

9码平刷一天赚10万

EMCm7DuGMf9lBRLV

9码平刷一天赚10万首个系统性工具使用奖励范式，ToolRL刷新大模型训练思路

钱成目前是伊利诺伊大学香槟分校 (UIUC) 一年级博士生，导师为季姮教授。本科就读于清华大学，导师为刘知远教授。其工作集中在大语言模型工具使用与推理以及人工智能体方向。曾在 ACL，EMNLP，COLM，COLING，ICLR 等多个学术会议发表论文十余篇，一作及共一论文十余篇，谷歌学术引用超 500，现担任 ACL Area Chair，以及 AAAI，EMNLP，COLM 等多个会议 Reviewer。

「工欲善其事，必先利其器。」如今，人工智能正以前所未有的速度革新人类认知的边界，而工具的高效应用已成为衡量人工智能真正智慧的关键标准。大语言模型凭借卓越的推理与规划能力，正在快速融入人类生产与生活，但传统的监督训练方法在面对复杂或全新的工具场景时，却常常显得捉襟见肘。如何帮助人工智能突破这一瓶颈，拥有真正自如运用工具的能力？ToolRL 的出现为我们带来了答案。

伊利诺伊大学香槟分校的研究团队率先提出了一项开创性的研究——ToolRL。不同于传统的监督式微调，ToolRL 首次系统性地探讨了强化学习范式下的工具使用训练方法，通过精细化的奖励设计，有效解决了工具推理中的泛化难题。

图 1: 主要 Benchmark 任务上不同训练策略效果对比。精细化奖励设计 ToolRL + GRPO 冷启动能够在不同模型上表现出最佳效果。观察右侧训练曲线，随着训练数据增加，奖励也呈现迅速攀升。

Tool-Integrated Reasoning：LLM 的「工具链式思维」

在 ToolRL 中，研究者将工具调用问题建模为 Tool-Integrated Reasoning (TIR) 的任务范式。这种任务不仅仅要求模型「用」工具，更要求它以合理顺序和逻辑调用多个工具，并基于中间结果灵活调整接下来的思维路径。

TIR 任务的关键特征包括：

图 2: SFT 在工具推理上难以泛化，可能造成过度推理等问题，而基于 RL 的方法具有更好的泛化能力。

设计的关键——不是「对」就够了

ToolRL 首次系统性地分析了工具使用任务中的奖励设计维度，包括：

研究表明，粗粒度、静态、或者仅以最终答案匹配为目标的奖励往往无法最有效地指导模型学习工具推理能力。为此，ToolRL 引入了一种结构化奖励设计，结合「格式规范」与「调用正确性」，确保模型不仅生成合理的工具链式思维，更能准确理解工具含义与调用语义，激发更好更精准的模型工具推理能力。

图 3: 工具推理中的 Rollout 示意图，以及精细化奖励设计示例。除了正确性外，奖励信号额外涉及「工具名称」，「参数名称」以及「参数内容」进行精细化匹配，以取得更好的工具推理奖励效果。

实验：从模仿到泛化，ToolRL 如何激发工具智能？

为了验证 ToolRL 在多工具推理任务中的有效性，研究团队在多个基准上进行了系统实验，涵盖从工具调用（Berkeley Function Calling Leaderboard）、API 交互（API-Bank）到问答任务（Bamboogle）的真实应用场景。

实验设置

核心结果

实验结果表明，ToolRL 不仅提升了语言模型的工具使用能力，更重要的是，它促使模型学会「何时该调用工具、如何调用工具」——这正是智能体走向自主智能的关键一步。

表 1-3: 在三个 Benchmark 上的测试结果，文章发现 GRPO 冷启动的方法往往能取得最好的效果

结语：ToolRL 不仅是一个方法，更是一套通用的奖励范式

结论：ToolRL 不仅是一种方法，更开创了基于工具调用的强化学习奖励新范式。通过大规模实验与深入对比分析，文章验证了三个核心发现：

表 4-5: TooRL 训练出的模型在不相关工具检测（BFCL 子任务）中表现出更好的泛化性与合理平衡工具调用以及自我知识的主动性。

相比于传统强化学习研究往往单纯以「结果正确性」为唯一优化目标，ToolRL 在奖励信号设计上引入了更丰富的维度，不仅量化了「是否正确」，还反映了「工具名称」、「参数规范」等多方面指标，弥补了现有方法对复杂工具链学习的欠缺。展望未来，ToolRL 所提出的奖励扩展框架不仅能适配更多样的任务类别，也为 LLM 与外部工具协同带来了更灵活、更可控的训练思路。我们期待基于这一范式的后续研究，进一步深化多模态工具交互、知识检索与规划生成等领域的智能化水平

正规分分彩app下载

澳洲幸运10官方是谁

168澳洲幸运5历史开奖记录查询

澳洲10专家推荐计划

幸运澳洲10怎么玩才能赢

幸运5全自动算账机器人

澳洲10最厉害三个方法网上玩

一天赚3000的黑方法

澳洲10全天人工年免费计划

幸运飞行艇官网开奖结果app下载

168飞艇开奖官网开奖

全天飞艇免费计划最新版

澳洲幸运10赛车公式怎么分析

澳洲10最厉害三个方法网上玩

飞艇稳赚不赔的打法

澳洲10专家推荐计划

飞艇pk10一天稳赚5000

ai人工智能计算彩票

快3赚钱平台推荐导师