YANCHEN LIU

University of Southern California, Los Angeles, CA, USA
+1 (213) 921-6117 |✉ liuyanch@usc.edu |🏠 https://jamesnulliu.github.io

EDUCATION HISTORY

M.Sc. in Computer Science **Sep. 2025 – Jun. 2027**
University of California, LA, CA, USA

B.Eng. in Computer Science **Sep. 2021 – Jun. 2025**
Shanghai University, Shanghai, China

HONORS AND AWARDS

1. [2024] *First Prize* and *Group Competition Award* in 2024 ASC Student Supercomputer Challenge Global Final.
2. [2022] *First-Class Academic Scholarship* for outstanding academic performance, Shanghai University.

PROFESSIONAL EXPERIENCE

Graduate Research Intern **Jun. 2025 - Present**
University of Southern California | INK Lab
Topic: RLHF and Reasoning for LLMs

1. Enhanced *Chain-of-Thought* data with *segment and token importance* evaluated by different methods; Utilized the data for *Supervised Fine-Tuning* of various reasoning models with *TRL*.
2. Modified *v1 engine of vllm* to implement and reproduce *soft-thinking* and *latent-thinking* for reasoning models.
3. Modified *verl* to selectively *filter out target tokens* (generated during inference process) based on *entropy aggregation*, which are later used for model updating.

Machine Learning Engineer Intern **Jul. 2024 - Jun. 2025**
Shanghai AI Laboratory
Topic: LLM Inference Engine, AI Compiler and Model Fine-tuning

1. Extended *vllm* for *LLM inference* on in-house TPUs, including researching and adapting *Speculative Decoding*, *Paged-Attention* and *Continuous-Batching*.
2. Developed *high-performance kernels* with *MLIR* to ensure seamless compatibility and optimal performance of LLMs on in-house TPUs.
3. Conducted research on *LLM knowledge injection and fine-tuning* for *kernel fusion and translation* across different hardware platforms.

Undergraduate Research Assistant **Mar. 2023 - Apr. 2025**
Shanghai University | Shanghai University Cyber Security Lab
Topic: Vehicle Modeling, Simulation, and Intrusion Detection

1. Combined *deep learning* and *traditional math modeling* to assess cyber security and functional safety in *in-vehicle communication system*.
2. Developed a novel efficient *gradient descent solver* for *Multi-Dimensional Hawkes Process*, and implemented algebraic simplification, vectorization and JIT compilation for optimization; Introduced a novel *MDHP-LSTM structure* for improved feature extraction in in-vehicle communication data of ECUs and related applications.

Team Leader of Shanghai University Super-Computing Team **Sep. 2023 - Jul. 2024**
Shanghai University | SHUSCT
Topic: Super-Computing

1. Participated in *2024 ASC Student Supercomputer Challenge*.
2. Assembled, benchmarked and optimized a *high-performance computing cluster* for running LLMs and super-computing programs.
3. Developed *a custom LLM inference engine* with various parallelism and scheduling policies; Integrated FlashAttention, quantization and pruning techniques to speed up inference.
4. Profiled and improved performance of *various super-computing programs* by applying vector instructions, OpenMP, loop unrolling, and MPI.

Undergraduate Research Intern **Jun.2023 – Jul. 2024**

Shanghai University | East China Air Traffic Control Bureau

Topic: Super Resolution for Meteorological Data

1. Deployed cutting-edge deep learning models including *PanGu* and *FourCastNet* to improve *meteorological monitoring, forecasting, and super-resolution* for the East China Air Traffic Control Bureau, replacing traditional numerical methods.
2. Utilized *key-point and semantic constraints*, combined with *feature map fusion and redesigned loss functions*, to improve the up-sampling process in *super-resolution* models, addressing the issue of semantic detail loss in certain areas.

PUBLICATIONS

1. [2025 | Preprint] J. Lv, X. He, **Y. Liu**, A. Shen, X. Dai, Y. Li, J. Hao, J. Ding, Y. Hu, S. Yin. "*HPCTransEval: A Benchmark of High-Performance GPU-to-CPU Transpilation with Pre-trained Large Language Models*".
2. [2025 | Preprint | Code] Q. Liu†, **Y. Liu†**, R. Li, C. Cao, Y. Li*, X. Li*, P. Wang, R. Feng, "*MDHP-Net: Detecting an Emerging Time-Exciting Threat in IVN*".
3. [2025 | Preprint] Z. Xu, A. Shen, D. Kong, X. Dai, J. Liu, **Y. Liu**, L. Wang, S. Wei, Y. Hu and S. Yin*, "*LLMEngine: Disaggregated Mapping and Memory Management Co-scheduling for Wafer-scale Chips*".
4. [2024 | IEEE Internet of Things Journal | Code] Q. Liu, X. Li, K. Sun, Y. Li* and **Y. Liu***, "*SISSA: Real-Time Monitoring of Hardware Functional Safety and Cybersecurity With In-Vehicle SOME/IP Ethernet Traffic*".
5. [2024 | MDPI Future Internet | Code] Li, X., R. Li, and **Y. Liu**. "*HP-LSTM: Hawkes Process–LSTM-Based Detection of DDoS Attack for In-Vehicle Network*".