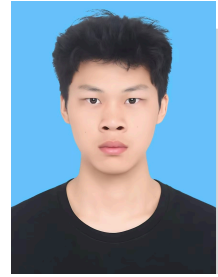


# 王司南

✉ sinanwang@hust.edu.cn    📞 150-0270-9274  
🌐 github.com/wsn18381187    📱 微信: 手机号  
🏠 个人主页    📍 武汉



## 教育经历

### 华中科技大学

2023.09 – 2027.06

本科生 计算机科学与技术专业 加权平均分 83.2 / 100

### 新加坡国立大学

2024.07

暑期交换项目 程序结构与解释课程 使用基于 Java 的 Unity 软件包开发游戏

## 技术技能

### 专业能力

熟练使用 C/C++、Python 进行开发；熟悉基于 Linux、Git 的工作流程；熟练掌握常见的数据结构与算法；熟悉使用 LaTeX、markdown 读写文档、报告；了解计算机领域的网络、系统学科知识

### 研究领域

LLM-as-a-Judge (LLM 自主评测), LLM 多模态/推理能力, Agent 框架开发

### 语言能力

英语: 雅思 6.5 分, 流利阅读英文文献与技术文档

## 项目经历

### 无需人工标注的 LLM-as-a-Judge 评测框架及应用

2025.08 – 2026.03

定义并验证大模型评估过程中的“情境偏好 (Situational Preference)”现象，提出一种“无标签 (Label-Free)”指标用来评估大模型的鲁棒性 (Robustness)

- 作为**共同第一作者**，完成项目的所有代码，参与项目实验设计，以及部分论文和制图工作。
- 基于“情境偏好”的现象，开发了 LLM-as-a-Judge 的评测框架，在该框架下可以通过自动化的评测流程清晰体现不同 LLM 在 Judge 时的稳定性差异。同时这种稳定性与其他主流 Benchmark 呈现高度相似性，验证了基于此框架训练并提升 LLM 性能的可行性。
- 自动化评测脚本使用 Python 进行开发，主要使用 OpenAI SDK 和 JSON 相关库实现。[项目仓库]

### LearnAgent —— AI 智能体框架

2026.04 – 现在

自主设计并开发的多功能 AI Agent 框架，支持多种底层 LLM 与外部工具无缝调用

- 设计并实现 **CodeAct 模式**，以 Python 代码生成替代 JSON function call，彻底解决参数转义问题。相比于直接通过 JSON function call，模型生成代码字符格式出错的概率大幅下降。
- 开发 **Condition Flow** 功能 (Plan-Execute-Judge-Conclude 架构)，搭建任务提取、拆分、分步执行、审查的自动化流程，优化了复杂任务执行中的信息链路传递，实现复杂任务的自动分解与迭代执行。
- 集成基于 RAG 历史检索和用户画像动态更新实现的双重记忆机制，实现 Agent 的长短期记忆功能；设计了多工具调度等模块，构建完整的 Agent 工具链。
- 项目代码使用 Python 进行开发，使用 OpenAI SDK 库实现模型调用相关功能、ChromaDB 库实现 RAG 数据库相关功能。项目已开源至 GitHub。[项目仓库]

### 多重参考的图像生成数据引擎

2025.03 – 2025.08

将真实图像、合成图像构成的数据集进行多维度拆分构建的数据引擎“RefBlend”，并基于此流程建立了高质量数据集“MultiRef-dataset”、基准测试“MultiRef-benchmark”

- 基于生成的目标图像的不同信号和指令，搭建 Pipeline 构建了多图源参考数据集“MultiRef-dataset”

和基准测试“MultiRef-benchmark”。

- 整理、开源数据集和基准测试至 Hugging Face 上，并负责后续的管理和维护工作。
- 参与 vllm 本地模型部署和数据标注工作。[项目仓库]

### 使用工具的强化视觉推理——大模型视觉推理训练框架

2025.01-2025.05

基于 GRPO 构建的多模态大模型使用外部工具辅助进行视觉思维链进行推理的训练框架

- 复现相关论文中类似使用大模型视觉思维链进行推理的基准测试数据
- 搭建 Pipeline 构建了部分用于训练的视觉思维链推理的数据集

## 论文 & 获奖

---

**MultiRef: Controllable Image Generation with Multiple Visual References**

2025.08

接收至 ACM MM 2025 [论文链接]

**Are We on the Right Way to Assess LLM-as-a-Judge?**

2025.12

在投于 EMNLP 2026 [论文链接]

**Reinforced Visual Perception with Tools**

2025.9

Arxiv Preprint [论文链接]

**美国大学生数学建模竞赛**

2025.05

Honorable Mention [二等奖]