

Yan Mi (密言)

✉ miyan26@mails.ucas.ac.cn | 📞 +86 155-8818-6051 | 📍 Beijing, China
🌐 Academic Homepage | 📝 Personal Blog | 🐙 GitHub: M1YAN | 🏠 Google Scholar

Research Interests: Agents · LLM4Rec · User Simulation · Agentic RL

🎓 Education

- **Institute of Computing Technology, Chinese Academy of Sciences**
M.S. in Artificial Intelligence State Key Laboratory of AI Safety Advisor: Qi Cao
Sep 2026 - Present
- **Beijing University of Posts and Telecommunications**
B.S. in Computer Science and Technology School of Future (Class Meta)
Sep 2022 - Jun 2026

💼 Internship Experience

- **Alibaba Group · Taobao and Tmall Group**
Research Intern, Recommendation Algorithms Taobao Homepage Feed
Dec 2025 - Aug 2026

RecVerse: Faithful E-commerce User Simulation WWW (Under Review) / Co-first Author

Agentic RL · User Simulation · Benchmark · GUI Agent · Hierarchical Memory

- **Core implementation:** Implemented the core user-simulation codebase, integrating a GUI environment, agents, training, and evaluation for multi-turn browsing, clicks, cart additions, and purchases.
- **Benchmark construction:** Built the USB environment from item exposures, layouts, and interface states, enabling multi-turn exploration beyond logged actions. Covered **5,274 trajectories, 69,842 actions, and 8 action types**.
- **Agent memory:** Implemented **working, episodic, and preference memory** for visual context, session history, and intent. Optimized memory updates as actions jointly with browsing to preserve cross-page preferences.
- **SFT data and training:** Synthesized user mindset and memory updates with a teacher model to create structured SFT data. Applied power-law smoothed resampling to mitigate action imbalance and warm-start **Qwen3.5-2B**.
- **Agentic RL:** Implemented multi-turn rollouts and trajectory-level **GRPO**, combining behavior-distribution, intent-consistency, and format rewards. Used **Megatron-LM** and **vLLM** for training and rollout inference.
- **Results:** Improved item-level F1 from **4.27% to 7.19%** and category-level HCO from **23.11% to 32.64%** over the GUI baseline STA; achieved **92% human preference** in anonymized trajectory comparisons against STA.

🔧 Research Experience

ARGUS: Misinformation Defense for Multi-Agent Systems ICLR 2026 / Second Author

Multi-Agent System · Dataset · Goal-Aware Reasoning

- **Dataset and threat modeling:** Co-developed **MisinfoTask: 108 complex tasks** with injection points and reference workflows. Evaluated robustness against prompt injection, RAG poisoning, and tool injection.
- **Two-stage defense:** Contributed to a training-free defense: graph-based channel localization using topology, semantic relevance, and communication frequency, followed by goal-aware, multi-round rectification.
- **Results:** On GPT-4o-mini, averaged across the three attack types, reduced misinformation toxicity from **5.22 to 3.43** and increased task success from **67.43% to 78.43%**.

📄 Selected Publications

Towards Faithful Simulation of Human Shopping Behavior

Jiakai Tang*, Yan Mi*, Jing Yu*, et al. WWW (Under Review) * Equal contribution

Goal-Aware Identification and Rectification of Misinformation in Multi-Agent Systems

Zherui Li, Yan Mi, et al. ICLR 2026 Second author

Omni-Safety under Cross-Modality Conflict: Vulnerabilities, Dynamics Mechanisms and Efficient Alignment

Kun Wang, Zherui Li, Zhenhong Zhou, Yitong Zhang, Yan Mi, et al. NeurIPS 2026

🔧 Technical Skills

Training & Inference SFT, GRPO, Agentic RL, Megatron-LM, vLLM

Agents & Evaluation GUI agents, hierarchical memory, user simulation, benchmarks, multi-agent safety

Engineering Python, PyTorch, Linux, Git, Shell, $\LaTeX$

🏆 Honors & Awards

National First Prize, Global Campus AI Algorithm Elite Competition (2024)

Beijing First Prize, RAICOM Robot Developer Competition (2024)

Academic Scholarship, BUPT (2025) Outstanding Student, BUPT (2022 - 2023)