

Hao Li

186-8687-6729 | haoli47@outlook.com | [Personal Page](#) | [LinkedIn](#) | [GitHub](#)

SUMMARY

Computer Science PhD specialising in post-training and reinforcement learning for large language models. At Baidu, I own RL training for multiple releases of the flagship ERNIE-5.1 model family. I have hands-on experience in multi-stage training recipe design, multi-GPU and distributed training environments experimentation, reward modelling, reasoning-efficiency optimisation, and long-horizon agentic data generation and trajectory evaluation. I developed methods that reduced reasoning length by 40% without measurable degradation on objective benchmarks and shipped reward-modelling systems to production, helping ERNIE-5.1-Preview rank first among Chinese models on LLM TextArena. I have authored 31 peer-reviewed publications with 400+ citations, including seven papers at NeurIPS, ICML, ACL, and KDD, four of them as first author.

EDUCATION

The University of Manchester

Ph.D. in Computer Science

Manchester, UK

Sep. 2021 – Dec. 2025

The University of Leeds

M.Sc. in Advanced Computer Science (AI)

Leeds, UK

Aug. 2019 – Dec. 2020

Northeast Petroleum University

B.Sc. in Information and Computing Science

Daqing, China

Sep. 2013 – July 2017

WORK EXPERIENCE

ERNIE Team, Baidu

Member of Technical Staff, Post-Training

Jan. 2026 – Present

Beijing, China

- **Flagship Model Training:** Hands on experience of end-to-end RL training for ERNIE-5.1, spanning multi-stage data-mixture and training-recipe design, large-scale experimentation, checkpoint selection, and release evaluation across chat, reasoning efficiency, and creative writing. Run concurrent multi-node experiments, with each training run using 768 NVIDIA H800 GPUs; delivered training recipes adopted in both official releases on schedule.
- **Cross-Team Technical Leadership:** Coordinate data, infrastructure, and evaluation teams to prioritise experiments, resolve training and evaluation bottlenecks, and support checkpoint-selection and release decisions under fixed production timelines.
- **General-Purpose Reward Modelling:** Designed and productionised a general-purpose reward framework that unifies heterogeneous reward signals and fusion mechanisms, reducing manual configuration across domains. Re-architected the verifier system to combine rubric-based and continuous rewards, and validated the framework on **20 benchmarks** spanning chat, question answering, long-form text, and mathematics. Deployed in flagship RL training, helping ERNIE-5.1-Preview rank **first among Chinese models on LLM TextArena**.
- **Reasoning Efficiency:** Developed two independently deployable RL objectives: a length penalty anchored to the shortest full-score rollout within each group, and a Bradley–Terry objective for pairwise length preferences. Reduced average reasoning-token usage by **40%** with no measurable degradation on objective benchmarks.
- **RSI-Based Agent Data Synthesis and Trajectory Evaluation:** Built a recursive self-improvement (RSI) data-synthesis pipeline that enables models to iteratively solve tasks and retains the resulting self-improvement trajectories as post-training data. Applied the pipeline to a sandboxed SWE-bench environment and GPQA/AIME evaluation workflows, generating **100+** long-horizon RSI trajectories. Developed a two-stage failure-attribution framework combining high-recall rule-based screening with LLM adjudication over a fine-grained error taxonomy to diagnose and filter trajectories for subsequent training iterations.

Microsoft Research Asia

Research Intern, Machine Learning Group

Aug. 2024 – May 2025

Beijing, China

- **MIRA — medical time-series foundation model:** Designed the architecture and pre-training scheme to handle irregular sampling in real-world clinical data, enabling cross-dataset zero- and few-shot forecasting and outperforming BASELINE by up to **8%** on average. First author, **NeurIPS 2025**.

- **BRIDGE — controllable generation under natural-language conditions:** Proposed a controllable generation paradigm combining multi-agent iterative optimisation with diffusion modelling, improving alignment between generated outputs and textual specifications by **12%** over BASELINE on MSE. First author, **ICML 2025**; open-sourced with **1.1k+ GitHub stars**.
- **TarDiff — target-aware synthetic data:** Co-developed a diffusion guidance method that steers generation using downstream-task feedback, raising the task utility of synthetic data. **KDD 2025**.

SELECTED PUBLICATIONS

31 peer-reviewed publications, **400+** citations, **14** at top-tier venues (**5 as first author**). Selected top-tier work below; full list and metrics on [Google Scholar](#).

1. **Hao Li**, Zhengkun Zhang, Gangqiang Hu, Zhen Zhang, Yude Gao, Dai Dai, Jing Liu. *RLVR²: Reinforcement Learning with Verifiable Rubric-based Ranking*. **Preprint, 2026**
 ↪ Group-wise ranking rewards driven by verifiable rubrics, improving the stability and scalability of RL training.
2. **Hao Li**, Yizheng Sun, Viktor Schlegel, Kailai Yang, Riza Batista-Navarro, Goran Nenadic. *Arg-LLaDA: Argument Summarization via Large Language Diffusion Models and Sufficiency-Aware Refinement*. **ACL 2026**
 ↪ Diffusion-based argument summarisation with sufficiency-aware iterative refinement, improving evidence completeness and factual consistency.
3. **Hao Li**, Bowen Deng, Chang Xu, Jiang Bian, et al. *MIRA: Medical Time Series Foundation Model for Real-World Health Data*. **NeurIPS 2025**
 ↪ Foundation-model architecture and pre-training scheme for irregularly sampled clinical series, enabling cross-dataset zero- and few-shot forecasting.
4. **Hao Li**, Yuhao Huang, Chang Xu, Jiang Bian, et al. *BRIDGE: Bootstrapping Text to Control Time-Series Generation via Multi-Agent Iterative Optimization and Diffusion Modelling*. **ICML 2025**
 ↪ Text-controllable time-series generation combining multi-agent iterative optimisation with diffusion modelling; **1.1k+ GitHub stars**.
5. **Hao Li**, Yuping Wu, Viktor Schlegel, et al. *Which Side Are You On? A Multi-Task Dataset for End-to-End Argument Summarisation and Evaluation*. **Findings of ACL 2024**
 ↪ Multi-task dataset providing a unified training and evaluation benchmark for end-to-end argument summarisation.
6. **Hao Li**, Viktor Schlegel, Riza Batista-Navarro, Goran Nenadic. *Do You Hear the People Singing? Key Point Analysis via Iterative Clustering and Abstractive Summarisation*. **ACL 2023**
 ↪ Iterative clustering plus abstractive summarisation to extract core viewpoints from large-scale comments and quantify their support.

TECHNICAL SKILLS

Languages: Python, L^AT_EX

Post-Training & RL: RLHF, RLVR, GRPO, PPO, DPO, SFT, Agent Trajectory Evaluation

Frameworks: PyTorch, verl, PaddlePaddle, Megatron-LM, DeepSpeed, Ray, Hugging Face Transformers

Inference & Infrastructure: SGLang, vLLM, Multi-node GPU training, Sandboxed Agent Environments, Docker, Git

AWARDS & PROFESSIONAL SERVICE

Service: Reviewer for NeurIPS, ICLR, ICML, ACL and EMNLP (since 2023)

Awards: Open Research Fund, University of Manchester; Academic Excellence Scholarship, University of Leeds; First Prize, National Big Data Competition (China)