Minyoung Choe

Ph.D. Candidate in Artificial Intelligence · *KAIST* +82-10-7212-7482 · minyoung.choe@kaist.ac.kr · github.com/young917

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST) — Seoul, South Korea
M.S. & Ph.D. in Artificial Intelligence — Mar 2021 – Present (Expected 2027.02)
Advisor: Prof. Kijung Shin

Sogang University — Seoul, South Korea
B.S. in Computer Science and Engineering — Mar 2017 – Feb 2021
GPA: 4.14 / 4.3 | Graduated Summa Cum Laude — Ranked 1st / 72

RESEARCH SUMMARY

- **Research Direction:** Structure analysis, scalable generation, and representation learning for graph and hypergraph data, with recent extensions to structured knowledge construction and retrieval-augmented AutoML agents.
- **Hypergraph Modeling:** Developed methods for analyzing, generating, and learning from hypergraph data that capture group-level relationships beyond simple pairwise links, resulting in publications at top data mining and AI venues including KDD, WWW, ICDM, IJCAI, and TKDD.
- **Structured Knowledge & AutoML Agents:** Built graph- and hypergraph-based representations from scientific literature, Kaggle notebooks, and GitHub code, capturing entities, relations, dependencies, evidence context, and reusable ML techniques for evidence reasoning and retrieval-augmented AutoML.
- **Core Strength:** Formulating new research problems from structural insights, translating empirical patterns and baseline failures into testable hypotheses, and executing them end-to-end through representation design, implementation, evaluation, and publication.

RESEARCH & INDUSTRY EXPERIENCE

NEC Laboratories America — Princeton, NJ, USA
Research Intern — Data Science & System Security — Jun – Sep 2025

- **Time-Series Foundation Models with Similar-Case Retrieval:** Used time-series foundation models such as VisionTS and Moirai as retrievers across numeric and visual views, and analyzed how retrieved evidence formats affect LLM-based prediction.
- **Retrieval-Augmented AutoML Agent:** Designed and implemented a Planner–Retriever–Coder–Debugger AutoML system that uses structured memory to retrieve reusable ML techniques from Kaggle competitions, notebooks, and GitHub repositories, generate executable pipelines, debug failures, and iteratively search for stronger model configurations.
- **Structured Memory for Agentic AutoML:** Constructed task-specific graph/hypergraph memory that supports the agent system by capturing preprocessing, feature engineering, modeling, ensembling, dependencies, execution order, and feedback from experimental runs as reusable ML knowledge.

Carnegie Mellon University — Pittsburgh, PA, USA
Visiting Researcher (IITP-funded CMU AI Concentrated Education Program) — Aug 2023 – Feb 2024

- **Realistic Hypergraph Data Generation:** Collaborated with *Prof. Christos Faloutsos* on a compact generative model for real-world hypergraphs, designed to reproduce realistic structural patterns and group relationships with few parameters.
- **WWW 2025 Hypergraph Generation Publication:** Developed research that resulted in a WWW 2025 publication on Kronecker generative models for power-law patterns in real-world hypergraphs.

RESEARCH PROJECTS

Scientific Literature QA and Knowledge Structuring — Industry Collaboration (Samsung) Jan 2026 – Present

- **Realistic Scientific QA Benchmarking:** Formulated condition-aware scientific QA tasks that extend standard fact-recall QA toward realistic R&D evidence reasoning, where answers depend on conditions, contexts, and evidence across multiple papers.
- **Multi-Agent QA Generation and Verification:** Designed and implemented paper-grounded generation and verification pipelines to improve the reliability of condition-aware QA construction with limited domain-expert annotation, filtering unsupported outputs through grounding, evidence-support, and recoverability checks.

AI-based Weather Forecast Support — Korea Meteorological Administration (KMA) Jul 2021 – Dec 2024

- **Operational Similar-Weather Retrieval:** Contributed to a deployed satellite video retrieval system that helps forecasters find historical weather situations analogous to current conditions.
- **Label-free Satellite Video Embedding:** Designed a self-supervised embedding pipeline that compresses high-resolution satellite videos into low-dimensional representations without manual similarity labels, enabling scalable retrieval over large meteorological archives.
- **End-to-End Retrieval Framework:** Built an integrated retrieval framework combining video compression, future-frame prediction, graph-based nearest-neighbor search, ranking, and visualization, leading to a KDD 2025 Industry Track publication as a co-first author.

PUBLICATIONS

(= equal contribution; underline = Minyoung Choe; [C] = Conference, [J] = Journal, [W] = Workshop)*

[W5] TacitFlow: Learning Workflow Representations for Tacit-Knowledge-Grounded Machine Learning Engineering Agents
Yushan Jiang, Wenchao Yu, Minyoung Choe, Dongjin Song, Jingchao Ni, Wei Cheng, Haifeng Chen
KDD 2026 AI Data Scientist Workshop

[C10] Identifying Group Anchors in Real-World Group Interactions Under Label Scarcity
Fanchen Bu, Geon Lee, Minyoung Choe, and Kijung Shin
ICDM 2025

[C9] Attributed Hypergraph Generation with Realistic Interplay Between Structure and Attributes
Jaewan Chun*, Seokbum Yoon*, Minyoung Choe, Geon Lee, and Kijung Shin
ICDM 2025 🏆 **Best Paper Award**

[C8] SkySearch: Satellite Video Search at Scale
Minyoung Choe*, Geon Lee*, Changhun Han*, Suji Kim, Woong Hu, Hyebeen Hwang, Geunseok Park, Byeongyeon Kim, Hyesook Lee, Ha-Myung Park, and Kijung Shin
KDD 2025 (Industry Track)

[C7] Kronecker Generative Models for Power-Law Patterns in Real-World Hypergraphs
Minyoung Choe, Jihoon Ko, Taehyung Kwon, Kijung Shin, and Christos Faloutsos
WWW 2025

[W4] Graphlets over Time: A New Lens for Temporal Network Analysis
Deukryeol Yoon, Dongjin Lee, Minyoung Choe, Kijung Shin
ECML/PKDD 2024 Workshop

[W3] GraphEHR: Heterogeneous Graph Neural Network for Electronic Health Records
Minyoung Choe*, Juho Jung*, Kushagra Agarwal*, Nivedhitha Dhanasekaran*
IJCAI 2024 Workshop

[J1] Representative and Back-In-Time Sampling from Real-world Hypergraphs
Minyoung Choe, Jaemin Yoo, Geon Lee, Woonsung Baek, U Kang, Kijung Shin
ACM Transactions on Knowledge Discovery from Data (TKDD), 2024

[W2] Temporal Graph Networks for Graph Anomaly Detection in Financial Networks
Yejin Kim*, Youngbin Lee*, Minyoung Choe, Sungju Oh, Yongjae Lee
AAAI 2024 Workshop

[C6] Classification of Edge-dependent Labels of Nodes in Hypergraphs
Minyoung Choe, Sunwoo Kim, Jaemin Yoo, and Kijung Shin
ACM KDD 2023

[C5] How Transitive Are Real-World Group Interactions? - Measurement and Reproduction
Sunwoo Kim, Fanchen Bu, Minyoung Choe, Jaemin Yoo, and Kijung Shin
ACM KDD 2023

[C4] Reciprocity in Directed Hypergraphs: Measures, Findings, and Generators
Sunwoo Kim, Minyoung Choe, Jaemin Yoo, and Kijung Shin
IEEE ICDM 2022

[C3] HashNWalk: Hash and Random Walk Based Anomaly Detection in Hyperedge Streams
Geon Lee, Minyoung Choe, and Kijung Shin
IJCAI 2022

[C2] MiDaS: Representative Sampling from Real-world Hypergraphs
Minyoung Choe, Jaemin Yoo, Geon Lee, Woonsung Baek, U Kang, and Kijung Shin
WWW 2022

[C1] How Do Hyperedges Overlap in Real-World Hypergraphs? - Patterns, Measures, and Generators
Geon Lee*, Minyoung Choe*, and Kijung Shin
WWW 2021

[W1] Pretraining Neural Architecture Search Controllers with Locality-based Self-Supervised Learning
Kwanghee Choi*, Minyoung Choe*, Hyelee Lee*
NeurIPS 2020 Workshop

AWARDS & HONORS

Best Paper Award, IEEE ICDM 2025	2025
Student Scholarship Award, The Web Conference (WWW 2021)	2021
Grand Prize, Open-Source Software Contribution (MSIT, Korea)	Nov 2019
National Science & Technology Scholarship (KOSAF)	2019 – 2021
Dean's Honor List — Top 1%, Sogang University	2018 – 2020
Silver Prize, Programming Competition, Sogang University	Nov 2018

TECHNICAL SKILLS

Research Areas: Graph & Hypergraph Mining and Representation Learning, Knowledge Graph/Hypergraph Construction, AutoML Agents, Retrieval-Augmented AI
ML / DL: PyTorch, PyTorch Geometric (PyG), DGL, Graph Neural Networks, Hypergraph Neural Networks, Self-supervised Learning, Representation Learning, Time-Series Foundation Models
LLM / KG / Retrieval Systems: AutoML Multi-agent Systems (Planner–Retriever–Coder–Debugger Agents), Knowledge Graph Construction, Graph/Hypergraph-Augmented Agents
Programming: Python, C/C++, Java
Languages: Korean (native), English (professional)

PROFESSIONAL SERVICE

Workshop Co-organizer — PAKDD 2025 Workshop on Graph Learning with Foundation Models, Sydney, Australia Jun 2025
Program Committee / Reviewer — ACM KDD, The Web Conference (WWW), ACM CIKM, PAKDD

TEACHING EXPERIENCE

KAIST — Teaching Assistant Seoul, South Korea
Data Mining and Search (Spring 2022, 2023, 2024) | Mathematics for AI (Fall 2022) Mar 2022 – Jun 2024

Sogang University — Teaching Assistant Seoul, South Korea
Computational Thinking Sep – Dec 2018