# 黄仁泓

+86-13620950268 | renh2@zju.edu.cn | renh2.github.io

## 教育背景

**浙江大学，博士研究生，计算机科学与技术** 2022.09 - 至今

- 导师：杨洋教授

**浙江大学，竺院 (荣誉学院)，计算机科学与技术** 2018.09 - 2022.06

- GPA：3.97/4.00 (**top 1%**)

**新加坡国立大学，访问学生，计算机科学与技术** 2021.08 - 2021.09

## 研究历程

**GNN(2022-2024,8 篇)->LLM(2025-2026,8 篇)->Embodied AI(2026)**。

## 主要发表论文

1. **Renhong Huang**, Ning Tang, Jiarong Xu, Yuxuan Cao, Qingqian Tu, Sheng Guo, Bo Zheng, Yang Yang. PolicySim: An LLM-Based Agent Social Simulation Sandbox for Proactive Policy Optimization. WWW 2026.
2. **Renhong Huang**, Dongdong Hua, Yifei Sun, Sitao Ding, Hanyang Yuan, Daixin Wang, Yang Yang. RAT: RunAnyThing via Fully Automated Environment Configuration. EMNLP Main 2026.
3. **Renhong Huang**, Yuxuan Cao, Yi Li, Junwei Hu, Zihua Xiong, Shuai Fang, Sheng Guo, Bo Zheng, Yang Yang. Trimming the Fat: Redundancy-Aware Acceleration Framework for DGNNs. AAAI 2026.
4. **Renhong Huang**, Jiarong Xu, Zhiming Yang, Xiang Si, Xin Jiang, Hanyang Yuan, Chunping Wang, Yang Yang. Extracting Training Data from Molecular Pre-trained Models. NeurIPS 2024.
5. **Renhong Huang**, Jiarong Xu, Xin Jiang, Ruichuan An, Yang Yang. Can Modifying Data Address Graph Domain Adaptation? KDD 2024.
6. **Renhong Huang**, Jiarong Xu, Xin Jiang, Chenglu Pan, Zhiming Yang, Chunping Wang, Yang Yang. Measuring Task Similarity and Its Implication in Fine-Tuning Graph Neural Networks. AAAI 2024.
7. Chaojian Shi*, **Renhong Huang***, Xulan Qiu, Jiawei Zhou, Yunsong Zhou. Modular Representation Steering for Multi-Attribute LLM Alignment. EMNLP Main 2026.
8. Zhengzhe Fu, **Renhong Huang**, Rongjie Jiang, Guandong Sun, Feng Zhao, Yang Yang. From Bag-of-Documents to Structure-Aware Synthesis: Resolving Conflicts in RAG. KDD 2026, under review.
9. Jiarong Xu, **Renhong Huang**, Xin Jiang, Yuxuan Cao, Carl Yang, Chunping Wang, Yang Yang. Better with Less: A Data-Centric Perspective on Pre-Training Graph Neural Networks. NeurIPS 2023.
10. Dongdong Hua, Yifei Sun, **Renhong Huang**, Feng Gao, Chunping Wang, Yang Yang. PTCG-Bench: Can LLM Agents Master Pokemon Trading Card Game?. EMNLP 2026.
11. Siwei Zhang, Yun Xiong, Xi Chen, Zi'an Jia, **Renhong Huang**, Jiarong Xu, Jiawei Zhang. RAPO: Expanding Exploration for LLM Agents via Retrieval-Augmented Policy Optimization. KDD 2026.

## 主要研究经历

**自动化数据合成流水线自进化** 合作者：郑波（蚂蚁网商银行总经理），杨洋教授 | 2026.06 - 至今

- LLM 业务后训练阶段普遍面临高质量推理（CoT/RL）数据稀缺问题，传统启发式数据合成方法缺乏迭代优化能力，导致模型早期探索能力弱、训练稳定性差、数据质量无法动态适配训练需求。
- 搭建可自进化的训练数据合成流水线，将数据生成模块封装为智能体，实现数据生产闭环迭代；摒弃固定生成策略，依托模型 SFT/RL 训练反馈、数据熵、多样性、场景覆盖率、自一致性增益等训练动力学指标，驱动数据合成智能体动态优化生成策略，实现数据质量的自进化。

**RAT**：跨语言全自动仓库环境配置 合作者：王岱鑫（蚂蚁百灵长文本负责人），杨洋教授 | 2025.12 - 2026.04

- 自动化运行环境配置对于自动化沙盒生成至关重要，真实开源仓库普遍存在多语言异构、依赖链路复杂、文档缺失、无标准化部署脚本等问题，导致模型生成的代码无法正常编译与运行。

- 首个提出 RAT(RunAnyThing) 跨语言全自动仓库环境配置框架，适配 Python、Java、Rust 等多种主流编程语言，构建跨语言统一抽象层，并利用镜像初始化，有效降低项目配置难度，同时设计环境配置记忆、经验总结等机制。
- 针对配置环境场景，构建包含 **2000+** 真实开源仓库的大规模多语言评测基准 RATBench，通过分层采样覆盖不同规模、热度的项目，以真实编译测试结果为评判标准，贴合工业界真实环境配置场景；环境搭建成功率 (ESSR) 平均提升 **29.6%**，在多语言场景下均表现优异，通用性与稳定性大幅领先现有方法，且成功率略优于人工水平。

**PolicySim：基于社会沙盒反馈优化** **合作者：郑波（蚂蚁网商银行总经理），杨洋教授** | 2024.12 - 2025.03

- 社交平台的推荐、内容曝光调控等平台策略会极大影响平台社区。业界主要依赖上线后 A/B 测试评估策略效果，不当策略极易导致观点极化等负面问题，且上线后评估存在严重滞后性。
- 提出多智能体社交仿真沙盒 PolicySim，通过 SFT+DPO 双阶段训练构建兼顾微观行为真实性与宏观舆论一致性的高保真环境，并首个基于仿真反馈信号实现推荐、曝光控制等平台策略的闭环优化链路。
- 在微观用户行为、宏观舆论动态、策略优化效果三个维度全面优于现有社交仿真平台；PolicySim 可精准模拟平台生态演化，**显著提升平台干预策略的预判与评估能力**。

**PTCG-Bench：长程策略智能体评测** **合作者：王春平（信也科技首席科学家），杨洋教授** | 2026.02 - 2026.05

- 现有 LLM 智能体评测难以统一衡量长程规划、博弈决策与自主进化能力，且难以区分模型能力与智能体框架贡献，缺乏支撑自进化智能体研究的标准化评测环境。
- 构建基于宝可梦集换式卡牌游戏的智能体评测基准 PTCG-Bench，打造双玩家对抗环境，覆盖推理、规划与战略决策能力。结合进化评测协议和模块化智能体框架，实现经验学习评估及模型能力与架构增益解耦。
- 基于多模型与自进化的大规模实验表明，PTCG-Bench 可**有效区分模型与智能体架构差异**，评估战略推理与持续学习水平，并揭示交互框架、上下文管理等 harness 对智能体性能的关键影响，**为自进化研究提供标准化评测范式**。

**RAPO：检索增强策略优化** 2025.12 - 2026.03

- Agentic RL 中，纯在线策略采样容易重复相似推理路径，难以突破固有推理范式。而离线采样增强仅利用完整 rollout 进行全局策略估计，无法在 rollout 过程中动态提升模型探索。
- 提出检索增强策略优化框架 RAPO，将离线步骤轨迹检索并注入智能体采样过程。为解决检索引入的训练噪声与梯度稀疏问题，提出熵驱动检索奖励和检索重要性塑形。
- **14 个数据集**平均性能提升 **5.0%**，在复杂数学推理、多跳问答、真实网页交互任务中均达到最优性能，同时实现了 **1.2**× 的训练效率提升。

**TrimDG：动态图模型加速** **合作者：郑波（蚂蚁网商银行总经理），杨洋教授** | 2024.12 - 2025.03

- 动态图广泛应用于金融、社交网络，但面临极高的计算与存储开销，难以工业落地。该问题归因于双重冗余：一是动态图本身存在大量静态冗余节点与高频重复交互事件；二是模型训练过程中存在大量运行时冗余采样与重复计算。
- 提出适配所有主流 DGNN 骨干模型的 TrimDG，基于动态节点影响力量化与图信息瓶颈消除静态冗余和运行时冗余，大幅降低无效采样频率与训练开销。
- TrimDG 可在基本不损失、甚至小幅提升模型预测性能的前提下，使各类 DGNN 模型训练耗时平均降低 **83.80%**，并在**十亿级工业金融动态图**部署减少 **45.5%** 开销。

**具身 VLA Harness 人形机器人系统** **合作者：开普勒人形团队，杨洋教授** | 2026.07 - 至今

- 人形机器人长程任务需要稳定的理解规划、执行反馈、人机交互与安全控制。单一 VLA 模型难以完成复杂链路。
- 首次提出面向人形的 VLA Harness 架构，将规划、推理、监控、记忆与工具调用解耦为独立模块，并打通开普勒人形全栈仿真、遥操、数采和部署；预期在仿真 benchmark **达到最优**，并在真机的长时间独立任务达到成功率 **90%**。

## 代表性奖项与荣誉

国家奖学金，2024。

优秀毕业论文，2022。

浙江大学竺可桢学院荣誉毕业生，2022。

美国数学竞赛 Meritorious Winner(前 7%)，2021。

# Renhong Huang

+86-13620950268 | renh2@zju.edu.cn | renh2.github.io

## Education

**Zhejiang University, PhD, Computer Science and Technology** 2022.09 - Current
• *Advisor: Professor Yang Yang*
**Zhejiang University, ZJU Honors College, Computer Science and Technology** 2018.09 - 2022.06
• GPA: 3.97/4.00 (**top 1%**)
**National University of Singapore, Visiting Student, Computer Science and Technology** 2021.08 - 2021.09

## Research Timeline

**GNN(2022-2024, 8 papers)->LLM(2025-2026, 8 papers)->Embodied AI(2026).**

## Selected Publications

1. **Renhong Huang**, Ning Tang, Jiarong Xu, Yuxuan Cao, Qingqian Tu, Sheng Guo, Bo Zheng, Yang Yang. PolicySim: An LLM-Based Agent Social Simulation Sandbox for Proactive Policy Optimization. WWW 2026.
2. **Renhong Huang**, Dongdong Hua, Yifei Sun, Sitao Ding, Hanyang Yuan, Daixin Wang, Yang Yang. RAT: RunAnyThing via Fully Automated Environment Configuration. EMNLP Main 2026.
3. **Renhong Huang**, Yuxuan Cao, Yi Li, Junwei Hu, Zihua Xiong, Shuai Fang, Sheng Guo, Bo Zheng, Yang Yang. Trimming the Fat: Redundancy-Aware Acceleration Framework for DGNNs. AAAI 2026.
4. **Renhong Huang**, Jiarong Xu, Zhiming Yang, Xiang Si, Xin Jiang, Hanyang Yuan, Chunping Wang, Yang Yang. Extracting Training Data from Molecular Pre-trained Models. NeurIPS 2024.
5. **Renhong Huang**, Jiarong Xu, Xin Jiang, Ruichuan An, Yang Yang. Can Modifying Data Address Graph Domain Adaptation? KDD 2024.
6. **Renhong Huang**, Jiarong Xu, Xin Jiang, Chenglu Pan, Zhiming Yang, Chunping Wang, Yang Yang. Measuring Task Similarity and Its Implication in Fine-Tuning Graph Neural Networks. AAAI 2024.
7. Chaojian Shi*, **Renhong Huang***, Xulan Qiu, Jiawei Zhou, Yunsong Zhou. Modular Representation Steering for Multi-Attribute LLM Alignment. EMNLP Main 2026.
8. Zhengzhe Fu, **Renhong Huang**, Rongjie Jiang, Guandong Sun, Feng Zhao, Yang Yang. From Bag-of-Documents to Structure-Aware Synthesis: Resolving Conflicts in RAG. KDD 2026, under review.
9. Jiarong Xu, **Renhong Huang**, Xin Jiang, Yuxuan Cao, Carl Yang, Chunping Wang, Yang Yang. Better with Less: A Data-Centric Perspective on Pre-Training Graph Neural Networks. NeurIPS 2023.
10. Dongdong Hua, Yifei Sun, **Renhong Huang**, Feng Gao, Chunping Wang, Yang Yang. PTCG-Bench: Can LLM Agents Master Pokemon Trading Card Game?. EMNLP 2026.
11. Siwei Zhang, Yun Xiong, Xi Chen, Zi'an Jia, **Renhong Huang**, Jiarong Xu, Jiawei Zhang. RAPO: Expanding Exploration for LLM Agents via Retrieval-Augmented Policy Optimization. KDD 2026.

## Research Experience

**Self-Evolving Automated Data Synthesis Pipeline** Bo Zheng, Prof. Yang Yang 2026.06 - Present

- Post-training for LLM applications often suffers from scarce high-quality reasoning data (CoT/RL), while heuristic synthesis methods lack iterative optimization, leading to weak early exploration, poor training stability, and data that cannot adapt to training needs.
- Built a self-evolving automated CoT&RL data synthesis pipeline by packaging data generation as an agent and closing the loop on data production; instead of a fixed generation policy, it uses SFT/RL feedback, data entropy, diversity, coverage, self-consistency gain, and other training-dynamics signals to dynamically optimize generation strategies and improve data quality through self-evolution.

Daixin Wang (Ant Group Bailing Long Context Lead), Prof. Yang Yang

**RAT: Fully Automated Environment Configuration** 2025.12 - 2026.04

- Automated runtime environment configuration is critical for automated sandbox generation, while real open-source repositories commonly involve heterogeneous languages, complex dependency chains, missing documentation, and non-standard deployment scripts, causing model-generated code to fail compilation and execution.
- Proposed RAT (RunAnyThing), a cross-language fully automated repository environment-configuration framework for Python, Java, Rust, and other mainstream languages. RAT builds a unified cross-language abstraction layer, uses image initialization to reduce setup difficulty, and introduces environment memory and experience summarization mechanisms.
- Built RATBench, a large-scale multilingual benchmark with **2000+** real open-source repositories for environment-configuration evaluation. The benchmark uses stratified sampling across project scale and popularity, and evaluates setup success with real compilation and test results to match industrial deployment scenarios; improved Environment Setup Success Rate (ESSR) by **29.6%** on average, substantially outperforming existing methods in generality and stability and slightly exceeding human-level success rate.

**PolicySim: Social Sandbox Feedback Optimization** Bo Zheng, Prof. Yang Yang 2024.12 - 2025.03

- Platform policies such as recommendation and content-exposure control can strongly shape platform communities. Industry practice mainly relies on post-deployment A/B testing, while improper policies can easily trigger negative effects such as opinion polarization and suffer from severe delayed feedback.
- Proposed PolicySim, a multi-agent social simulation sandbox for proactive evaluation and adaptive optimization of intervention policies. Through two-stage SFT+DPO training, PolicySim builds a high-fidelity environment that balances microscopic behavioral realism and macroscopic opinion consistency, and uses simulation feedback signals to close the loop for recommendation and exposure-control optimization.
- Outperformed existing social simulation platforms across microscopic user behavior, macroscopic opinion dynamics, and policy optimization effectiveness; PolicySim accurately simulates platform ecosystem evolution and significantly improves the ability to anticipate and evaluate intervention policies.

**PTCG-Bench: Strategic LLM-Agent Benchmark** Chunping Wang, Prof. Yang Yang 2026.02 - 2026.05

- Existing LLM-agent benchmarks struggle to jointly evaluate long-horizon planning, game-theoretic decision-making, and autonomous evolution, while also failing to separate model capability from agent-framework contributions, leaving self-evolving agent research without a standardized evaluation environment.
- Built PTCG-Bench, an agent benchmark based on the Pokemon Trading Card Game, with a two-player adversarial environment covering reasoning, planning, and strategic decision-making. Combined an evolution evaluation protocol with a modular agent framework to evaluate experience learning and disentangle model capability from architectural gains.
- Large-scale experiments with multiple models and self-evolving algorithms show that PTCG-Bench effectively distinguishes model and agent-architecture capabilities, evaluates strategic reasoning and continual learning, and reveals the critical impact of interaction frameworks, context management, and harness design on agent performance, providing a standardized evaluation paradigm for self-evolving research.

**RAPO: Retrieval-Augmented Policy Optimization for LLM Agents** 2025.12 - 2026.03

- In agentic RL, pure on-policy sampling often repeats similar reasoning paths and struggles to break out of fixed reasoning patterns. Off-policy augmentation, by contrast, only uses complete rollouts for global policy estimation and cannot dynamically improve exploration during rollouts.
- Proposed RAPO, a retrieval-augmented policy optimization framework that retrieves off-policy step traces and injects them into agent rollouts. To address training noise and gradient sparsity introduced by retrieval, we introduced entropy-driven retrieval rewards and retrieval importance shaping.
- Achieved a **5.0%** average performance gain across **14 datasets**, reaching **SOTA** on complex math reasoning, multi-hop QA, and real web interaction tasks, while also delivering a **1.2**× training efficiency improvement.

**TrimDG: Acceleration for Dynamic Graph Models** Bo Zheng, Prof. Yang Yang 2024.12 - 2025.03

- Dynamic graphs are widely used in finance and social networks, but their high computation and storage costs make industrial deployment difficult. The bottleneck comes from dual redundancy: static redundant nodes and high-frequency repeated interactions in dynamic graphs, plus runtime redundant sampling and repeated computation during model training.
- Proposed TrimDG for mainstream DGNN backbones, eliminating static and runtime redundancy through quantified dynamic node influence and the graph information bottleneck, substantially reducing invalid sampling frequency and training cost.
- TrimDG reduces training time by **83.80%** on average across DGNN models with little to no prediction loss and sometimes slight performance gains, and reduces deployment cost by **45.5%** on billion-scale industrial financial dynamic graphs.

**Humanoid VLA Harness for Embodied Robotic Systems** Kepler Team, Prof. Yang Yang 2026.07 - Present

- Long-horizon humanoid tasks require stable understanding and planning, execution feedback, human-robot interaction, and safety control. A single VLA model struggles to complete complex task chains.
- First proposed a VLA Harness architecture for humanoids, decoupling planning, reasoning, monitoring, memory, and tool use into independent modules, while connecting full-stack Kepler humanoid simulation, teleoperation, data collection, and deployment; expected to reach the best simulation benchmark performance and achieve a **90%** success rate on long-duration independent real-robot tasks.

## Selected Awards and Honors

National Scholarship, 2024.
Outstanding Graduate Thesis, 2022.
Award of Honor for Graduate at Chu Kochen Honors College of Zhejiang University, 2022.
Meritorious Winner for American Mathematical Competition (top 7%), 2021.