Tianle Li

tianleli@berkeley.edu — +1 (657) 395-9520 — www.linkedin.com/in/tianleli — Personal Website

RESEARCH INTEREST

Intersection of Large Model Evaluation and Post-Training focusing on improving model capability and reliability.

EDUCATION

University of California-Berkeley,   2021 — 05/2025
Bachelor of Science: Electrical Engineering and Computer Science   Cumulative GPA: 3.75/4.0
Relevant Courses: Machine Learning, Deep Neural Networks, Deep Reinforcement Learning, Computer Architecture, Convex Optimization, Probability Theories

PUBLICATION

G Google Scholar

Under Review

1. **Tianle Li***, Wei-Lin Chiang*, Evan Frick, Lisa Dunlap, Tianhao Wu, Banghua Zhu, Joseph E. Gonzalez, Ion Stoica. *From Crowdsourced Data to High-Quality Benchmarks: Arena-Hard and BenchBuilder Pipeline* 2024. arXiv: 2406.11939 [cs.LG]. https://arxiv.org/abs/2406.11939.

2. Evan Frick*, Connor Chen*, Joseph Tennyson*, **Tianle Li***, Wei-Lin Chiang*, Anastasios N. Angelopoulos*, Ion Stoica. *Prompt-to-Leaderboard* 2025. arXiv: 2502.14855 [cs.LG]. https://arxiv.org/abs/2502.14855.

3. Evan Frick, **Tianle Li**, Connor Chen, Wei-Lin Chiang, Anastasios N. Angelopoulos, Jiantao Jiao, Banghua Zhu, Joseph E. Gonzalez, Ion Stoica. *How to Evaluate Reward Models for RLHF* 2024. arXiv: 2410.14872 [cs.LG]. https://arxiv.org/abs/2410.14872.

4. Lucas Spangher, **Tianle Li**, William F. Arnold, Nick Masiewicki, Xerxes Dotiwalla, Rama Parusmathi, Peter Grabowski, Eugene Ie, Dan Gruhl. *Project MPG: towards a generalized performance benchmark for LLM capabilities* 2024. arXiv: 2410.22368 [cs.SE]. https://arxiv.org/abs/2410.22368.

Conference Proceeding

5. Wei-Lin* Chiang, Lianmin* Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, **Tianle Li**, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E. Gonzalez, Ion Stoica. *Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference* in *Proceedings of the 41st International Conference on Machine Learning* (2024).

6. Lianmin Zheng*, Wei-Lin Chiang*, Ying Sheng, **Tianle Li**, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zhuohan Li, Zi Lin, Eric Xing, Joseph E. Gonzalez, Ion Stoica, Hao Zhang. *LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset* in *The Twelfth International Conference on Learning Representations* (2024). https://openreview.net/forum?id=BOfDKxfwt0.

7. Jonathan Pei, Zeeshan Patel, Karim El-Refai, **Tianle Li**. *SWAG: Storytelling With Action Guidance* in *Findings of the Association for Computational Linguistics: EMNLP 2024* (Association for Computational Linguistics, Nov. 2024), 14086–14106. https://aclanthology.org/2024.findings-emnlp.824.

TECHNICAL BLOG

8. **Tianle Li**, Wei-Lin Chiang, Yifan Song, Naman Jain, Lisa Dunlap, Dacheng Li, Evan Frick, Anastasios N. Angelopoulos. *Chatbot Arena Categories: Definitions, Methods, and Insights* Oct. 2024. https://blog.lmarena.ai/blog/2024/arena-category/.

9. **Tianle Li***, Anastasios Angelopoulos*, Wei-Lin Chiang*. *Does Style Matter? Disentangling style and substance in Chatbot Arena* Aug. 2024. https://blog.lmarena.ai/blog/2024/style-control/.

* means equal contribution.

10. Evan Frick*, Peter Jin*, **Tianle Li***, Karthik Ganesan, Jian Zhang, Jiantao Jiao, Banghua Zhu. *Athene-70B: Redefining the Boundaries of Post-Training for Open Models* July 2024. https://huggingface.co/Nexusflow/Athene-70B.

11. **Tianle Li**, Wei-Lin Chiang, Lisa Dunlap. *Introducing Hard Prompts Category in Chatbot Arena* May 2024. https://lmsys.org/blog/2024-05-17-category-hard/.

12. Lisa Dunlap, Evan Frick, **Tianle Li**, Isaac Ong, Joseph E. Gonzalez, Wei-Lin Chiang. *What's up with Llama 3? Arena data analysis* May 2024. https://lmsys.org/blog/2024-05-08-llama3/.

EXPERIENCE

Berkeley Sky Computing Lab   Berkeley, CA
Research Lead   07/2023 — Present

- **Chatbot Arena**: An open platform for evaluating LLMs via human preference with over 1M monthly visits.
- Lead research on automatic offline evaluation, human preference research, and data pipeline and analysis.
- Created Style Control, Hard-Prompts, Math, Instruction-Following, and Creative Writing, category leaderboards on Chatbot Arena.

NexusFlow   Palo, Alto
Research Engineer   05/2024 — Present

- LLM post-training team, co-trained **Athene-70B** and **Athene-V2-Chat**.
- Lead targeted data curation and design model evaluation, and explores various RLHF and RLAIF methods.

AMD   San, José
Software Development Intern   05/2023 — 08/2023

- **Vitis Model Composer**: An AI powered digital signal processing and optimization library tool in MATLAB.
- Architected and developed a new example infrastructure for DSP algorithms in VMC 2023.2 release update.

Ingram Micro   Irvine, California
Software Engineer Intern   05/2022 — 08/2022

- Programmed an API deployment specification CI/CD pipeline, integrating local repo with proxy platform.
- Conducted 200+ code reviews of web API repositories to ensure compliance with REST Architecture.

TEACHING

EECS 127: Convex Optimization for Machine Learning   09/2023 — 05/2024
Teaching Assistant   Department of Electrical Engineering and Computer Sciences at UC Berkeley

- This upper division course offers the theories behind optimization models and their applications, ranging from machine learning and statistics to decision-making and control, with emphasis on numerically tractable problems, such as linear, quadratic, conic, or constrained least-squares optimization.

OPEN SOURCE PROJECT

FastChat *Core Contributor*   A system for training, serving, and evaluating LLM-based chatbots (37K+ stars).

Arena-Hard-Auto *Lead Contributor*   An automatic evaluation tool for instruction-tuned LLMs (650+ stars).

Athene-V2-Chat-72B *Core Contributor* Best open-weights LLM, on par with GPT-4o (5K+ Monthly Downloads).

Athene-70B *Co-Lead*   SOTA open-weights LLM post-trained from Llama-3 (8K+ Monthly Downloads).

ACADEMIC SERVICE

Served as an **ICLR 2025** Reviewer.

---

* means equal contribution.