YUXIN YANG

Department of Computer Science, University of Southern California
@ yyang393@usc.edu — yang-yuxin.github.io — +1 323-401-8402

RESEARCH INTERESTS

Large Language Models, Agentic AI Systems, Generative Recommendation, Retrieval-Augmented Generation, Test-time Scaling, Graph Machine Learning

EDUCATION

Ph.D. in Computer Science, University of Southern California — Sep 2023 – present
Advisor: Prof. Viktor Prasanna; GPA: 3.83/4.00

B.Eng. in Automation, Tsinghua University — Sep 2019 – Jun 2023
GPA: 3.84/4.00

INDUSTRY EXPERIENCE

ByteDance — May – Aug 2026
Summer Intern, LLM-native Generative Recommendation

- Built training and offline evaluation infrastructure for predicting next-item **semantic IDs (SIDs)** from interaction histories; designed pretraining + SFT experiments and implemented daily fine-tuning, **next-N prediction** with a shared history context, and its training kernel.
- Studied model-size and context-length scaling, with modest gains from larger models; **SID-text alignment** initialization before recommendation pretraining outperformed direct initialization from pretrained Qwen weights.
- Designed **tree decoding** and contributed to **BF16-to-FP8** conversion and validation; evaluation metric differences were **below 0.0002**.
- Subsequent online experiments showed gains of approximately **1.5× the internal shipping bar**; this does not imply launch-review approval.

SELECTED AWARDS AND HONORS

- Capital One PhD Fellow — 2024–2025, 2025–2026, 2026–2027
- Changtong Scholarship (top 1% in the Department) — 2022
- Comprehensive Scholarship of University (top 5% in University) — 2020, 2021

PUBLICATIONS AND MANUSCRIPTS

Yuxin Yang, Gangda Deng, Ömer Faruk Akgül, Nima Chitsazan, Yash Govilkar, Akasha Tigalappanavara, Shi-Xiong Zhang, Sambit Sahu, Viktor Prasanna. "SPARC-RAG: Adaptive Sequential-Parallel Scaling with Context Management for Retrieval-Augmented Generation." EMNLP 2026, Main Conference (accepted).

Gangda Deng, Zhaoling Chen, Zhongming Yu, Haoyang Fan, Yuhong Liu, **Yuxin Yang**, Dhruv Parikh, Rajgopal Kannan, Le Cong, Mengdi Wang, Qian Zhang, Viktor Prasanna, Xiangru Tang, Xingyao Wang. "SWE-Milestone: Evaluating AI Agents on Continuous Software Evolution." ICML 2026.

Abdulla Alshabanah, **Yuxin Yang**, Murali Annavaram. "SAGERec: Sampling and Gating for Enhanced Long-Tail Item Recommendations." WSDM 2026.

Gangda Deng*, **Yuxin Yang***, Ömer Faruk Akgül*, Hanqing Zeng, Yinglong Xia, Rajgopal Kannan, Viktor Prasanna. "Training Diverse Graph Experts for Ensembles: A Systematic Empirical Study." Preprint, 2025.

Ömer Faruk Akgül*, Feiyu Zhu*, **Yuxin Yang**, Rajgopal Kannan, Viktor Prasanna. "RECIPE-TKG: From Sparse History to Structured Reasoning for LLM-based Temporal Knowledge Graph Completion." EACL 2026.

Yuxin Yang, Hongkuan Zhou, Rajgopal Kannan, Viktor Prasanna. "Towards Ideal Temporal Graph Neural Networks: Evaluations and Conclusions after 10,000 GPU Hours." VLDB 2025.

Yuxin Yang, Yitao Liang, and Muhan Zhang. "PA-GNN: Parameter-Adaptive Graph Neural Networks." ICML DyNN Workshop 2022, oral presentation.

COMPUTER AND LANGUAGE SKILLS

- **Programming**: Python, C/C++, JavaScript
- **Tools**: PyTorch, MATLAB, CMake, LabVIEW, Linux, Git, LaTeX
- **Language**: TOEFL 114/120 (Reading 30, Listening 29, Speaking 25, Writing 30); GRE 327/340+3.5/6.0 (Verbal 157, Quantitative 170, Analytical Writing 3.5)