Yucheng Shi

Homepage | LinkedIn | GitHub | Google Scholar

Email: mailofsyc@gmail.com
Mobile: +1-706-765-5574

Summary

Research Scientist specializing in **self-evolving AI**, **agentic RL**, and **long-horizon agent training**. I build systems where foundation models generate or select their own training environments, act through real tools and runtime harnesses, and improve with verifiable or process-level feedback. Recent work spans self-evolving terminal-agent RL, verifiable environment synthesis for reasoning RL, online RL for GUI agents, and production-scale recommendation agents. Published first/co-first author work at ICLR, ICML, NeurIPS.

Education

University of Georgia
Ph.D. in Computer Science (Advisor: Ninghao Liu) *Jan 2022 – Jan 2026*

- Outstanding Graduate Student Award, 2025; Dissertation Completion Award Fellowship, 2025.

North China Electric Power University
B.Eng. and M.S. in Electrical Engineering *Sep 2014 – Jun 2021*

- China National Scholarship, 2020.

Research Experience

Tencent Hunyuan / AI Lab (Seattle) Feb 2026 – Present
Research Scientist *Agentic RL, terminal agents, self-evolving environments*

- **Terminal-agent RL:** Lead a multi-intern project training coding agents through real shell interaction, sandboxed execution, and verifiable runtime feedback; built rollout/training pipelines across **SLIME**, Terminus/Harbor, Apptainer/Daytona, PPO/GRPO, and Qwen3.5-class models.
- **Long-horizon Task Synthesis:** Created a closed-loop pipeline that converts seed environments and failed terminal trajectories into SFT and RL-ready tasks; Qwen3.5-27B runs improved Terminal-Bench 2 from **40.4%** to **46.1%** without stronger teacher models.
- **Self-Evolving RLVR:** Engineered a generator-solver RL framework for models to autonomously synthesize Python training environments. Used the target model itself to curate **840 verified environments** via rigorous execution pipelines, boosting overall RLVR accuracy (**72.4%** to **80.4%**) and zero-shot transfer on GPQA Diamond (**65.2%** to **70.2%**) and LiveCodeBench (**59.0%** to **63.2%**).
- **Long-horizon Agent Evaluation:** Coordinated **46** validated terminal tasks with **85 min**/**200-step** average runs; built Handbook tooling that improved behavior-to-harness localization from **60%** to **90%**.
- **Self-evolving Vision RL:** Co-led **TRON**, building 520 rule-verifiable **visual reasoning RL environments**; improved three VLMs on 10 vision reasoning benchmarks. Developed quality-aware **self-distillation for GUI grounding**, raising Qwen3.5-9B GUI grounding macro accuracy to **72.23%**, +**6.37%** over GRPO baseline.

Netflix Core Recommendation Team Sep 2025 – Dec 2025
Research Scientist Intern *LLM recommendation agents*

- Engineered a **Self-evolving Recommendation Agent** to learn optimal log-to-language transformations, significantly enhancing zero-shot reasoning-based recommendations at industry scale.

Tencent AI Lab (Seattle) May 2025 – Aug 2025
Research Scientist Intern *GUI agents, visual reasoning RL*

- Built **MobileGUI-RL** / **MobGRPO**, an online RL framework that trains mobile GUI agents through real environment interaction and verifiable feedback.

Harvard Medical School May 2024 – Sep 2024
Student Researcher *Medical AI*

Developed a 70B Radiology LLM by further pre-training on **6.5M+** radiology reports with **DeepSpeed ZeRO-3**.

Selected Publications (Full List)

Agentic RL & Self-evolving AI

[1] *"Learning to Build the Environment: Self-Evolving Reasoning RL via Verifiable Environment Synthesis."*
– **Yucheng Shi**, Zhenwen Liang, Kishan Panaganti, Dian Yu, Wenhao Yu, Haitao Mi.
• **arXiv 2026**.

[2] *"Harness Handbook: Making Evolving Agent Harnesses Readable, Navigable, and Editable."*
– Ruhan Wang, **Yucheng Shi**, Zongxia Li, Zhongzhi Li, Kishan, Haitao Mi, Dongruo Zhou, Leoweiliang.
• **arXiv**.

[3] *"TRON: Targeted Rule-Verifiable Online Environments for Visual Reasoning RL."*
– Tianze Yang*, **Yucheng Shi***, Ruitong Sun, Jingyuan Huang, Ninghao Liu, Jin Sun.
• **arXiv 2026**.

[4] *"Reasoning or Memorization? Direction-Aware Diversity Exploration in LLM Reinforcement Learning."*
– Jiangnan Xia, **Yucheng Shi**, Yu Yang, Kishan Panaganti, Zhenwen Liang, Ninghao Liu.
• **arXiv 2026**.

[5] *"Trust the Right Teacher: Quality-Aware Self-Distillation for GUI Grounding."*
– Jingyuan Huang, Zuming Huang, **Yucheng Shi**, Tianze Yang, Xiaoming Zhai, Wei Chu, Ninghao Liu.
• **Preprint 2026**.

[6] *"MobileGUI-RL: Advancing Mobile GUI Agent through Reinforcement Learning in Online Environment."*
– **Yucheng Shi***, Wenhao Yu*, Zaitang Li, Yonglin Wang, Hongming Zhang, Ninghao Liu, Haitao Mi, Dong Yu.
• **Preprint 2025**.

[7] *"Self-improving Small Object Localization for Large Vision Language Models."*
– Tianze Yang, **Yucheng Shi**, Ruitong Sun, Ninghao Liu, Jin Sun.
• **Preprint 2025**.

[8] *"From Logs to Language: Learning Optimal Verbalization for LLM-Based Recommendation in Production."*
– **Yucheng Shi**, Ying Li, Yu Wang, Yesu Feng, Arjun Rao, Rein Houthooft, Shradha Sehgal, Jin Wang, Hao Zhen, Ninghao Liu.
• **Preprint 2025**.

[9] *"Enhancing Cognition and Explainability of Multimodal Foundation Models with Self-Synthesized Data."*
– **Yucheng Shi**, Quanzheng Li, Jin Sun, Xiang Li, Ninghao Liu.
• **ICLR 2025**.

Retrieval-Augmented Generation

[10] *"Retrieval-enhanced Knowledge Editing for Multi-hop Question Answering in Language Models."*
– **Yucheng Shi**, Qiaoyu Tan, Xuansheng Wu, Shaochen Zhong, Kaixiong Zhou, Ninghao Liu.
• **CIKM 2024**.

[11] *"MKRAG: Medical Knowledge Retrieval Augmented Generation for Medical Question Answering."*
– **Yucheng Shi**, Shaochen Xu, Tianze Yang, Zhengliang Liu, Tianming Liu, Quanzheng Li, Xiang Li, Ninghao Liu.
• **AMIA 2024, Distinguished Paper Award**.

Trustworthy Foundation Models

[12] *"CORTEX: Concept-Centric Token Interpretation for Vector-Quantized Generative Models."*
– Tianze Yang*, **Yucheng Shi* (co-first author)**, Mengnan Du, Xuansheng Wu, Qiaoyu Tan, Jin Sun, Ninghao Liu.
• **ICML 2025**.

[13] *"Black-box Backdoor Defense via Zero-shot Image Purification."*
– **Yucheng Shi**, Mengnan Du, Xuansheng Wu, Zihan Guan, Jin Sun, Ninghao Liu.
• **NeurIPS 2023**.

Technical Skills

- **Post-training** / **RL:** PPO, GRPO, DAPO-style RL, async RL, RLVR, process rewards, verifiable rewards, synthetic curriculum generation, online environment design.
- **Agent Systems:** Terminal agents, GUI agents, tool-use harnesses, executable environments, sandboxed rollout, benchmark/evaluation pipelines.
- **LLM/VLM Infrastructure:** PyTorch, Transformers, TRL, PEFT, vLLM, SGLang, VERL, SLIME, Ray, DeepSpeed, Accelerate, Docker, Apptainer, Slurm.