

Runpeng Dai

919-9515584 | Website
runpeng@unc.edu

Researcher specializing in **Reinforcement Learning** and **Efficient LLMs**, with expertise spanning classical RL, RLVR, and large-scale RL training using Verl.

EDUCATION

University of North Carolina at Chapel Hill
PhD Candidate in Biostatistics

Chapel Hill, NC
Aug 2023 - Exp. May 2027

INDUSTRY EXPERIENCE

Apple Ads — Machine Learning Research Intern May 2026 - Aug 2026

- Developed CoGR, a co-evolving RL framework that jointly optimizes query- and item-side LLM generators for ads retrieval.
- Outperformed strong internal and 10 external baselines on App Store retrieval.

Tencent AI Lab (Seattle) — Research Intern May 2025 - Aug 2025

- Developed CDE, a curiosity-driven RLVR framework that improves exploration via actor uncertainty and critic disagreement.
- Improved PPO and GRPO performance on AIME by 3 points, with the resulting paper accepted to ICLR 2026.

Baidu — LLM Research Intern May 2024 - July 2024

- Developed a novel sensitivity metric to identify vulnerable dimensions in LLMs and VLMs, with applications to model merging and quantization.

SELECTED PUBLICATIONS (see full list on Google Scholar)

Reinforcement Learning

- **Runpeng Dai**, Kaili Huang, Changsung Kang, Ciya Liao. [It Takes Two to Match: Co-Evolving Generative Retriever with Reinforcement Learning](#)
ArXiv 2026. Work done at Apple Ads.
- **Runpeng Dai**, Linfeng Song, Haolin Liu, Zhenwen Liang, Dian Yu, Haitao Mi, Zhaopeng Tu, ..., Dong Yu. [CDE: Curiosity-Driven Exploration for Efficient Reinforcement Learning in Large Language Models](#)
ICLR 2026. Work done at Tencent AI Lab.
- **Runpeng Dai**, Tong Zheng, Run Yang, Kaixian Yu, Hongtu Zhu. [R1-RE: Cross-Domain Relation Extraction with RLVR](#)
ACL 2026 main (Oral)
- Guohao Shen, **Runpeng Dai**, Guojun Wu, Shikai Luo, Chengchun Shi, Hongtu Zhu. [Deep Distributional Learning with Non-crossing Quantile Network](#)
ArXiv 2025, Major revision, JASA
- **Runpeng Dai**, Jianing Wang, Fan Zhou, Shikai Luo, Zhiwei Qin, Chengchun Shi, Hongtu Zhu. [Causal Deepsets for Off-policy Evaluation under Spatial or Spatio-temporal Interferences](#)
ArXiv 2024, Major revision, JRSSB

Efficient LLM

- **Runpeng Dai**, Tong Zheng, Rui Liu, Chengsong Huang, Hongtu Zhu. [Small RL Controller, Large Language Model: RL-Guided Adaptive Sampling for Test-Time Scaling](#)
EMNLP 2026 main
- Tong Zheng*, Chengsong Huang*, **Runpeng Dai***, Yun He, Rui Liu, Xin Ni, Huiwen Bao, Kaishen Wang, Hongtu Zhu, Jiaxin Huang, Furong Huang, Heng Huang. [Parallel-Probe: Towards Efficient Parallel Thinking via 2D Probing](#)
ICML 2026
- Tong Zheng, Hongming Zhang, Wenhao Yu, Xiaoyang Wang, **Runpeng Dai** et al. [Parallel-R1: Towards Parallel Thinking via Reinforcement Learning](#)
ICLR 2026

LLM Reliability

- **Runpeng Dai**, Run Yang, Fan Zhou, Hongtu Zhu. [Breach in the Shield: Unveiling the Vulnerabilities of Large Language Models](#)
EACL 2026 main