Tianming (Tommy) Sha

Research Engineer / Scientist · LLM Post-Training & Agentic RL · Verifiable Data & Environment Synthesis · Agent Runtime & Eval

Tianming.Sha@stonybrook.edu | +1 934-223-9561 | shatianming5.github.io | GitHub | Google Scholar

EDUCATION

Stony Brook University · B.S. in Applied Mathematics & Statistics — Stony Brook, NY · Aug 2025 – May 2027 (expected)

Sun Yat-sen University · Computer Science & Management (transferred) — Guangzhou, China · Sep 2023 – Aug 2025

EXPERIENCE

Octopus · Project Lead & First Author (Project Page · Tech Report · Code · Env & Corpus) — Stony Brook, NY · Sep 2026 – Present

- **Multi-agent RL (DAG-GRPO).** Tree-packed branching rollouts and shared-KV gradients in one pass (**8.1**× token throughput, **32**× lower KL), scaling Qwen3.5-4B to **49.6%** on BrowseComp-Plus (**+17.5** pts) and **51.2%** on OpenResearcher (matching o3, beating **250**× larger models).
- **Branching SGLang runtime.** Built a multi-agent serving engine over radix KV trees (*fork*, *rewind*, *wait*, *done*) with 32K branch budgets, achieving **89.6%** GPU KV cache hit rate across 450K H200 batches, **98.3%** less communication overhead, and **249**× faster 128K context assembly.
- **Zero-copy cross-agent KV sharing.** Mounted sibling KV caches selectively into full-attention layers with monotonic RoPE offsets (skipping recurrent states), retaining **92.2%** of full-text answer quality at **0** extra KV memory and cutting off-task hijacking by **21**× (57.9% → 2.7%).

XYZ AI Lab · Research Intern (Mentors: Xiao Yang, Jiang Bian, Lei Song) — Shanghai, China · Apr 2026 – Aug 2026

- **Long-horizon search agents (XYZ-Aquila).** Post-trained **XYZ-Aquila-mini / pro** (35B / 397B MoE) over 256K-token tool-use trajectories (search, web scraping, Python): mini ranks **#1** among sub-40B open-weight models (BrowseComp **78.8**, GAIA **97.1**); pro leads sub-400B models on BrowseComp **84.8**, BrowseComp-ZH **85.1**, LiveBrowseComp **53.7**, and WideSearch **81.2**.
- **Long-context agent runtime & data synthesis.** Co-built the **AxisAgentic** runtime (preserving exact per-turn agent context across memory compaction and rollback) and a multi-hop QA synthesis pipeline, training Aquila on trajectories averaging **102.7K** tokens (72 tool calls) and open-sourcing **7,000** verified bilingual trajectories.
- **Executable RL environments at scale.** Built an automated pipeline converting 100K GitHub repositories into **2,000+** containerized (rootless udocker) environments and **50K+** unit-test-verified coding and tool-use tasks, scaling multi-node RL post-training across **100K** GPU-hours on **2,000** H200 GPUs.

Microsoft Research Asia · Research Intern (Mentors: Xiao Yang, Weiqing Liu, Jiang Bian) — Beijing, China · Jan 2026 – Apr 2026

- **Agentic RL evaluation (Agent^2 RL-Bench).** Co-first author (NeurIPS 2026 Evaluations); built the sandboxed Docker evaluation harness in Microsoft's **RD-Agent** (**14.8K** GitHub stars) where autonomous agents configure SFT warm-up and GRPO pipelines, lifting ALFWorld success from **4.85%** → **93.28%** within a 12h budget.
- **Benchmark redundancy analysis (BenchScope).** Co-first author; evaluated **8,400+** model runs across 22 LLM benchmarks, demonstrating that the six Open LLM Leaderboard tasks span an effective dimension of only **1.7** (with BBH and MMLU-Pro correlating at $\rho = \mathbf{0.96}$).

Florida State University (RAI Lab) · Research Assistant (Advisor: Yushun Dong) — Tallahassee, FL (Remote) · Aug 2025 – Aug 2026

- **Retrieval-augmented agent skills (SkillCenter).** First author; built **LangSkills** (100+ stars), extracting **119K+** source-grounded procedural skills into lightweight SQLite indexes so LLM agents retrieve verified workflows rather than hallucinating tool steps.
- **Autonomous ML experimentation (PaperFarm).** Built an open-source multi-agent research workflow (**350+** GitHub stars, PyPI) that plans hypotheses, runs parallel experiments in isolated Git worktrees, critiques results, and automatically rolls back failed runs.

Sun Yat-sen University (HCP Lab) · Research Assistant (Advisors: Keze Wang, Liang Lin) — Guangzhou, China · Sep 2023 – Jul 2025

- **Difficulty-aware RLVR sampling (ThinkPrior).** First author; mitigated zero-advantage rollout waste in GRPO (where 39% of sampled groups are all-correct or all-wrong) by estimating prompt difficulty prior to generation, cutting DAPO training rollouts by **10.6%** on Qwen2.5-Math-7B.

SELECTED PUBLICATIONS

- **Octopus: One Brain, Many Branches for Latent-KV Multi-Agent Reinforcement Learning** · **Sha, T.**, Li, J., Zhao, S. · **Tech Report 2026** · **Code**
- **Agent^2 RL-Bench: Can LLM Agents Engineer Agentic RL Post-Training?** · Chen, W.*, Yang, X.*, Yang, X.*, **Sha, T.***, et al. · **NeurIPS 2026**
- **AI4AI at Scale: A Full-Pipeline System for Enhancing LLM Agentic Capabilities** · XYZ Agentic Team (incl. **Sha, T.**) · **Tech Report 2026**
- **ThinkPrior: Zero-Rollout Difficulty Priors for RLVR** · **Sha, T.**, Zhai, S., Zhao, S. · **arXiv 2026**
- **BenchScope: How Many Independent Signals Does Your Benchmark Provide?** · **Sha, T.**, Zhao, S. · **arXiv 2026**
- **SkillCenter** · **Sha, T.**, Zhao, Y., Sun, L., Dong, Y. · **arXiv 2026**
- **FAST-CAD** · **Sha, T.**, Cheng, Z., Zhai, H., et al. · **AAAI 2026 (Oral)**

*equal contribution

SKILLS & HONORS

Post-training RLVR (GRPO, DAPO), multi-agent & multi-turn RL, long-context SFT, MoE routing replay (R3), verifiable data & environment synthesis

Systems Python, PyTorch, SGLang, AReaL, Megatron (AxisRL), Docker / udocker, multi-node GPU training, C++

Honors URECA Award (2026) · AAAI-26 Scholarship · Dean's List (2025) · Reviewer: ICLR, ICML, AAAI, KDD, CVPR