

Yingru Li

szrlee@gmail.com • +65 8747 5669

Curriculum Vitae

EDUCATION

The Chinese University of Hong Kong (CUHK) 2025

Ph.D. in Computer Science and Information Engineering.

Advisor: Zhi-Quan (Tom) Luo.

Committee: Xinyun Chen, Baoxiang Wang, John C.S. Lui, Benjamin Van Roy (Stanford & DeepMind).

Huazhong University of Science and Technology (HUST) 2020

B.E./M.S. in Computer Science. Ranked 1st/134 overall; 1st/26 in Software and Theoretical CS.

PROFESSIONAL EXPERIENCE

xAI - SpaceX Singapore
Member of Technical Staff December 2025 – Present

- **Core contribution to large-scale (“big-run”) reinforcement learning for the Grok models.**
 - *Science* — stability, efficiency, and the core algorithms, recipes, optimizer and scaling laws; helped identify and resolve critical training instabilities that benefit every stage of the RL pipeline, contributing to one of the longest stable RL runs to date.
 - *Long Horizon* — contributed to training recipes for long-horizon agentic RL.
 - *Infrastructure* — built and helped scale these training recipes in the large-scale distributed RL systems.

ByteDance Singapore
LLM Algorithm Research Scientist 2025 – December 2025

- **Led research on RL training stability and efficiency for LLM reasoning and agents.**
 - Diagnosed and formalized the *RL Training Collapse* phenomenon, tracing instability to implicit off-policyyness caused by the training–inference mismatch; developed Trust Region Masking (ICML’26), Optimal Token Baseline (ICML’26), Dynamic Vocabulary Pruning, and a simple LR-scheduling fix.
 - Co-led project on end-to-end RL training for multi-turn tool-integrated reasoning (SimpleTIR, ICLR’26) and Triton kernel code generation agent (Dr. Kernel, ICML’26).

Tencent AI & Robotics X Shenzhen, China
Student Research Staff Member, Agent Center 2019 – 2022

- Led research for *scalable exploration in deep RL*: HyperAgent (ICML’24) and HyperDQN (ICLR’22), and *stable off-policy policy optimization*: Divergence-Augmented Policy Optimization (NeurIPS’19).

Microsoft Research Beijing, China
Research Intern, Theory Center 2016

- Research on combinatorial optimization and reinforcement / bandit learning algorithms.

PUBLICATIONS

Full list at [Google Scholar](#). ★ denotes equal contribution; † denotes corresponding author.

Peer-Reviewed Conference & Journal Publications

1. **The Optimal Token Baseline: Variance Reduction for Long-Horizon LLM-RL**
Yingru Li, Jiawei Xu, Ziniu Li, Jiakai Liu, Wei Liu, Yuxuan Tong, Longtao Zheng, Zhenghai Xue, Yaxiang Zhang, Tianle Cai, Ge Zhang, Qian Liu, Baoxiang Wang. *ICML 2026*.
2. **Trust Region Masking for Long-Horizon LLM Reinforcement Learning**
Yingru Li, Jiakai Liu, Jiawei Xu, Yuxuan Tong, Ziniu Li, Qian Liu, Baoxiang Wang. *ICML 2026*.
3. **Optimistic Thompson Sampling for No-Regret Learning in Unknown Games**
Yingru Li, Liangqi Liu, Wenqiang Pu, Hao Liang, Zhi-Quan Luo. *IEEE Transactions on Signal Processing*, 2026.

4. **Dr. Kernel: Reinforcement Learning Done Right for Triton Kernel Generations**
Wei Liu, Jiawei Xu, **Yingru Li**, Longtao Zheng, Tianjian Li, Qian Liu, Junxian He. *ICML 2026*.
5. **Harnessing Uncertainty: Entropy-Modulated Policy Gradients for Long-Horizon LLM Agents**
Jiawei Wang, Jiakai Liu, Yuqian Fu, **Yingru Li**, Xintao Wang, Yuan Lin, Yu Yue, Lin Zhang, Yang Wang, Ke Wang. *ICML 2026*.
6. **SimpleTIR: End-to-End Reinforcement Learning for Multi-Turn Tool-Integrated Reasoning**
Zhenghai Xue, Longtao Zheng, Qian Liu, **Yingru Li**, Xiaosen Zheng, Zejun Ma, Bo An. *ICLR 2026*.
7. **Scalable Exploration via Ensemble++**
Yingru Li, Jiawei Xu, Baoxiang Wang, Zhi-Quan Luo. *NeurIPS 2025*.
8. **Owl: Optimized Workforce Learning for General Multi-Agent Assistance in Real-World Task Automation**
Meng kang Hu, Yuhang Zhou, Wendong Fan, Yuzhou Nie, Bowei Xia, Tao Sun, Ziyu Ye, Zhaoxuan Jin, **Yingru Li**, Qiguang Chen, Zeyu Zhang, Yifeng Wang, Qianshuo Ye, Bernard Ghanem, Ping Luo, Guohao Li. *NeurIPS 2025*.
9. **Q-Star Meets Scalable Posterior Sampling: Bridging Theory and Practice via HyperAgent**
Yingru Li, Jiawei Xu, Lei Han, Zhi-Quan Luo. *ICML 2024*.
10. **Prior-Dependent Analysis of Posterior Sampling RL with Function Approximation**
Yingru Li, Zhi-Quan Luo. *AISTATS 2024*.
11. **HyperDQN: A Randomized Exploration Method for Deep Reinforcement Learning**
Ziniu Li, **Yingru Li**[†], Yushun Zhang, Tong Zhang, Zhi-Quan Luo. *ICLR 2022*.
12. **Divergence-Augmented Policy Optimization**
Qing Wang*, **Yingru Li***, Jiechao Xiong, Tong Zhang. *NeurIPS 2019*.

Preprints

1. **When Speed Kills Stability: Demystifying RL Collapse from the Training-Inference Mismatch**
Jiakai Liu*, **Yingru Li***[†], Yuqian Fu, Jiawei Wang, Qian Liu, Yu Shen. *Technical Blog*, 2025.
2. **Beyond Precision: Training-Inference Mismatch is an Optimization Problem and Simple LR Scheduling Fixes It**
Yaxiang Zhang*, **Yingru Li***[†], Jiakai Liu, Jiawei Xu, Ziniu Li, Qian Liu, Haoyuan Li. *arXiv 2026*.
3. **Learning an Opponent-Aware Anti-Jamming Strategy via Online Convex Optimization**
Liangqi Liu, Wenqiang Pu, **Yingru Li**, Zhi-Quan Luo. *Under review (minor revision), IEEE Transactions on Signal Processing; arXiv 2026*.
4. **Taming the Tail: Stable LLM Reinforcement Learning via Dynamic Vocabulary Pruning**
Yingru Li, Jiawei Xu, Jiakai Liu, Yuxuan Tong, Ziniu Li, Tianle Cai, Ge Zhang, Qian Liu, Baoxiang Wang. *arXiv 2025*.
5. **A Note on Hybrid Online RL and Imitation Learning for LLMs: Formulations and Algorithms**
Yingru Li, Ziniu Li, Jiakai Liu. *arXiv 2025*.
6. **Logit Dynamics in Softmax Policy Gradient Methods**
Yingru Li. *arXiv 2025*.
7. **Scaling Flaws of Verifier-Guided Search in Mathematical Reasoning**
Fei Yu, **Yingru Li**, Benyou Wang. *arXiv 2025*.
8. **Uncertainty-Aware Search and Value Models: Mitigating Search Scaling Flaws in LLMs**
Fei Yu, **Yingru Li**, Benyou Wang. *arXiv 2025*.
9. **Probability Tools for Sequential Random Projection**
Yingru Li. *arXiv 2024*.
10. **Simple, Unified Analysis of Johnson-Lindenstrauss with Applications**
Yingru Li. *arXiv 2024*.

PATENTS

- Liangqi Liu, Wenqiang Pu, **Yingru Li**, Zhi-Quan Luo. *Game-Strategy Update Method, Device, and Medium for Unknown Matrix Games*. Chinese invention patent application, CN117787419A (Shenzhen Research Institute of Big Data; published Mar. 2024).

RESEARCH IMPACT & ADOPTION

The training–inference mismatch / RL-collapse line of work (*When Speed Kills Stability*) has been adopted across major open-source RL training stacks and frontier models:

- **DeepSeek-V3.2** adopts the geometric sequence masking to address the training–inference mismatch and off-policy training instability ([arXiv:2512.02556](#)).
- Integrated into **veRL** (v0.6.0+): geometric-mean importance-sampling masking / rejection (Geo-MIS / Geo-RS) and **rollout correction** for general off-policy training; Trust Region Masking and the Optimal Token Baseline are also available in veRL.
- Integrated into **Slime** and **Miles**. Discussed in a **vLLM engineering blog** (bitwise-consistent on-policy RL) and a **SGLang Blog**.
- Used by **Cognition’s SWE-grep** and discussed in the **Qwen team technical report**.

AWARDS & HONORS

- **Best Paper Award**, 3rd Doctoral and Postdoctoral Daoyuan Academic Forum, 2024.
- **Best Student Paper**, IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2024.
- **SRIBD Ph.D. Fellowship (Gold Class)**, Shenzhen Research Institute of Big Data, 2023.
- **Presidential Ph.D. Fellowship**, The Chinese University of Hong Kong, 2019–2023.
- **Tencent AI Ph.D. Fellowship**, Tencent & The Chinese University of Hong Kong, 2018.
- **National Scholarship (China)** — highest national academic honor, top 0.2% nationwide, 2018.
- **Qiming Star Award** — 1 of 5 recipients from 7,112 undergraduates, HUST, 2016.
- **First Prize**, Parallel Computation and Application Contest (PAC), Intel & CCF, 2015.
- **First Prize**, China National Mathematics Olympiad (Province-level Math League), 2012.

PEER REVIEW & PROFESSIONAL SERVICE

Reviewer for the leading peer-reviewed venues in machine learning and artificial intelligence. NeurIPS, ICML, and ICLR are the three top-tier conferences in the field; AISTATS is a leading venue in machine-learning theory and statistics; TMLR is a flagship ML journal.

Venue	Role	Papers reviewed
NeurIPS (2023, 2024, 2025)	Reviewer	6 + 6 + 5 = 17
ICLR (2024, 2025, 2026)	Reviewer	2 + 3 + 5 = 10
ICML (2025, 2026)	Reviewer [†]	5 + 6 = 11
AISTATS (2025)	Reviewer	2
TMLR (ongoing)	Reviewer	3
Total (2023–2026)		43 papers

- † **ICML 2026 Gold Reviewer** — recognized by the ICML Program Chairs as among the top reviewers for the conference, based on area-chair evaluations of the quality of reviews submitted; awarded complimentary registration.
- Admitted to the NeurIPS 2026 and AISTATS 2026 reviewer pools (paper assignments pending).
- **28 additional papers** reviewed across 9 affiliated workshops, including ICLR Workshop “Bridging the Gap Between Practice and Theory in Deep Learning,” ICML Workshop “Aligning RL Experimentalists and Theorists,” and NeurIPS Workshops on Bayesian Decision-making & Uncertainty and RL4RealLife.
- **Session Chair**, 2 sessions on “Integrating Generative AI with Sequential Decision-making,” INFORMS Annual Meeting 2024.
- **Organizer / Chair**, Reinforcement Learning Seminar, CUHK (8 semesters, 2019–2022).

INVITED TALKS & ORAL PRESENTATIONS

- **Revisiting Trust Region Policy Optimization in the Era of LLMs** — Invited talk, QingkeAI (Dec. 2025).
- **LLM-RL Instability and the Training–Inference Mismatch** — Invited talks at SLAI and the Bay Area RL community (Nov. 2025).
- **Tackling Data Scarcity for Trustworthy Agents** — Invited talk, ETH Zurich (Nov. 2024).
- **Exploration at Scale: Theory, Algorithms & Applications** (*a.k.a. Scalable Uncertainty Quantification for Exploration and LLM Reasoning*) — Invited talks at MIT (Jul. 2024), Princeton University (May 2024), INFORMS Annual Meeting (Seattle, Oct. 2024), International Symposium on Mathematical Programming / ISMP (Montréal, Jul. 2024), INFORMS Optimization Society Conference at Rice University (Mar. 2024), and RLChina (Jun. 2024); contributed talk at the 3rd Daoyuan Academic Forum (Jan. 2024).
- **No-Regret Learning in Unknown Games with Applications** — Invited talk, RL Theory Student Workshop, Nanjing University (Aug. 2022); contributed talk, 2nd Daoyuan Academic Forum (Aug. 2022).
- **HyperDQN: A Randomized Exploration Method for Deep RL** — Contributed talk, NeurIPS Workshop on Ecological Theory of RL (Dec. 2021).

SELECTED MEDIA COVERAGE

- Featured for the Qiming Star Award in the HUST campus newspaper and HUST Online, 2016.

TEACHING

Teaching Assistant, The Chinese University of Hong Kong (Shenzhen) — duties delivered in English (tutorials, grading, laboratory sessions): Stochastic Processes (Fall 2018), Optimization II (Spring 2019), Distributed and Parallel Computing (Fall 2019), Reinforcement Learning (Fall 2020), Matrix Analysis (Spring 2021), and Deep Learning and Their Applications (Spring 2022).