

# Minjoon Jung

📍 Seoul, South Korea   ✉ Email   🌐 Homepage   in LinkedIn   🐙 Github

## Research Interests

I have a broad interest in AI systems that interact with **multimodal data (Language + X, where X includes vision, video, and robotics)**. My long-term research goal is to develop fine-grained, trustworthy multimodal AI agents that are both interpretable and highly efficient. Currently, I am focusing on how AI systems comprehend, learn from, and interact with videos — particularly by advancing methods for temporal video understanding and conducting fine-grained evaluations to enhance the reliability and trustworthiness of video comprehension.

## Education

- |            |   |                     |
|------------|---|---------------------|
| <b>PhD</b> | <b>Seoul National University</b> , Interdisciplinary Program in Artificial Intelligence | Mar 2021 – Present  |
|            | • Advisor: <a href="#">Prof. Byoung-Tak Zhang</a>                                       |                     |
| <b>BS</b>  | <b>Chung-Ang University</b> , Software Engineering                                      | Mar 2015 – Sep 2021 |
|            | • Mentored by <a href="#">Prof. Junseok Kwon</a>  |                     |
|            | • 2 years of absence for military service   |                     |

## Experience

- |  |                     |
|--|---------------------|
| <b>National University of Singapore @ CVML</b> , Research Intern             | Singapore, SG       |
| • Host: <a href="#">Prof. Angela Yao</a> and <a href="#">Dr. Junbin Xiao</a> | Feb 2024 – Feb 2025 |
| <b>SNU-NAVER Hyperscale AI Center</b> , Student Researcher                   | Seoul, KR           |
| • Host: <a href="#">Dr. Jin-Hwa Kim</a>                                      | Sep 2021 – Sep 2023 |
| <b>Daewoong AI &amp; BIG DATA</b> , Student Researcher                       | Seoul, KR           |
|  | Mar 2021 – Feb 2022 |

## Publications

C: conference, W: workshop, J: journal, P: preprint / \*: equal contribution, †: corresponding author / 🔗: Link

### [W4] Exploring Ego-Exo View-Invariant Temporal Understanding in Video LLMs

*NeurIPS Workshop on Multimodal Algorithmic Reasoning*, 2025

[Minjoon Jung](#), [Junbin Xiao](#)<sup>†</sup>, [Junghyun Kim](#), [Byoung-Tak Zhang](#), [Angela Yao](#)

### [C5] Confidence-guided Refinement Reasoning for Zero-shot Question Answering

*Empirical Methods in Natural Language Processing (EMNLP)*, 2025

[Youwon Jang](#), [Woo-Suk Choi](#), [Minjoon Jung](#), [Minsu Lee](#), [Byoung-Tak Zhang](#)<sup>†</sup>

### [C4] On the Consistency of Video Large Language Models in Temporal Comprehension 🔗

*Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025

[Minjoon Jung](#), [Junbin Xiao](#)<sup>†</sup>, [Byoung-Tak Zhang](#), [Angela Yao](#)

### [W3] Exploring Ordinal Bias in Action Recognition for Instructional Videos 🔗

*ICLR Workshop on Spurious Correlation and Shortcut Learning*, 2025

[Joochan Kim](#), [Minjoon Jung](#), [Byoung-Tak Zhang](#)<sup>†</sup>

### [C3] Background-aware Moment Detection for Video Moment Retrieval 🔗

*Winter Conference on Applications of Computer Vision (WACV)*, 2025

[Minjoon Jung](#), [Youwon Jang](#), [Seongho Choi](#), [Joochan Kim](#), [Jin-Hwa Kim](#)<sup>†</sup>, [Byoung-Tak Zhang](#)<sup>†</sup>

### [W2] Can Video Large Language Models Comprehend Language in Videos? 🔗

*NeurIPS Workshop on Video-Language Models*, 2024

[Minjoon Jung](#), [Junbin Xiao](#)<sup>†</sup>, [Byoung-Tak Zhang](#), [Angela Yao](#)

## [C2] PGA: Personalizing Grasping Agents with Single Human-Robot Interaction [↗](#)

International Conference on Intelligent Robots (IROS), 2024 **(Oral)**

Junghyun Kim, Gi-Cheon Kang, Jaein Kim, Seoyun Yang, Minjoon Jung, Byoung-Tak Zhang<sup>†</sup>

## [C1] Modal-specific Pseudo Query Generation for Video Corpus Moment Retrieval [↗](#)

Empirical Methods in Natural Language Processing (EMNLP), 2022

Minjoon Jung, Seongho Choi, Joochan Kim, Jin-Hwa Kim<sup>†</sup>, Byoung-Tak Zhang<sup>†</sup>

## [J1] Stagemix Video Generation using Face and Body Keypoints Detection [↗](#)

Multimedia Tools and Applications (MTAP), 2022

Minjoon Jung, Seunghyun Lee, Eun-Seon Sim, Min-Ho Jo, Yu-Jin Lee, Hye-Bin Choi, Junseok Kwon<sup>†</sup>

## [W1] Toward a Human-Level Video Understanding Intelligence [↗](#)

AAAI Fall Symposium Series, 2021

Yu-Jung Heo, Minsu Lee, Seongho Choi, Woo-Suk Choi, MinJung Shin, Minjoon Jung, Jeh-Kwang Ryu, Byoung-Tak Zhang<sup>†</sup>

## Honors & Awards

---

<b>Honorable Mention</b> , Seoul National University Graduate School of AI Research Paper Contest	Jun 2025
<b>Grand Prize</b> , Seoul National University Graduate School of AI Research Paper Contest	Dec 2024
<b>Full Scholarship</b> , Seoul National University Graduate School of AI	Mar 2021 - Feb 2023
<b>Scholarship</b> , Daewoong Foundation	Mar 2021 - Feb 2022
<b>Full Scholarship</b> , Chung-Ang University School of Software Engineering	Mar 2015 - Feb 2021

## Invited Talks

---

Korea Software Congress (KSC)	Jul 2025
AI Seoul	Feb 2024
Seoul National University Graduate School of AI, School Symposium	Aug 2023
Seoul National University Graduate School of AI, Achievement Conference	Jul 2023
Seoul National University Artificial Intelligence Institute, Retreat	Nov 2022

## Academic Services

---

**Conference Reviewer:** NeurIPS, CVPR, WACV, EMNLP, ACL

**Journal Rewviewer:** MTAP

**Teaching Assistant:**

- [SNU AI773] Multimodal Generative AI Theories and Applications, 2025 [↗](#)

## References

---

**Junseok Kwon** [↗](#)

Associate Professor, CTO HKLABS

Chung-Ang University

**Byoung-Tak Zhang** [↗](#)

Associate Professor, Chairman of AI Institute

Seoul National University

**Jin-Hwa Kim** [↗](#)

Leader of Generation Research and Neural 3D TF

Naver AI Lab

**Angela Yao** [↗](#)

Associate Professor

National University of Singapore

