

# Tianle Li

*My interests are training large models, building more capable and reliable AI, and solving superintelligence.*

## EDUCATION

### UNIVERSITY OF CALIFORNIA, BERKELEY

Berkeley, CA

**B.S. Electrical Engineering and Computer Science;** GPA: 3.7

2021-2025

Data Structure, Algorithms, Computer Architecture, Convex Optimization, Machine Learning, Deep Neural Networks, Deep Reinforcement Learning, Natural Language Processing.

Advisor: [Jon Stoica](#)

## EXPERIENCE

### xAI

PALO ALTO, CA

#### Member of Technical Staff

May 2025 - Present

- AI Researcher & Engineer: Grok 3.5 post-training.
- Creating generalizable reward models and developing novel reinforcement learning algorithms.
- Improving data mixtures by building data collection pipelines and developing data generation techniques.
- Designing and implementing large-scale model training frameworks.
- Collaborating with pre-training, reasoning, multimodal, applied, product efforts to push the frontiers of model capability.

### BERKELEY SKY COMPUTING LAB

BERKELEY, CA

#### Undergrad Researcher

JULY 2023 - May 2025

[Chatbot Arena](#) (Core Contributor)

- An open platform for evaluating LLMs by human preference with millions of monthly users.
- We privately tested GPT-4o, Grok 3, Gemini Flash and Pro, Meta Llama 3.2, and more.
- I lead research on human preference and data pipeline / analysis.
- I built the categories: Hard Prompt, Style Control, Instruction-Following, Math, Creative Writing, and more.
- I'm advising ongoing projects including Search Arena, PDFChat, User leaderboard, and more.

### NEXUSFLOW

PALO ALTO, CA

#### Machine Learning Engineer

MAY 2024 - May 2025

[Athene-V2-Chat-72B](#): co-trained the best open weight LLM; post-trained from Qwen 2.5-72b-chat.

- On par with GPT-4o and Claude 3.5 Sonnet on Chatbot Arena (Rank 5), surpass GPT-4o and Llama-3.1-405B on LiveCodeBench, Aider, GPQA, MATH, and more (Nov 2024).

[Athene-70B](#): co-trained the best open weight chat LLM post-trained using Llama-3-70b base model.

- Rank 8 on Chatbot Arena Overall, Rank 5 on Chatbot Arena Hard Prompt (July 2024).
- I co-trained the 70B reward model for aligning Athene using PPO using 32 H100 GPUs.
- I led data curation and evaluation, successfully improved Llama-3 on technical and multilingual queries.

[Starling-LM-7B-beta](#): the world's best 7B chat LLM post-trained from OpenChat.

- Best 7B model on Chatbot Arena, on par with Llama-2-70B and Vicuna-33B (Nov 2023). I led the evaluation part.

### GOOGLE AI

MOUNTAIN VIEW, CA

#### Student Researcher

FEB 2025 - May 2025

- Improve reasoning in LLMs. Our evaluation method was published in NAACL.

### AMD

San Jose, CA

#### Software Development Intern

MAY 2023 - August 2023

- Built Xilinx's Digital Signal Processing library.

## SELECTED PUBLICATION ([Google Scholar](#))

---

1. [From Crowdsourced Data to High-Quality Benchmarks: Arena-Hard and BenchBuilder](#) (ICML 2025)  
Tianle Li\*, Wei-Lin Chiang\*, Evan Frick, Lisa Dunlap, Tianhao Wu, Banghua Zhu, Joseph E. Gonzalez, Ion Stoica.
2. [Prompt-to-Leaderboard](#) (ICML 2025)  
Evan Frick\*, Connor Chen\*, Joseph Tennyson\*, Tianle Li\*, Wei-Lin Chiang\*, Anastasios N. Angelopoulos\*, Ion Stoica.
3. [How to Evaluate Reward Models for RLHF](#) (ICLR 2025)  
Evan Frick, Tianle Li, Connor Chen, Wei-Lin Chiang, Anastasios N. Angelopoulos, Jiantao Jiao, Banghua Zhu, Joseph E. Gonzalez, Ion Stoica.
4. [Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference](#) (ICML 2024)  
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.
5. [LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset](#) (ICLR 2024 Spotlight)  
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.
6. [Project MPG: towards a generalized performance benchmark for LLM capabilities](#) (NAACL 2025)  
Lucas Spangher, Tianle Li, William F. Arnold, Nick Masiewicki, Xerxes Dotiwalla, Rama Parusmathi, Peter Grabowski, Eugene Ie, Dan Gruhl.

## TECHNICAL BLOG

---

1. [Chatbot Arena Categories: Definitions, Methods, and Insights](#)  
Tianle Li, Wei-Lin Chiang, Yifan Song, Naman Jain, Lisa Dunlap, Dacheng Li, Evan Frick, Anastasios N. Angelopoulos.
2. [Does Style Matter? Disentangling style and substance in Chatbot Arena](#)  
Tianle Li\*, Anastasios Angelopoulos\*, Wei-Lin Chiang\*.
3. [Athene-70B: Redefining the Boundaries of Post-Training for Open Models](#)  
Evan Frick\*, Peter Jin\*, Tianle Li\*, Karthik Ganesan, Jian Zhang, Jiantao Jiao, Banghua Zhu.
4. [Introducing Hard Prompts Category in Chatbot Arena](#)  
Tianle Li, Wei-Lin Chiang, Lisa Dunlap.
5. [What's up with Llama 3? Arena data analysis](#)  
Lisa Dunlap, Evan Frick, Tianle Li, Isaac Ong, Joseph E. Gonzalez, Wei-Lin Chiang.
6. [Chatbot Arena: New models & Elo system update](#)  
Wei-Lin Chiang, Tianle Li, Joseph E. Gonzalez, Ion Stoica.
7. [Introducing Athene-V2: Advancing Beyond the Limits of Scaling with Targeted Post-training](#)  
The Nexusflow Team.

## OPEN SOURCE PROJECT

---

1. [FastChat](#) (Contributor) 38K+ Stars  
An open infra for training, serving, and evaluating large language models. Release repo for Vicuna and Chatbot Arena.
2. [Arena-Hard-Auto](#) (Lead) 800+ Stars  
An automatic evaluation tool for instruction-tuned LLMs, highly correlated with Chatbot Arena.

## TEACHING

---

### EECS 127: Convex Optimization for Machine Learning

UC BERKELEY

#### Teaching Assistant

SEPTEMBER 2023 - MAY 2024

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.

\* means equal contribution.

## PERSONAL PROJECT

---

### Speaking in Chess

[https://github.com/CodingWithTim/Speaking\\_in\\_Chess](https://github.com/CodingWithTim/Speaking_in_Chess)

- Pretrained and supervised fine-tuned GPT-2 128M on over 20 million chess games using a custom chess tokenizer.
- Evaluated 6 RL strategies, including 3 novel algorithms: Fictitious Self-Play, Past-Present Q-Iteration, Funnel Searching.
- Achieve over 95% draw rate against StockFish 3000 elo chess engine with gameplay accuracy averaging 90%.