

Geewook Kim (김기욱)

Team Manager (Multimodal Alignment, Vision Understanding, Drone AI) and Research Scientist at NAVER Cloud

✉ gwkim.rsrch@gmail.com 🏠 <https://geewook.kim> 🌐 <https://linkedin.com/in/gwkim-rsrch>

Research Interests

My research builds **resource-efficient multimodal AI**: models that perceive across text, image, audio, and video under real-world compute and data limits. I study what shapes their capabilities and limits, and turn these findings into reliable systems and the perceptual groundwork for agentic and embodied AI.

Research directions:

- **Efficient multimodal and omni-modal architectures.** Models that perceive across modalities under tight compute, including OCR-free document understanding [C7, ECCV22], state-space models for hour-long video [C17, AAAI26], and cost-efficient document parsing [C11, ICDAR24].
- **Evaluation and benchmarking across Language, Vision, Audio, and Agents.** Benchmarks, model-as-judge methods, and audits of what multimodal models truly understand, including a Korean educational benchmark for multimodal models [C15, NAACL25], vision-language judges [C12, ACL-F24], an audio-visual benchmark audit for Video-LLMs [C20, Interspeech26], a Korean web-browsing agent benchmark [C21, EMNLP-F26], and a drone-control benchmark for vision-language-action agents [O9, arXiv26].
- **Foundation model training methods and analyses.** Learning methods and analyses of how training reshapes model behavior, including decentralized instruction tuning via weight merging [C19, ICML26], the safety cost of vision-language adaptation [C14, ICLR25], the text capability it erodes and how to predict that loss [C24, EMNLP26], and what drives efficient vision-language assistants [C13, EMNLP24].

My work bridges fundamental research and real-world systems, from open-source tools like [Donut](#) (6.9K ★), which is widely adopted for OCR-free document AIs, to the production multimodal models behind NAVER's [HyperCLOVA X](#)^{[Vision][Video][Omni]}, turning research insights into models that are efficient and broadly useful.

Education

Korea Advanced Institute of Science & Technology (KAIST)

Sep. 2023–Aug. 2026

Ph.D. in Artificial Intelligence, [Language & Knowledge Lab.](#) ^[Link]

Kim Jaechul Graduate School of Artificial Intelligence

(Supervisor: [Prof. Minjoon Seo](#))

Dissertation: *Resource-Efficient Architecture and Comprehensive Evaluation of Multimodal Large Language Models*

Kyoto University

Apr. 2018–Mar. 2020

Master of Informatics, [Statistical Intelligence Lab.](#) ^[Link]

Graduate School of Informatics

(Supervisor: [Prof. Hidetoshi Shimodaira](#))

Kyoto University

Apr. 2014–Mar. 2018

Bachelor of Engineering, [Applied Mathematics and Physics Course](#) ^[Link]

School of Informatics and Mathematical Science

(Supervisor: Prof. Hidetoshi Shimodaira)

Professional Appointments

Team Manager (Leader) / Applied Research Scientist, **NAVER Cloud Corp.**, Korea May 2023–Present

- Lead applied research and engineering for NAVER's LLM-based multimodal models (HyperCLOVA X Vision, Video & Omni, <https://clova.ai/hyperclova>), driving omni-modal LLM post-training. Appointed Leader in 2026, heading the Multimodal Alignment, Vision Understanding, and Drone AI teams and mentoring up to 14 researchers and engineers. (See *Research Leadership & Industrial Impact* and *Student Mentoring*.)

Applied Research Scientist, **NAVER Corp.**, Korea Apr. 2020–Apr. 2023

- Conducted research and developed systems for NAVER's Document AI product family, including OCR [\[Link\]](#) and Information Extraction [\[Link\]](#). Initiated and led several research and open-source projects, including [Donut](#) and [Webvicob](#).

Research part-timer/trainee, Mathematical Statistics Team, **RIKEN Center for Advanced Intelligence Project**, Japan Sep. 2017–Feb. 2020
Supervisor: Prof. Hidetoshi Shimodaira

Research Intern, **CLOVA OCR, NAVER Corp.** [\[Link\]](#), Korea Aug. 2018–Oct. 2018, Aug. 2019–Sep. 2019
Supervisor: Dr. Hwalsuk Lee

ML Engineering Intern, **Recruit Holdings Co., Ltd.**, Tokyo, Japan 2017 (2 Months)

Software Engineering Intern, **Abeja, Inc.**, Tokyo, Japan 2016 (2 Months)

Planning Intern, **SoftBank Corp.**, Tokyo, Japan 2015 (1 Month)

Publications

Google Scholar: <https://scholar.google.com/citations?user=1a2QbgEAAAAJ>

26 peer-reviewed international conference papers, 4 workshop papers, 4 technical reports, and 8 patents. Among the conference papers, 14 are first/co-first author. Total citations 2,500+ (Google Scholar, September 2026); open-source projects total 10K+ GitHub stars.

* equal contribution; † corresponding author.

Selected Publications

- **Geewook Kim**[†] et al. (first and corresponding author), “*OCR-free Document Understanding Transformer (Donut)*”, **ECCV 2022**. 1,200+ citations, 6.9K+ GitHub stars.
Introduced an OCR-free paradigm for visual document understanding.
- **Geewook Kim** (first author) and Minjoon Seo, “*On Efficient Language and Vision Assistants for Visually-Situated Natural Language Understanding: What Matters in Reading and Reasoning*”, **EMNLP 2024**.
Studies the design factors behind efficient language-and-vision assistants for visually-situated NLU.
- Seongyun Lee*, **Geewook Kim*** (co-first author), Jiyeon Kim* et al., “*How Does Vision-Language Adaptation Impact the Safety of Vision Language Models?*”, **ICLR 2025**.
Examines how vision-language adaptation affects model safety and reliability.

- **Geewook Kim** (first author) and Minjoon Seo, “*State-Space Hierarchical Compression with Gated Attention and Learnable Sampling for Hour-Long Video Understanding in Large Multimodal Models*”, **AAAI 2026. Oral (4.6%)**.
State-space hierarchical compression for hour-long video understanding.
- Minsik Choi* and **Geewook Kim***[†] (co-first and corresponding author), “*Decentralized Instruction Tuning: Conflict-Aware Splitting and Weight Merging*”, **ICML 2026**.
Conflict-aware decentralized instruction tuning with weight merging.

Peer-Reviewed International Conferences (26 papers)

- | | | |
|-------|---|---------------------|
| [C26] | Hodong Lee, Sanghee Park, Dohoon Ryu, Jungwhan Kim, Junyeob Kim, Soyoon Kim, and Geewook Kim [†] (corresponding author), “ <i>A Composable Evaluation System for Reproducible Omni-Modal Foundation Model Evaluation</i> ” ^[Code] , Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2026 (to appear). | EMNLP Demo 2026 |
| [C25] | <u>Sukmin Seo</u> * and Geewook Kim * [†] (co-first and corresponding author), “ <i>Natural-Language Temporal Grounding in Hour-Long Videos is a Search Problem: A Benchmark and Empirical Decomposition</i> ”, Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, 2026 (to appear). | EMNLP 2026 |
| [C24] | <u>Minsik Choi</u> * and Geewook Kim * (co-first author), and Young Geun Kim, “ <i>Text Capability Loss in Vision-Language Adaptation: An Attention-Sink Diagnosis</i> ” ^[Code] , Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, 2026 (to appear). | EMNLP 2026 |
| [C23] | <u>Sanghee Park</u> *, Geewook Kim * [†] (co-first and co-corresponding author), and Kee-Eung Kim [†] , “ <i>KCSAT-ML: Probing Reasoning Models with Nationwide-Cohort Human Difficulty</i> ”, Findings of the Association for Computational Linguistics, 2026 (to appear). | EMNLP Findings 2026 |
| [C22] | <u>Juan Yeo</u> * and Geewook Kim * [†] (co-first and corresponding author), “ <i>Compliance, Capability, and Conflict: Benchmarking Multimodal LLMs under System Messages</i> ”, Findings of the Association for Computational Linguistics, 2026 (to appear). | EMNLP Findings 2026 |
| [C21] | Nahyun Lee, Dongkeun Yoon, Guijin Son, Geewook Kim , Dayoon Ko, Jeonghun Park, Haneul Yoo, Jaewon Cho, Junghun Park, Changyoon Lee, Kyochul Jang, Jaeyeon Kim, Eunsu Kim, Woojin Cho, and Seungone Kim, “ <i>K-BrowseComp: A Web Browsing Agent Benchmark Grounded in Korean Contexts</i> ”, Findings of the Association for Computational Linguistics, 2026 (to appear). | EMNLP Findings 2026 |
| [C20] | Geewook Kim and Minjoon Seo, “ <i>Do Modern Video-LLMs Need to Listen? A Benchmark Audit and Scalable Remedy</i> ”, Proceedings of the 27th Annual Conference of the International Speech Communication Association, 2026 (to appear). | Interspeech 2026 |

- [C19] Minsik Choi* and Geewook Kim*[†] (co-first and corresponding author), “*Decentralized Instruction Tuning: Conflict-Aware Splitting and Weight Merging*”, Proceedings of the Forty-Third International Conference on Machine Learning, 2026. ICML 2026
- [C18] Hyunji Lee, Seunghyun Yoon, Yunjae Won, Hanseok Oh, **Geewook Kim**, Trung Bui, Franck Dernoncourt, Elias Stengel-Eskin, Mohit Bansal, and Minjoon Seo, “*Instruction Tuning with and without Context: Behavioral Shifts and Downstream Impact*”, Proceedings of the 2026 Conference of the European Chapter of the Association for Computational Linguistics, 2026. EACL 2026
- [C17] **Geewook Kim** and Minjoon Seo, “*State-Space Hierarchical Compression with Gated Attention and Learnable Sampling for Hour-Long Video Understanding in Large Multimodal Models*”, Proceedings of the Fortieth AAAI Conference on Artificial Intelligence, 2026. AAI 2026
Oral Presentation (1089/23678=4.60%)
- [C16] Gio Paik, **Geewook Kim**, and Jinbae Im, “*MMRefine: Unveiling the Obstacles to Robust Refinement in Multimodal Large Language Models*”, Findings of the Association for Computational Linguistics, 2025. ACL Findings 2025
- [C15] Sanghee Park*, Geewook Kim*[†] (co-first and corresponding author), “*Evaluating Multimodal Generative AI with Korean Educational Standards*”, Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, 2025. NAACL-HLT 2025
- [C14] Seongyun Lee*, Geewook Kim* (co-first author), Jiyeon Kim* (co-first author), Hyunji Lee, Hyeon Chang, Sue Hyun Park, and Minjoon Seo, “*How Does Vision-Language Adaptation Impact the Safety of Vision Language Models?*”, Proceedings of the Thirteenth International Conference on Learning Representations, 2025. ICLR 2025
- [C13] **Geewook Kim** and Minjoon Seo, “*On Efficient Language and Vision Assistants for Visually-Situated Natural Language Understanding: What Matters in Reading and Reasoning*”, Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 2024. EMNLP 2024
- [C12] Seongyun Lee, Seungone Kim, Sue Hyun Park, **Geewook Kim**, and Minjoon Seo, “*Prometheus-Vision: Vision-Language Model as a Judge for Fine-Grained Evaluation*”, Findings of the Association for Computational Linguistics, 2024. ACL Findings 2024
- [C11] Yamato Okamoto, Youngmin Baek, **Geewook Kim**, Ryota Nakao, DongHyun Kim, Moon Bin Yim, Seunghyun Park, and Bado Lee, “*CREPE: Coordinate-Aware Cost-Efficient Document Parsing End-to-End Model*”, Proceedings of the International Conference on Document Analysis and Recognition, 2024. ICDAR 2024

- [C10] **Geewook Kim**[†], Hodong Lee, Daehee Kim, Haeji Jung, Sanghee Park, Yoonsik Kim, Sangdoo Yun, Taeho Kil, Bado Lee, and Seunghyun Park, “*Cream: Visually-Situated Natural Language Understanding with Contrastive Reading Model and Frozen Large Language Models*”, Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023. EMNLP 2023
- [C9] Daehee Kim, Yoonsik Kim, DongHyun Kim, Yumin Lim, **Geewook Kim**, and Taeho Kil, “*SCOB: Universal Text Understanding via Character-wise Supervised Contrastive Learning with Online Text Rendering for Bridging Domain Gap*”, IEEE/CVF International Conference on Computer Vision, 2023. ICCV 2023
- [C8] Donghyun Kim, Teakgyu Hong, Moonbin Yim, Yoonsik Kim, and **Geewook Kim**[†] (corresponding author), “*On Web-based Visual Corpus Construction for Visual Document Understanding*”, International Conference on Document Analysis and Recognition, 2023. ICDAR 2023
- [C7] **Geewook Kim**[†], Teakgyu Hong, Moonbin Yim, Jeongyeon Nam, Jinyoung Park, Jinyeong Yim, Wonseok Hwang, Sangdoo Yun, Dongyoon Han, and Seunghyun Park, “*OCR-free Document Understanding Transformer*”, European Conference of Computer Vision, 2022. ECCV 2022
1,200+ citations, 6.9K+ stars at GitHub
- [C6] Wonseok Hwang, Hyunji Lee, Jinyeong Yim, **Geewook Kim**, and Minjoon Seo, “*Cost-effective End-to-end Information Extraction for Semi-structured Document Images*”, Empirical Methods in Natural Language Processing, 2021. EMNLP 2021
- [C5] Sungrae Park, **Geewook Kim**, Junyeop Lee, Junbum Cha, Ji-Hoon Kim, and Hwalsuk Lee, “*Scale down Transformer by Grouping Features for a Lightweight Character-level Language Model*”, International Conference on Computational Linguistics, 2020. COLING 2020
- [C4] Jeonghun Baek, **Geewook Kim**, Junyeop Lee, Sungrae Park, Dongyoon Han, Sangdoo Yun, Seong Joon Oh, Hwalsuk Lee, “*What is wrong with scene text recognition model comparisons? dataset and model analysis*”, International Conference on Computer Vision, 2019. ICCV 2019
900+ citations, 3.9K+ stars at GitHub
Oral Presentation (187/4304=4.3%)
- [C3] **Geewook Kim**, Akifumi Okuno, Kazuki Fukui, and Hidetoshi Shimodaira, “*Representation Learning with Weighted Inner Product for Universal Approximation of General Similarities*”, International Joint Conference on Artificial Intelligence, 2019. (Presented as both **oral** and **poster**). IJCAI 2019
- [C2] **Geewook Kim**, Kazuki Fukui, Hidetoshi Shimodaira, “*Segmentation-free Compositional n-gram Embedding*”, North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019. NAACL-HLT 2019

- [C1] Akifumi Okuno, **Geewook Kim**, and Hidetoshi Shimodaira, “*Graph Embedding with Shifted Inner Product Similarity and Its Improved Approximation Capability*”, International Conference on Artificial Intelligence and Statistics, 2019. AISTATS 2019

Peer-Reviewed Workshop Papers

- [W4] **Geewook Kim**^{*}, Shuhei Yokoo^{*} (co-first author), Sukmin Seo, Atsuki Osanai, Yamato Okamoto, and Youngmin Baek, “*On Text Localization in End-to-End OCR-Free Document Understanding Transformer Without Text Localization Supervision*”, International Conference on Document Analysis and Recognition Workshops, 2023. ICDAR-W 2023
- [W3] **Geewook Kim**, Wonseok Hwang, Minjoon Seo, and Seunghyun Park, “*Semi-Structured Query Grounding for Document-Oriented Databases with Deep Retrieval and Its Application to Receipt and POI Matching*”, AAAI Workshop on Knowledge Discovery from Unstructured Data in Financial Services, 2022. AAAI-W 2022
- [W2] Masahiro Naito, Sho Yokoi, **Geewook Kim**, Hidetoshi Shimodaira, “*Revisiting Additive Compositionality: AND, OR and NOT Operations with Word Embeddings*”, ACL-IJCNLP 2021 Student Research Workshop, 2021. ACL-SRW 2021
- [W1] **Geewook Kim**, Kazuki Fukui, and Hidetoshi Shimodaira, “*Word-like Character n-gram Embedding*”, EMNLP Workshop on Noisy User-generated Text, 2018. EMNLP-W 2018

Tech Reports

- [T4] NAVER Cloud HyperCLOVA X Team, “*HyperCLOVA X 32B Think Technical Report*”, 2026. Tech Report 2026
- [T3] NAVER Cloud HyperCLOVA X Team, “*HyperCLOVA X 8B Omni Technical Report*”, 2026. Tech Report 2026
- [T2] HyperCLOVA X Team, “*HyperCLOVA X THINK Technical Report*”, 2025. Tech Report 2025
- [T1] HyperCLOVA X Team, “*HyperCLOVA X Technical Report*”, 2024. Tech Report 2024

Preprints, Tech Blogs, and Domestic Publications

- [O9] Jaewoo Park^{*}, Minyoung Lee^{*}, Sukmin Seo^{*} (co-first authors), Moonbin Yim, Hyunwook Yoon, Dohoon Ryu, Daehee Kim, Myungseo Song, Jihyuk Byun, Seunggyu Chang, Taeho Kil, Jiseob Kim, Bado Lee, and **Geewook Kim**[†] (corresponding author), “*Evaluating Multimodal LLMs as Generalist Vision-Language-Action Agents for Drone Control: Commanding, Approaching, Tracking and Searching*”, arXiv preprint, 2026. Preprint 2026

- | | | |
|------|---|----------------|
| [O8] | NAVER Cloud (CLOVA AI Tech Blog), “ Split the conflict, merge the gains: MERIT, an efficient post-training strategy for LLMs ”, CLOVA AI Tech Blog, 2026. | Tech Blog 2026 |
| [O7] | NAVER Cloud (CLOVA AI Tech Blog), “ HyperCLOVA X OMNI: Toward an Omni-Model (국가대표 AI, 옴니모델을 향한 여정) ”, CLOVA AI Tech Blog, 2025. | Tech Blog 2025 |
| [O6] | NAVER, “ If AI Took the College Entrance Exam (AI가 수능을 본다면): A Six-Month Thought Experiment ”, 2025. | Media 2025 |
| [O5] | NAVER Cloud (CLOVA AI Tech Blog), “ HyperCLOVA X Video: Seeing through motion ”, CLOVA AI Tech Blog, 2025. | Tech Blog 2025 |
| [O4] | NAVER Cloud (CLOVA AI Tech Blog), “ Introducing HyperCLOVA X Vision ”, CLOVA AI Tech Blog, 2024. | Tech Blog 2024 |
| [O3] | Morihiro Mizutani, Akifumi Okuno, Geewook Kim , Hidetoshi Shimodaira, “ Stochastic Neighbor Embedding of Multimodal Relational Data for Image-Text Simultaneous Visualization ”, arXiv preprint, 2020. | Preprint 2020 |
| [O2] | Geewook Kim , Sho Yokoi, and Hidetoshi Shimodaira, “ 単語埋め込みの二種類の加法構成性 (Two Types of Additive Compositionality in Word Embedding) ”, Annual Meeting of the Association for Natural Language Processing (in Japanese), 2020. | ANLP 2020 |
| [O1] | Geewook Kim , Akifumi Okuno, and Hidetoshi Shimodaira, “ 擬ユークリッド空間への単語埋め込み (Embedding Words into Pseudo-Euclidean Space) ”, Annual Meeting of the Association for Natural Language Processing (in Japanese), 2019. (Selected to receive both awards: Young Researcher Award ^[Link] and Best Poster Award). | ANLP 2019 |

Invited Talks & Seminars

- | | |
|---|-----------|
| HyperCLOVA X Vision: Open Your Eyes, CLOVA X!
TEAM NAVER CONFERENCE DAN 24 ^[Session Link] ^[Slide] | Nov. 2024 |
| HyperCLOVA X Vision: Open Your Eyes, CLOVA X!
NAVER ENGINEERING DAY | Oct. 2024 |
| Vision-Language Models for Context-Rich Image Understanding Tasks
University of Seoul ^[Slide] ^[Session Link] | Apr. 2024 |
| Fine-Grained Evaluation of Vision-Language Models through VLM as a Judge
NAVER Tech Meetup ^[Slide] | Feb. 2024 |
| Recent Advances in Document AI
Korea University | Mar. 2023 |
| Recent Advances in Document AI
Kookmin University | Dec. 2022 |

OCR-Free Document Understanding Transformer Microsoft [Slide]	Nov. 2022
Identifying a store from a receipt image DEVVIEW Conference [Video] [Slide] [Session Link]	2021
Representation Learning with Weighted Inner Product (WIP) Michinoku Communication Science Seminar, Tohoku University [Session Link]	2019

Academic Service

Journal Reviewer:

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (since 2025),
Pattern Recognition (PR) (since 2025), *International Journal of Computer Vision (IJCV)* (since 2026),
ACM Computing Surveys (CSUR) (since 2025), *Expert Systems with Applications (ESWA)* (since 2025),
IEEE Access (since 2023), *Journal of Medical Internet Research (JMIR)* (since 2026)

Conference Reviewer:

ICML (since 2026), *NeurIPS* (since 2026), *AAAI* (since 2025),
ICCV (since 2025), *CVPR* (since 2024), *BMVC* (since 2026),
ACL ARR (ACL, EMNLP, NAACL, COLING) (since 2024),
 Industry Track of *ACL, EMNLP, NAACL, EACL, COLING* (since 2022)

Selected Honors and Awards

Top Reviewer (Gold), ICML 2026 (Top 25% of reviewers)	2026
Seiwa International Students Scholarship [Link]	2019
Korea-Japan Joint Government Scholarship [Link]	2013–2018
Admission and tuition fees, and living costs covered for a year of preliminary education and four years of Bachelor's studies	
Young Researcher Award at Annual Meeting of the Association for Natural Language Processing [Link]	2019

Patents

[P8] Geewook Kim , Hodong Lee, Daehee Kim, Sanghee Park, Yoonsik Kim, Sangdoo Yun, Bado Lee, and Seunghyun Park, " METHOD AND SYSTEM FOR ENHANCING IMAGE UNDERSTANDING ", KR Patent KR1020230117309A, filed in September 2023.	Sep. 2023
[P7] Geewook Kim and Youngmin Baek, " METHOD, COMPUTER DEVICE, AND COMPUTER PROGRAM FOR TEXT LOCALIZATION IN END-TO-END DOCUMENT UNDERSTANDING MODEL ", KR Patent KR1020230075133A, filed in June 2023.	Jun. 2023
[P6] Daehee Kim, Yoonsik Kim, DongHyun Kim, Yumin Lim, Geewook Kim , and Taeho Kil, " METHOD AND SYSTEM FOR PERFORMING OCR USING CHARACTER-WISE SUPERVISED CONTRASTIVE LEARNING MODEL ", KR Patent KR1020230037945A, filed in March 2023.	Mar. 2023

- [P5] **Geewook Kim**, Teakgyu Hong, Moonbin Yim, and Seunghyun Park, “[METHOD AND SYSTEM FOR EXTRACTING STRUCTURED INFORMATION FROM SEMI-STRUCTURED DOCUMENTS THROUGH DEEP-LEARNING](#)”, KR Patent KR102784722B1, filed in January 2022 and granted in March 2025. Jan. 2022
- [P4] **Geewook Kim**, Wonseok Hwang, and Minjoon Seo, “[METHOD AND SYSTEM FOR DATA SEARCHING](#)”, KR Patent KR102684423B1, filed in August 2021 and granted in July 2024 (also granted as JP Patent JP7367139B2, filed in July 2022 and granted in October 2023). Aug. 2021
- [P3] **Geewook Kim**, Seung Shin, Youngmin Baek, Hyosun Wang, Jungun Kim, and Seungbeom Choi, “[CHARACTER RECOGNITION METHOD AND SYSTEM ROBUST TO ERRORS OF CHARACTER RECOGNITION THAT RECOGNIZE INFORMATION INCLUDED IN TABLES](#)”, KR Patent KR102697516B1, filed in July 2021 and granted in August 2024 (also granted as JP Patent JP7398526B2, filed in July 2022 and granted in December 2023). Jul. 2021
- [P2] Wonseok Hwang, Jinyeong Yim, **Geewook Kim**, Minjoon Seo, and Hyunji Lee, “[METHOD AND SYSTEM FOR EXTRACTING INFORMATION FROM SEMI-STRUCTURED DOCUMENTS](#)”, KR Patent KR102649429B1, filed in May 2021 and granted in March 2024. May 2021
- [P1] Seonghyeon Kim, **Geewook Kim**, Jaeheung Surh, Daehyun Nam, Seungbeom Choi, Seung Shin, Youngmin Baek, and Hyosun Wang, “[METHOD AND SYSTEM FOR RECOGNIZING TABLES](#)”, KR Patent KR102699224B1, filed in March 2021 and granted in August 2024. Mar. 2021

Research Leadership & Industrial Impact

At NAVER Cloud, I lead applied research and engineering for HyperCLOVA X's multimodal models (Vision, Video, and Omni), guiding projects from research direction through publication and deployment.

NAVER's AI foundation-model leadership (HyperCLOVA X). Technical lead for the HyperCLOVA X SEED Omni model line, NAVER Cloud's omni-modal foundation models that read, see, and listen across text, image, audio, and video, including the open-weight [SEED-Omni-8B](#) and [SEED-Think-32B](#). Developed amid Korea's national push for sovereign AI foundation models (독자 AI 파운데이션 모델) ^[news], this model line targets strong multimodal capability at low compute, the efficiency principle at the core of my research.

Open-source and open models. Initiated and led [Donut](#), a widely adopted OCR-free document understanding framework (6.9K+ GitHub stars), and released open models and tools alongside peer-reviewed papers.

Research to production. Translated research outcomes into production multimodal systems in NAVER's products, documented in technical reports, open models, and engineering blogs for HyperCLOVA X Vision, Video, and Omni.

Teaching

Instructor of Record, AI Engineering in Production (산업AI공학)

Spring 2025

University of Seoul, *Graduate course*, 3 credits, 28 students

Designed the full course (syllabus, lectures, assignments, and final project) covering large language models, multimodal models, retrieval-augmented generation, evaluation, deployment, and responsible AI, and supervised student projects. Taught in Korean.

Course Website: <https://geewook.kim/lecture/uos25spring-91035>

Student Mentoring

I have mentored nine interns and graduate-student residents at NAVER Cloud, guiding their research from problem formulation through publications, technical reports, and model releases. Mentees who have co-authored peer-reviewed work include [Haeji Jung](#) (Cream, EMNLP 2023), [Minsik Choi](#) (MERIT, ICML 2026; EMNLP 2026), [Juan Yeo](#) (EMNLP Findings 2026), [Jinyoung Park](#) (Donut, ECCV 2022), and [Hyunwoo Kim](#) (HyperCLOVA X THINK Technical Report). Through NAVER Cloud's industry-academia residency program (산학협력 레지던시 프로그램) ^[news], I also supervised three Seoul National University students, [Inju Ha](#), [Byung Hyun Lee](#), and [Juan Yeo](#), who contributed to HyperCLOVA X SEED Omni and SEED 32B Think. I have also mentored [Minseo Kang](#) and [Jiwoo Lee](#) on multimodal research projects.

References

Prof. Minjoon Seo (서민준), Ph.D. Advisor

Associate Professor, Kim Jaechul Graduate School of AI, KAIST

Dr. Sang-doo Yun (윤상두), Research Collaborator at NAVER

Head of NAVER AI Lab

Prof. Hidetoshi Shimodaira, M.S. Advisor

Professor, Graduate School of Informatics, Kyoto University

Contact details available upon request.