

Xuan (Lily) Yang

xuan.yang@duke.edu | xuanyang19.github.io | Durham, NC, USA

EDUCATION

Duke University

PhD in Computer Science (expected 2027); **Advisor: Prof. Jian Pei**

Durham, NC, USA

Sep. 2023 – Present

Zhejiang University

M.S. in Computer Science; **GPA: 3.98/4.00**

Hangzhou, China

Sep. 2020 – Mar. 2023

Zhejiang University

B.S. in Digital Media Technology; **GPA: 3.96/4.00**

Hangzhou, China

Sep. 2016 – Jun. 2020

SELECTED PUBLICATIONS

Data-Centric AI: Data Valuation & Curation

1. **Xuan Yang**, Hsi-Wen Chen, Ming-Syan Chen, Jian Pei. Local Shapley: Model-Induced Locality and Optimal Reuse in Data Valuation. **VLDB 2026**. [\[code\]](#)
2. **Xuan Yang**, Hsi-Wen Chen, Ming-Syan Chen, Jian Pei. D-Shap: Dynamic Shapley Computation. *Manuscript under review*, 2026.

LLM Reasoning & Agent Evaluation

3. **Xuan Yang**, Furong Jia, Roy Xie, Xiong Xi, Hengwei Bian, Jian Li, Monica Agrawal. Batch-of-Thought: Cross-Instance Learning for Enhanced LLM Reasoning. **EMNLP 2026 Main Conference**. [\[code\]](#)
4. M. Zhang, Y. Yang, Z. Jia, **Xuan Yang**, J. Pei, Y. Zang, X. Deng, X. Chen. MPCEval: A Benchmark for Multi-Party Conversation Generation. **KDD 2026 (Oral)**.
5. Shoufa Chen, Luyuan Wang, **Xuan Yang**, et al. (Meta AI · Duke · Stanford). TUA-Bench: A Benchmark for General-Purpose Terminal-Use Agents. *Manuscript under review*, 2026. [\[site\]](#) [\[code\]](#)

Graph Learning & Data Mining

6. Taoran Fang, Zhiqing Xiao, Chunping Wang, Jiarong Xu, **Xuan Yang**, Yang Yang. DropMessage: Unifying Random Dropping for Graph Neural Networks. **AAAI 2023 (Distinguished Paper Award)**.
7. **Xuan Yang**, Yang Yang, Chenhao Tan, Yinghe Lin, Zhengzhe Fu, Fei Wu, Yueting Zhuang. Unfolding and Modeling the Recovery Process after COVID Lockdowns. **Scientific Reports** (Nature Portfolio), 2023.
8. **Xuan Yang**, Yang Yang, Jintao Su, Yifei Sun, Shen Fan, Zhongyao Wang, Jun Zhan, Jingmin Chen. Who's Next: Rising Star Prediction via Diffusion of User Interest in Social Networks. **IEEE TKDE**, 2022.

RESEARCH EXPERIENCE

Data-Centric AI: Efficient Data Valuation

Apr. 2024 – Present

Duke University

- **Local Shapley (VLDB 2026)**: Formalized *model-induced locality* through support sets and designed LSMR, proven optimal in the number of model-training runs, plus an unbiased reuse-aware estimator; achieved up to $10^6 \times$ lower time-to- ϵ while preserving data-selection utility.
- **D-Shap**: Formulated dynamic data valuation as maintaining a player-by-task value matrix and designed structure-aware updates, enabling **millisecond** task updates and up to **three orders of magnitude** faster player updates at competitive fidelity.
- **Current research**: Extending data valuation to LLM data curation and agent memory, studying which training examples and interaction traces should be retained as models, tasks, and environments evolve.

LLM Reasoning & Agent Evaluation

Jun. 2025 – Present

Duke University

- **Batch-of-Thought (EMNLP 2026)**: Proposed a training-free framework that jointly processes related queries to share reasoning patterns and detect errors through cross-instance consistency; improved reasoning accuracy and calibration while reducing inference cost by up to **61%** across three model families and six benchmarks.
- **TUA-Bench**: Co-developed a benchmark with **Meta AI** and Stanford containing **120 tasks across five families** for evaluating general-purpose terminal-use agents in real environments with deterministic setup and execution-based scoring.

- **DropMessage (AAAI 2023, Distinguished Paper Award):** Unified random-dropping methods as operations on the message matrix of message-passing GNNs; established sample-variance reduction and information-diversity guarantees.
- **Graph Modeling (Nature Scientific Reports 2023):** Built a graph-learning framework over spatiotemporal electricity data to quantify sector-level post-lockdown recovery and policy impacts; featured on the front page of *Zhejiang Daily*.
- **RiseNet (TKDE 2022):** Modeled the dynamic diffusion of user interest with a multi-task GNN for rising-star prediction, validated on real-world Taobao data and recommendation system.

INDUSTRY EXPERIENCE

Google

Kirkland, WA

PhD SWE Intern, Google Cloud

May 2026 – Aug. 2026

- **Autonomous performance optimization:** Designed, built, and deployed an end-to-end multi-agent system for Google Cloud Service Control (**100M+ requests/s**), combining adaptive anomaly zooming, call-graph-preserving resource attribution, targeted code-fix generation, test/microbenchmark validation, persistent memory, and reviewer-feedback learning.
- **Efficient embedding model distillation:** Studied token-level representation stability under compression to guide model distillation, structured pruning, and representation reuse; reduced embedding computation by **45.94%** while achieving **0.95 NDCG@10**.

TikTok

Bellevue, WA

Research Intern, Risk Control Team

May 2025 – Apr. 2026

- Pioneered agentic workflows for risk identification and led the replacement of legacy pipelines with multi-agent systems, launched on E-commerce Risk Control: increased pre-GMV fraud recall by **26%** while cutting manual review effort by **96%**.
- Diagnosed reliability issues in deployed agents and redesigned the multi-agent architecture for traceability and explainability, cutting token cost by **60%**.

Alibaba Group

Hangzhou, China

Research Intern, Data Assets and Algorithm Team

Oct. 2020 – Dec. 2021

- Built a billion-scale Taocode recommendation dataset and deployed ranking models for Taobao product surfaces, boosting CTR by **2.3%** over the production baseline.

SELECTED HONORS

Distinguished Paper Award, AAAI	2023
Tencent Technology Excellence Scholarship	2022
Zhejiang University Excellent Postgraduate Student	2023

TEACHING EXPERIENCE

Teaching Assistant, Duke University: COMPSCI 526, Applied Machine Learning and Data Science (Fall 2026); COMPSCI 316, Databases (Fall 2024)

ACADEMIC SERVICE

Reviewer: NeurIPS 2026; ACM Transactions on Knowledge Discovery from Data (TKDD)

TECHNICAL SKILLS

Programming: Python (PyTorch, JAX), C++ (production), SQL
Research & Systems: data valuation and curation; dense retrieval and representation learning; LLM agents and evaluation; distributed-service profiling and microbenchmarking; large-scale data processing