

Tianyang Liu

 University of California, San Diego, La Jolla, CA, USA  (858) 531-7715  til040@ucsd.edu
 Personal Profile  GitHub  LinkedIn  Google Scholar  Twitter

EDUCATION

University of California, San Diego <i>Ph.D. in Computer Science</i> Advisor: Julian McAuley	San Diego, CA, USA 2024–Present
University of California, San Diego <i>M.Sc. in Computer Science</i> Advisors: Zhiting Hu, Julian McAuley	San Diego, CA, USA 2022–2024 GPA: 4.00/4.00
Wuhan University <i>B.Eng. in Software Engineering</i> Advisors: Peng Liang, Chong Wang	Wuhan, Hubei, China 2018–2022 GPA: 3.85/4.00

WORK EXPERIENCE

Snowflake <i>Part-time Research Scientist</i>	Remote Sep 2025–March 2026
Amazon AWS, Developer Q Team <i>Applied Scientist Intern</i>	Santa Clara, CA, USA June 2025– Sep 2025
NVIDIA <i>Summer Intern</i>	Santa Clara, CA, USA April 2024–Sep 2024

RESEARCH INTERESTS

I am interested in understanding the capabilities and limitations of LLMs and AI agents, including their behavior across diverse real-world settings. A core part of this is developing evaluations that go beyond narrow benchmark metrics and better reflect practical utility and user experience.

PUBLICATIONS (* EQUAL CONTRIBUTION)

----- 2026 -----

- [1] **Decompose to Compose: Multidimensional Evaluation of LLM-Generated UIs**
Tianyang Liu, Xiaoyang Wang, Zijian Wang, Murali Krishna Ramanathan, Varun Kumar
Submitted to COLM 2026
- [2] **CocoaBench: Evaluating Unified Digital Agents in the Wild**
Shibo Hao*, Zhining Zhang*, Zhiqi Liang*, Tianyang Liu*, Zilong Wang*, Kun Zhou, Yuheng Zha, Qiyue Gao, Jixuan Chen, Zhoujun Cheng, Yu Wang, Feng Yao, and 8 more authors
arXiv preprint 2026; submitted to COLM 2026 [[arXiv](#)] [[blog](#)] [[code](#)] [[leaderboard](#)]
- [3] **Spider 2.0-AIFunc: Extending Real-World Text-to-SQL to AI-Native SQL Workflows**
Tianyang Liu, Canwen Xu, Fangyu Lei, Nikki Lijing Kuang, Jixuan Chen, Tao Yu, Julian McAuley, Zhewei Yao, Yuxiong He
Submitted to COLM 2026

- [4] **BigCodeArena: Unveiling More Reliable Human Preferences in Code Generation via Execution**
Terry Yue Zhuo, Xiaolong Jin, Hange Liu, Juyong Jiang, **Tianyang Liu**, Chen Gong, Bhupesh Bishnoi and 33 more authors
Preprint 2025 [[arXiv](#)][[arena](#)] [[code](#)] [[collections](#)]
- [5] **You Don't Know Until You Click: Automated GUI Testing for Production-Ready Software Evaluation**
Yutong Bian, Xianhao Lin, Yupeng Xie, **Tianyang Liu**, Mingchen Zhuge, Siyuan Lu, Haoming Tang, Jinlin Wang, Jiayi Zhang, Jiaqi Chen, Xiangru Tang, Yongxin Ni, Sirui Hong, Chenglin Wu
Preprint 2025 [[arXiv](#)] [[code](#)]
- [6] **Revisiting Reinforcement Learning for LLM Reasoning from A Cross-Domain Perspective**
Zhoujun Cheng*, Shibo Hao*, **Tianyang Liu***, Fan Zhou, Yutao Xie, Feng Yao, Yuexin Bian, Yonghao Zhuang, Nilabjo Dey, Yuheng Zha, Yi Gu, Kun Zhou, and 12 more authors
NeurIPS 2025 [[arXiv](#)] [[blog](#)] [[code](#)] [[dataset](#)] [[model](#)]
- [7] **Decentralized Arena via Collective LLM Intelligence: Building Automated, Robust, and Transparent LLM Evaluation for Numerous Dimensions**
Yanbin Yin, Zhen Wang, Kun Zhou, Xiangdong Zhang, Shibo Hao, Yi Gu, Jieyuan Liu, Somanshu Singla, **Tianyang Liu**, Eric P. Xing, Zhengzhong Liu, Haojian Jin, Zhiting Hu
Preprint 2025 [[arXiv](#)] [[blog](#)] [[leaderboard](#)]
- [8] **Code to Think, Think to Code: A Survey on Code-Enhanced Reasoning and Reasoning-Driven Code Intelligence in LLMs**
Dayu Yang*, **Tianyang Liu***, Daoan Zhang*, Antoine Simoulin, Xiaoyi Liu, Yuwei Cao, Zhaopu Teng, Xin Qian, Grey Yang, Jiebo Luo, Julian McAuley
EMNLP 2025 (main) [[arXiv](#)] [[project](#)]
- [9] **Symbolic Representation for Any-to-Any Generative Tasks**
Jiaqi Chen, Xiaoye Zhu, Yue Wang, **Tianyang Liu**, Xinhui Chen, Ying Chen, Chak Tou Leong, Yifei Ke, Joseph Liu, Yiwen Yuan, Julian McAuley, Li-jia Li
CVPR 2025 [[arXiv](#)]
- [10] **Imitate Before Detect: Aligning Machine Stylistic Preference for Machine-Revised Text Detection**
Jiaqi Chen*, Xiaoye Zhu*, **Tianyang Liu***, Ying Chen, Xinhui Chen, Yiwen Yuan, Chak Tou Leong, Zuchao Li, Long Tang, Lei Zhang, Chenyu Yan, Guanghao Mei, Jie Zhang, Lefei Zhang
AAAI 2025 (Oral) [[arXiv](#)] [[project](#)] [[code](#)] [[demo](#)]

- [11] **Dynamic Rewarding with Prompt Optimization Enables Tuning-free Self-Alignment of Language Models**
Somanshu Singla*, Zhen Wang*, **Tianyang Liu**, Abdullah Ashfaq, Zhiting Hu, Eric P. Xing
EMNLP 2024 [[arXiv](#)] [[code](#)]
- [12] **LLM Reasoners: New Evaluation, Library, and Analysis of Step-by-Step Reasoning with Large Language Models**
Shibo Hao*, Yi Gu*, Haotian Luo*, **Tianyang Liu**, Xiyang Shao, Xinyuan Wang, Shuhua Xie, Haodi Ma, Adithya Samavedhi, Qiyue Gao, Zhen Wang, Zhiting Hu
COLM 2024 & ICLR 2024 Workshop on LLM Agents [[arXiv](#)] [[blog](#)] [[Github](#)]
- [13] **StarCoder 2 and The Stack v2: The Next Generation**
Anton Lozhkov, Raymond Li, Loubna Ben Allal, Federico Cassano, Joel Lamy-Poirier, Nouamane Tazi, Ao

Tang, Dmytro Pykhtar, Jiawei Liu, Yuxiang Wei, **Tianyang Liu**, Max Tian, and 54 more authors
Technical Report 2024 [[arXiv](#)] [[blog](#)] [[models](#)]

[14] **Rethinking Tabular Data Understanding with Large Language Models**

Tianyang Liu, Fei Wang, Muhao Chen

NAACL 2024 [[arXiv](#)] [[code](#)]

[15] **RepoBench: Benchmarking Repository-Level Code Auto-Completion Systems**

Tianyang Liu, Canwen Xu, Julian McAuley

ICLR 2024 [[arXiv](#)] [[code](#)] [[OpenReview](#)]

----- 2023 -----

[16] **ToolkenGPT: Augmenting Frozen Language Models with Massive Tools via Tool Embeddings**

Shibo Hao, **Tianyang Liu**, Zhen Wang, Zhiting Hu

NeurIPS 2023 (Oral) [[arXiv](#)] [[code](#)]

Best Paper Award at SoCal NLP 2023

[17] **Architecture Decisions in AI-based Systems Development: An Empirical Study**

Beiqi Zhang, **Tianyang Liu**, Peng Liang, Chong Wang, Mojtaba Shahin, Jiaxin Yu

SANER 2023

[18] **RoseMatcher: Identifying the Impact of User Reviews on App Updates**

Tianyang Liu, Chong Wang, Kun Huang, Peng Liang, Beiqi Zhang, Maya Daneva, Marten van Sinderen

Information and Software Technology 2023

----- 2022 -----

[19] **The Role of User Reviews in App Updates: A Preliminary Investigation on App Release Notes**

Chong Wang*, **Tianyang Liu***, Peng Liang, Maya Daneva, Marten van Sinderen

APSEC 2022

 **PROFESSIONAL SERVICES**

Professional Service

Reviewer ACL ARR, NeurIPS, ICLR, ICML, COLM, AAAI, AISTATS, NLPCC
2023 – 2026

Invited Speaker LlamaIndex Seminar on Tabular Data Understanding

 **AWARDS**

- Honorable Mention for Excellence in Research (CSE Department Award), UCSD, 2024
- Best Paper Award, SoCal NLP Symposium, 2023
- Second Level Scholarship, Wuhan University, 2021