

Murong Ma

Ph.D. Student in LLM Post-Training, Agentic RL, Repo-level Code Generation, Data Curation | Research Intern, Microsoft Research Asia

murongma@u.nus.edu | +86 186 3611 6115 | [Website](#) | [Google Scholar](#) | [GitHub](#) | [ORCID](#)

RESEARCH PROFILE

Computer Science Ph.D. student building post-training systems for long-horizon agents, with research in process supervision, repository generation, agentic RL, and verified data curation. Hands-on experience training 1.5B–32B models and running training and serving experiments on 8 to over 100 GPUs. Additional research spans multimodal learning, video understanding, and speech.

TECHNICAL EXPERTISE

- **LLM Post-Training:** supervised fine-tuning (SFT), agentic RL, process supervision, and reward design.
- **Agent Data & Evaluation:** repository-level task construction, trajectory synthesis and curation, sandboxed verification.
- **Training & Serving Systems:** slime, SWIFT-Megatron, Megatron-LM, SGLang, and vLLM; distributed GPU training.
- **Multimodal Research:** video understanding and event spotting, speech-driven 3D facial animation, speech recognition.

RESEARCH & ENGINEERING EXPERIENCE

Microsoft Research Asia (MSRA)

Jul 2025–Present

Research Intern, working with [Dr. Yeyun Gong](#)

- **Privileged supervision for agent post-training (project lead).** Using reference artifacts to make hard tasks learnable through guided trajectory construction and verifiable intermediate rewards.
 - **P2T: Process-supervised trajectory generation for SFT.**
Developed a multi-agent framework that uses reference patches to recover successful trajectories on issue-resolution tasks where unguided teacher rollouts fail. Outcome-only filtering discards these hard tasks and cannot establish the validity of intermediate reasoning. Introduced **fact-DAG distillation** to turn privileged information into actionable guidance and a **neural-symbolic critic** to preserve trajectory realism without exposing reference solutions to the agent. Served Qwen3-Coder-480B and GLM-5.1 teachers on 32 H200 GPUs and trained Qwen2.5-Coder 7B/14B/32B students on approximately 2,000 tasks using SWIFT-Megatron. Achieved up to 10.8 percentage points higher Pass@1 across SWE-bench Verified/Lite and DeepSWE, with approximately 15% lower per-instance inference cost. [NeurIPS 2026](#).
 - **Milestone-based RL for long-horizon tasks.**
Extending privileged supervision from offline trajectory construction to RL (GRPO) by decomposing reference solutions and terminal goals into intermediate, verifiable milestones. These milestones make partial progress measurable, providing learning signals for hard tasks even when a rollout does not achieve end-to-end success. Training Qwen3.5-30B with slime on 64 MI300X GPUs over 4,000 SWE tasks, with 16 rollouts per task. Increased the fraction of rollouts receiving a learning signal from **18% to 82%**.
- **Complex task construction and teacher distillation (collaborator).** Expanding agent training beyond issue fixing to full-repository generation and multi-step terminal workflows.
 - **Full-repository generation.**
Co-collected 22,000 repositories from PyPI and GitHub, including larger projects requiring architectural design and cross-file implementation. Constructed tasks through feature-to-test mapping and feature recombination, tying functional requirements to executable checks while increasing task diversity and difficulty. SFT on Qwen3.8-Next-Flash teacher trajectories improved Qwen3.5-27B from **27.3% to 38.0% on NL2Repo**.
 - **Terminal-agent task synthesis and evolution.**
Co-developed a pipeline that grounds Stack Exchange QA pairs in real GitHub issues and repositories and automatically constructs executable task environments. Iteratively evolved seed tasks through algorithmic upgrades, fault diagnosis and repair, and toolchain integration, introducing new capability requirements rather than merely concatenating subtasks. Grounding tasks in existing code requires agents to inspect project structure and modify real implementations; successive evolution rounds deepen reasoning and broaden tool use. SFT on GLM-5.2 teacher trajectories improved Qwen3.5-27B from **36.15% to 49.44% Pass@1 on Terminal-Bench 2.1**.
- **Repository-level mid-training (collaborator).** Co-developed a large-scale pipeline to curate and verify pull requests as training signals for repository-level code editing. [ICML 2026](#).

ShowLab, National University of Singapore

Aug 2021–Sep 2022

Research Assistant, working with [Prof. Mike Zheng Shou](#)

- Developed methods for subject-adaptive, speech-driven 3D facial animation. Built a 4K VoxCeleb variant for face generation, talking-head synthesis, and cross-modal research.
- Helped organize the [Ego4D Challenge 2022](#) and served as a contact for its [Talking to Me Challenge](#).

SMIIP Lab, Duke Kunshan University

Jun 2020–Feb 2021; Jul–Sep 2019

Research Intern, working with [Prof. Ming Li](#) (currently at CUHK-Shenzhen)

- Researched speech recognition, bilingual spoken-term detection, keyword spotting, and multimodal facial animation.
- Helped organize the 2020 Personalized Voice Trigger Challenge. The team placed second in the Interspeech 2021 Auto-KWS Challenge.

EDUCATION

National University of Singapore (NUS)

Aug 2023–Expected 2027

Ph.D. in Computer Science

- Advisors: [Prof. Jin Song Dong](#) and [Prof. Yun Lin](#); research in code-agent post-training, data curation, and explainable AI.
- GPA: 4.92/5.00.
- NUS Research Scholarship.

National University of Singapore (NUS)

2021–2023

M.S. in Computer Science, Artificial Intelligence specialization

- GPA: 4.70/5.00.

Beijing University of Posts and Telecommunications (BUPT)

2016–2020

B.E. in Computer Science and Technology

- Excellent Bachelor Thesis, ranked among the top 10 of 630 students; GPA: 3.70/4.00.

SELECTED PUBLICATIONS & MANUSCRIPTS

LLM, Agents, and Data

[From Patches to Trajectories: Privileged Process Supervision for Software-Engineering Agents](#)

Murong Ma, Tianyu Chen, Yun Lin, Shuai Lu, Qinglin Zhu, Yeyun Gong, Zhiyong Huang, Peng Cheng, Yan Lu, and Jin Song Dong
Neural Information Processing Systems (NeurIPS), 2026.

[Pull Requests as a Training Signal for Repo-Level Code Editing](#)

Qinglin Zhu, Tianyu Chen, Shuai Lu, Lei Ji, Runcong Zhao, **Murong Ma**, Xiangxiang Dai, Yulan He, Lin Gui, Peng Cheng, and Yeyun Gong
International Conference on Machine Learning (ICML), 2026.

[TrainRef: Curating Data with Label Distribution and Minimal Reference for Accurate Prediction and Reliable Confidence](#)

Murong Ma, Ruofan Liu, Yun Lin, Zhiyong Huang, and Jin Song Dong
International Conference on Learning Representations (ICLR), 2026.

[Revisiting the Conflict-Resolving Problem from a Semantic Perspective](#)

Jinhao Dong, Jun Sun, Yun Lin, Yedi Zhang, **Murong Ma**, Jin Song Dong, and Dan Hao
IEEE/ACM International Conference on Automated Software Engineering (ASE), 2024.

[CausalCoder: Causality-Guided Data Curation for Robust Code Generation](#)

Murong Ma, Yilu Cao, Ruofan Liu, Wenhao Zhang, Jin Song Dong, and Yun Lin
Submitted to FSE, 2027; under review.

Multimodal, Vision, and Speech

[Few-Shot Precise Event Spotting via Unified Multi-Entity Graph and Distillation](#)

Zhaoyu Liu, Kan Jiang, **Murong Ma**, Zhe Hou, Yun Lin, and Jin Song Dong
AAAI Conference on Artificial Intelligence, 2026.

[F³Set: Towards Analyzing Fast, Frequent, and Fine-grained Events from Videos](#)

Zhaoyu Liu, Kan Jiang, **Murong Ma**, Zhe Hou, Yun Lin, and Jin Song Dong
International Conference on Learning Representations (ICLR), 2025.

[Novel View Synthesis for High-Fidelity Headshot Scenes](#)

Satoshi Tsutsui, Weijia Mao, Sijing Lin, Yunyi Zhu, **Murong Ma**, and Mike Zheng Shou
arXiv preprint arXiv:2205.15595, 2022.

[Acoustic Word Embedding System for Code-Switching Query-by-Example Spoken Term Detection](#)

Murong Ma, Haiwei Wu, Xuyang Wang, Lin Yang, Junjie Wang, and Ming Li
International Symposium on Chinese Spoken Language Processing (ISCSLP), 2021.

[The DKU System Description for the Interspeech 2021 Auto-KWS Challenge](#)

Yechen Wang, Yan Jia, **Murong Ma**, Zexin Cai, and Ming Li
arXiv preprint arXiv:2104.04993, 2021.

[Training Wake Word Detection with Synthesized Speech Data on Confusion Words](#)

Yan Jia, Zexin Cai, **Murong Ma**, Zeqing Zhao, Xuyang Wang, Junjie Wang, and Ming Li
arXiv preprint arXiv:2011.01460, 2020.