

## EDUCATION

### Ph.D. Candidate in Artificial Intelligence

2022 - Present

Department of Artificial Intelligence, Ajou University

Suwon, South Korea

- Advisor: Prof. Kyung-Ah Sohn
- Research Areas: Natural Language Processing, AI Safety, Medical AI
- research focus: Robust AI, Multi-Agent Systems, LLM Bias, Explainable AI

### B.S. in Applied Artificial Intelligence & B.A. in Sociology

2018 - 2022

Department of Applied Artificial Intelligence,  
College of Social Sciences, Ajou University

Suwon, South Korea

## PUBLICATIONS

1. Jaesik Kim\*, **Byoungchan Lee\***, Kyung-Ah Sohn, Dokyoon Kim, Young Chan Lee. Evaluation of Multi-Agent LLMs in Multidisciplinary Team Decision-Making for Challenging Cancer Cases. *Machine Learning for Healthcare (MLHC)*, 2025.
2. Hyeonwoo Jung, **Byoungchan Lee**, Kyung-Ah Sohn. Counter Narrative Generation through Knowledge-Injected Prompt Learning. *Korea Software Congress (KSC2024)*, Yeosu, South Korea, 2024.
3. **Byoungchan Lee**, Jaesik Kim, Junekeyu Park, Kyung-Ah Sohn. Improving Unsupervised Out-of-domain Detection through Pseudo Labeling and Learning. *Findings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Dubrovnik, Croatia, 2023.
4. Yunseo Choi, **Byoungchan Lee**, Kyung-Ah Sohn. Korean Hate Speech Detection Using Hate-Rationale Word Prediction. *Korea Software Congress (KSC2023)*, Busan, South Korea, 2023.
5. Kyungju Choi, **Byoungchan Lee**, Kyung-Ah Sohn. Efficient Privacy-Preserving Chatbot Learning through Prefix-Learning Based Federated Learning. *Conference on Information Security and Cryptography-Winter (CISC-W 2023)*, Seoul, South Korea, 2023.
6. Jiyun Kim, **Byoungchan Lee**, Kyung-Ah Sohn. Why Is It Hate Speech? Masked Rationale Prediction for Explainable Hate Speech Detection. *International Conference on Computational Linguistics (COLING)*, Gyeongju, South Korea, 2022.
7. **Byoungchan Lee**, Kyung-Ah Sohn. Disaster Detection Using Text Anomaly Detection in SNS. *Korea Next Generation Computing Conference (KINGPC)*, South Korea, 2022.

## RESEARCH

## EXPERIENCE

### Robust Machine Learning for Low-Quality and Heterogeneous Text Data Environments

Robust AI | National Research Foundation of Korea

2022 - Present

- Developed novel pseudo-labeling techniques for out-of-domain detection
- Published research findings at EACL 2023 (Findings)

### Understanding and Mitigating Societal Biases Embedded in Large Language Models

LLM bias

2023 - Present

- Assessed multi-agent LLM systems for multidisciplinary medical decisions
- Evaluated LLM performance in challenging cancer case scenarios

### Safe and Explainable AI Systems for Content Safety and Moderation

AI Safety in NLP | National Research Foundation of Korea

2022 - 2024

- Implemented explainable hate speech detection using masked rationale prediction
- Created knowledge-injected prompt learning for counter narrative generation

TEACHING EXPERIENCE	<b>Teaching Assistant - Machine Learning</b>		
	<i>Department of Software and Computer Engineering, Ajou University</i>		Mar 2025, 2024
	<ul style="list-style-type: none"> <li>• Designed and graded ML programming assignments</li> <li>• Provided technical support and conducted Q&amp;A sessions for students</li> <li>• Led tutorials on machine learning algorithms and implementation</li> </ul>		
	<b>Teaching Assistant - Web System Project</b>		
	<i>Department of Software and Computer Engineering, Ajou University</i>		Sep 2024
	<ul style="list-style-type: none"> <li>• Supervised tutorials and provided technical guidance for web development projects</li> <li>• Conducted Q&amp;A sessions and troubleshooting support</li> </ul>		
	<b>Teaching Assistant - Machine Learning Fundamentals</b>		
	<i>Department of Applied Artificial Intelligence, Ajou University</i>		Mar 2022 - Sep 2023
	<ul style="list-style-type: none"> <li>• Developed and graded programming assignments and final projects</li> <li>• Coached students in transforming field problems into AI-driven solutions</li> <li>• Led tutorials on ML algorithms and implementation</li> </ul>		
SKILLS	<b>Machine Learning &amp; Data Science:</b> PyTorch, HuggingFace, vLLM, Scikit-learn, NumPy, Pandas, Matplotlib, Seaborn		
	<b>Programming &amp; Development:</b> Python, R, Shell Scripting, Git, Docker, Linux, OpenAI API, REST API		
	<b>Languages:</b> Korean (Native), English (Proficient)		
PRESENTATIONS	<b>Poster Presentation:</b> <i>Outstanding Research Presentation, AI Graduate School Symposium, 2025.</i>		
	<b>Oral Presentation:</b> <i>Large Language Models and Human Collaboration in Hate Speech Response, Public Good Convergence Forum, 2024.</i>		
	<b>Poster Presentation:</b> <i>Outstanding Research Presentation, AI Graduate School Symposium, 2023.</i>		
	<b>Oral Presentation:</b> <i>Online Hate Speech and Extremism: How Should Public-Private-Academic Networks Respond?, Public Good Convergence Forum, 2023.</i>		
AWARDS AND HONORS	<b>Grand Prize, MZ AI Hackathon - Medical Q&amp;A System (KRW 5,000,000)</b>		2021
	<b>Grand Prize, AJOU Valley Festa ONtact, Ajou University (KRW 3,000,000)</b>		2021
	<b>Excellence Award, SoftCon Competition, Ajou University</b>		2021