WHANHEE CHO

Email: whanhee@cs.utah.edu GitHub: https://github.com/whnhch LinkedIn: linkedin.com/in/whanhee-cho/

RESEARCH INTERESTS

AI-Enhanced Database Management, Data Usability, Data Interpretability and Explanation

EDUCATION

UNIVERSITY OF UTAH, Graduate School, Salt Lake City, UT, the United States of America Present
Doctor of Philosophy in Computer Science

HANYANG UNIVERSITY, Graduate School, Seoul, Republic of Korea
Master of Science in Computer Science
Summa Cum Laude

HANYANG UNIVERSITY, Seoul, Republic of Korea
Bachelor of Computer Science and Software Engineering
Summa Cum Laude

PUBLICATIONS

- Whanhee Cho, Anna Fariha, "Data-Semantics-Aware Recommendation of Diverse Pivot Tables", *SIGMOD, 2026*
- Whanhee Cho, Anna Fariha. "UTOPIA: Automatic Pivot Table Assistant", *PVLDB, 2024*
- Minji Kim, Whanhee Cho, Soohyeong Kim, and Yong Suk Choi. "Simple Data Transformations for Mitigating the Syntactic Similarity to Improve Sentence Embeddings at Supervised Contrastive Learning", *Advanced Intelligent Systems, 2024*
- Soohyeong Kim, Whanhee Cho, Minji Kim, and Yong Suk Choi. "Bidirectional Masked Self-attention and N-gram Span Attention for Constituency Parsing", *EMNLP Findings, 2023*
- Whanhee Cho, and Yong Suk Choi. "LMGAN: Linguistically Informed Semi-Supervised GAN with Multiple Generators", *Sensors 22, no. 22: 8761 2022.*

PREPRINT

- Whanhee Cho "ACCIO: Table Understanding Enhanced via Contrastive Learning with Aggregations." Preprint, 2024

DATABASE MANAGEMENT RESEARCH EXPERIENCE

University of Utah DBLAB Salt Lake City, USA
Research Assistant

- Personalized Extractive Summarization
  - Developed a novel algorithmic framework for detecting complex user intent for summarization without explicit prompts, resolving combinatorial explosion via Integer Linear Programming (ILP) with PaQL. Demonstrated superior performance to large language models on extractive summarization tasks.
- Diversified Interesting Query Result Recommendation (Accepted at SIGMOD 2026)
  - Designed a recommendation system that evaluates the usability of query results and leverages Large Language Models (LLMs) to capture human common sense, successfully increasing the surprisingness and utility of the output.
  - Solved the combinatorial explosion of candidate sets with pruning strategies, reducing candidate evaluation cost by 80% while maintaining quality. We evaluated our work and novel metrics via a formal user study.
- Semantic aware data post-processing in the pivot table (Published at VLDB 2024)
  - Designed a system that leverages language models and clustering algorithms to automatically group related entities and correct typos, delivering semantically accurate pivot tables in spreadsheet.
- Table Understanding with Contrastive Learning
  - Introduced novel contrastive learning paradigm using related table pairs, achieving 91.1% micro-F1 on colu mn type annotation. Demonstrated competitive performance approaching state-of-the-art methods (within 4 % on micro-F1) while using conceptually simpler approach without data augmentation.

AI RESEARCH EXPERIENCE

Hanyang University Artificial Intelligence LAB Seoul, South Korea
Full-time Student Researcher

- Proposed and implemented a method to syntactically augment a dataset to improve sentence embedding performance. Our work demonstrates that simple transformations can have a significant impact on model performance, addressing the limitations of existing semantic-only approaches
- Leverage the prior works' exploration that the internal layers of language encoders learn different aspects of a sentence. We demonstrated that using these varied outputs as input for a GAN produces a more meaningful representation than using random noise, which often fails in low-resource settings.
- Data preprocessing, and data filtering for pre-training large-scale A.I. model of KT AI One Team*. Led the development of a toxic dataset and classifier for pre-training large-scale AI models, approaching data filtering as a classification task to protect the model from training with toxic data. (KT AI One Team*: A national AI cooperative organization of 12 industry-university-research institutes.)
- Investigated relations and impacts of next token probability onto positional embeddings. The purpose of this project was to validate the effect of GPT-2 next token prediction probability on positional embedding at the relation extraction task using the TACRED dataset. According to the next token probability, set the positional embedding values of the next gold token to be closer or further from the current token.

TECHNICAL SKILLS

- Programming Languages: C/C++, Python, Java, SQL
- Database Systems: PostgreSQL, MySQL, SQL Server, MongoDB
- ML/AI Frameworks: PyTorch, TensorFlow, scikit-learn, Huggingface
- Tools & Technologies: Git, CMake, Linux kernel programming, CUDA, Docker

DATABASE PROJECTS

- Database Management | Built B+ tree with insert, delete, and range query operations in C++
- Advanced Database Systems | Implemented concurrent B^{ε} Skiplist Tree with lock-free operations

CONFERENCE & PRESENTATION

- University of Utah, School of Computing, "Data Science Research," Salt Lake City, UT, USA, April 2025
- VLDB 2024, "UTOPIA: Automatic Pivot Table Assistant", Guangzhou, China, August 2024
- NWDS 2023, "UTOPIA: Automatic Pivot Table Assistant," Seattle, WA, USA, August 2023

TEACHING EXPERIENCE

Teaching Assistant | University of Utah
- CS6530: Advanced Database Systems | Fall 2025

Teaching Assistant | Hanyang University
- ITE4053: Deep Learning Methods and Applications | Fall 2022
- CSE4007: Artificial Intelligence | Fall 2021
- CSE0092: Introduction to Artificial Intelligence | Fall 2021, Spring 2022

WORK EXPERIENCE

DAYONE COMPANY, SAMSUNG DS Seoul, South Korea
Lecturer April 2023 – June 2023

- Lectured about basic python programming, data preprocessing, and probability and statistics.

AWARDS & HONORS

Fellowship, Kahlert School of Computing, University of Utah
Brain Korea 21 Fellowship, NRF Korea
Need Based Scholarship, Hanyang University
Merit Based Scholarship, Hanyang University
Honors student in the 1st semester of 2018, Hanyang University
3rd Prize, University-wide Group Study Competition
National Grant, Korea Student Aid Foundation (KOSAF)