

# Daeun Jung

✉ daeunj@umd.edu | 🌐 [Homepage](#) | [LinkedIn](#) | 📍 College Park, MD

## EDUCATION

|                                                                                                          |                                             |
|----------------------------------------------------------------------------------------------------------|---------------------------------------------|
| <b>University of Maryland, College Park</b><br><i>Ph.D. Student in Computer Science; Advisor: Ang Li</i> | Aug. 2022 – Present<br>College Park, MD     |
| <b>Ewha Womans University</b><br><i>M.S. in Electronic and Electrical Engineering</i>                    | Mar. 2019 – Aug. 2021<br>Seoul, South Korea |
| <b>Ewha Womans University</b><br><i>B.S. in Electronics Engineering</i>                                  | Mar. 2014 – Feb. 2019<br>Seoul, South Korea |

## RESEARCH EXPERIENCE

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                         |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|
| <b>University of Maryland</b><br><i>Graduate Research Assistant</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Aug. 2022 – Present<br>College Park, MD |
| <ul style="list-style-type: none"><li>Develop machine-unlearning methods and representation-level audits for multilingual large language models, with emphasis on distinguishing knowledge deletion from behavioral suppression.</li><li>Study personalized federated optimization for reasoning models under heterogeneous client rewards and data distributions.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                         |
| <b>Runara.ai</b><br><i>Research Intern</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | May 2026 – Jul. 2026<br>San Jose, CA    |
| <ul style="list-style-type: none"><li>Built reproducible GPT-OSS-120B SpecPrefill benchmarks on vLLM for 16K–32K contexts; measured <b>3.70–3.88× target-model TTFT reduction</b> and isolated scorer overhead and answer-quality degradation as the main end-to-end bottlenecks.</li><li>Designed CUDA LayerNorm/RMSNorm+Linear fusion kernels for transformer inference on NVIDIA H100, achieving up to <b>2.17× single-operation speedup</b> and <b>1.24× end-to-end speedup</b> on OPT-6.7B; ported the implementation to HIP for AMD MI300X.</li><li>Extended RMSNorm+QKV fusion to Qwen3-Coder-480B with custom CUDA kernels, reaching <b>1.47× peak layer speedup</b> while preserving cosine similarity above <b>0.99998</b>.</li><li>Evaluated GQA KV-cache reduction on Qwen3-30B and identified 8-bit quantization as the viable design, delivering <b>49–50% cache savings</b> at approximately <b>97% top-1 agreement</b>; ruled out inaccurate K↔V reconstruction through layer-wise and output-equivalence tests.</li></ul> |                                         |
| <b>Carnegie Mellon University</b><br><i>Visiting Scholar</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Jan. 2020 – Jul. 2020<br>Pittsburgh, PA |
| <ul style="list-style-type: none"><li>Developed GAN-based multimedia generation pipelines and NLP-based conversational systems using PyTorch and AWS EC2.</li><li>Processed large-scale multimodal datasets for deep-learning experiments.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                         |

## SELECTED PUBLICATIONS

- [1] **Exposure Leaves a Trace: A Provenance Audit for Machine Unlearning.**  
Daeun Jung and Ang Li.  
*International Conference on Learning Representations (ICLR)*, 2027. Under review.
- [2] **Privacy Risks of Federated Genomic Foundation Models for Multi-institutional Digital Medicine.**  
Daeun Jung, Ziyao Wang, Jiaheng Guo, Song Wang, Tianlong Chen, and Ang Li.  
*npj Digital Medicine*, 2026. Under review.
- [3] **FedMOA: Federated GRPO for Personalized Reasoning LLMs under Heterogeneous Rewards.**  
Ziyao Wang\*, Daeun Jung\*, Yexiao He, Guoheng Sun, Zheyu Shen, Myungjin Lee, and Ang Li.  
*ARR - Oct.2026*. Under review. [[pdf](#)]
- [4] **Impact of Input Data Randomness on Training Performance of Autoencoder.**  
Joohong Rhee\*, Daeun Jung, and Hyunggon Park.  
*KICS Summer Conference*, 2021. Best Paper Award.

[5] **Traffic Data Classification using Machine Learning Algorithms in SDN Networks.**

Jungmin Kwon\*, **Daeun Jung\***, and Hyunggon Park.

*International Conference on ICT Convergence (ICTC)*, 2020. [[pdf](#)]

[6] **An Iterative Algorithm of Key Feature Selection for Multi-class Classification.**

**Daeun Jung\*** and Hyunggon Park.

*International Conference on Ubiquitous and Future Networks (ICUFN)*, 2019. [[pdf](#)]

## HONORS

**Best Paper Award**, KICS

2021

**Research Assistant Scholarship**, Ewha Womans University

2020

**Dean's List**, Ewha Womans University

2015, 2017, 2018

## SKILLS

**Languages:** Python, C/C++, CUDA, HIP, Java, MATLAB, R

**ML/LLM:** PyTorch, Hugging Face Transformers, PyTorch Lightning, vLLM, SGLang, TensorRT-LLM

**Systems:** Linux, Git, AWS, NVIDIA H100, AMD MI300X, GPU profiling and kernel optimization