# Geonmo Gu

gmgu@theory.snu.ac.kr https://github.com/gmgu

## RESEARCH INTERESTS

**Deep Learning**: At LG Electronics, I am developing an AI coding assistant using large language models (LLMs). I have successfully trained LLMs in the distributed settings, and have deployed LLMs to hundreds of users. Recently, I am conducting research on fast and accurate LLM inference.
**Algorithm Engineering**: My primary research efforts have been devoted to developing fast algorithms. I developed fast algorithms for graph isomorphism, graph isomorphism query processing, and multiple pattern Cartesian tree matching during my Ph.D. studies.

## WORK EXPERIENCE

**LG Electronics** **Seoul, Korea**
*Senior Researcher* *Apr. 2022 – Present*
Artificial Intelligence Lab

- Jan. 2024 – Present: **Development of AI Coding Assistant using Large Language Model**
  - Conducting research on domain adaptive continual pretraining code LLMs.
  - Maintaining custom benchmark dataset for offline evaluation.
  - Analyzing user data and feedback for online evaluation.
  - Constructing instruction dataset and conducting instruction-tuning.
- Aug. 2022 – Dec. 2023: **Development of AI Coding Assistant using Large Language Model**
  - Conducted distributed training of LLMs based on decoder-only transformer.
  - Filtered and deduplicated terabytes of source code data.
  - Developed a fast LLM inference server in terms of latency and throughput.
- Apr. 2022 – Dec. 2022: **Development of Coding Education Program Utilizing AI**
  - Constructed training data for generating Python code from natural language instruction.
  - Trained an encoder-decoder transformer from scratch.
  - Developed a web client that inputs prompt, prints AI-generated code, and executes Python code.
  - Created a inference server that runs on multiple GPUs, loads multiple copies of the model, and offers dynamic batching for increased throughput.

**Seoul National University** **Seoul, Korea**
*Post-Doctoral Assistant* *Jan. 2022 – Mar. 2022*
Institute of Computer Technology

- Jan. 2022 – Mar. 2022: Algorithm Development for Graph Isomorphism Query Processing
  - Developed a fast graph isomorphism query processing algorithm that runs orders of magnitude faster than state-of-the-art algorithms.

**NAVER** **Gyeonggi-do, Korea**
*Internship* *Oct. 2021*
AI Dev2

- Oct. 2021: Analyzing Conversion Tracking Data
  - Conducted exploratory data analysis on glad for advertisement data to find meaningful trends.
  - Handled hundred gigabytes of (raw) conversion tracking data.
  - Solved optimization problem of maximizing conversion rate using linear programming.

## EDUCATION

**Seoul National University** **Seoul, Korea**
Ph.D. in Computer Science and Engineering *Mar. 2014 – Aug. 2021*

- Thesis: Fast Graph Isomorphism using Pairwise Color Refinement and Efficient Backtracking
- Advisor: Prof. Kunsoo Park
- GPA: 3.99/4.3

**Incheon National University** **Incheon, Korea**
B.S. in Computer Science and Engineering *Mar. 2010 – Feb. 2014*

- GPA: 4.4/4.5 (summa cum laude)

## PUBLICATIONS

**Geonmo Gu**, Yehyun Nam, Kunsoo Park, Zvi Galil, Giuseppe F. Italiano, and Wook-Shin Han. "Efficient Graph Isomorphism Query Processing using Degree Sequences and Color-Label Distributions." *IEEE International Conference on Data Engineering*, 2022.

**Geonmo Gu**, Yehyun Nam, Kunsoo Park, Zvi Galil, Giuseppe F. Italiano, and Wook-Shin Han. "Scalable Graph Isomorphism: Combining Pairwise Color Refinement and Backtracking via Compressed Candidate Space." *IEEE International Conference on Data Engineering*, 2021.

Siwoo Song, **Geonmo Gu**, Cheol Ryu, Simone Faro, Thierry Lecroq, and Kunsoo Park. "Fast Algorithms for Single and Multiple Pattern Cartesian Tree Matching." *Theoretical Computer Science*, 2020.

**Geonmo Gu**, Siwoo Song, Simone Faro, Thierry Lecroq, and Kunsoo Park. "Fast Multiple Pattern Cartesian Tree Matching." *International Conference and Workshop on Algorithms and Computation*, 2020.

Myoungji Han, Hyunjoon Kim, **Geonmo Gu**, Kunsoo Park, and Wook-Shin Han. "Efficient Subgraph Matching: Harmonizing Dynamic Programming, Adaptive Matching Order, and Failing Set Together." *ACM SIGMOD International Conference on Management of Data*, 2019.

Myoungji Han, Munseong Kang, Sukhyeun Cho, **Geonmo Gu**, Jeong Seop Sim, and Kunsoo Park. "Fast Multiple Order-Preserving Matching Algorithms." *International Workshop on Combinatorial Algorithms*, 2015.

Seongi Hong, **Geonmo Gu**, Hyunjoon Kim, Kunsoo Park. "Performance Comparison of Adaptive Matching Orders for the Subgraph Isomorphism Problem." *KIISE Transactions on Computing Practices*, 26.1:38-43. 2020.

Seongi Hong, **Geonmo Gu**, Hyunjoon Kim, Kunsoo Park. "Performance Comparison of Candidate-Size Ordering and Path-Size Ordering for Subgraph Isomorphism Problem." *Korea Computer Congress*, 2019

## PROJECTS

**Framework of Practical Algorithms for NP-hard Graph Problems** **Seoul National University**
*Funded by the Korea government (Ministry of Science and ICT)* *Apr. 2018 – Aug. 2021*

- Algorithm development for fast subgraph isomorphism, graph isomorphism, and graph isomorphism query processing.
- Open source contribution for practical graph algorithms (`https://github.com/SNUCSE-CTA`).

**Algorithm Development for Scanner/Stage Path Generation** **Seoul National University**
*Supported by JASTECH* *Jul. 2014 – Jun. 2017*

- Sophisticated algorithm that can synchronize Scanner and Stage.
- Development of path simplification method based on chain stabbing (computational geometry).
- Efficient path generation methods by solving the traveling salesman problem (NP-complete).

**NIPA-PURDUE Capstone Program** **Purdue University**
*Center for Robotic Innovation, Commercialization and Education* *Jan. 2014 – Feb. 2014*

- Robot programming (Robotis Bioloid) in collaboration with students of Purdue University.

## PROFESSIONAL ACTIVITIES

Seminar about Distributed Training Large Language Models **Dankook University**
*Apr. 2024*

Reviewer of Information Processing Letters **ELSEVIER**
*Dec. 2020 – Sep. 2023*

Seminar about Distributed Training Techniques for Large AI Models **LG Electronics**
*Jul. 2023*

Invited talk at STARLAB Meeting **Korea Computer Congress**
*Jun. 2023*

Invited talk at 2023 TOPCIT Workshop **IITP**
*Mar. 2023*

## HONORS

**2023 Innovation Awards of CTO Division** **LG Electronics**
*Awarded a Grand Prize* *Jan. 2024*

**The 14th Open SW Developer Contest** **Ministry of Science and ICT**
*Awarded a Gold Prize* *Nov. 2020*

**The 2nd Test of Practical Competency in IT (TOPCIT)** **Ministry of Science and ICT**
*Awarded a Silver Prize* *Sep. 2013*

**The 1st Test of Practical Competency in IT (TOPCIT)** **Ministry of Knowledge Economy**
*Awarded a Grand Prize* *Oct. 2012*

## SKILLS

**Competitive Programming.**

- BAEKJOON: `https://www.acmicpc.net/user/gmgu`

**Programming Languages.** C/C++, Python, CUDA C++, Rust, C#, Java, Shell Script, LaTeX

- C++: `https://github.com/gmgu/GI`
- CUDA C++: `https://github.com/gmgu/study-cuda`
- Rust: `https://github.com/gmgu/study-rust`

**Libraries.** PyTorch, TensorFlow, HuggingFace Transformers, DeepSpeed, Triton (NVIDIA), FasterTransformer, FastAPI, Triton (OpenAI), Seaborn, Pandas, PySpark, gtest

- Triton: `https://github.com/gmgu/study-trident`

**Others.** AWS (SageMaker, EC2, Lustre, S3)