In Gim

Ph.D Student
Department of Computer Science
Yale University, New Haven, CT 06511

in.gim@yale.edu
https://ingim.org

Academic Interests

I am a second-year Ph.D. student in the Department of Computer Science at Yale University. I am interested in building efficient systems for generative AI through system design, formal methods, and prompt programming languages.

Education

Yale University

- Ph.D., Computer Science, expected graduation: 2028.
- Advisor: Prof. Lin Zhong

Yonsei University

- B.S., Integrated Technology (Computer Science), 2021.
- Advisor: Prof. JeongGil Ko

Publications and Preprints

4. In Gim, Binghong Li, Guojun Chen, Seung-Seob Lee, Nikhil Sarda, Anurag Khandelwal, and Lin Zhong: **"Scaling Language Model Serving with Multiplex Inference,"** in submission, December 2023. (pdf)

3. In Gim, Guojun Chen, Seung-Seob Lee, Nikhil Sarda, Anurag Khandelwal, and Lin Zhong: **"Prompt Cache: Modular Attention Reuse for Low-Latency Inference,"** to appear in Proc. Conf. Machine Learning and Systems (MLSys), May 2024. (pdf)

2. In Gim and JeongGil Ko: **"Memory-Efficient DNN Training on Mobile Devices,"** in Proc. ACM Int. Conf. Mobile Systems, Applications and Services (MobiSys), July 2022. (pdf)

1. Hyunjun Kim (In Gim) and JeongGil Ko: **"Fast Monte-Carlo Approximation of the Attention Mechanism,"** in Proc. of the AAAI Conference on Artificial Intelligence (AAAI), January 2022. (pdf)

Services

- Graduate Student Advisory Committee, Dept. of Computer Science (2023-2024).

Awards

- Yale University Graduate Fellowship (2022).

---

[1]Updated March 1, 2024