JISEON KIM

Ph.D candidate

✉ jiseon_kim@kaist.ac.kr   🌐 hikoseon12.github.io

in jiseon-kim-8ab574136   ○ github.com/hikoseon12

SUMMARY

I'm a PhD candidate advised by Alice Oh at KAIST. My research interests lie in natural language processing (NLP) and computational social science (CSS), with a focus on **1) AI alignment with human and societal values** and **2) AI for social good**. In particular, I work on the following topics:

- **LLM-Human Alignment & Evaluation:** I explore LLM alignment with human values and society, examining their behaviors and limitations (e.g., moral decision [C9], cultural bias [C5], social reasoning [C7]).

- **AI for Science & Social Impact:** I develop AI frameworks to process large-scale, expertise-driven data, particularly in political science (e.g., legislative processes[C4], lobbying[W2]), to uncover hidden dynamics, enhance transparency, and assess societal impact.

Keywords: AI Alignment, LLM Evaluation, AI for Policy & Governance, AI for Social Good, NLP, Computational Social Science

EXPERIENCE

3/2023 - 6/2023   **Research Intern** @ NAVER AI Lab   **NAVER AI Lab**
- Constructed a Korean bias benchmark dataset to make safer and trustworthy Korean LLM
- Work published at TACL 2024 - "KoBBQ: Korean Bias Benchmark for Question Answering"
- Advised by Hwaran Lee

5/2024 - 5/2024   **Visiting Researcher** @ MIT   **MIT**
7/2022 - 8/2022
6/2019 - 8/2019
- Conducted interdisciplinary research with political science to understand the US legislative process
- Work published at EMNLP 2021 - "Learning Bill Similarity with Annotated and Augmented Corpora of Bills"
- Collaborated with Elden Griggs and In Song Kim

3/2019 – 2/2020   **Researcher** @ KAIST   **KAIST**
- Researched multimodal NLP utilizing text and color
- Advised by Alice Oh

6/2015 – 8/2015   **Visiting Student** @ UC Berkeley   **UC Berkeley**
- Completed Computer Science 61A, the structure and Interpretation of Computer Programs
- Received support for the UC Berkeley summer session program from the Sookmyung Women's University

EDUCATION

3/2020-Present   **Korea Advanced Institute of Science and Technology**   **Daejeon, Korea**
Ph.D. in School of Computing
Advised by Alice Oh

3/2017-2/2019   **Korea Advanced Institute of Science and Technology**   **Daejeon, Korea**
Master in School of Computing
Thesis: Color Generation for Paragraph Level of Text
Advised by Alice Oh

3/2013-2/2017   **Sookmyung Women's University**   **Seoul, Korea**
B.S. Student in Computer Science
Graduated with the highest honor (1/68)

PUBLICATION

EMNLP 2024   **[C7] Perceptions to Beliefs: Exploring Precursory Inferences for Theory of Mind in Large Language Models**
Chani Jung, Dongkwan Kim, Jiho Jin, <u>Jiseon Kim</u>, Yeon Seonwoo, Yejin Choi, Alice Oh, Hyunwoo Kim

Technical Report 2024   **[C6] HyperCLOVA X Technical Report**   **NAVER AI Lab**
Kang Min Yoo, Jaegeun Han, Sookyo In, Heewon Jeon, ... <u>Jiseon Kim</u>*...(additional authors)

TACL 2024, present at ACL 2024   **[C5] KoBBQ: Korean Bias Benchmark for Question Answering**   **NAVER AI Lab**
Jiho Jin*, <u>Jiseon Kim</u>*, Nayeon Lee*, Hanual Yoo*, Alice Oh, Hwaran Lee [*]equal contribution

EMNLP 2021 Long paper   **[C4] Learning Bill Similarity with Annotated and Augmented Corpora of Bills**   **MIT**
<u>Jiseon Kim</u>, Elden Griggs, In Song Kim, Alice Oh

EMNLP 2021 Short paper	**[C3] Efficient Contrastive Learning via Novel Data Augmentation and Curriculum Learning** Seonghyeon Ye, <u>Jiseon Kim</u>, Alice Oh	
EMNLP 2021 Long paper	**[C2] Dimensional emotion detection from categorical emotion** Sungjoon Park, <u>Jiseon Kim</u>, Seonghyeon Ye, Jaeyeol Jeon, Hee Young Park, Alice Oh	
IEEE transactions on intelligent transportation systems 2020	**[C1] Denoising recurrent neural networks for classifying crash-related events** Sungjoon Park, Yeon Seonwoo, <u>Jiseon Kim</u>, Jooyeon Kim, Alice Oh	

WORKSHOP

WiML@ NeurIPS 2024	**[W2] Understanding Lobbying Strategies in Legislative Process: Bill Position Dataset and Lobbying Analysis** <u>Jiseon Kim</u>, Dongkwan Kim, Joohye Jeong, In Song Kim, Alice Oh	
C3NLP, Co-located with ACL 2024	**[W1] KoBBQ: Korean Bias Benchmark for Question Answering** Jiho Jin*, <u>Jiseon Kim</u>*, Nayeon Lee*, Hanual Yoo*, Alice Oh, Hwaran Lee [*]equal contribution	**NAVER AI Lab**

AWARDS, SCHOLARSHIPS & FUNDINGS

10/2024	**2024 KAIST Graduate Student Outstanding Paper Award** Awarded for KoBBQ: Korean Bias Benchmark for Question Answering	**KAIST**
12/2019 - 8/2024	**MISTI Global Seed Funds** MIT's Global Seed Funds facilitate international collaborations for addressing global challenges	**MIT**
3/2020 - Present	**KAIST Support Scholarship (Ph.D.)**	**KAIST**
3/2017 - 2/2019	**KAIST Support Scholarship (M.S.)**	**KAIST**
2/2016	**Naver Open API Awards in Hackathon** IT community United Hackathon	**Unithon**
3/2015 - 3/2017	**Korea National Science & Technology Scholarship (B.S.)**	**Sookmyung Women's University**

TEACHING EXPERIENCE

Fall 2021 Spring 2021	**Machine Learning for NLP** Teaching Assistant	**KAIST**
Fall 2021	**Advanced Data Mining** Teaching Assistant	**KAIST**
Spring 2020	**Artificial Intelligence and Machine Learning** Head Teaching Assistant	**KAIST**
Fall 2018 Spring 2018 Fall 2017	**Data Structure** Teaching Assistant, Developed assignments	**KAIST**

ACADEMIC SERVICE

Reviwer Feb ARR 2025, ICLR 2025 Bi-Align Workshop, Feb/Apr/June ARR 2024

Volunteer FAccT 2022, COLING 2022

Undergraduate Research Program @ KAIST Spring 2024 (Received an Encouragement Award)

Individual Research Mentoring @ KAIST Summer/Fall 2020, Spring/Fall 2021, Spring 2022, Fall 2023 Spring/Fall 2024

SKILL

Language Python, Latex, PostgreSQL

Framework Pytorch, Docker, Git

LANGUAGE

English Professional

Korean Native

REFERENCE

Alice Oh Professor in School of Computing, KAIST (alice.oh@kaist.edu)