

Jaehong Kim

Ph.D. Candidate at KAIST (Expected Graduation: February 2027)
luke.4.18@kaist.ac.kr | <https://jaehong-k.github.io/>

RESEARCH INTERESTS

My research centers on **moral emotions**, the emotions at the core of the **attention economy** and **affective polarization**. I (1) develop models that **measure** them across languages and modalities, (2) trace their impacts, from exploitation in suspected **influence operations** to spillover into **offline protest**, and (3) study how to control them at the root by locating their representations inside **large language models**.

PUBLICATIONS

First-Author Publications (* denotes co-first authorship, † denotes international collaboration)

- **J. Kim***, H. Kim*, J. Kim, A. Oh, T. Holz, W. Lee, M. Cha. (2026). Cross-National Information Attacks: A Two-Decade Analysis of Troll Behavior in Korea. *In Proc. of the 35th USENIX Security Symposium (USENIX Security '26)*. [\[Zenodo\]](#) †
- S. Park*, **J. Kim***, H. Kim, H. Bin, S. Moon, W. Lee. (2026). Moral Outrage Shapes Commitments Beyond Attention: Multimodal Moral Emotions on YouTube in Korea and the US. *In Proc. of The Web Conference (WWW 2026)*. [\[GitHub\]](#) †
- W. Dong*, M. Sundriyal*, S. Park*, **J. Kim***, M. Cha, T. Chakraborty, W. Lee. (2025). Parallel Communities Across the Surface Web and the Dark Web. *In Findings of Empirical Methods in Natural Language Processing (EMNLP Findings 2025)*. *Best Resource Paper Nomination* †
- **J. Kim***, C. Jeong*, S. Park*, M. Cha, W. Lee. (2024). How Do Moral Emotions Shape Political Participation? A Cross-Cultural Analysis of Online Petitions Using Language Models. *In Findings of the Association for Computational Linguistics (ACL Findings 2024)*. [\[GitHub\]](#) †
- **J. Kim**, C. Jeong, M. Cha, W. Lee. (2023). KOMA (Korean Online Moral Emotion): A Korean Dataset of Fine-Grained Moral Emotions. *Korea Computer Congress 2023 (KCI)*. [\[GitHub\]](#)

Co-Authored Publications

- J. Kim, J. Kwon, L. F. Vecchietti, W. Dong, **J. Kim**, M. Cha. (2026). Machine Behavior in Relational Moral Dilemmas: Moral Rightness, Predicted Human Behavior, and Model Decisions. *In Findings of the Association for Computational Linguistics (ACL Findings 2026)*. [\[GitHub\]](#) †
- J. Jeon, B. Kim, **J. Kim**, R. Kim. (2023). Opening the Blackbox of Computational Social Science Research Process: A Case of Comparative Study of Social Science Academia. *Korean Journal of Sociology (KCI)*
- J. Park, Y. Cho, **J. Kim**, H. Choi. (2021). A Commercial District Analysis Service: A Service Reflecting the Current COVID-19's Situation in the Seocho-gu District. *Proceedings of HCI Korea (KCI)*
- J. Lee, **J. Kim**, Y. Cho, M. Lee, H. Choi. (2021). Study on Effective Extraction of New Coined Vocabulary from Political Domain Articles and News Comments. *The Journal of the Convergence on Culture Technology (KCI)*
- H. Choi, **J. Kim**, J. Lee, M. Lee. (2020). Political Information Filtering on Online News Comment. *The Journal of the Convergence on Culture Technology (KCI)*

Working Papers

- **J. Kim**, Junsol Kim, J. Kwon, H. Kim, M. Cha, W. Lee. The Concept of Moral Outrage in Large Language Models and Its Function in Moral Decision-Making. *Working paper*. †
- **J. Kim**, S. Park, W. Dong, A. Zhunis, M. Cha, W. Lee. From Online Outrage to Offline Protest: Evidence from the Israel-Palestine and Russia-Ukraine Conflicts. *Working paper*. †

RESEARCH EXPERIENCE

Max Planck Institute for Security and Privacy (MPI-SP)

Research Intern, Data Science for Humanity (Advisor: Meeyoung Cha)
Research Intern, Data Science for Humanity (Advisor: Meeyoung Cha)

Bochum, Germany
2025.09 - 2026.02
2024.09 - 2024.11

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST) Daejeon, Korea
Ph.D., Graduate School of Culture Technology, Social Computing Lab (Advisor: Wonjae Lee) 2023.08 - Present
Dissertation: Measuring, Understanding, and Controlling Moral Emotions: From Online Platforms to Large Language Models
(Expected: February 2027)

M.S., Graduate School of Culture Technology, Social Computing Lab (Advisor: Wonjae Lee) 2021.08 - 2023.08
Thesis: How moral emotions shape the commitment of political support in the online petition

Handong Global University Pohang, Korea
B.S., Communication, Psychology, ICT Convergence (Minor) 2015.03 - 2021.08
Cum Laude

INVITED TALKS

NYU Center for Conflict and Cooperation 2026

Max Planck Institute for Security and Privacy, Data Science for Humanity Group 2026

Reichman University, AI/NLP Seminar 2026

Korea Computer Congress Top Conference Session 2026

Max Planck Institute for Security and Privacy, Workshop on the Societal Impact of AI Systems 2025

FELLOWSHIPS & AWARDS

KAIST Fake News Challenge Daejeon, Korea
Best Prize (KRW 2,000,000) 2026

Global Advanced Human Resource Development Program in Information Security, IITP Daejeon, Korea
Competitive national fellowship for overseas research training; funded research visit to MPI-SP (KRW 18,000,000) 2025

Hyundai Motor Chung Mong-Koo Global Scholarship Seoul, Korea
(KRW 28,000,000) 2025

Adventure Research Program, KAIST Daejeon, Korea
(KRW 1,000,000) 2024

Young-Han Kim Global Leader Scholarship, KAIST Daejeon, Korea
(KRW 4,000,000) 2023

TEACHING ASSISTANT

Natural Language Processing for the Humanities and Social Sciences (HSS510) 2025.03 – 2025.06
Led lab sessions on LLM fine-tuning, covering core concepts and hands-on implementation

Social Network Analysis (CTP471) 2024.03 – 2024.06
Assisted student projects on network theory and analysis in R, providing feedback and technical support

Social Computing (GCT576) 2023.09 – 2023.12
Mentored student research applying core sociological concepts to computational social science

Big Data (MIP801) 2022.09 – 2026.06
Led hands-on tutorials on web crawling and data collection

Computational Social Science (HSS312). Selected as Best TA 2021.09 – 2021.12
Led coding tutorials on data collection methods (e.g., web scraping, APIs)

ACADEMIC SERVICE

CSCW 2025, FinNLP-FNP-LLMFinLegal-2024 Shared Task