

Sunghwan Kim

📍 Seoul, Republic of Korea ✉ kimsh8564@yonsei.ac.kr 🔗 kimsh0507.github.io 📄 Google Scholar

RESEARCH INTEREST

I aim to build human-like intelligent systems that can autonomously learn, reason, and adapt to diverse environments. My recent research interests include: (1) **Reinforcement Learning (RL)** to solve long-horizon tasks (2) Developing intelligent systems that **learn through interaction with the environment**, and (3) **Investigating language models (LMs)** to identify their limitations and enhance their performance across diverse applications. My **ongoing projects** are below:

- Reward Model for Web Agent
- Personalized Embodied Agents with Memory System

EDUCATION

- M.S. in Artificial Intelligence**, Yonsei University Sep. 2024 - Present
- Advisor: Prof. Jinyoung Yeo
- B.S. in Materials Science & Engineering**, Yonsei University Mar. 2018 - Aug. 2024
- GPA: 3.89/4.30
 - Military Service (Aug. 2020 - Feb. 2022)

EXPERIENCE

- Language & AGI Lab**, Yonsei University Jul. 2023 - Aug. 2024
Undergraduate Research Intern (Advisor: Prof. Jinyoung Yeo) Seoul, Korea
- Analyzed LLMs' capabilities in providing emotional support and developed open-source datasets and models for a counseling chatbot.
- HOLIX Growth Team** Feb. 2022 - Jun. 2022
Intern (Mentor: Rankyung Kim) Seoul, Korea
- Analyzed large-scale user data to identify trends and extract key business insights.
 - Developed data-driven strategies to optimize user engagement and product performance.

PUBLICATIONS

Conference Papers

- [1] **Web Agents with World Models: Learning and Leveraging Environment Dynamics in Web Navigation**
Hyunjoo Chae, Namyoung Kim, Kai Tzu-iunn Ong, Minju Gwak, Gwanwoo Song, Jihoon Kim, **Sunghwan Kim**, Dongha Lee, Jinyoung Yeo
ICLR 2025 (*Sys2Reasoning at Scale Workshop* at NeurIPS 2024), [pdf](#)
- [2] **CACTUS: Towards Psychological Counseling Conversations using Cognitive Behavioral Theory**
Suyeon Lee*, **Sunghwan Kim***, Minju Kim*, Dongjin Kang, Dongil Yang, Harim Kim, Minseok Kang, Dayi Jung, Min Hee Kim, Seungbeen Lee, Kyoung-Mee Chung, Youngjae Yu, Dongha Lee, Jinyoung Yeo
EMNLP 2024 Findings, [pdf](#)
- [3] **Language Models as Compilers: Simulating Pseudocode Execution Improves Algorithmic Reasoning in Language Models**
Hyunjoo Chae, Yeonghyeon Kim, Seungone Kim, Kai Tzu-iunn Ong, Beong-woo Kwak, Moohyeon Kim, **Sunghwan Kim**, Taeyoon Kwon, Jiwan Chung, Youngjae Yu, Jinyoung Yeo
EMNLP 2024, [pdf](#)

- [4] **Can Large Language Models be Good Emotional Supporter? Mitigating Preference Bias on Emotional Support Conversation**
 Dongjin Kang*, **Sunghwan Kim***, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, Jinyoung Yeo
 ACL 2024 (**Outstanding Paper**), [pdf](#)

Preprints

- [1] **Rethinking Reward Model Evaluation Through the Lens of Reward Overoptimization**
Sunghwan Kim*, Dongjin Kang*, Taeyoon Kwon, Hyungjoo Chae, Dongha Lee, Jinyoung Yeo
 Under Review at ACL 2025. (Preprint available upon request)
- [2] **LLM Meets Scene Graph: Can Large Language Models Understand and Generate Scene Graphs? A Benchmark and Empirical Study**
 Dongil Yang, Minjin Kim, **Sunghwan Kim**, Beongwoo Kwak, Minjun Park, Jinseok Hong, Woontack Woo, Jinyoung Yeo
 Under Review at ACL 2025. (Preprint available upon request)

AWARDS AND HONORS

Outstanding Paper Award ACL, 2024	2024
Scholarship for Bachelor's Degree Shilla Cultural Scholarship Foundation (Shilla Holdings)	2019 - 2023
Grade Excellence Award Yonsei University	2018-1,2, 2019-2, 2022-2
Creativity Award, Army Startup Competition Awarded by the Republic of Korea Army	2021
Dean's List Yonsei University	2019-1

TEACHING ASSISTANT

Introduction to Artificial Intelligence	Sep. 2024 - Dec. 2024
Natural Language Processing and ChatGPT	Sep. 2024 - Dec. 2024