# Pengbo Li

**Machine Learning Systems | LLM Reinforcement Learning | Training Infrastructure**

plibt@connect.ust.hk | +86 181 0864 4095 | Hong Kong
github.com/Azency | azency.github.io/pengboli.github.io

## EDUCATION

**Hong Kong University of Science and Technology** Sep 2024 - Present
Ph.D. student, Computer Science | Research: Machine Learning Systems | CGA: 3.843 / 4.3

**Huazhong University of Science and Technology** Sep 2021 - Jun 2024
Master's degree, School of Mathematics and Statistics

- Average: 90.64 / 100 (top 10%); National Scholarship for Graduate Students; Outstanding Graduate Student award.

**Huazhong University of Science and Technology** Sep 2017 - Jun 2021
Bachelor's degree, School of Mathematics and Statistics

- Average: 85.16 / 100; Outstanding Graduate; provincial third prize in a mathematics competition.

## INTERNSHIP

**Tencent Hunyuan | RL Infrastructure Intern** Mar 2026 - Present

- **Training recovery:** Develop checkpoint save/restore capabilities to resume interrupted RL training from the latest checkpoint and recover training state as completely as possible.
- **Numerical correctness:** Investigate and validate numerical differences in a dedicated RL training engine to maintain correctness across engine updates.
- **Prefix reuse:** Optimize shared-prefix computation through training-time KV caching and gradient accumulation; this work led to the first-author paper PrefixFlow (ATC 2026, accepted).

## SELECTED RESEARCH & SYSTEMS PROJECTS

**PrefixFlow | Prefix reuse for LLM RL training** Dec 2025 - Present
ATC 2026, accepted | First author | Paper

- Design a three-phase schedule: prefix forward, suffix training, and delayed prefix backward. Share K/V across microbatches and accumulate their gradients so each trajectory group computes the prefix only once in each direction.
- Offload dormant prefix activations and prefetch them before backward; adapt context, pipeline, and data parallelism while supporting tensor and expert parallelism. Preserve BF16-level equivalence on dense and MoE models with Rollout Routing Replay.
- Achieve **1.90x end-to-end RL speedup** and **6.1x actor-update speedup** on long-prefix RAG in verl; reduce peak suffix-training memory by up to **59.1%** in controlled experiments.

**Canopy | Tree-aware scheduling for agent RL**
NeurIPS 2026, accepted | Second author | Paper

- Build tree-aware rollout scheduling on verl and vLLM. Co-locate sibling branches and retain shared prefixes; use partial preemption to reclaim private suffixes, and choose KV transfer or recomputation based on cost.
- Accelerate rollout generation by up to **2.11x** on long-horizon, multi-turn agent tasks. Reduce full shared-prefix eviction events by **51.1% on average** and up to **94.1%** under memory pressure.

## SELECTED RESEARCH & SYSTEMS PROJECTS (CONTINUED)

### Scaling batch size in LLM reinforcement learning

arXiv preprint, 2026 | Paper

- Separate sample efficiency from system throughput in GRPO/PPO; compare learning curves by cumulative samples and use square-root learning-rate scaling for Adam to study time-to-target improvements.
- Increase generation throughput by up to **2.29x** on fixed hardware. With larger batches and learning-rate retuning, reduce GRPO time-to-target by up to **29%**.

### Centrifuge | Efficient token filtering for LLM training

Jan - Oct 2025

ICLR 2026, accepted | Second author | Paper

- Filtered non-critical token activations inside attention backward kernels and automatically transformed irregular sparse GEMM into dimension-reduced dense GEMM using standard optimized operators.
- Reduced end-to-end training time by up to **34.7%** across 1.1B-40B models; enabled one-line integration for frameworks already using token filtering.

## ADDITIONAL ENGINEERING EXPERIENCE

### HaaS Compute Platform | Lab-incubated startup project

Jan - Jun 2025

- Built Python REST APIs over Kubernetes/CRDs for training-job submission, GPU/CPU/memory allocation, and lifecycle management. Enabled direct SSH debugging of pods through dynamic network mappings.

### Privasea | Homomorphic-encryption backend development

Dec 2022 - Jun 2023

- Built a client-server backend for private inference; ported LightGBM to PALISADE (C++) and optimized encrypted operators. Implemented privacy-preserving face recognition with TFHE-rs (Rust).

## PUBLICATIONS

[1] **Pengbo Li**, Feiyuan Zhang, et al. PrefixFlow: Training-Time KV Caching for Schedule-Level Prefix Reuse in LLM RL Training. *ATC 2026, accepted.*

[2] Feiyuan Zhang, **Pengbo Li**, et al. Canopy: Tree-Aware Rollout Scheduling for Agent Reinforcement Learning. *NeurIPS 2026, accepted.*

[3] Ziniu Li, Jinbo Wang, Guanhua Huang, Feiyuan Zhang, **Pengbo Li**, Alex Chen. When Do Larger Batches Help Scale LLM Reinforcement Learning? *arXiv:2608.29296, 2026.*

[4] Di Chai, **Pengbo Li**, et al. Unlocking Full Efficiency of Token Filtering in Large Language Model Training. *ICLR 2026, accepted.*

[5] Qijing Li, Xinyang Huang, Bowen Liu, **Pengbo Li**, Junxue Zhang, Kai Chen. Cache-Aware I/O Rate Control for RDMA. *APNET 2025.*

[6] Huifang Huang, Zhuo Jin, **Pengbo Li**, Fuke Wu, Hailiang Yang. Optimizing portfolios with surrender variable annuities: A deep reinforcement learning approach. *Insurance: Mathematics and Economics, 2026.*

[7] **Pengbo Li** et al. Privacy Preserving Discretized Spiking Neural Network with TFHE. *ICNC 2023.*

[8] Huifang Huang, Ting Gao, **Pengbo Li**, Jin Guo, Peng Zhang, Nan Du, Jinqiao Duan. Model-based reinforcement learning with non-Gaussian environment dynamics and its application to portfolio optimization. *Chaos, 2023.*