

## AFFILIATION

---

### Carnegie Mellon University

Visiting Scholar, Computer Science

- Host Professor: Sherry Tongshuang Wu

Pittsburgh, PA, USA

Sep. 2025- Mar. 2026

### Korea Advanced Institute of Science & Technology (KAIST)

M.S. Computer Science

- Advisor: Alice Oh

Daejeon, Korea

Aug. 2023 – Present

### Korea Advanced Institute of Science & Technology (KAIST)

B.S. Electronic and Electrical Engineering

- GPA: 4.02/4.3 (Major GPA: 4.11/4.3) *Summa Cum Laude*
- Rank: 3rd in the department, 9th among all graduates

Daejeon, Korea

Feb. 2019 – Aug. 2023

## PUBLICATIONS

---

\* indicates equal contribution

### FIRST-AUTHORED

---

1. **World in a Frame: Understanding Culture Mixing as a New Challenge for Vision-Language Models**  
*Eunsu Kim\**, Junyeong Park\*, Namin Ahn\*, Junseong Kim, Hitesh Laxmichand Patel, Jiho Jin, Julia Kruk, Amit Agarwal, Srikant Panda, Fenal Ashokbhai Ilasariya, Hyunjeong Shim<sup>†</sup>, Alice Oh<sup>†</sup>  
CVPR, 2026 [Arxiv](#)
2. **Are they lovers or friends? Evaluating LLMs' Social Reasoning in English and Korean Dialogues**  
*Eunsu Kim*, Junyeong Park, Juhyun Oh, Kiwoong Park, Seyoung Song, A. Seza Doğruöz, Najoung Kim\*, Alice Oh\*  
ACL (main), 2026 [Arxiv](#)
3. **BenchHub: A Unified Benchmark Suite for Holistic and Customizable LLM Evaluation**  
*Eunsu Kim\**, Haneul Yoo\*, Guijin Son, Hitesh Patel, Amit Agarwal, Alice Oh  
NeurIPS EM Workshop / Under Review in ICLR, 2025 [Arxiv](#)
4. **LLM-as-an-Interviewer: Beyond Static Testing Through Dynamic LLM Evaluation**  
*Eunsu Kim*, Juyoung Suk, Seungone Kim, Niklas Muennighoff, Dongkwan Kim, Alice Oh  
ACL (Findings), 2025 [Arxiv](#)
5. **Uncovering Factor Level Preferences to Improve Human-Model Alignment**  
Juhyun Oh\*, *Eunsu Kim\**, Jiseon Kim, Wenda Xu, Inha Cha, William Yang Wang, Alice Oh  
EMNLP (Findings), 2025 [Arxiv](#)
6. **CLiCK: A Benchmark Dataset of Cultural and Linguistic Intelligence in Korean**  
*Eunsu Kim*, Juyoung Suk, Philhoon Oh, Haneul Yoo, James Thorne, Alice Oh  
LREC-COLING, 2024 [Arxiv](#)
7. **The Generative AI Paradox on Evaluation: What It Can Solve, It May Not Evaluate**  
Juhyun Oh\*, *Eunsu Kim\**, Inha Cha\*, Alice Oh  
EACL SRW, 2024 [Arxiv](#)

### CO-AUTHORED

---

1. **RoleConflictBench: A Benchmark of Role Conflict Scenarios for Evaluating LLMs' Contextual Sensitivity**  
Jisu Shin, Hoyun Song, Juhyun Oh, Changgeon Ko, *Eunsu Kim*, Chani Jung, Alice Oh  
ACL (Findings), 2026 [Arxiv](#)
2. **Spotting Out-of-Character Behavior: Atomic-Level Evaluation of Persona Fidelity in Open-Ended Generation**  
Jisu Shin, Juhyun Oh\*, *Eunsu Kim\**, Hoyun Song, Alice Oh  
ACL (Findings), 2025 [Arxiv](#)
3. **MUG-Eval: A Proxy Evaluation Framework for Multilingual Generation Capabilities in Any Language**  
Seyoung Song, Seogyong Jeong, *Eunsu Kim*, Jiho Jin, Dongkwan Kim, Jay Shin, Alice Oh  
EMNLP (Findings), 2025 [Arxiv](#)
4. **Flex-TravelPlanner: A Benchmark For Flexible Planning with Language Agents**  
Juhyun Oh, *Eunsu Kim*, Alice Oh  
ICLR 2025 Reasoning and Planning Workshop, 2025 [Arxiv](#)

5. **BLEnD: A Benchmark for LLMs on Everyday Knowledge in Diverse Cultures and Languages**  
*Junho Myung\*, Nayeon Lee\*, Yi Zhou\*, Jiho Jin, Rifki Afina Putri, Dimosthenis Antypas, Hsuvas Borkakoty, Eunsu Kim, Carla Perez Almendros, Abinew Ali Ayele, Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, Hwaran Lee, Shamsuddeen Hassan Muhammad, Kiwoong Park, Anar Sabuhi Rzayev, Nina White, Seid Muhie Yimam, Mohammad Taher Pilehvar, Nedjma Ousidhoum, Jose Camacho-Collados, Alice Oh*  
 NeurIPS D&B, 2024 [Arxiv](#)

## MENTORING

*Projects where I served as a mentor for undergrads*

1. **ML-IAM: Emulating Integrated Assessment Models With Machine Learning**  
*Yen Shin, Changyoon Lee, Eunsu Kim, Junho Myung, Kiwoong Park, Min-Young Choi, Bomi Kim, Jiheun Ha, Hyun W. Ka, Jung-Hun Woo, Alice Oh, Haewon McJeon*  
 NeurIPS TCCML Workshop, 2025 / Under Review in GMD, 2025 [Poster](#)
2. **When Tom Eats Kimchi: Evaluating Cultural Awareness of Multimodal Large Language Models in Cultural Mixture Contexts**  
*Jun Seong Kim, Kyaw Ye Thu, Javad Ismayilzada, Junyeong Park, Eunsu Kim, Huzama Ahmad, Na Min An, James Thorne, Alice Oh*  
 NAACL C3NLP (Outstanding Paper 🏆), 2025 [Arxiv](#)
3. **Diffusion Models Through a Global Lens: Are They Culturally Inclusive?**  
*Zahra Bayramli, Ayhan Suleymanzade, Na Min An, Huzama Ahmad, Eunsu Kim, Junyeong Park, James Thorne, Alice Oh*  
 ACL, 2025 (Oral) / NAACL C3NLP, 2025 [Arxiv](#)
4. **Multi-Fact: Assessing Factuality of Multilingual LLMs using FactScore**  
*Sheikh Shafayat, Eunsu Kim\*, Juhyun Oh\*, Alice Oh*  
 COLM, 2024 [Arxiv](#)

## WORK EXPERIENCE

**Hyundai Motors Company, NLP Team**  
*Internship*

Seoul, Korea  
 Mar 2022 – Aug 2022

## TEACHING EXPERIENCE

### TEACHING ASSISTANT

#### AI Ethics

*Assisted in lectures and assignments for the undergraduate course on AI Ethics*

KAIST  
 Spring 2024

### COUNSELING ASSISTANT

#### Leader of Counseling Assistants

*Served as the leader of Counseling Assistants in the CS Dept.*

KAIST CS  
 Spring 2024

#### Counseling Assistants

*One of three Counseling Assistants in the CS Dept, Counseling 80+ KAIST students.*

KAIST CS  
 Fall 2023–Spring 2024

## MENTOR

- Yen Shin (KAIST, 2024 Fall–Present)
- Jinwoong Park (KAIST, 2025 Fall–Present)
- Junseong Kim (KAIST, 2024 Fall–Present)
- Kyaw Ye Thu, Javad Ismayilzada, Zahra Bayramli, Ayhan Suleymanzade (KAIST, 2024 Fall–2025 Spring)
- Eva Goldie (MIT, 2024 Summer)
- Sheikh Shafayat (KAIST, 2023 Fall–2024 Spring)

## TUTOR

### General Physics 1

*Provided one-on-one tutoring for undergraduate students in physics*

KAIST  
 Mar 2022 – Dec 2022

### Math & Science

*Mentored middle school students in advanced math and science topics*

KAIST Center for Gifted Education  
 Apr 2021 – Dec 2024

### AI

*Designed AI-related educational content for gifted middle school students*

KAIST Center for Gifted Education  
 Sep 2023 – Dec 2023

## ACADEMIC SERVICE

### ORGANIZER

#### International NLP Workshop at KAIST

*Organized sessions and coordinated logistics for the workshop*

Seoul, Korea  
 Aug 2024

REVIEWER

---

ARR Feb 2025, CVPR DemoDiv 2025, NeurIPS EM 2025, ICLR 2026

TALKS

---

**Hanbat National University**

*Evaluating Large Language Models: A Perspective of Building Evaluation Benchmark Datasets*

- Invited talk on evaluation benchmarks for large language models

Daejeon, Korea

Jun 2024

SCHOLARSHIPS & AWARDS

---

**National Science and Technology Scholarship**

*Scholarship Awardee*

Korea

2020 – 2022

**Global Information Security Overseas Research Fellowship**

*Scholarship Awardee*

KIISC

2025