

## Research Interest

My research centers on multimodal understanding and generation, with a particular focus on Multimodal Large Language Models (MLLMs). I explore how multimodal context and text can be more effectively connected in AI systems, spanning both architectural innovations and evaluative methodologies.

## Education

2023-26*	<b>Ph.D.</b> , Artificial Intelligence, Yonsei University <i>Advisor: Seon Joo Kim (present), Youngjae Yu (~ August 2025)</i>
2019-23	<b>M.S.</b> , Computer Engineering, Seoul National University <i>Advisor: Gunhee Kim</i>
2014-18	<b>B.Sc.</b> , Computer Science and Philosophy, Yonsei University <i>Double Major</i>

## Research Experience

2025 Summer	Research Intern, Microsoft Research AI Frontiers
2024 Spring	Research Intern, LG AI Research
2023 Fall	Research Intern, Naver

## Conference Papers

 [Google Scholar](#)

† → Equal contribution

- . **Explain with Visual Keypoints Like a Real Mentor! A Benchmark for Multimodal Solution Explanation**  
Jaewoo Park, Jungyang Park, Dongju Jang, **Jiwan Chung**, Byungwoo Yoo, Jaewoo Shin, Seon-joon Park, Taehyeong Kim, Youngjae Yu  
*The 40th Annual AAAI Conference on Artificial Intelligence: AAAI 2026*
- . **Are Any-to-Any Models More Consistent Across Modality Transfers Than Specialists?**  
**Jiwan Chung**, Janghan Yoon, Junhyeong Park, Sangeyl Lee, Joowon Yang, Sooyeon Park, Young-jae Yu  
*Proceedings of the Association for Computational Linguistics: ACL 2025*
- . **Speaking Beyond Language: A Large-Scale Multimodal Dataset for Learning Nonverbal Cues from Video-Grounded Dialogues**  
**Jiwan Chung**†, Youngmin Kim†, Jisoo Kim, Sunghyun Lee, Sangkyu Lee, Junhyeok Kim, Cheoljong Yang, Youngjae Yu  
*Proceedings of the Association for Computational Linguistics: ACL 2025*

---

\*Expected.

- VAGUE: Visual Contexts Clarify Ambiguous Expressions**  
 Heejeong Nam, Jinwoo Ahn, Keummin Ka, **Jiwan Chung**, Youngjae Yu  
*Proceedings of the IEEE/CVF International Conference on Computer Vision: ICCV 2025*
- VisEscape: A Benchmark for Evaluating Exploration-driven Decision-making in Virtual Escape Rooms**  
 Seungwon Lim, Sungwoong Kim, Jihwan Yu, Sungjae Lee, **Jiwan Chung**, Youngjae Yu  
*Proceedings of the Conference on Empirical Methods in Natural Language Processing: EMNLP 2025*
- EgoSpeak: Learning When to Speak for Egocentric Conversational Agents in the Wild**  
 Junhyeok Kim, Minsoo Kim, **Jiwan Chung**, Jungbin Cho, Jisoo Kim, Sungwoong Kim, Gyeongbo Sim, Youngjae Yu  
*Findings of the Association for Computational Linguistics: NAACL Findings 2025*
- Do LLMs Have Distinct and Consistent Personality? TRAIT: Personality Testset designed for LLMs with Psychometrics**  
 Seungbeen Lee, Seungwon Lim, Seungju Han, Giyeong Oh, Hyungjoo Chae, **Jiwan Chung**, Minju Kim, Beong-woo Kwak, Yeonsoo Lee, Dongha Lee, Jinyoung Yeo, Youngjae Yu  
*Findings of the Association for Computational Linguistics: NAACL Findings 2025*
- MASS: Overcoming Language Bias in Image-Text Matching**  
**Jiwan Chung**, Seungwon Lim, Sangkyu Lee, Youngjae Yu  
*The 39th Annual AAAI Conference on Artificial Intelligence: AAAI 2025*
- CANVAS: Commonsense-Aware Navigation System for Intuitive Human-Robot Interaction**  
 Suhwan Choi, Yongjun Cho, Minchan Kim, Jaeyoon Jung, Myunchul Joe, Yubeen Park, Minseo Kim, Sungwoong Kim, Sungjae Lee, Hwiseong Park, **Jiwan Chung**, Youngjae Yu  
*IEEE International Conference on Robotics and Automation: ICRA 2025*
- Towards Visual Text Design Transfer Across Languages**  
**Jiwan Chung**†, Yejin Choi†, Sumin Shim, Giyeong Oh, Youngjae Yu  
*Advances in Neural Information Processing Systems, Datasets and Benchmarks: NeurIPS D&B 2024*
- Selective Vision is the Challenge for Visual Reasoning: A Benchmark for Visual Argument Understanding**  
**Jiwan Chung**†, Sungjae Lee†, Minseo Kim, Seungju Han, Ashkan Yousefpour, Jack Hessel, Youngjae Yu  
*Proceedings of the Conference on Empirical Methods in Natural Language Processing: EMNLP 2024*
- Can visual language models resolve textual ambiguity with visual cues? Let visual puns tell you!**  
**Jiwan Chung**, Seungwon Lim, Jaehyun Jeon, Seungbeen Lee, Yu Youngjae  
*Proceedings of the Conference on Empirical Methods in Natural Language Processing: EMNLP 2024*
- Language models as compilers: Simulating pseudocode execution improves algorithmic reasoning in language models**  
 Hyungjoo Chae, Yeonghyeon Kim, Seungone Kim, Kai Tzu-iunn Ong, Beong-woo Kwak, Moohyeon Kim, Seonghwan Kim, Taeyoon Kwon, **Jiwan Chung**, Youngjae Yu  
*Proceedings of the Conference on Empirical Methods in Natural Language Processing: EMNLP 2024*

- . **VLIS: Unimodal Language Models Guide Multimodal Language Generation**  
**Jiwan Chung**, Youngjae Yu  
*Proceedings of the Conference on Empirical Methods in Natural Language Processing: EMNLP 2023*
- . **Reading Books is Great, But Not if You Are Driving! Visually Grounded Reasoning about Defeasible Commonsense Norms**  
Seungju Han, Junhyeok Kim, Jack Hessel, Liwei Jiang, **Jiwan Chung**, Yejin Son, Yejin Choi, Youngjae Yu  
*Proceedings of the Conference on Empirical Methods in Natural Language Processing: EMNLP 2023*
- . **Long Story Short: a Summarize-then-Search Method for Long Video Question Answering**  
**Jiwan Chung**, Youngjae Yu  
*British Machine Vision Conference: BMVC 2023*
- . **Fusing pre-trained language models with multimodal prompts through reinforcement learning**  
**Jiwan Chung†**, Youngjae Yu†, Heeseung Yun, Jack Hessel, Jae Sung Park, Ximing Lu, Rowan Zellers, Prithviraj Ammanabrolu, Ronan Le Bras, Gunhee Kim  
*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: CVPR 2023*
- . **Acav100m: Automatic curation of large-scale datasets for audio-visual video representation learning**  
**Jiwan Chung†**, Sangho Lee†, Youngjae Yu, Gunhee Kim, Thomas Breuel, Gal Chechik, Yale Song  
*Proceedings of the IEEE/CVF International Conference on Computer Vision: ICCV 2021*
- . **Transitional adaptation of pretrained models for visual storytelling**  
**Jiwan Chung†**, Youngjae Yu†, Heeseung Yun, Jongseok Kim, Gunhee Kim  
*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: CVPR 2021*
- . **Character grounding and re-identification in story of videos and text descriptions**  
Youngjae Yu, Jongseok Kim, Heeseung Yun, **Jiwan Chung**, Gunhee Kim  
*European Conference on Computer Vision: ECCV 2020*

## Preprints

- . **What MLLMs Learn about When they Learn about Multimodal Reasoning: Perception, Reasoning, or their Integration?**  
**Jiwan Chung**, Neel Joshi, Pratyusha Sharma, Youngjae Yu, Vibhav Vineet  
*arXiv preprint arXiv:2510.01719 2025*
- . **v1: Learning to Point Visual Tokens for Multimodal Grounded Reasoning**  
**Jiwan Chung**, Junhyeok Kim, Siyeol Kim, Jaeyoung Lee, Min Soo Kim, Youngjae Yu  
*arXiv preprint arXiv:2505.18842 2025*
- . **Teaching Metric Distance to Discrete Autoregressive Language Models**  
**Jiwan Chung**, Saejin Kim, Yongrae Jo, Jaewoo Park, Dongjun Min, Youngjae Yu  
*arXiv preprint arXiv:2503.02379 2025*

- . **GuideDog: A Real-World Egocentric Multimodal Dataset for Blind and Low-Vision Accessibility-Aware Guidance**  
Junhyeok Kim, Jaewoo Park, Junhee Park, Sangeyl Lee, **Jiwan Chung**, Jisung Kim, Ji Hoon Joung, Youngjae Yu  
*arXiv preprint arXiv:2503.12844* 2025
- . **SEAL: Entangled White-box Watermarks on Low-Rank Adaptation**  
Giyeong Oh, Saejin Kim, Woohyun Cho, Sangkyu Lee, **Jiwan Chung**, Dokyung Song, Youngjae Yu  
*arXiv preprint arXiv:2501.09284* 2025
- . **HyperCLOVA X Technical Report**  
Kang Min Yoo, Jaegeun Han, Sookyo In, Heewon Jeon, Jisu Jeong, Jaewook Kang, Hyunwook Kim, Kyung-Min Kim, Munhyong Kim, Sungju Kim, HyperCLOVA X Team  
*arXiv preprint arXiv:2404.01954* 2024

---

Last updated: November 10, 2025