

# Kaiyan Zhang

📍 Beijing    ✉ zhang-ky22@mails.tsinghua.edu.cn    🌐 iseesaw.github.io    📧 iseesaw

## Vision towards AGI


- **Self-Evolving Intelligence.** Pursue continual/online RL with streaming data, retrieval-augmented memory, and meta-learning for rapid adaptation. This follows the experience-centric view that agents should learn from ongoing interaction and leverage past experience to solve previously unseen tasks.
- **Scalable Environments for RL.** Build RL environments for large reasoning/agent/coding models that deliver dynamic feedback by combining static datasets, rule- and code-based checks, interactive games/simulators, and model-based evaluators, enabling experience-driven and scalable lifelong reinforcement learning.
- **Role of Multi-Agent Systems.** AGI may emerge as a self-evolving system rather than a single model. Scalable multi-agent architectures can both amplify reasoning (via specialization, coordination, and tool use) and accelerate foundation-model learning through oversight and self-improvement.

## Education

- Tsinghua University** *PhD in Electronic Engineering* *Sept 2022 – Jun 2026*
- Advisor: [Bowen Zhou](#) 📧
  - **Interest:** Large Language Models, Test-Time Scaling, Reinforcement Learning, Multi-Agent Systems
- Harbin Institute of Technology** *MS in Computer Science* *Sept 2020 – Jun 2022*
- Advisor: [Weinan Zhang](#) 📧 & [Ting Liu](#) 📧
- Harbin Institute of Technology** *BS in Computer Science* *Sept 2016 – Jun 2020*

## Highlights


- TTRL: Test-Time Reinforcement Learning.** (700+ Stars on [GitHub](#) 📧) *Preprint*
- Yuxin Zuo\*, **Kaiyan Zhang\*** (Project Lead), Shang Qu, Li Sheng, Xuekai Zhu, Biqing Qi, Youbang Sun, Ganqu Cui, Ning Ding, and Bowen Zhou.
  - As project lead, proposed a label-free test-time RL method that uses majority-vote rewards—consistently beating test-time scaling upper bounds under equal compute (e.g., about 211% pass@1 gain on AIME-24 for Qwen-2.5-Math-7B using only unlabeled test data).
- MARTI: A Framework for Multi-Agent LLM Systems Reinforced Training and Inference.** (About 200 Stars on [GitHub](#) 📧) *Preprint*
- **Kaiyan Zhang\***, Runze Liu\*, Xuekai Zhu\*, Kai Tian\*, Sihang Zeng\*, Guoli Jia\*, Yuchen Fan\*, Xingtai Lv\*, Yuxin Zuo\*, Che Jiang\*, Jianyu Wang, Yuru Wang, Ruotong Zhao, Ermo Hua, Shijie Wang, Junqi Gao, Xinwei Long, Youbang Sun, Zhiyuan Ma, Ganqu Cui, Lei Bai, Ning Ding, Biqing Qi, Bowen Zhou.
  - Project lead for an open-source multi-agent RL/inference stack: graph workflows (debate/chain/MoA), heterogeneous agents, async tool-use and workflows, and RL plugins (PPO/GRPO/REINFORCE++/TTRL); preliminary results show multi-agent RL surpasses single-agent under the same budget.
- SSRL: Self-Search Reinforcement Learning** *Preprint*
- Yuchen Fan\*, **Kaiyan Zhang\*** (Project Lead), Heng Zhou\*, Yuxin Zuo, Yanxu Chen, Yu Fu, Xinwei Long, Xuekai Zhu, Che Jiang, Yuchen Zhang, Li Kang, Bingning Wang, Lei Bai, Ning Ding, Bowen Zhou.
  - Lead project defining an inference-time scaling law for agentic search and a “self-search” RL algorithm; exploring the capability of policy model as textual world model; excellent sim-to-real evaluations.
- OpenPRM: Building Open-domain Process-based Reward Models with Preference Trees.** *ICLR 2025*
- **Kaiyan Zhang**, Jiayuan Zhang, Haoxin Li, Xuekai Zhu, Ermo Hua, Xingtai Lv, Ning Ding, Biqing Qi, Bowen Zhou.
  - Extend outcome-based RMs to process-based via sentence-level preference trees derived from ORMs; unified pairwise training yields +3–5% on RewardBench and better scaling for inference-time compute than ORMs in open-domain tasks (best-of- $N$ ).

**UltraMedical: Building Specialized Generalists in Biomedicine.** (More than 30K downloads on [Huggingface](#) ) *NeurIPS 2024 (Spotlight)*

- **Kaiyan Zhang**, Sihang Zeng, Ermo Hua, Ning Ding, Zhang-Ren Chen, Zhiyuan Ma, Haoxin Li, Ganqu Cui, Biqing Qi, Xuekai Zhu, Xingtai Lv, Hu Jinfang, Zhiyuan Liu, Bowen Zhou.
- Released a 410K-instruction medical dataset (with about 100K preference pairs) and open Llama-3-based medical LLMs (8B/70B) trained with SFT + preference learning with more 100+ A100 GPUs.

## Selected Publications

---

\*indicates co-first authors, full paper list on [Google Scholar](#) 

- **Kaiyan Zhang**, Jianyu Wang, Ermo Hua, Biqing Qi, et al. *CoGenesis: A Framework Collaborating Large and Small Language Models for Secure Context-Aware Instruction Following*. (ACL 2024)
- **Kaiyan Zhang**, Jianyu Wang, Ning Ding, Biqing Qi, Ermo Hua, et al. *Fast and Slow Generating: An Empirical Study on Large and Small Language Models Collaborative Decoding*. (ICML@MAS 2025)
- Biqing Qi\*, **Kaiyan Zhang\***, Kai Tian, Haoxiang Li, Zhang-Ren Chen, Sihang Zeng, Ermo Hua, et al. *Large Language Models as Biomedical Hypothesis Generators: A Comprehensive Evaluation*. (COLM 2024)
- **Kaiyan Zhang**, Ning Ding, Biqing Qi, Xuekai Zhu, Xinwei Long, Bowen Zhou. *CRaSh: Clustering, Removing, and Sharing Enhance Fine-tuning without Full Large Language Model*. (EMNLP 2023)
- **Kaiyan Zhang\***, Jianyu Wang\*, Xiang Xu\*, Runze Liu, Kai Tian, Jiayuan Zhang, Youbang Sun, Biqing Qi, et al. *ReSpecT: Reinforced Speculative Thinking for Large Reasoning Models*. (UnderReview)
- **Kaiyan Zhang**, Biqing Qi, Bowen Zhou. *Towards Building Specialized Generalist AI with System 1 and System 2 Fusion*. (UnderReview)
- Sihang Zeng\*, Kai Tian\*, **Kaiyan Zhang\***, Yuru wang, Junqi Gao, Runze Liu, Sa Yang, Jingxuan Li, Xinwei Long, et al. *ReviewRL: Towards Automated Scientific Review with RL*. (UnderReview)
- Ermo Hua, Biqing Qi, **Kaiyan Zhang**, Yue Yu, Ning Ding, Xingtai Lv, Kai Tian, Bowen Zhou. *Intuitive Fine-Tuning: Towards Simplifying Alignment into a Single Process*. (ACL 2025)
- Ermo Hua, Che Jiang, Xingtai Lv, **Kaiyan Zhang**, Ning Ding, Youbang Sun, Biqing Qi, et al. *Fourier Position Embedding: Enhancing Attention's Periodic Extension for Length Generalization*. (ICML 2025)
- Xuekai Zhu, Daixuan Cheng, Hengli Li, **Kaiyan Zhang**, Ermo Hua, Xingtai Lv, Ning Ding, Zhouhan Lin, Zilong Zheng, Bowen Zhou. *How to Synthesize Text Data without Model Collapse?* (ICML 2025)
- Biqing Qi, Junqi Gao, **Kaiyan Zhang**, Dong Li, Jianxing Liu, Ligang Wu, Bowen Zhou. *SMR: State Memory Replay for Long Sequence Modeling*. (ACL 2024)
- Xuekai Zhu, Biqing Qi, **Kaiyan Zhang**, Xinwei Long, Zhouhan Lin, Bowen Zhou. *PaD: Program-aided Distillation Specializes Large Models in Reasoning*. (NAACL 2024)
- Xingtai Lv\*, Ning Ding\*, **Kaiyan Zhang**, Ermo Hua, Ganqu Cui, Bowen Zhou. *Scalable Efficient Training of Large Language Models with Low-dimensional Projected Attention*. (EMNLP 2024)

## Honors & Awards

---

- Tsinghua University 2024 Huiyan First-Class Scholarship (Top 5%)
- Harbin Institute of Technology Outstanding Master's Thesis Award, Class of 2022 (Top 5%)
- Lenovo Scholarship (Top 5%)
- CETC 14th Research Institute (NRIET) Guorui Scholarship (Top 5%)
- CCKS 2021: Medical Dialogue Entity Generation (Finals: 3rd Place, Team Leader)
- 2019 Future Cup Collegiate AI Challenge (Northeast Region: 2nd Place, Team Leader)

## Services

---

**Reviewer:** ICLR, NeurIPS, ACL, COLM, EMNLP, AACL, ICCV