Geewook Kim (김기욱)

Team Manager (Multimodal Alignment, Vision Understanding, Drone AI) and Applied 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 foundation models**: architectures that do more with less compute, and training that extracts more supervision from the same data.

Research directions:

- **Efficient architectures for long and omni-modal inputs:** OCR-free document understanding and state-space token compression for hour-long video [C7, ECCV22; C10, EMNLP23; C17, AAAI26].
- **Data-efficient training:** synthetic data generation and web-based corpora [C8, ICDAR23; C9, ICCV23], read-and-reason prompting [C13, EMNLP24], and conflict-aware data splitting with weight merging [C19, ICML26].
- **Evaluation and model feedback:** judges, self-refinement, and Korean-grounded benchmarks for multimodal and agentic models [C12, ACL-F24; C16, ACL-F25; C15, NAACL25; C21, EMNLP-F26], and the safety and text-capability costs of vision-language adaptation [C14, ICLR25; C24, EMNLP26].
- **Deployable multimodal models and agents:** on-device omni-modal models [O8, Blog26] and multimodal LLM agents for closed-loop control of real drones and multi-drone coordination [O10, arXiv26].

Research vision: make capable multimodal models cheaper to train and run. Next, I aim to compress long audio-visual streams into compact tokens, turn evaluation signals into training supervision, and make small on-device models reliable agents. This work already reaches practice through Donut (6.9K★), NAVER's HyperCLOVA X models (SEED 4B, 8B Omni, and 32B Think) [O8, T3, T4], and DroneCATS [O10].

Education

Korea Advanced Institute of Science & Technology (KAIST) — Aug. 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 multimodal LLMs (HyperCLOVA X Vision, Video, and Omni) [Link] and serve as technical lead for the open-weight HyperCLOVA X SEED 4B and SEED-Omni-8B and for the vision capabilities of SEED-Think-32B, directing omni-modal pre- and post-training (supervised fine-tuning, reinforcement learning, and distillation) on clusters of thousands of GPUs; contribute to NAVER Cloud's role in NVIDIA's open-model coalitions, including Nemotron. Appointed Leader in 2026, heading three teams and mentoring up to 14 researchers and engineers.

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- or co-first-authored. Total citations 2,800+ (Google Scholar, November 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**. **1,500+ 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), "*A Composable Evaluation System for Reproducible Omni-Modal Foundation Model Evaluation*" [Code], Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2026.	EMNLP Demo 2026
[C25]	Sukmin Seo* and **Geewook Kim***† (co-first and corresponding author), "*Natural-Language Temporal Grounding in Hour-Long Videos is a Search Problem: A Benchmark and Empirical Decomposition*" [Press] [News], Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, 2026.	EMNLP 2026
[C24]	Minsik Choi* and **Geewook Kim*** (co-first author), and Young Geun Kim, "*Text Capability Loss in Vision-Language Adaptation: An Attention-Sink Diagnosis*" [Code] [Press], Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, 2026.	EMNLP 2026
[C23]	Sanghee Park*, **Geewook Kim***† (co-first and co-corresponding author), and Kee-Eung Kim†, "*KCSAT-ML: Probing Reasoning Models with Nationwide-Cohort Human Difficulty*", Findings of the Association for Computational Linguistics, 2026.	EMNLP Findings 2026
[C22]	Juan Yeo* and **Geewook Kim***† (co-first and corresponding author), "*Compliance, Capability, and Conflict: Benchmarking Multimodal LLMs under System Messages*", Findings of the Association for Computational Linguistics, 2026.	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, "*K-BrowseComp: A Web Browsing Agent Benchmark Grounded in Korean Contexts*", Findings of the Association for Computational Linguistics, 2026.	EMNLP Findings 2026

[C20] **Geewook Kim** and Minjoon Seo, "*Do Modern Video-LLMs Need to Listen? A Benchmark Audit and Scalable Remedy*", Proceedings of the 27th Annual Conference of the International Speech Communication Association, 2026. — 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 19th 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. — AAAI 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, Hoyeon 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 End-to-End Document Parser*", 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, "*Visually-Situated Natural Language Understanding with Contrastive Reading Model and Frozen Large Language Models*" (**Cream**), 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 on Computer Vision, 2022.
1,500+ citations, **6.9K+ stars at GitHub** — ECCV 2022

[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.
900+ citations, **3.9K+ stars at GitHub**
Oral Presentation (187/4304=4.3%) — ICCV 2019

[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*", 2026. — Tech Report 2026

[T3] NAVER Cloud HyperCLOVA X Team, "*HyperCLOVA X 8B Omni*", 2026. — Tech Report 2026

[T2] NAVER Cloud HyperCLOVA X Team, "*HyperCLOVA X THINK Technical Report*", 2025. — Tech Report 2025

[T1] NAVER Cloud HyperCLOVA X Team, "*HyperCLOVA X Technical Report*", 2024. — Tech Report 2024

Preprints, Tech Blogs, and Domestic Publications

[O10] 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*" (**DroneCATS**) [Project] [Code] [Model], arXiv preprint, 2026. — Preprint 2026

[O9] 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

[O8] NAVER Cloud (CLOVA AI Tech Blog), "*Seeing and Hearing the Field for Ourselves: Building Our Own Encoders and HyperCLOVA X SEED 4B* (우리의 눈과 귀로 현장을 이해하다 : 독자 인코더와 HyperCLOVA X SEED 4B 개발기)", 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

Multimodal Large Language Models: Long-Video Understanding, Omni-Modality, and Physical AI — Oct. 2026
University of Seoul [Slide]

HyperCLOVA X Vision: Open Your Eyes, CLOVA X! — Nov. 2024
TEAM NAVER CONFERENCE DAN 24 [Session Link][Slide]

HyperCLOVA X Vision: Open Your Eyes, CLOVA X! — Oct. 2024
NAVER ENGINEERING DAY

Vision-Language Models for Context-Rich Image Understanding Tasks — Apr. 2024
University of Seoul [Slide][Session Link]

Fine-Grained Evaluation of Vision-Language Models through VLM as a Judge — Feb. 2024
NAVER Tech Meetup [Slide]

Recent Advances in Document AI — Mar. 2023, Dec. 2022
Korea University; Kookmin University

OCR-free Document Understanding Transformer — Nov. 2022
Microsoft [Slide]

Identifying a Store from a Receipt Image — 2021
DEVIEW Conference [Slide][Session Link]

Representation Learning with Weighted Inner Product (WIP) — 2019
Michinoku Communication Science Seminar, Tohoku University [Session Link]

Academic Service

Journal Reviewer:
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (since 2025),
Transactions of the Association for Computational Linguistics (TACL) (since 2026),
International Journal of Computer Vision (IJCV) (since 2026), *Pattern Recognition (PR)* (since 2025),
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), *ICLR* (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 KR103024971B1, filed in September 2023 and granted in September 2026. — 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 KR103026156B1, filed in June 2023 and granted in September 2026. — Jun. 2023

[P6] Daehee Kim, Yoonsik Kim, Yumin Lim, Donghyun Kim, **Geewook Kim**, and Taeho Kil, "*METHOD AND SYSTEM FOR PERFORMING OCR USING CHARACTER-WISE SUPERVISED CONTRASTIVE LEARNING MODEL*", KR Patent Application KR20240143101A, filed in March 2023 (pending); international application PCT/KR2024/003483 (WO2024196144A1), filed in March 2024, followed by applications in China (CN120937061A) and the United States (US20260017967A1, pending). 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 (also granted as JP Patent JP7393509B2, filed in November 2022 and granted in December 2023). 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 (also granted as JP Patent JP7453731B2, filed in April 2022 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

NAVER's AI foundation-model leadership (HyperCLOVA X). Technical lead for NAVER Cloud's open-weight HyperCLOVA X SEED multimodal models, including SEED 4B and SEED-Omni-8B, an omni-modal model that reads, sees, and listens across text, image, audio, and video, and for the vision capabilities of SEED-Think-32B, a reasoning model. Developed amid Korea's national push for sovereign AI foundation models (독자 AI 파운데이션 모델) [news], the line targets strong multimodal capability at low compute.

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 into NAVER products, from Document AI services such as CLOVA OCR [Link] and receipt-verified reviews on NAVER Place [DEVIEW] to the HyperCLOVA X Vision, Video, and Omni models, documented in technical reports, open models, and engineering blogs.

External collaborations. Co-authored research with Korea University, Kyoto University, RIKEN AIP, Adobe Research, and UNC Chapel Hill; supervised Seoul National University students through NAVER Cloud's residency program (see *Student Mentoring*); and currently contribute to NAVER Cloud's participation in NVIDIA's open-model coalitions, including the Nemotron Coalition [Link], a collaboration on open frontier models and the next-generation HyperCLOVA X.

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 and NAVER Cloud, guiding their research from problem formulation through publications, technical reports, and model releases. Mentees who have co-authored papers or technical reports 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-8B and SEED-Think-32B. I have also mentored Minseo Kang and Jiwoo Lee on multimodal research projects.

References

Contact details available upon request.

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

❖ Last updated on November 26, 2026 ❖