JIYOUNG LEE

https://github.com/jiyounglee-0523 https://jiyounglee-0523.github.io jiyounglee0523@kaist.ac.kr

RESEARCH INTEREST

MAIN INTEREST: Evaluate AI models to ensure they align with human values and perception
KEYWORDS: benchmarks, AI-human alignment, safety, social ethics

My research focuses on assessing AI models before deployment to ensure they align with human values and perception across multiple dimensions, including uncertainty, social ethics, and linguistic diversity. I primarily investigate whether (1) AI models exhibit robustness to human variations, such as dialects, and (2) they can recognize and interpret the world in a manner similar to humans.
Through my work in constructing benchmarks, I have extensive experience in conducting human surveys involving over 6,000 participants, actively collaboration with experts from different fields, and applying statistical knowledge to build robust datasets.

EDUCATION

Ph.D in Artifical Intelligence *Sep 2022 - Present*
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
Advisor: Prof. Edward Choi

M.Sc. in Artifical Intelligence *Sep 2020 - Aug 2022*
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
Advisor: Prof. Edward Choi

B.Sc. in Statistics *March 2016 - Aug 2020*
Sookmyung Women's University, Seoul, South Korea
GPA: 4.08 / 4.3 (Ranked 1st in the Department of Statistics)

WORK EXPERIENCE

Machine Translation Research Intern *Feb 2022 - Aug 2022*
Naver Corporation - Papago Team
Specializing Multi-domain NMT via Penalizing Low Mutual Information (EMNLP 2022)

PROFESSIONAL SERVICE

Advisory Committee Member *Sep 2023 - Feb 2024*
SelectStar - LLM Dataset Construction

PUBLICATIONS

Conference Proceedings

- Soyoung Yang, Hojun Cho, **Jiyoung Lee**, Sohee Yoon, Edward Choi, Jaegul Choo, and Won Ik Cho. Single Ground Truth Is Not Enough: Add Linguistic Variability to Aspect-based Sentiment Analysis Evaluation. In *Proc. of Conference on Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL) 2025*
- **Jiyoung Lee**, Minwoo Kim, Seungho Kim, Junghwan Kim, Seunghyun Won, Hwaran Lee, and Edward Choi. KorNAT: LLM Alignment Benchmark for Korean Social Values and Common Knowledge. In *Findings in Association for Computational Linguistics (ACL) 2024*
- **Jiyoung Lee**, Seungho Kim, Seunghyun Won, Joonseok Lee, Marzyeh Ghassemi, James Thorne, Jaeseok Choi, O-Kil Kwon, and Edward Choi. VisAlign: Dataset for Measuring the Degree of Alignment between AI and Humans in Visual Perception. In *Proc. of Neural Information Processing Systems (NeurIPS) 2023 Datasets and Benchmarks*

- Woncheol Shin, Gyubok Lee, **Jiyoung Lee**, Eunyi Lyou, Joonseok Lee, and Edward Choi. Translation-equivariant Image Quantizer for Bi-directional Image-Text Generation. In *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2023*, *(Oral Presentation)*
- **Jiyoung Lee**, Hantae Kim, Hyunchang Cho, Edward Choi, and Cheonbok Park. Specializing Multi-domain NMT via Penalizing Low Mutual Information. In *Proc. of Conference on Empirical Methods in Natural Language Processing (EMNLP) 2022*
- Kyunghoon Hur*, **Jiyoung Lee***, Jungwoo Oh, Wesley Price, Young-Hak Kim, and Edward Choi. Unifying Heterogenous Electronic Health Records Systems via Text-Based Code Embedding. In *Proc. of Conference on Health, Inference, and Learning (CHIL) 2022*

Workshops, Preprints, and Domestic Conferences

- Radhika Dua, **Jiyoung Lee**, Joon-myung Kwon, and Edward Choi. Automatic Detection of Noisy Electrocardiogram Signals without Explicit Noise Labels. In *International Workshop on Pattern Recognition in Healthcare Analytics (PRHA) 2022*
- **Jiyoung Lee**, Wonjae Kim, Daehoon Gwak, and Edward Choi. Conditional Generation of Periodic Signals with Fourier-Based Decoder. In *Deep Generative Models and Downstream Applications Workshop at NeurIPS 2021*

TECHNICAL SKILLS

Program Language	Python
Packages	Pandas, Numpy, NLTK, Pytorch, HuggingFace, PyTorchLightning
Language	Korean (Native), English (Fluent)

EXPERIENCES

Teaching Assistant	AI504: Programming for AI, KAIST, *2024 Fall*
	AI504: Programming for AI, KAIST, *2023 Fall*
	AI612: Machine Learning for Healthcare, KAIST, *2023 Spring*
	AI504: Programming for AI, KAIST, *2021 Fall*
	AI612: Machine Learning for Healthcare, KAIST, *2021 Spring*
	AI504: Programming for AI, KAIST, *2020 Fall*
Academic Service	ACCV 2022 Reviewer
	ACL 2025 Reviewer

AWARDS & SCHOLARSHIPS

NeurIPS 2023 Travel Award
Google Conference Scholarship 2022
National Science & Technology Scholarship (2018 - 2020)