

Jaywon Koo

✉ jk125@rice.edu ☎ 346-475-0948 🏠 jaywonkoo17.github.io
in jaywonkoo 🐦 JaywonKoo 🎓 Google Scholar

Research Interest

Self-improving AI agents and multimodal reasoning, with a focus on hypothesis-driven discovery, tool use, and reliable evaluation.

Education

- | | |
|---|---------------------|
| Rice University , USA | Aug 2023 - Present |
| <ul style="list-style-type: none">• Ph.D. in Computer Science• Advisor: Vicente Ordóñez Román | |
| Columbia University , USA | Sep 2021 - Dec 2022 |
| <ul style="list-style-type: none">• M.S. in Computer Science (Natural Language Processing Track) | |
| Ewha Womans University , South Korea | Mar 2016 - Aug 2021 |
| <ul style="list-style-type: none">• B.S. in Computer Science and Engineering, <i>Magna Cum Laude</i>• B.S. in Self-Designed Major (Scranton Honors Program-Convergence of Science) | |

Publications

- [C8] “[Agentic Discovery with Active Hypothesis Exploration for Visual Recognition](#)”
Jaywon Koo, Jefferson Hernandez, Ruozhen He, Hanjie Chen, Chen Wei, Vicente Ordonez
The Conference on Empirical Methods in Natural Language Processing (**EMNLP**), Findings, 2026 [Paper] [↗](#)
- [C7] “[Beyond Referring Expressions: Scenario Comprehension Visual Grounding](#)”
Ruozhen He, Nisarg A Shah, Qihua Dong, Zilin Xiao, **Jaywon Koo**, Vicente Ordonez
The Conference on Empirical Methods in Natural Language Processing (**EMNLP**), Main, 2026 [Paper] [↗](#)
- [C6] “[SAGE: Learning to Reason using Softmax Advantage Group Estimation](#)”
Jefferson Hernandez, **Jaywon Koo**, Zilin Xiao, Chen Wei, Vicente Ordonez
Conference on Language Modeling (**COLM**), 2026
- [C5] “[ProxyThinker: Test-Time Guidance through Small Visual Reasoners](#)”
Zilin Xiao, **Jaywon Koo**, Siru Ouyang, Jefferson Hernandez, Yu Meng, Vicente Ordonez
The International Conference on Learning Representations (**ICLR**), 2026 [Paper] [↗](#)
- [C4] “[Evaluating Text-to-Image and Text-to-Video Synthesis with a Conditional Fréchet Distance](#)”
Jaywon Koo*, Jefferson Hernandez*, Moayed Haji-Ali, Ziyang Yang, Vicente Ordonez (*Joint First Authors)
Winter Conference on Applications of Computer Vision (**WACV**), 2026 [Paper] [↗](#)
- [C3] “[PropTest: Automatic Property Testing for Improved Visual Programming](#)”
Jaywon Koo, Ziyang Yang, Paola Cascante-Bonilla, Baishakhi Ray, Vicente Ordonez. The Conference on Empirical Methods in Natural Language Processing (**EMNLP**), Findings, 2024 [Webpage] [↗](#) [Paper] [↗](#)
- [C2] “[Multi-Modality Multi-Loss Fusion Network](#)”
Zehui Wu, Ziwei Gong, **Jaywon Koo**, Julia Hirschberg.
The North American Chapter of the Association for Computational Linguistics (**NAACL**), Oral, 2024 [Paper] [↗](#)
- [C1] “[Beyond Grounding: Extracting Event Hierarchies from Multimodal Content](#)”
Hammad A. Ayyubi, Christopher Thomas, Lovish Chum, Rahul Lokesh, Yulei Niu, Xudong Lin, Long Chen, **Jaywon Koo**, Sounak Ray and Shih-Fu Chang.
Proceedings of the AAAI Conference on Artificial Intelligence (**AAAI**), 2024 [Paper] [↗](#)

Experiences

Music Foundation Model Team, Sony AI

Research Intern

New York, USA

May 2026 - Sep 2026

- Build an audio-language agent that automates hierarchical music mixing through tool planning and learned parameter refinement.
- Developed an LLM-as-a-judge method to evaluate music post-production quality.

Integrated Systems Team, NEC LABs America

Research Intern

Princeton, USA

May 2025 - Aug 2025

- Developed a multi-vector late-interaction retrieval system to optimize cross-modal retrieval and representation learning for dense, multimodal radiology reports.

Vision, Language, and Learning Lab, Rice University

Graduate Research Assistant (Advisor: Prof. Vicente Ordóñez Román)

Houston, USA

Aug 2023 - Present

- Developing a self-evolving, agentic visual monitor that checks robot task progress and detects execution failures.

Speech Lab, Columbia University

Graduate Research Assistant (Advisor: Prof. Julia Hirschberg)

New York, USA

Jan 2022 - Dec 2022

- Developed a multimodal model through a DAPRA funded project (Cross-Cultural Harmony through Affect and Response Mediation) for detecting social norms, emotion, and communication failure
- Proposed a multi-person, in-context multimodal emotion recognition model using hierarchical fusion method

DVMM Lab, Columbia University

Graduate Research Assistant (Advisor: Prof. Shih-Fu Chang)

New York, USA

Sep 2021 - Dec 2021

- Created a new multimodal task, event-event relations detection where the model detects events in text and video and predicts relations among them
- Implemented two ways to utilize commonsense knowledge features in event-event relations detection using ConceptNet and CLIP
- Built a large event-event relationship dataset and test benchmark containing 100k video-news article pairs

Projects

DARPA CCU:Cross-Cultural Harmony through Affect and Response Mediation

Jan 2022 - Dec 2022

- Led the development of multilingual emotion recognition
- **Ranked 1st** at the official evaluations on Text emotion recognition model for Mandarin
- Presented our work on speech and text emotion recognition at DARPA

Awards and Honors

Graduate Student Spring Travel Grant, Rice Engineering Alumni

Oct 2024

Naver Best Paper Award (top-5 papers), IEIE Conference

July 2021

Second Place, Graduation Project Contest, Ewha Womans University

Fall 2020

Scholarship of Academic Excellence, Ewha Womans University

Fall 2020, Spring 2021

Future Competent Scholarship, Ewha Womans University

Summer 2020

Technical Skills

Advanced

Python, C, Java, Pytorch, Tensorflow, Keras, MySQL, Git, OpenCV, Sckit-learn

Moderate

LangChain, LangGraph, C++, Matlab, SPSS, LaTeX, OpenGL

Novice

Swift