Donghyun Kim

GDC, 2317 Speedway, Austin, TX 78712
donghyun@utexas.edu | donghyun-kim.com

Research Interests

ML for systems, Cloud Systems, Synthetic trace generation

Education

The University of Texas at Austin (UT Austin) *Austin, TX, USA*
Ph.D. in Computer Science *Aug 2023 - Present*
- Advised by Professor Aditya Akella and Professor Daehyeok Kim

Seoul National University (SNU) *Seoul, South Korea*
M.S in Artificial Intelligence *Aug 2022*
- Advised by Professor Byung-Gon Chun

Pohang University of Science and Technology (POSTECH) *Pohang, South Korea*
B.S. in Convergence IT Engineering, Minor in Computer Science *Aug 2020*

Publications

1. **Omega: A Low-Latency GNN Serving System for Large Graphs,** Geon-Woo Kim, Donghyun Kim, Jeongyoon Moon, Henry Liu, Tarannum Khan, Anand Iyer, Daehyeok Kim, Aditya Akella. Preprint, 2025.
2. **Large Language Models as Realistic Microservice Trace Generators,** Donghyun Kim, Sriram Ravula, Taemin Ha, Alex Dimakis, Daehyeok Kim, Aditya Akella. **EMNLP**, 2025 (Acceptance rate=22.16%).
3. **StitchLLM: Serving LLMs, One Block at a Time,** Bodun Hu, Shuozhe Li, Saurabh Agarwal, Myungjin Lee, Akshay Jajoo, Jiamin Li, Le Xu, Geon-Woo Kim, Donghyun Kim, Hong Xu, Amy Zhang, Aditya Akella. **ACL**, 2025 (Acceptance rate=20.3%).
4. **On a Foundation Model for Operating Systems,** Divyanshu Saxena, Nihal Sharma, Donghyun Kim, Rohit Dwivedula, Jiayi Chen, Chenxi Yang, Sriram Ravula, Zichao Hu, Aditya Akella, Sebastian Angel, Joydeep Biswas, Swarat Chaudhuri, Isil Dillig, Alex Dimakis, Brighten P Godfrey, Daehyeok Kim, Christopher Rossbach, and Gang Wang. **MLSys@NeurIPS**, 2023.
5. **Band: Coordinated Multi-DNN Inference on Heterogeneous Mobile Processors,** Joo Seong Jeong*, Jingyu Lee*, Donghyun Kim, Changmin Jeon, Changjin Jeong, Youngki Lee, Byung-Gon Chun. **MobiSys**, 2022 (Acceptance rate=21.6%).

* equal contribution

Professional Experience

UT Networked Systems Lab, UT Austin *Austin, TX, USA*
Graduate Research Assistant *Aug 2023 - Present*
- Synthetic computer system trace generation using large language models.
- Synthetic application generation that emulates system resource usage patterns using generative models.

FriendliAI *Seoul, South Korea*
Software Engineer *Jan 2022 - Jun 2023*
- Developed an LLM serving service for automated deployment and resource scaling on cloud platforms (e.g., AWS, GCP, Azure).

Software Platform Lab, SNU *Seoul, South Korea*
Graduate Research Assistant *Jan 2020 - Aug 2022*
- *Band*: a coordinated Multi-DNN Inference system for mobile devices with heterogeneous processors.
- *Crane*: a GPU cluster management system for DL training and AutoML workloads.

Data Systems Lab, POSTECH *Pohang, South Korea*
Undergraduate Research Intern *Jun 2019 - Aug 2019*
- Developed a video query processing pipeline leveraging deep neural networks.

Teaching

Computer Networks *UT Austin*
Teaching Assistant *Spring 2025*

Software Development Principles and Practices *SNU*
Teaching Assistant *Fall 2020, Fall 2021*

Honors and Awards

National Scholarship from ICT Creative Consilience Program *Mar 2014 - Aug 2019*

Technical Skills

Programming Language	Python, C++, C, CUDA, Go, Verilog
Library/Framework	PyTorch, Tensorflow Lite, vLLM
Cloud Platform	AWS, GCP, Azure
Tool	Kubernetes, Docker, Docker Swarm, Terraform, Elasticsearch, Fluent Bit, Prometheus