

Email: [jinheon.baek@kaist.ac.kr](mailto:jinheon.baek@kaist.ac.kr)Homepage: <https://jinheonbaek.github.io>Google Scholar: [/JinheonBaek](#)RESEARCH  
INTEREST

My research interest lies in machine learning for language, knowledge, and their intersections at scale. Previous work includes modeling interconnected data structures (texts and graphs), and retrieving them to augment language models, while verifying and personalizing their outputs, for practical natural language applications.

## EDUCATION

**KAIST**

Ph.D. in Artificial Intelligence

Seoul, South Korea

M.S. in Artificial Intelligence

Mar 2022 – Present

Thesis: Toward Accurate Learning of Graph Representations, GPA: 4.22/4.30

Mar 2020 – Feb 2022

**Korea University**

Seoul, South Korea

B.S. in Computer Science and Engineering

Mar 2016 – Feb 2020

B.E. in Software Technology and Enterprise Program (Interdisciplinary Program)

GPA: 4.40/4.50, GPA of Computer Science and Engineering: 4.48/4.50

## EMPLOYMENT

**Research Assistant, KAIST** (Advisor: Prof. Sung Ju Hwang)

Jul 2019 - Present

- Conducting research on machine learning for graphs (representation, reconstruction, generation, and retrieval), and their applications to natural language (knowledge graphs) and computer vision (object graphs).

**Research Intern, Google DeepMind (Gemini)**

Sep 2024 - Nov 2024

- Conducted research on in-context learning with many examples for frontier long context language models.

**Research Intern, IBM Research**

May 2024 - Aug 2024

- Conducted research on augmenting large language models for text to SQL with knowledge base construction.

**Research Intern, Microsoft Research**

Jun 2023 - Sep 2023

- Conducted research on augmenting large language models for personalization for search query suggestions.

**Applied Scientist II Intern, Amazon (Alexa AI)**

Aug 2022 - Nov 2022

- Conducted research on knowledge-augmented large language models with retrieval from knowledge graphs.

**Chief Technology Officer (CTO), Jobshopper**

Jul 2018 - Feb 2020

- Co-founded an education-technology startup for student careers, led a research & development group, exited the big-data-based university information platform (Majormap), and now only advising technical directions.

**Undergrad. Research Assistant, Korea University** (Advisor: Prof. Jaewoo Kang)

Mar 2018 - Jun 2019

- Conducted research on representing interconnected knowledge (food ingredients and academic majors) for their recommendations, and research on investigating the spatial reasoning ability of neural models.

**Research Intern, Smart Data Research Group, ETRI**

Jan 2019 - Feb 2019

- Researched an open data distribution platform with the international standard (DCAT), and utilized DevOps tools (Ansible, AWX, and Traefik) to manage nodes on a cloud, in a National AI Research Institute in Korea.

## PUBLICATIONS

**International Publications** (\*: equal contribution; †: corresponding author)

- [43] Revisiting In-Context Learning with Long Context Language Models  
[Jinheon Baek](#), Sun Jae Lee, Prakhar Gupta, Geunseob (GS) Oh, Siddharth Dalmia, and Prateek Kolhar  
arXiv preprint
- [42] Knowledge Base Construction for Knowledge-Augmented Text-to-SQL  
[Jinheon Baek](#), Horst Samulowitz, Oktie Hassanzadeh,  
Dharmashankar Subramanian, Sola Shirai, Alfio Gliozzo, and Debarun Bhattacharjya  
arXiv preprint
- [41] VideoRAG: Retrieval-Augmented Generation over Video Corpus  
Soyeong Jeong\*, Kangsan Kim\*, [Jinheon Baek\\*](#), and Sung Ju Hwang  
arXiv preprint
- [40] Unified Multi-Modal Interleaved Document Representation for Information Retrieval  
Jaewoo Lee\*, Joonho Ko\*, [Jinheon Baek\\*](#), Soyeong Jeong, and Sung Ju Hwang  
arXiv preprint
- [39] Paper2Code: Automating Code Generation from Scientific Papers in Machine Learning  
Minju Seo, [Jinheon Baek](#), Seongyun Lee, and Sung Ju Hwang  
arXiv preprint

- [38] Sketch-of-Thought: Efficient LLM Reasoning with Adaptive Cognitive-Inspired Sketching  
Simon A. Aytes, [Jinheon Baek](#), and Sung Ju Hwang  
arXiv preprint
- [37] Efficient Long Context Language Model Retrieval with Compression  
Minju Seo, [Jinheon Baek](#), Seongyun Lee, and Sung Ju Hwang  
arXiv preprint
- [36] Real-time Verification and Refinement of Language Model Text Generation  
Joonho Ko, [Jinheon Baek](#), and Sung Ju Hwang  
arXiv preprint
- [35] Database-Augmented Query Representation for Information Retrieval  
Soyeong Jeong, [Jinheon Baek](#), Sukmin Cho, Sung Ju Hwang, and Jong C. Park  
arXiv preprint
- [34] Towards Better Understanding of Program-of-Thought Reasoning in Cross-Lingual and Multilingual Environments  
Patomporn Payoungkhamdee, Pume Tuchinda, [Jinheon Baek](#), Samuel Cahyawijaya, Can Udomcharoenchaikit, Potsawee Manakul, Peerat Limkonchotiwat, Ekapol Chuangsuwanich, and Sarana Nutanong  
arXiv preprint
- [33] ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models  
[Jinheon Baek](#), Sujay Kumar Jauhar, Silviu Cucerzan, and Sung Ju Hwang  
Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (**NAACL**), 2025 (**Oral**)
- [32] The BiGGen Bench: A Principled Benchmark for Fine-grained Evaluation of Language Models with Language Models  
Seungone Kim, Juyoung Suk, Ji Yong Cho, Shayne Longpre, Chaeun Kim, Dongkeun Yoon, Guijin Son, Yejin Cho, Sheikh Shafayat, [Jinheon Baek](#), ..., Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo  
Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (**NAACL**), 2025 (**Best Paper**)
- [31] CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark  
David Romero, Chenyang Lyu, Haryo Akbarianto Wibowo, Teresa Lynn, Injy Hamed, Aditya Nanda Kishore, ..., [Jinheon Baek](#), ..., Soyeong Jeong, ..., Thamar Solorio, and Alham Fikri Aji  
Conference on Neural Information Processing Systems (**NeurIPS**), 2024 (**Oral**)
- [30] Retrieval-Augmented Data Augmentation for Low-Resource Domain Tasks  
Minju Seo\*, [Jinheon Baek\\*](#), James Thorne, and Sung Ju Hwang  
Adaptive Foundation Models Workshop at NeurIPS (**AFM @ NeurIPS**), 2024
- [29] An Empirical Study of Multilingual Reasoning Distillation for Question Answering  
Patomporn Payoungkhamdee, Peerat Limkonchotiwat, [Jinheon Baek](#), Potsawee Manakul, Can Udomcharoenchaikit, Ekapol Chuangsuwanich, and Sarana Nutanong  
Empirical Methods in Natural Language Processing (**EMNLP**), 2024
- [28] Rethinking Code Refinement: Learning to Judge Code Efficiency  
Minju Seo, [Jinheon Baek](#), and Sung Ju Hwang  
Findings of Empirical Methods in Natural Language Processing (**Findings of EMNLP**), 2024
- [27] Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity  
Soyeong Jeong, [Jinheon Baek](#), Sukmin Cho, Sung Ju Hwang, and Jong C. Park  
Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**), 2024
- [26] Knowledge-Augmented Large Language Models for Personalized Contextual Query Suggestion  
[Jinheon Baek](#), Nirupama Chandrasekaran, Silviu Cucerzan, Allen Herring, and Sujay Kumar Jauhar  
The Web Conference (**WWW**), 2024
- [25] Knowledge-Augmented Language Model Verification  
[Jinheon Baek](#), Soyeong Jeong, Minki Kang, Jong C. Park, and Sung Ju Hwang  
Empirical Methods in Natural Language Processing (**EMNLP**), 2023
- [24] Test-Time Self-Adaptive Small Language Models for Question Answering  
Soyeong Jeong, [Jinheon Baek](#), Sukmin Cho, Sung Ju Hwang, and Jong C. Park  
Findings of Empirical Methods in Natural Language Processing (**Findings of EMNLP**), 2023
- [23] Knowledge-Augmented Reasoning Distillation for Small Language Models in Knowledge-Intensive Tasks  
Minki Kang, Seanie Lee, [Jinheon Baek](#), Kenji Kawaguchi, and Sung Ju Hwang  
Conference on Neural Information Processing Systems (**NeurIPS**), 2023

- [22] Direct Fact Retrieval from Knowledge Graphs without Entity Linking  
Jinheon Baek<sup>†</sup>, Alham Fikri Aji, Jens Lehmann, and Sung Ju Hwang  
Annual Meeting of the Association for Computational Linguistics (**ACL**), 2023
- [21] Phrase Retrieval for Open-Domain Conversational Question Answering  
with Conversational Dependency Modeling via Contrastive Learning  
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park  
Findings of the Association for Computational Linguistics (**Findings of ACL**), 2023
- [20] Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering  
Jinheon Baek<sup>†</sup>, Alham Fikri Aji, and Amir Saffari  
Natural Language Reasoning and Structured Explanations Workshop at ACL (**NLRSE @ ACL**), 2023 (**Best Paper**)  
Matching from Unstructured and Structured Data Workshop at ACL (**MATCHING @ ACL**), 2023 (**Oral**)
- [19] Personalized Subgraph Federated Learning  
Jinheon Baek\*, Wonyong Jeong\*, Jiongdao Jin, Jaehong Yoon, and Sung Ju Hwang  
International Conference on Machine Learning (**ICML**), 2023
- [18] Realistic Conversational Question Answering with Answer Selection  
based on Calibrated Confidence and Uncertainty Measurement  
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park  
Conference of the European Chapter of the Association for Computational Linguistics (**EACL**), 2023
- [17] Graph Self-supervised Learning with Accurate Discrepancy Learning  
Dongki Kim\*, Jinheon Baek\*, and Sung Ju Hwang  
Conference on Neural Information Processing Systems (**NeurIPS**), 2022
- [16] Object Detection in Aerial Images with Uncertainty-Aware Graph Network  
Jongha Kim, Jinheon Baek, and Sung Ju Hwang  
Visual Object-oriented Learning meets Interaction Workshop at ECCV (**VOLI @ ECCV**), 2022
- [15] Knowledge Graph-Augmented Language Models for Knowledge-Grounded Dialogue Generation  
Minki Kang\*, Jin Myung Kwak\*, Jinheon Baek\*, and Sung Ju Hwang  
Knowledge Retrieval and Language Models Workshop at ICML (**KRLM @ ICML**), 2022
- [14] KALA: Knowledge-Augmented Language Model Adaptation  
Minki Kang\*, Jinheon Baek\*, and Sung Ju Hwang  
Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**), 2022 (**Oral**)
- [13] Augmenting Document Representations for Dense Retrieval with Interpolation and Perturbation  
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park  
Annual Meeting of the Association for Computational Linguistics (**ACL**), 2022 (**Oral**)
- [12] Edge Representation Learning with Hypergraphs  
Jaehyeong Jo\*, Jinheon Baek\*, Seul Lee\*, Dongki Kim, Minki Kang, and Sung Ju Hwang  
Conference on Neural Information Processing Systems (**NeurIPS**), 2021
- [11] Task-Adaptive Neural Network Retrieval with Meta-Contrastive Learning  
Wonyong Jeong\*, Hayeon Lee\*, Geon Park\*, Eunyoung Hyung, Jinheon Baek, and Sung Ju Hwang  
Conference on Neural Information Processing Systems (**NeurIPS**), 2021 (**Spotlight**)
- [10] Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation  
Soyeong Jeong, Jinheon Baek, ChaeHun Park, and Jong C. Park  
Scholarly Document Processing Workshop at NAACL (**SDP @ NAACL**), 2021 (**Oral**)
- [9] Accurate Learning of Graph Representations with Graph Multiset Pooling  
Jinheon Baek\*, Minki Kang\*, and Sung Ju Hwang  
International Conference on Learning Representations (**ICLR**), 2021
- [8] Exploring The Spatial Reasoning Ability of Neural Models in Human IQ Tests  
Hyunjae Kim\*, Yookyung Koh\*, Jinheon Baek, and Jaewoo Kang  
**Neural Networks**, 2021
- [7] Learning to Extrapolate Knowledge: Transductive Few-shot Out-of-Graph Link Prediction  
Jinheon Baek, Dong Bok Lee, and Sung Ju Hwang  
Conference on Neural Information Processing Systems (**NeurIPS**), 2020

**Domestic Publications**, mostly written in Korean

- [6] Artificial Intelligence-Based High School Course and University Major Recommendation System for Course-Related Career Exploration  
Jinheon Baek, Hayeon Kim, and Kiwon Kwon  
KIPS Transactions on Software and Data Engineering (KTSDE), 2021
- [5] Graph Representation Learning with Attention-based Set Pooling  
Jinheon Baek, Minki Kang, and Sung Ju Hwang  
Conference of Korean Artificial Intelligence Association (CKAIA), 2020 (**Best Paper**)
- [4] Embedding Academic Majors and Lectures for Analyzing Departments in University  
Jinheon Baek, Gwanghoon Jang, Soyeong Jeong, Donghyeon Park, Kiwon Kwon, and Jaewoo Kang  
Korea Computer Congress (KCC), 2019 (**Oral**)
- [3] Embedding Food Ingredients Based on Chemical Combination in Dense Vector Space  
Donghyeon Park, Yonggyu Park, Buru Chang, Jinheon Baek, and Jaewoo Kang  
Korea Computer Congress (KCC), 2018 (**Best Paper**)
- [2] Multimodal Approach for Blocking Obscene and Violent Contents  
Jinheon Baek, Dakyeong Lee, Chaeyeon Hong, and Byeongtae Ahn  
Journal of Convergence for Information Technology (JCIT), 7(6), 2017

**Thesis**

- [1] Toward Accurate Learning of Graph Representations  
Jinheon Baek  
Master's Thesis, KAIST, 2022

**PATENTS**

- [6] Apparatus and Method for Providing Interactive Question and Answer Service  
Jong C. Park, Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Sukmin Cho  
Korean patent number: 10-2024-0053617, filed on Apr 22, 2024
- [5] Contextualization of Generative Language Models Based on Entity Resource Identifiers  
Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, Nirupama Chandrasekaran, and Allen Herring  
USSN: 18/529,910, filed on Dec 05, 2023
- [4] Entity-Based Personalization of Generative Language Models  
Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, Nirupama Chandrasekaran, and Allen Herring  
USSN: 63/545,909, filed on Nov 03, 2023
- [3] Electronic Apparatus and Method for Controlling Thereof  
Sung Ju Hwang, Jinheon Baek, Wonyong Jeong, Hyeonuk Kim, Jiongdoo Jin, and Jaehong Yoon  
Korean patent number: 10-2022-0175165, filed on Dec 14, 2022
- [2] Method for Building an Artificial Intelligence based Research Topics Database  
Kiwon Kwon, and Jinheon Baek  
Korean patent number: 10-2021-0017765, filed on Feb 08, 2021, and registered on Apr 19, 2022
- [1] Method for Consulting on Exploration Activities based on Target Departments and Curriculum Achievement Standards  
Kiwon Kwon, and Jinheon Baek  
Korean patent number: 10-2021-0017764, filed on Feb 08, 2021, and registered on Jun 17, 2022

**HONORS AND  
AWARDS**

Awarded the Presidential Science Scholarship for Graduate Study	2024-2026
Received the 3rd Place Poster Presentation Award at Citadel   Citadel Securities PhD Summit	2025
Awarded the Travel Grant from KAIST-Google Partnership Program for WWW 2024	2024
Received the Best Poster Presentation Award at Samsung AI Forum 2023	2023
Received the Best Paper Award at NLRSE Workshop in ACL 2023	2023
Awarded the ICML Travel Grant for ICML 2023	2023
Awarded the Google Travel Grant for NeurIPS 2022	2022
Selected as One of the Top Reviewers (Top 10%) of NeurIPS 2022	2022
Selected as One of the Highlighted Reviewers (Top 10%) of ICLR 2022	2022
Selected as One of the Best Reviewers (Top 10%) of ICML 2021	2021
Received the Best Paper Award at CKAIA 2020	2020
Awarded the Samsung Dream Scholarship	2016-2020
Received the First Prize in the Graduation Project Competition at Korea University	2019
Received the Academic Excellence Award (highest GPA) at Korea University	2019
Received the Second Prize for Excellence in the Microsoft Student Partners Activities	2018
Nominated as the Representative of Korean for Excellence in Microsoft Student Partners Activities	2017

ACADEMIC SERVICES	Reviewer of <b>NeurIPS</b> (Conference on Neural Information Processing Systems)	2021-2025
	Reviewer of <b>ICML</b> (International Conference on Machine Learning)	2021-2025
	Reviewer of <b>ICLR</b> (International Conference on Learning Representations)	2022-2025
	Reviewer of <b>ACL ARR</b> (Association for Computational Linguistics Rolling Review)	2023-2025
	Reviewer of <b>WWW</b> (The Web Conference)	2024-2025
	Reviewer of <b>AAAI</b> (The Association for the Advancement of Artificial Intelligence)	2025
	Reviewer of <b>TMLR</b> (Transactions on Machine Learning Research)	2022-2025
	Reviewer of <b>COLM</b> (Conference on Language Modeling)	2024-2025
	Reviewer of <b>LoG</b> (Learning on Graphs Conference)	2022-2024
	Reviewer of <b>EACL</b> (Conference of the European Chapter of the Association for Computational Linguistics)	2023
	Reviewer of <b>Pattern Recognition</b>	2025
	Reviewer of <b>Internet of Things</b> (IEEE Internet of Things Journal)	2024
	Reviewer of <b>TKDE</b> (IEEE Transactions on Knowledge and Data Engineering)	2023-2024
	Reviewer of <b>SMCA</b> (IEEE Transactions on Systems, Man and Cybernetics: Systems)	2024
	Reviewer of <b>T-IFS</b> (IEEE Transactions on Information Forensics & Security)	2024
	Reviewer of <b>TPAMI</b> (IEEE Transactions on Pattern Analysis and Machine Intelligence)	2023
	Reviewer of <b>TNNLS</b> (IEEE Transactions on Neural Networks and Learning Systems)	2023
	Reviewer of <b>TETCI</b> (IEEE Transactions on Emerging Topics in Computational Intelligence)	2023
	Reviewer of <b>Big Data Research</b>	2023
	Reviewer of <b>AISD</b> (AI and Scientific Discovery: Directions and Opportunities Workshop)	2025
	Reviewer of <b>NLRSE</b> (Natural Language Reasoning and Structured Explanations Workshop)	2023-2024
VOLUNTEERING EXPERIENCE	<b>Oxford COVID-19 Government Response Tracker (OxCGRT)</b>	Jan 2022 - Mar 2023
	• Collected systematic data on policy measures that governments took to tackle COVID-19, aiding decision-makers and citizens in understanding governmental responses over 180 countries. Research papers: [a, b].	
	<b>WoonWha - Education Volunteer Club, Korea University</b>	Jul 2016 - Jul 2017
	• Volunteered for 10 middle school students vulnerable to education twice a week, about 300 hours in total.	
	<b>5th Volunteering Program, Samsung Dream Scholarship Foundation</b>	Jan 2017
	• Volunteered for children in Cambodia by teaching students and exchanging cultures for two weeks.	
TALKS AND SEMINARS	Augmenting Language Models with External Knowledge	
	LG TECH CONFERENCE	Apr 2025
	Citadel   Citadel Securities PhD Summit	Mar 2025
	ResearchAgent: Iterative Research Idea Generation over Scientific Literature with LLMs	
	Korea AI Frontier Lab	Mar 2025
	Knowledge-Augmented (Large) Language Models	
	Vidyasirimedhi Institute of Science and Technology	May 2024
	ETRI	Oct 2023
	Microsoft Research	Aug 2023
	Knowledge-Augmented Language Model Verification	
	Korea Software Congress (KSC)	Dec 2023
	Samsung AI Forum	Nov 2023
	Direct Fact Retrieval from Knowledge Graphs & Personalized Subgraph Federated Learning	
	Korea Software Congress (KSC)	Dec 2023
	POSTECH ML Lab	Jul 2023
	Representation Learning on Graphs with Their Real-World Applications	
	Modulabs	Jul 2022
	Edge Representation Learning with Hypergraphs	
	KAIST AI Workshop	Jan 2022
	Accurate Learning of Graph Representations with Graph Multiset Pooling	
	KAIST AI Workshop	Jan 2022
	ICLR Social - ML in Korea	May 2021
	Introduction of Graph Neural Networks	
	Modulabs	Jan 2022
	Special Talk in Artificial Intelligence as an Alumnus	
	Korea University	Aug 2021
	Learning to Extrapolate Knowledge: Transductive Few-shot Out-of-Graph Link Prediction	
	Korean Conference on Computer Vision (KCCV)	Aug 2021
	NeurIPS Social - ML in Korea	Dec 2020

Basic Algorithms of Machine Learning and Deep Learning with Python Code Challenges	Jul 2018
Microsoft Coding Education Volunteer Group Seminar	May 2018
Microsoft Student Partner Newcomer Tech Seminar	
IaaS (Infrastructure as a Service) in Microsoft Azure	Apr 2018
Microsoft Student Partner Newcomer Tech Seminar	
Phishing Website Detection with Microsoft Azure Machine Learning Studio	Feb 2018
KUSISWALL: Joint Seminar of Korea, Sookmyung and Hanyang Universities	
Basic Algorithms of Machine Learning with Azure Machine Learning Studio	Nov 2017, Nov 2016
Microsoft Student Partner Evangelism	
Azure Cognitive Services with Serverless Cloud Computing using JavaScript and Python	Nov 2017
Microsoft Student Partner Evangelism	
Windows Hello: Microsoft Face Recognition System in Windows 10	Oct 2016
Microsoft Student Partner Inner Tech Seminar	

#### EXTRA- CURRICULAR ACTIVITIES

- Lead & Representative of Korea, Microsoft Student Partner** Jul 2016 - Jul 2020
- As a leader, promoted Microsoft products, recruited members, planned programs, and reviewed projects.
  - As a representative of Korea on a global summit program, invited to Microsoft Build in 2017 which is the biggest annual conference held by Microsoft, and conducted a role as a technical ambassador.
- Chairman, Inc0gnito Hacking Conference** Mar 2018 - Feb 2019
- Organized the information security conference in Korea held by undergraduate students over 10 universities, which covers techs, news, and trends on information security areas.
- 8th Trainee, SW Maestro** Jul 2017 - Dec 2017
- Researched and developed systems that automatically block obscene and violent movie content with machine learning, within the SW Maestro program that aims to train talented students on software in Korea.

#### SKILLS

Languages: Korean (mother tongue), English (fluent)  
 Programming (Coding): Python, JavaScript, Java, and C