

RYO KAMOI

ryokamoi@psu.edu

<https://ryokamoi.github.io/>**EDUCATION**

- Pennsylvania State University** - Ph.D. Student in Computer Science Aug 2023 -
Advised by Dr. Rui Zhang
- University of Texas at Austin** - Master of Science in Computer Science Aug 2020 - Dec 2022
Advised by Dr. Greg Durrett
- Keio University**, Japan - Bachelor of Engineering in Statistics Apr 2016 - Mar 2020
Advised by Dr. Kei Kobayashi, Top student in the Department of Mathematics (Keio Engineering Foundation Award)

RESEARCH INTERNSHIPS

- Amazon**, Cambridge, UK - Applied scientist intern in Alexa team (NLP) Jul - Dec 2021
- Developed an interpretable answer quality evaluation metric for chatbot outputs.
- SenseTime Japan**, Tokyo, Japan - Research internship in computer vision for autonomous driving Feb 2020 - Jan 2021
- Developed a system for unknown instance detection on a monocular camera for autonomous driving.
- Datasection Inc**, Tokyo, Japan - Research internship in NLP May 2017 - Aug 2018
- Research in text generation models (e.g., text VAEs) with small training datasets.

SELECTED PAPERS<https://scholar.google.com/citations?user=4OWTLKAAAAAJ>

My research interests lie in the broad area of natural language processing, including the detection and mitigation of errors in responses from language models, evaluation of language models, fact verification, and text summarization.

- Ryo Kamoi**, Yusen Zhang, Nan Zhang, Jiawei Han, Rui Zhang. 2024. When Can LLMs Actually Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs. *arXiv preprint arXiv:2406.01297*.
- Ryo Kamoi**, Sarkar Snigdha Sarathi Das, Renze Lou, Jihyun Janice Ahn, Yilun Zhao, Xiaoxin Lu, Nan Zhang, Yusen Zhang, Ranran Haoran Zhang, Sujeeth Reddy Vummanthala, Salika Dave, Shaobo Qin, Arman Cohan, Wenpeng Yin, Rui Zhang. 2024. Evaluating LLMs at Detecting Errors in LLM Responses. *In COLM*.
- Ryo Kamoi**, Tanya Goyal, Juan Diego Rodriguez, and Greg Durrett. 2023. WiCE: Real-world Entailment for Claims in Wikipedia. *In EMNLP (main)*.
- Ryo Kamoi**, Tanya Goyal, Juan Diego Rodriguez, and Greg Durrett. 2023. Shortcomings of Question Answering Based Factuality Frameworks for Error Localization. *In EACL (main)*.
- Ryo Kamoi** and Kei Kobayashi. 2020. Why is the Mahalanobis Distance Effective for Anomaly Detection? *arXiv preprint arXiv:2003.00402*.

SERVICE

Reviewer - ARR (2023-present), COLM (2024), AISTATS (2021), BMVC (2020)

Teaching Assistant, Penn State University

CMPSC 442: Artificial Intelligence

Spring 2024

CMPSC 448: Machine Learning and AI

Fall 2023

HONORS AND AWARDS

Keio University Global Fellowship - Scholarships for graduate study at UT Austin	2020
Keio Engineering Foundation Award - Graduation with highest honors (First place in the Dept. of Mathematics)	2020
Japan Student Services Organization (JASSO) Exchange Student Scholarship	2018

INVITED TALKS

NLP Colloquium	2023
Nagoya NLP Seminar at Nagoya University	2023

OTHER PAPERS

- Yusen Zhang, Nan Zhang, Yixin Liu, Alexander Fabbri, Junru Liu, **Ryo Kamoi**, Xiaoxin Lu, Caiming Xiong, Jieyu Zhao, Dragomir Radev, Kathleen McKeown, Rui Zhang. 2024. Fair Abstractive Summarization of Diverse Perspectives. *In NAACL 2024*.
- Yilun Zhao, Yitao Long, Hongjun Liu, Linyong Nan, Lyuhao Chen, **Ryo Kamoi**, Yixin Liu, Xiangru Tang, Rui Zhang, Arman Cohan. 2023. DocMath-Eval: Evaluating Numerical Reasoning Capabilities of LLMs in Understanding Long Documents with Tabular Data. *arXiv preprint arXiv:2311.09805*.
- Ryo Kamoi** and Kaname Tomite. 2021. Efficient Unknown Object Detection with Discrepancy Networks for Semantic Segmentation. *In the NeurIPS 2021 Workshop on Machine Learning for Autonomous Driving*.
- Ryo Kamoi** and Kei Kobayashi. 2020. Out-of-Distribution Detection with Likelihoods Assigned by Deep Generative Models Using Multimodal Prior Distributions. *In The AAAI's Workshop on Artificial Intelligence Safety*.
- Ryo Kamoi**, Kei Kobayashi. 2019. Likelihood Assignment for Out-of-Distribution Inputs in Deep Generative Models is Sensitive to Prior Distribution Choice. *arXiv preprint arXiv:1911.06515*.