Heeseung Yun

heeseung.yun@kaist.ac.kr | heeseung.research@gmail.com | heeseungyun.com
Heeseung | hs-yn | Google Scholar | Full CV available upon request

EDUCATION

Seoul National University — **Seoul, South Korea**
Ph.D. in Computer Science and Engineering — *Mar 2020 - Feb 2026*
Thesis: Building Multisensory Egocentric Intelligence: from Comprehensive to Proactive Perception
- Next Generation Engineering Leader Award (3 recipients among all engineering PhDs in the class of 2026)
- Computer Science and Engineering Best Thesis Award

Seoul National University, College of Liberal Studies — **Seoul, South Korea**
B.S. in Computer Science and Engineering & Value Science and Engineering — *Mar 2015 - Feb 2020*
- College of Liberal Studies Valedictorian (Dean's award)
- Computer Science and Engineering Best Thesis Award

Technical University of Munich — **Munich, Germany**
Department of Informatics (Exchange Student) — *Oct 2018 - Mar 2019*

RESEARCH INTEREST

My research builds egocentric and situated omnimodal intelligence for human-centric agents, with applications in extended reality, immersive telepresence, and robotic embodiment. My active research spans (i) egocentric multisensory perception across audio, vision, and behavior (*e.g.*, gaze and pose) [ECCV24, ICLR25, NeurIPS25, IJCV26], (ii) panoramic and 3D scene understanding [ICCV21, ECCV22, ICCV23], (iii) situated agents with embodiment [MobileHCI19, EMNLP25], and (iv) efficient multimodal adaptation underlying these systems [CVPR21, CVPR23, CVPR25].

EXPERIENCE

InnoCORE LLM Research Center, KAIST — **Seoul, South Korea**
Postdoctoral Research Fellow — *Mar 2026 - Present*
- Selected as a postdoctoral fellow of InnoCORE, South Korea's most competitive postdoctoral program in STEM
- Building Large Language Model (LLM) and Large Multimodal Model (LMM) agents with multisensory grounding, interacting in full-duplex with multiple humans and agents in a shared 3D environment

Vision and Learning Lab, Seoul National University — **Seoul, South Korea**
Graduate Research Assistant (Advisor: Gunhee Kim) — *Mar 2020 - Feb 2026*
Research Intern — *Jun 2019 - Feb 2020*
- Seven first-author projects on multimodal representation learning (audio-visual, egocentric, 3D) at top-tier venues
- Led the LLM subgroup in accessibility-focused AI docent consortium (2024-2025): managed five researchers, deployed RAG-based docent systems in three major museums in Seoul (Featured in Korea Metaverse Festival)

Reality Labs Research, Meta — **Redmond, WA**
External Research Collaborator (Remote) — *Oct 2023 - Jan 2024*
Research Scientist Intern (On-site / Host: Calvin Murdock, Vamsi Krishna Ithapu) — *May 2023 - Aug 2023*
- Developed and shipped Spherical World-Locking for audio-visual and behavioral learning, published at ECCV 2024
- Built pipelines over egocentric audio, video, gaze, and IMU; contributed to in-house data curation (Project Aria)

Intel-NTU Connected Context Computing Center — **Taipei, Taiwan**
Research Intern (Advisor: Chih-Yuan Yang) — *Jul 2018 - Aug 2018*
- Built Docker-based server-client control and interface systems for social robots deployed in the wild

PUBLICATIONS

Full publication list (15+) on Google Scholar (* equal contribution)

[15] WoW-Bench: Evaluating Fine-Grained Acoustic Perception in Audio-Language Models via Marine Mammal Vocalization
Jaeyeon Kim, **Heeseung Yun**, Sang Hoon Woo, Chao-Han Huck Yang, Gunhee Kim
Findings of the Association for Computational Linguistics (**ACL Findings 2026**)
- First benchmark for low-level acoustic perception in LALMs: below 50% vs. 97% for humans (*In collaboration with NVIDIA*)

[14] Towards Scene-Aware Video-to-Spatial Audio Generation
Jaeyeon Kim*, **Heeseung Yun***, Gunhee Kim
International Journal of Computer Vision (**IJCV 2026**)
- Extending ViSAGe with 4× more efficient framework & 3× larger dataset for video to ambisonics generation

[13] Gaze Beyond the Frame: Forecasting Egocentric 3D Visual Span
Heeseung Yun, Joonil Na, Jaeyeon Kim, Calvin Murdock, Gunhee Kim
Advances in Neural Information Processing Systems (**NeurIPS 2025 Spotlight**)
- Introducing SLAM-based visual attention lifting to anticipate where we look (before we leap) in the 3D world

[12] FlashAdventure: A Benchmark for GUI Agents Solving Full Story Arcs in Diverse Adventure Games
Jaewoo Ahn*, Junseo Kim*, **Heeseung Yun**, Jaehyeon Son, Dongmin Park, Jaewoong Cho, Gunhee Kim
Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2025**)
- Assessing lateral thinking and long-term memory of GUI agents with adventure games (*In collaboration with Krafton*)

[11] Can LLMs Deceive CLIP? Benchmarking Adversarial Compositionality of Pre-trained Multimodal Representation via Text Updates
Jaewoo Ahn*, **Heeseung Yun***, Dayoon Ko, Gunhee Kim
Proceedings of the Association for Computational Linguistics (**ACL 2025**)
- Exposing compositional vulnerabilities in multimodal models *(e.g., image, video, audio)* via LLMs with diversity-promoting self-training

[10] ReSpec: Relevance and Specificity Grounded Online Filter for Learning on Video-Text Data Streams
Chris Dongjoo Kim*, Jihwan Moon*, Sangwoo Moon, **Heeseung Yun**, Sihaeng Lee, Aniruddha Kembhavi, Soonyoung Lee, Gunhee Kim, Sangho Lee, Christopher Clark
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2025**)
- Training state-of-the-art video-language model with as little as *5% of pretraining data* (*In collaboration with AI2 & LG AI Research*)

[9] ViSAGe: Video to Spatial Audio Generation
Jaeyeon Kim, **Heeseung Yun**, Gunhee Kim
International Conference on Learning Representations (**ICLR 2025**)
- Proposing the first end-to-end framework & benchmark for generating immersive ambisonics from silent video

[8] Spherical World-Locking for Audio-Visual Localization in Egocentric Videos
Heeseung Yun, Ruohan Gao, Ishwarya Ananthabhotla, Anurag Kumar, Jacob Donley, Chao Li, Gunhee Kim, Vamsi Krishna Ithapu, Calvin Murdock
European Conference on Computer Vision (**ECCV 2024**)
- Leveraging egocentric self-motion to better stabilize & localize *any* user-centric multisensory signals (*Work done during internship at Meta*)

[7] Dense 2D-3D Indoor Prediction with Sound via Aligned Cross-Modal Distillation
Heeseung Yun*, Joonil Na*, Gunhee Kim
Proceedings of the IEEE/CVF International Conference on Computer Vision (**ICCV 2023**)
- Predicting 2D depth/semantics and 3D indoor structure from binaural audio alone via spatially-aligned distillation

[6] Fusing Pre-trained Language Models with Multimodal Prompts through Reinforcement Learning
Youngjae Yu*, Jiwan Chung*, **Heeseung Yun**, Jack Hessel, Jae sung Park, Ximing Lu, Rowan Zellers, Prithviraj Ammanabrolu, Ronan Le Bras, Gunhee Kim, Yejin Choi
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2023**)
- Aligning text-only models to multimodal prompts *without* paired image/audio supervision (*In collaboration with AI2*)

[5] Panoramic Vision Transformer for Saliency Detection in 360° Videos
Heeseung Yun, Sehun Lee, Gunhee Kim
European Conference on Computer Vision (**ECCV 2022**)
- Enabling ViT variants to process omnidirectional imagery with minimal distortion via a *single step* geometric approximation

[4] Pano-AVQA: Grounded Audio-Visual Question Answering on 360° Videos
Heeseung Yun, Youngjae Yu, Wonsuk Yang, Kangil Lee, Gunhee Kim
Proceedings of the IEEE/CVF International Conference on Computer Vision (**ICCV 2021**)
- Establishing the first large-scale benchmark & framework for audio-visual QA with panoramic videos (*In collaboration with Hyundai*)

[3] Transitional Adaptation of Pretrained Models for Visual Storytelling
Youngjae Yu*, Jiwan Chung*, **Heeseung Yun**, Jongseok Kim, Gunhee Kim
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2021**)
- Proposing a *transitional* stage to better align pretrained vision encoders and language models prior to fine-tuning

[2] 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 Spotlight**)
- Unifying movie character identity matching across video, text, and story context into a single end-to-end loop

[1] A Mobile Robot Generating Video Summaries of Seniors' Indoor Activities
Chih-Yuan Yang, **Heeseung Yun**, Srenivas Varadarajan, Jane Yung-jen Hsu
ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (**MobileHCI 2019**)
- Developing a social robot framework that actively tracks seniors to synthesize long-term indoor footage into glanceable summaries

Patents

[T1] Method and System of Providing Interface for Visual Question Answering
Kang Il Lee, **Heeseung Yun**, Youngjae Yu, Gunhee Kim, Wonsuk Yang (US12373483B2)

HONORS & AWARDS

CVPR 2026 Doctoral Consortium *CVPR, 2026*
Next Generation Engineering Leader Award (3 PhDs in class of 2026) *Seoul National University, 2026*
Best Thesis Award *SNU Dept. of Computer Science and Engineering, 2026*
Qualcomm Innovation Fellowship 2023 *Qualcomm, 2023*
Yulchon AI Star Scholarship 2023 *Yulchon, 2023*
NAVER Ph.D. Fellowship Award 2022 *NAVER, 2022*
1st Place of DramaQA Challenge 2020 *VTT Workshop @ ECCV, 2020*
1st Place of Large Scale Movie Description Challenge 2019 *CLVL Workshop @ ICCV, 2019*
National Scholarship for Science and Engineering *Ministry of Science and ICT, 2017–2019*

TECHNICAL SKILLS

Deep Learning: PyTorch + Slurm + Wandb
Modalities: Language, Image, Video (Normal, Egocentric, 360°), Audio (Mono, Binaural, Ambisonics, Array, Speech), Behavior (IMU, Gaze), 3D (SLAM, Mesh, Voxel, Simulation)

INVITED TALKS

How to Build a Docent LLM from Scratch *Peopulley, 2025*
Dense Indoor Prediction with Sound via Aligned Cross-Modal Distillation *Samsung AI Forum, 2023*
Panoramic Vision Transformers for 360° Video Understanding *CJ AI Center, 2023*
Panoramic Vision Transformer for Saliency Detection in 360° Videos *Korea AI Summit, 2022*
Multimodal Grounding for Panoramic Video Understanding *SNU Graduate Seminar, 2022*
Grounded Audio-Visual Question Answering on 360° Videos *Hyundai AI Research Lab, 2022*

ACADEMIC SERVICES

Reviewer

- **Honors:** Outstanding Reviewer (CVPR 2025); contributed 50+ reviews for major venues (2022–Present).
- **Journals:** TPAMI, IJCV, TVCG, TMLR
- **Conferences:** CVPR, NeurIPS, ICCV, ICLR, ECCV, ACM MM, ACCV, WACV, BMVC
- **Program Committee:** ACM MM 2026, WorldModelBench Workshop @ CVPR 2025
- **Workshop Organizer:** Bridging Gaps for Underrepresented Researchers in ML @ ICML 2026

Mentoring

Jaeyeon Kim (Undergrad @ SNU): 2024-2025, next - PhD student @ CMU
Cheyon Jin (Undergrad @ SNU): 2024-2025, best thesis award with AI docent project

Teaching Assistant

Computer Vision (M1522.001000) *Fall 2022*
Probabilistic Graphical Models (M1522.001300) *Spring 2022*
Project Mentor for 10+ DL Projects on Semiconductor Manufacturing in SK Hynix *2021–2022*
Dept. of Engineering ML Engineer Course (Computer Vision) *2021–2022*
Samsung DS^2 Course (Reinforcement Learning) *2020–2021*
Theory and Lab of IoT, AI and Big Data (M2177.004900) *Spring 2019 / Fall 2021*