Eunsu Kim

✉ kes0317@kaist.ac.kr | 🌐 eunsu-k1m.github.io

AFFILIATION

Carnegie Mellon University   Pittsburgh, PA, USA
Visiting Scholar, Computer Science   *Sep. 2025- Mar. 2026*
- Host Professor: Sherry Tongshuang Wu

Korea Advanced Institute of Science & Technology (KAIST)   Daejeon, Korea
M.S. Computer Science   *Aug. 2023 – Present*
- Advisor: Alice Oh

Korea Advanced Institute of Science & Technology (KAIST)   Daejeon, Korea
B.S. Electronic and Electrical Engineering   *Feb. 2019 – Aug. 2023*
- GPA: 4.02/4.3, *Summa Cum Laude*
- Major GPA: 4.11/4.3

PUBLICATIONS   ** indicates equal contribution*

FIRST-AUTHORED

1. **SCRIPTS: Evaluating LLMs' Interpersonal Social Reasoning in Conversational Context**
   __Eunsu Kim__, Junyeong Park, Juhyun Oh, Kiwoong Park, Seyoung Song, A. Seza Doğruöz, Najoung Kim, Alice Oh
   Under Review in TACL, 2025

2. **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

3. **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

4. **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

5. **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

6. **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

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
   Under Review in ICLR, 2025

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

3. **MUG-Eval: A Proxy Evaluation Framework for Multilingual Generation Capabilities in Any Language**
   *Seyoung Song, Seogyeong Jeong, **Eunsu Kim**, Jiho Jin, Dongkwan Kim, Jay Shin, Alice Oh*
   EMNLP (Findings), 2025

4. **Flex-TravelPlanner: A Benchmark For Flexible Planning with Language Agents**
   *Juhyun Oh, **Eunsu Kim**, Alice Oh*
   ICLR 2025 Reasoning and Planning for LLMs Workshop, 2025

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

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, and Haewon McJeon*
   Neurips TCCML Workshop/ Under Review in GMD journal, 2025

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

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 (Oral)/ NAACL C3NLP, 2025

4. **Multi-FAct: Assessing Factuality of Multilingual LLMs using FActScore**
   *Sheikh Shafayat, **Eunsu Kim***, Juhyun Oh*, Alice Oh*
   COLM, 2024

TEACHING EXPERIENCE

TEACHING ASSISTANT

AI Ethics   KAIST
Assisted in lectures and assignments for the undergraduate course on AI Ethics   *Spring 2024*

COUNSELING ASSISTANT

Leader of Counseling Assistants   KAIST CS
Served as the leader of Counseling Assistants in the CS Dept.   *Spring 2024*

Counseling Assistants   KAIST CS
One of three Counseling Assistants in the CS Dept.   *Fall 2023*

TUTOR

General Physics 1   KAIST
Provided one-on-one tutoring for undergraduate students in physics   *Mar 2022 – Dec 2022*

Math & Science	KAIST Center for Gifted Education
Mentored middle school students in advanced math and science topics	*Apr 2021 – Dec 2024*
AI	KAIST Center for Gifted Education
Designed AI-related educational content for gifted middle school students	*Sep 2023 – Dec 2023*

WORK EXPERIENCE

Hyundai Motors Company, NLP Team	Seoul, Korea
Internship	*Mar 2022 – Aug 2022*

- Worked on natural language processing projects in the NLP Team

ACADEMIC SERVICE

ORGANIZER

International NLP Workshop at KAIST	Seoul, Korea
Organized sessions and coordinated logistics for the workshop	*Aug 2024*

REVIEWER

ARR Feb 2025, CVPR DemoDiv 2025, ICLR 2026

TALKS

Hanbat National University	Daejeon, Korea
Evaluating Large Language Models: A Perspective of Building Evaluation Benchmark Datasets	*Jun 2024*

- Invited talk on evaluation benchmarks for large language models

SKILLS

- Programming — Python, Pytorch, Matlab, C

SCHOLARSHIPS & AWARDS

National Science and Technology Scholarship	Korea
Scholarship Awardee	*2020 – 2022*