Subhajit Chaudhury***, Elliot Nelson, Igor Melnyk, Sarath Swaminathan and *other authors*, **Larimar: Large Language Models with Episodic Memory Control**, **ICML** 2024 (* denotes equal contribution)
- Kinjal Basu, Ibrahim Abdelaziz, **Subhajit Chaudhury**, Soham Dan, Maxwell Crouse, Asim Munawar, Sadhana Kumaravel, Vinod Muthusamy, Pavan Kapanipathi, Luis Lastras, **API-BLEND: A Comprehensive Corpora for Training and Benchmarking API LLMs**, **ACL** 2024
- **Subhajit Chaudhury**, Sarath Swaminathan, *and other authors*, **Learning Symbolic Rules over Abstract Meaning Representations for Textual Reinforcement Learning**, **ACL** 2023
- Heshan Devaka Fernando, Han Shen, Miao Liu, **Subhajit Chaudhury**, Keerthiram Murugesan, Tianyi Chen, **Mitigating Gradient Bias in Multi-objective Learning: A Provably Convergent Approach**, **ICLR**, 2023 (accepted as notable top 5% paper).
- **Subhajit Chaudhury**, Sarathkrishna Swaminathan, Chulaka Gunasekara, Maxwell Crouse, Srinivas Ravishankar and *other authors*, **X-FACTOR: A Cross-metric Evaluation of Factual Correctness in Abstractive Summarization**, **EMNLP** 2022.
- Keerthiram Murugesan, **Subhajit Chaudhury**, Kartik Talamadupula, **Eye of the Beholder: Improved Relation Generalization for Text-based Reinforcement Learning Agents**, **AAAI**, 2022.
- **Subhajit Chaudhury**, Prithviraj Sen, Masaki Ono, Daiki Kimura, Michiaki Tatsubori and Asim Munawar, **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

Languages: Python, C++, Java.
ML Tools: PyTorch, transformers, scikit-learn.
Other Tools: MATLAB, ROS, Gazebo, OpenCV, CUDA, OpenGL.

SELECTED INVITED TALKS

- Invited Talk on *"Larimar: Large Language Models with Episodic Memory Control"* at **Amazon Search Research Talk Series**, Oct 2024.

- Invited Talk, Subhajit Chaudhury, "X-FACTOR: A Cross-metric Evaluation of Factual Correctness in Abstractive Summarization", **IBM Neuro-Symbolic AI Workshop 2023**, New York, USA.

- Invited Talk, Subhajit Chaudhury, "Neuro-symbolic reinforcement learning for text-based games" in the **IBM Neuro-Symbolic AI Workshop 2022**, New York, USA.

- Invited Talk on *"Visual Imitation Learning for Autonomous Control"* at **NASA, Jet Propulsion Laboratory (JPL)**, Dec 2019.

PROFESSIONAL ACTIVITIES

- **Senior Program Committee member** (AAAI 2023);
- **Program Committee member** (IJCAI 2023, IJCAI 2020, AAAI 2021, KBRL workshop at IJCAI 2020, AAAI 2022);
- **Reviewer** for ICLR 2025, ICRA 2018/2020, IROS 2018/2019, ICML 2022, CVPR 2021/2022, IEEE GlobeComm 2021/2022, IEEE Access 2021, IEEE Signal Processing Letters 2018.