Yuanzhe Hu

📞 +1-619-937-1812
✉ yuh127@ucsd.edu

 GitHub Profile
 LinkedIn Profile

EDUCATION

University of California, San Diego *Sep. 2024 – Mar. 2026*
MS. in Computer Science Engineering

Huazhong University of Science and Technology *Sep. 2020 – Jun. 2024*
BEng. in Artificial Intelligence (Innovation Experimental Honor Class) GPA: 3.91/4.0

PUBLICATIONS AND PREPRINTS

1. Y. Hu, K. Goel, V. Killiakov, Y. Yang. "Eigenspectrum Analysis of Weight Matrices without Aspect Ratio Bias", ICML 2025, Paper Link.

2. Y. Hu*, Y. Wang*, K. Lin, J. McAuley. "Evaluating Memory in LLM Agents via Incremental Multi-Turn Interactions", Submitted to NeurIPS 2025 (D&B Track).

3. Z. Liu*, **Y. Hu***, T. Pang, Y. Zhou, P. Ren, Y. Yang. "Model Balancing Helps Low-data Training and Fine-tuning", EMNLP 2024 (Oral) (168/6105), Paper Link.

4. Y. Wang, D. Krotov, **Y. Hu**, Y. Gao, W. Zhou, J. McAuley, D. Gutfreund, R. Feris, Z. He. "M+: Extending MemoryLLM with Scalable Long-Term Memory", ICML 2025, Paper Link.

* Equal Contribution

ACADEMIC EXPERIENCE

Large-Scale Reasoning LLM Training *Jun. 2025 – Now*
*Research Intern, Advised by **Prof. Zhiting Hu** and **PhDs. Shibo Hao / Zhoujun Cheng*** La Jolla, CA
  - Research Topics: Reasoning Model, Supervised Fine-Tuning, RL.

  - Performed Supervised Fine-Tuning of Qwen2.5-32B and LLaMA-3.1-70B models using LLaMA Factory on AM-Thinking-v1-Distilled dataset.

Constructing Benchmarks on LLMs' Long Context Understanding Ability *Oct. 2024 – May. 2025*
*Research Intern, Advised by **Prof. Julian McAuley** and **PhDs. Yu Wang*** La Jolla, CA
  - Research Topics: Long Context LLM, Memory Agent.

  - Designed a long context modeling benchmark focused on global temporal reasoning for novel-length texts, and validated the benchmark's effectiveness across the LLMs like LLaMA-3.1-8B and MemoryLLM. **(ICML 2025)**

  - Established a comprehensive long context language understanding benchmark for memory agents by summarizing and reformulating existing benchmarks and datasets. Implemented our benchmark to evaluate current memory agent systems such as Letta, Mem0 and Cognee. **(Submitted to NeurIPS 2025)**

Model Diagnosis Based on Random Matrix Theory *Jul. 2023 – May. 2025*
*Research Intern, Advised by **Prof. Yaoqing Yang*** Hanover, NH
  - Research Topics: LLM Supervised Fine-Tuning, LLM Pruning, Scientific ML, Random Matrix Theory.

  - Modified a model weight matrices analysis method based on Marchenko–Pastur distribution and evaluated our method on downstream tasks like layer-wise LLM Pruning and SciML Fine-tuning. **(ICML 2025)**

  - Researched a modified layer-wise optimization algorithm for low-data training and supervised fine-tuning in domains such as SciML and NLP. For example, it improves the performance for language models such as LLaMA-7B or RoBerta-Base by about 2% - 10% in various tasks and low data source settings. **(EMNLP 2024)**

  - Contributed to the rebuttal phase of the TempBalance paper by supplementing additional experimental data, validating the method's effectiveness across downstream tasks including object detection (YOLO-v8) and language modeling.

AWARDS

17th National College Students Intelligent Car Race, Chinese Automation Association *Sep. 2022*
• Second Prize at National Level (Top 7% among 2771 teams)

SKILLS AND INTERESTS

English Skills: IELTS 7.5
Programming Language: Python, C/C++, SQL, MATLAB
Field of Interest: Random Matrix Theory, Model Diagnosis, SciML, Long Context LLM
Service: Reviwer for ICML 2025 / ICLR 2025 / NeurIPS 2024 Workshop