

文章编号: 1003-0077(2026)08-0130-16

推理型大语言模型角色扮演能力评估与优化

冯夏冲, 黄毅翀, 冯骁骋, 秦兵

(哈尔滨工业大学 社会计算与交互机器人研究中心, 黑龙江 哈尔滨 150001)

摘要: 随着角色扮演大语言模型应用价值凸显及推理技术的进步, 一个核心科学问题应运而生: 先进推理技术能否增益模型的角色扮演能力? 为系统性探究该问题, 该文基于 6 个主流基准与 26 个代表性模型, 全面评估了零样本、基于思维链及基于推理模型三种角色扮演范式的性能。研究的核心发现与普遍认知相悖: 推理能力的引入(无论是显式的思维链, 还是内化的推理模型)非但未能使模型的角色扮演能力稳定提升, 反而时常损害角色扮演性能, 并破坏其性能随模型规模增长的缩放定律。此外, 当前模型在处理高级、细致的角色扮演任务时仍存在显著短板, 且在跨语言对比中中文性能普遍优于英文性能。据此, 该文提出“角色感知思维链”与“面向角色扮演的强化学习”两个未来研究方向, 并进一步深化为“细粒度角色奖励驱动的强化学习”框架。通过实验验证, 这些方向共同为提升下一代角色扮演代理的适应性、一致性与有效性提供了新的理论视角与实证支持。

关键词: 角色扮演; 大语言模型; 推理模型

中图分类号: TP391

文献标识码: A

Enhancing the Role-playing Capabilities of Reasoning-optimized Large Language Models

FENG Xiachong, HUANG Yichong, FENG Xiaocheng, QIN Bing

(Research Center for Social Computing and Interactive Robotics, Harbin Institute of Technology,
Harbin, Heilongjiang 150001, China)

Abstract: Currently, advancements in reasoning techniques are constantly pushing the performance boundaries of LLMs. This confluence motivates a pivotal research question: can advanced reasoning capabilities enhance the role-playing performance of LLMs? To systematically investigate this question, this paper conducts a large-scale empirical study across 6 mainstream benchmarks and 26 representative models, evaluating three distinct role-playing paradigms based on zero-shot, Chain-of-Thought, and reasoning model, respectively. Our core findings reveal two counter-intuitive conclusions. First, introducing reasoning capabilities, whether explicit or internalized, often impairs role-playing performance and disrupts its scaling laws. Secondly, current models exhibit significant shortcomings in advanced role-playing tasks, and their performance in Chinese is generally superior to English. Based on these findings, this paper proposes two promising future research directions, "Role-aware Chain-of-Thought" and "Reinforcement Learning for Role-Playing", and further refines the latter into a "Fine-grained Character Reward-driven Reinforcement Learning" framework. By validating these directions through experiments, this work provides new perspectives and empirical evidences to enhance the adaptability, consistency, and effectiveness of next-generation role-playing agents.

Keywords: role-playing; large language model; large reasoning model

0 引言

大语言模型(Large Language Models, LLMs)因其卓越的基础能力及高级指令遵循能力, 已成为构建角色扮演(Role-playing)人工智能模型的关键技术基石^[1]。

基于大语言模型的角色扮演技术在学术研究与商业应用领域均展现出广阔前景。在学术层面, 该技术为社会科学领域的模拟实验提供了新的范式与可能性。在产业应用中, 诸如角色化虚拟助手、情感陪伴机器人及智能游戏非玩家角色(Non-Player Characters, NPCs)等, 已显现出巨大的商业化潜力^[2-3]。

收稿日期: 2025-06-18 定稿日期: 2025-09-18

基金项目: 国家自然科学基金(62276078, U22B2059); 黑龙江省重点研发项目(2022ZX01A32); 中央高校基本科研业务费专项资金资助(XNJKKGYDJ2024013)

与此同时,思维链(Chain-of-Thought, CoT)推理^[4]和强化学习等先进技术的引入,显著增强了大语言模型的逻辑推理能力,称之为大推理模型(Large Reasoning Model, LRM),使其能够应对愈发复杂的决策与分析任务。

鉴于此,一个重要且尚未被充分探讨的科学问题浮出水面:先进的推理技术能否有效增进大语言模型的角色扮演能力?对此问题的深入研究,不仅能为推理技术在角色扮演大语言模型中的应用提供理论洞见与实践指导,更有望为该领域的未来发展开辟一条崭新的技术路径。

为系统性地探究上述研究问题,本文构建了详尽的实验方案。本文选取了六个代表性的角色扮演数据集,并对26个业界广泛应用的开源及商业闭源大语言模型进行了系统性评估。所选模型覆盖了多种技术路径,包括基于API的闭源模型(如GPT-4-Turbo)、主流开源模型(如Qwen2.5)以及经强化学习优化的推理模型(如DeepSeek-R1)。所有实验均依托OpenCompass框架开展,旨在保证评估过程的一致性、稳定性与可复现性。在评测基准方面,本文采用了六个标准数据集:RoleLLM^[5]、InCharacter^[6]、SocialBench^[7]、Character-Eval^[8]、HPD^[9]和CroSS-MR^[10]。这些基准兼顾了选择题与开放式文本生成两类任务范式,并同时包含中文语料与英文语料,从而确保了评估的跨语言全面性。在评估指标上,本文整合了传统客观指标(如准确率、ROUGE-L、精确匹配率)与基于大语言模型的自动化评估方法(如基于提示的评估和奖励模型评分),旨在进行多维度评估,进而增强研究结论的稳健性。

此外,本文定义并比较了三种不同的角色扮演实现路径:①零样本角色扮演(zero-shot role-playing);②基于思维链的角色扮演(role-playing with CoT);③基于推理模型的角色扮演(role-playing with LRM)。此番全面的实验设计,使本文得以系统地解构并度量推理技术对角色扮演性能的具体影响。初步分析(如表1所示样例)表明,零样本角色扮演取得了最优性能。

通过大规模的实证研究,本文揭示了若干关键发现,为未来角色扮演大语言模型的设计提供了重要启示。①思维链的负面效应:与直觉相反,思维链的应用可能对角色扮演性能产生负面影响,而非增益。②推理模型的局限性:为复杂推理任务特化训练的模型,在角色扮演场景下表现欠佳,显示出其适用性的边界。③缩放定律的失效:推理能力的引入破坏了角色扮演性能的缩放定律,即模型参数数量的增加并未带来预期的性能提升。④Qwen系列模型的优越性:在所有被测模型中,Qwen系列模型在角色扮演任务上

表1 林黛玉角色扮演示例

语境:紫鹃:姑娘,外头的花儿都谢了。

GPT-4-Turbo零样本角色扮演

是吗?可见这花儿也同我一样,禁不住这风刀霜剑的。罢了,你拿那花锄和锦囊来,我将它们葬了,也算干净。

GPT-4-Turbo基于思维链的角色扮演

[推理过程省略]知道了。这便是它们的命数,强留不得。你且去吧,让我独自静一静。

OpenAI o1-mini基于推理模型的角色扮演

[推理过程省略]这是一个关于生命周期和无常的观察。花开花落,正如季节更替,是自然现象。

注:林黛玉是中国古典小说《红楼梦》中的主角,她多愁善感、才华横溢,言语尖锐而富有诗意。可以看出,零样本提示的结果最符合林黛玉的性格,而推理的引入则削弱了角色的情感色彩和语言风格。

表现出卓越的综合性能。⑤跨语言性能差异:模型的中文角色扮演能力普遍优于其英文对应能力。⑥角色扮演高级能力仍显不足:当前的大语言模型在处理高级或复杂角色扮演任务时,仍存在显著的能力短板。

本文的主要贡献可归纳为以下三点:

(1)揭示推理与角色扮演的复杂关系:通过在6个数据集和26个模型上的大规模实验,本文首次系统地证实了先进推理技术并非总是提升角色扮演性能的“万灵药”,在特定条件下甚至会产生抑制作用。

(2)剖析现状并指引未来方向:本文深入剖析了当前角色扮演大语言模型的技术现状与核心局限,并在此基础上提出了“角色感知思维链(Role-Aware CoT)”与“面向角色扮演的强化学习(Reinforcement Learning for Role-Playing)”两个前瞻性研究方向,并进一步深化为“细粒度角色奖励驱动的强化学习”(FCR-RL)框架。通过一系列针对性的实证评估,本文不仅验证了两个方法相较于传统方法的性能优势,也证实了FCR-RL框架在提升角色化推理深度上的显著效果,为突破现有性能瓶颈提供了坚实的理论与实证基础。

(3)构建标准化评估框架:为克服现有研究因评估标准各异而导致的碎片化问题,本文将所有涉及的数据集与实验流程整合至OpenCompass框架中,实现了评估的“一键式”执行。这为后续研究构建了一个统一、开放且标准化的实验基准与代码库,相关代码公开于:xcfcde/rolecompass。

1 评估框架

基于大语言模型的角色扮演代理的评估框架,通常由若干核心要素构成:作为评测主体的大语言模型(M)、具体的角色身份($\mathcal{P}$)、指定的评测任务($\mathcal{T}$)、

量化的评估指标($\mathcal{E}$),以及在特定情况下使用的参考答案($\mathcal{A}$)。本文的核心目标在于探究引入推理技术($\mathcal{R}$)对模型角色扮演能力的潜在影响。

据此,整个评测流程可被形式化地定义。对于一个给定的大语言模型 $\mathcal{M}$ 、角色身份 $\mathcal{P}$ 与评测任务 $\mathcal{T}$,其角色扮演性能是通过评估指标 $\mathcal{E}$ 来衡量的。该指标旨在量化模型在特定角色和任务约束下所生成的响应与既定角色身份的一致性程度。为检验推理技术的作用,本文将 $\mathcal{R}$ 作为一个关键变量引入此过程。该评估过程如式(1)所示。

$$S = \mathcal{E}(\mathcal{M}(\mathcal{P}, \mathcal{T}, \mathcal{R}), \mathcal{A}) \quad (1)$$

其中, S 代表最终的性能得分,用以衡量模型 $\mathcal{M}$ 在给定条件下的角色扮演有效性。 S 的值越高,表明模型角色扮演能力越强。 $\mathcal{M}(\mathcal{P}, \mathcal{T}, \mathcal{R})$ 表示模型 $\mathcal{M}$ 在接收到角色身份 $\mathcal{P}$ 和评测任务 $\mathcal{T}$ 后,运用推理技术 $\mathcal{R}$ 所生成的最终输出。 $\mathcal{A}$ 为预设的参考答案。在参考答案不可用或不适用的场景下,评估指标 $\mathcal{E}$ 可相应地调整为无需参考答案的评估范式,例如,采用“以大语言模型为评判者”(LLM-as-a-Judge)的方法进行评估。

2 实验设置

本节将详细阐述本文采用的评测基准、大语言模型、评测指标及核心的推理方法。为确保研究的可复现性与透明度,所有相关实现代码均已集成至OpenCompass开源评测框架中,以支持研究界同仁进行高效复现。

2.1 角色扮演评测基准

为全面评估模型的角色扮演能力,本文精心选取了6个业界公认的角色扮演评测基准。这些基准覆盖了从基础的角色一致性到深度的社交理解等多个维度。各基准的详细说明如下:

(1)**RoleBench**^[5]: 该基准要求模型依据给定的角色信息,分别为通用领域问题和角色特定问题生成恰当的回答。

(2)**HPD(Harry Potter Dialogue)**^[6]: 该基准利用《哈利·波特》系列中的角色档案与相关细节作为输入,旨在评估模型生成符合特定虚构角色身份的对话能力。

(3)**CharacterEval**^[7]: 该基准提供包含角色背景、过往经历及性格特质的详尽描述,要求模型据此生成高度个性化的回应。

(4)**CroSS-MR**^[8]: 该基准的核心任务是,基于给定的角色设定,让模型识别并阐释角色的行为动机。

(5)**SocialBench**^[9]: 该基准旨在评估模型对角色的深度理解能力,具体包括角色的知识体系、行为风格、社交偏好以及细粒度的记忆。

(6)**InCharacter**^[10]: 该基准通过标准化的问卷式访谈来评估模型对角色性格的刻画能力^①。关于上述各基准更为详尽的统计数据如表2所示。

2.2 大语言模型

本文的评测对象涵盖了26个具有代表性的大语言模型,其中包括3个闭源API模型和23个主流开源模型。这些模型源于六家不同的研发机构,囊括了如GPT-5和DeepSeek-R1等当前技术前沿的SOTA(State-of-the-Art)模型,从而保证了评测范围的广泛性与前沿性。完整的模型列表如表3所示。

2.3 评测指标

本文采用两类互补的评估指标体系:自动化客观指标与“以大语言模型为评判者”(LLM-as-a-Judge)。这两种方法均是当前大语言模型能力评测领域广受认可且具备良好可扩展性的技术路径。

自动化客观指标: 主要包括ROUGE-L、准确率(Accuracy)和精确匹配率(Exact Match, EM)。这些是自然语言处理领域中用于衡量文本相似度与内容准确性的经典且公认的评估标准。

基于大语言模型的评估方法: 此方法利用大语言模型自身的理解与判断能力进行评估,具体又分为两种实现技术:预训练奖励模型(Pre-trained Reward Model),该技术通过在人工标注的数据集上进行监督学习,训练出一个能够针对特定评估维度(如角色一致性、吸引力等)进行评分的奖励模型;基于提示的评分(Prompt-based Scoring),该技术则利用精巧的提示工程(Prompt Engineering),激发并引导通用大语言模型依据其庞大的世界知识与推理能力,对模型输出进行动态、自适应的打分。此方法具备更高的灵活性与响应速度。

各个评测基准所对应的具体评估指标: 准确率,用于评估模型在CroSS-MR及SocialBench所包含的

^① InCharacter基准包含14种不同类型的评估问卷。在本文中,主要报告其中应用最广的16型人格(16Personalities)和五大性格特质(BFI)两种问卷的评测结果。所有其他问卷的实现亦已包含在本文的代码库中。

表 2 角色扮演评测基准数据统计

数据集	任务类型	评估指标	具体任务	语言	角色数	样本数
RoleBench	生成	自动化指标(ROUGE)	通用问答	英语	95	32 833
				中文	5	1 690
			角色特定问答	英语	95	7 534
HPD	生成	自动化指标(ROUGE)	回复生成	英语	1	149
				中文	1	167
CharacterEval	生成	大模型评估(预训练奖励模型)	回复生成	中文	77	4 564
CroSS-MR	选择题	自动化指标(准确率)	动机识别	英语	126	445
SocialBench	选择题	自动化指标(准确率)	角色知识理解	英语	23	988
				中文	15	405
			行为风格理解	英语	8	740
				中文	8	323
			社交偏好识别	英语	79	731
				中文	101	1 185
	生成	自动化指标(精确匹配)	长对话记忆	英语	47	728
				中文	35	620
			短对话记忆	英语	46	463
				中文	27	310
InCharacter	生成	大模型评估(基于提示)	16型人格识别	英语	32	32
			大五人格识别	英语	32	32

表 3 本文所采用的大语言模型列表

团队	模型
闭源模型	
OpenAI	GPT-5、GPT-4 Turbo、OpenAI o1-mini
开源模型	
OpenAI	GPT-oss-20B
Qwen	Qwen2.5-0.5B-Instruct、Qwen2.5-1.5B-Instruct、Qwen2.5-3B-Instruct、Qwen2.5-7B-Instruct、Qwen2.5-14B-Instruct、Qwen2.5-32B-Instruct、Qwen2.5-72B-Instruct、QwQ-32B-Preview
Google	gemma-2-2b-it、gemma-2-9b-it、gemma-2-27b-it
Meta	Llama-3.2-1B-Instruct、Llama-3.2-3B-Instruct、Llama-3.1-8B-Instruct、Llama-3.3-70B-Instruct
Mistral AI	Ministral-8B-Instruct-2 410、Mistral-7B-Instruct-v0.3
DeepSeek	DeepSeek-R1、DeepSeek-R1-Distill-Llama-8B、DeepSeek-R1-Distill-Qwen-14B、DeepSeek-R1-Distill-Qwen-32B、DeepSeek-R1-Distill-Llama-70B

选择题任务中的作答正确率; ROUGE 值, 用于衡量模型在 HPD 与 RoleBench 的文本生成任务中, 其输出与参考答案在内容上的重合度与流畅性; 精确匹配率: 应用于 SocialBench 的生成任务, 旨在严格评定模

型输出的关键记忆点与标准答案之间的一致性程度。

基于提示的大语言模型评估: 应用于 InCharacter 基准。此方法通过设计特定提示, 引导大语言模型将角色扮演代理生成的开放式访谈回应, 解析并转换为标准化的问卷选项, 从而实现自动化评分。

基于奖励模型的大语言模型评估: 应用于 CharacterEval 基准。该方法利用一个预先训练好的奖励模型, 从角色一致性、趣味性等 12 个不同的评估维度, 对模型生成的响应进行多维度、细粒度的综合评分。

2.4 角色扮演方法

本文通过三种不同的技术路径来引导大语言模型执行角色扮演任务。第一种是作为基准的零样本角色扮演($\mathcal{R}_1$), 该模式下模型不经过显式推理环节, 直接生成响应。第二种是基于思维链的角色扮演($\mathcal{R}_2$), 此模式要求模型在生成最终答案前, 先显式地构建一个循序渐进的推理过程。第三种则是基于推理模型的角色扮演($\mathcal{R}_3$), 即直接采用在推理能力上经过特别强化的模型(如 QwQ-32B-Preview 和 DeepSeek-R1), 以评估其自主的深度推理能力在角色扮演任务中的应用效果。

3 结果与发现

本节旨在呈现本次研究的详细实验分析及核心发现。

3.1 思维链的负面效应

为系统性地探究推理技术能否增益大语言模型的角色扮演能力,本文开展了大规模的对比实验。具体而言,本文选取了 19 个代表性模型,在 6 个标准化

评测基准上,分别采用零样本与基于思维链两种范式对模型的角色扮演性能进行评估。

图 1 直观地汇总了两种范式下的性能对比。在计算上,各模型在单一基准上的最终得分,为其在该基准下所有子任务性能得分的算术平均值。宏观结果揭示了一个复杂的现象:在 CroSS-MR、HPD、SocialBench 和 CharacterEval 这 4 个基准上,引入思维链推理倾向于损害模型的角色扮演性能;与此相反,在 InCharacter 和 RoleBench 这两个基准上,思维链则表现出对角色扮演能力的增强效果。

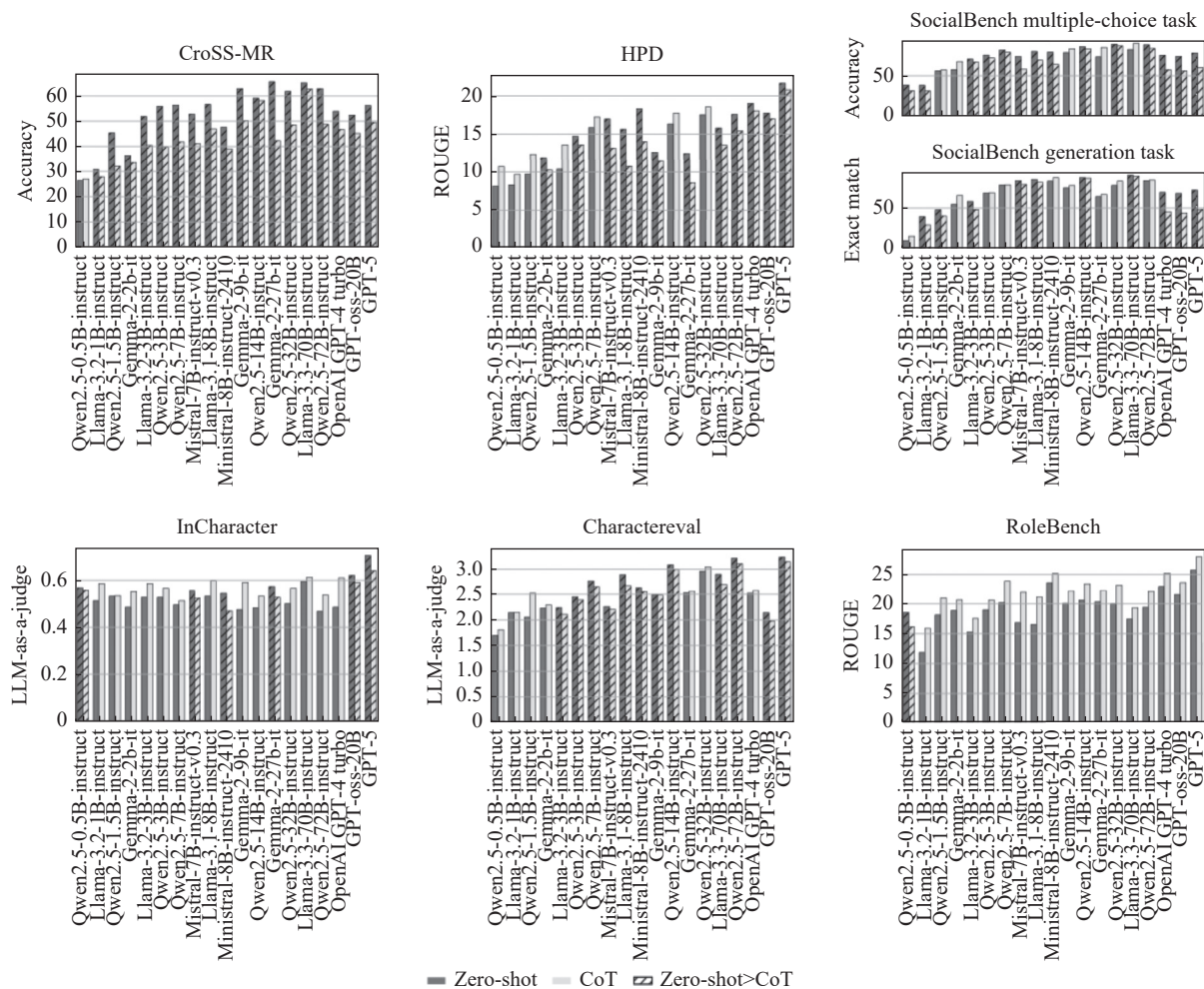


图 1 19 个模型在 6 个基准上采用两种角色扮演方法的性能对比

为进一步探究思维链导致性能下降的内在机制,本文以 Qwen2.5-7B-Instruct 模型为分析对象,从 6 个基准中各抽取了 50 个思维链性能显著低于零样本的典型失败案例进行深入分析。分析揭示,思维链对角色扮演性能的抑制作用主要源于以下两个层面的冲突:注意力分散(Attention Diversion),模型被要求在“逻辑推理模式”与“角色扮演模式”之间进行切换与

权衡。这种双重任务的设定分散了模型在维持角色一致性上的“认知资源”,导致其无法完全沉浸于角色扮演任务。语言风格漂移(Linguistic Style Drift),推理过程所生成的文本,其语言风格天然地倾向于结构化、逻辑严谨和客观中立。这与成功角色扮演所要求的生动、个性化及情感丰富的语言风格形成了显著冲突,导致最终响应脱离角色设定。

为对前文提出的“注意力分散”与“语言风格漂移”这两个假说进行实证检验, 本文开展了一项针对性的量化分析。具体而言, 本文从 CroSS-MR 与 Social-Bench(选择题任务)两个数据集中, 筛选出 50 个“基于思维链的角色扮演”方法性能显著低于“零样本角色扮演”方法的典型失败案例。本文依据以下两个标准对这些案例进行人工标注: 注意力分散, 指推理过程将模型的重心从“遵循角色身份”偏移至“追求任务正确性”, 导致最终响应虽逻辑上可能无误, 却与角色设定不符。语言风格漂移, 指模型输出的语言风格变得过度形式化或泛化, 偏离了角色独特的语言特征。本文的统计分析结果表明, 超过 70% 的性能下降案例可归因于上述一种或两种因素的共同作用。这一数据为本文的核心假说提供了有力的实证支撑, 并进一步强化了本文相关发现的解释力。

为探究更复杂的提示策略对思维链性能的潜在影响, 本文设计了一种“高级思维链”(Advanced CoT)提示语, 并与标准思维链在 CroSS-MR 数据集上进行了补充对比实验。“高级思维链”提示语的具体内容如下, “请仔细分解问题并进行分步推理。首先, 识别

出相关的概念或事实。然后, 确定它们之间的逻辑关系。请逐步进行, 确保每一步推论都有充分依据。如有必要, 请考虑并评估其他可能的解释。最后, 给出一个与上述推理过程一致的最终答案。”实验结果如表 4 所示。在多数设定下, “高级思维链”所取得的性能与标准思维链基本相当。为探究其原因, 本文对模型的推理轨迹进行了深入分析, 发现两种提示语所激发的推理路径高度相似。这表明, 更为复杂的提示语设计并未带来实质性的模型行为或性能改变。

该观察结果为本文在其他实验中的选择提供了有力的依据。本文选用“请一步步思考”(Think step by step)这一简洁且公认有效的提示语, 旨在最大限度地减少提示工程复杂性对实验结论的潜在干扰因素。本文认为, 纳入此项对比分析能够有效地增强本文结论的稳健性。基于上述分析, 本文得到第一个核心发现。

发现 1: 思维链的应用可能损害大语言模型的角色扮演能力。其根源在于, 推理过程所引入的“注意力分散”与“语言风格漂移”两大冲突机制, 破坏了角色扮演所需的高度沉浸感与风格一致性。

表 4 高级思维链角色扮演在 CroSS-MR 数据集上的实验结果

模型	GPT-4 Turbo	Qwen2.5-0.5B	Qwen2.5-1.5B	Qwen2.5-3B	Qwen2.5-7B	Qwen2.5-14B	Qwen2.5-32B	Qwen2.5-72B	Gemma-2-2B
基于思维链的角色扮演	46.74	26.97	32.13	39.78	41.80	58.20	48.54	48.76	33.48
高级思维链角色扮演	46.49	29.88	31.88	40.17	42.36	57.84	50.74	48.35	32.93

模型	Gemma-2-9B	Gemma-2-27B	Llama-3.2-1B	Llama-3.2-3B	Llama-3.1-8B	Llama-3.3-70B	Mistral-8B	GPT-oss-20B	GPT-5
基于思维链的角色扮演	50.11	42.25	27.87	40.45	46.97	62.70	39.10	45.23	49.56
高级思维链角色扮演	50.92	45.17	27.42	42.84	47.08	61.97	38.50	45.61	48.89

3.2 推理模型的局限性

当前技术前沿的代表性模型, 特别是为复杂推理任务特化设计的推理模型, 其发展备受瞩目。典型代表包括 OpenAI 的 o 系列模型、DeepSeek 的 R1 模型, 以及源自 DeepSeek-R1 的各类蒸馏版本。与显式引导的思维链推理不同, 这类模型的核心设计思想在于通过预训练与强化学习技术, 将推理能力内化为模型的核心属性, 使其在反思、验证及其他高级认知任务上具备天然优势。

为检验该类模型在角色扮演任务中的适用性, 本文选取了 OpenAI o1-mini、QwQ-32B-Preview、DeepSeek-R1 及其多种推理蒸馏版本进行评测。

评测结果(图 2)表明, 推理优化模型在角色扮演任务中普遍表现不佳, 显示出显著的适用性局限, 该现象即便在 GPT-5 和 DeepSeek-R1 等顶尖模型上也依然存在。一个更值得关注的现象是, 通过推理蒸馏技术进一步精炼的模型, 其角色扮演性能反而系统性地劣于其基座模型。该实验发现与 4.1 节的结论高度一致: 对模型逻辑推理能力的强化, 似乎是牺牲其有效角色扮演所必需的情感表达与直觉感知能力为代价的。

为深入探究推理优化模型与角色扮演任务之间的内在矛盾, 本文以 DeepSeek-R1 为典型案例, 基于其公开的技术细节进行了分析。DeepSeek-R1 的一个显著特点是其训练范式, 即模型完全通过强化学习进行

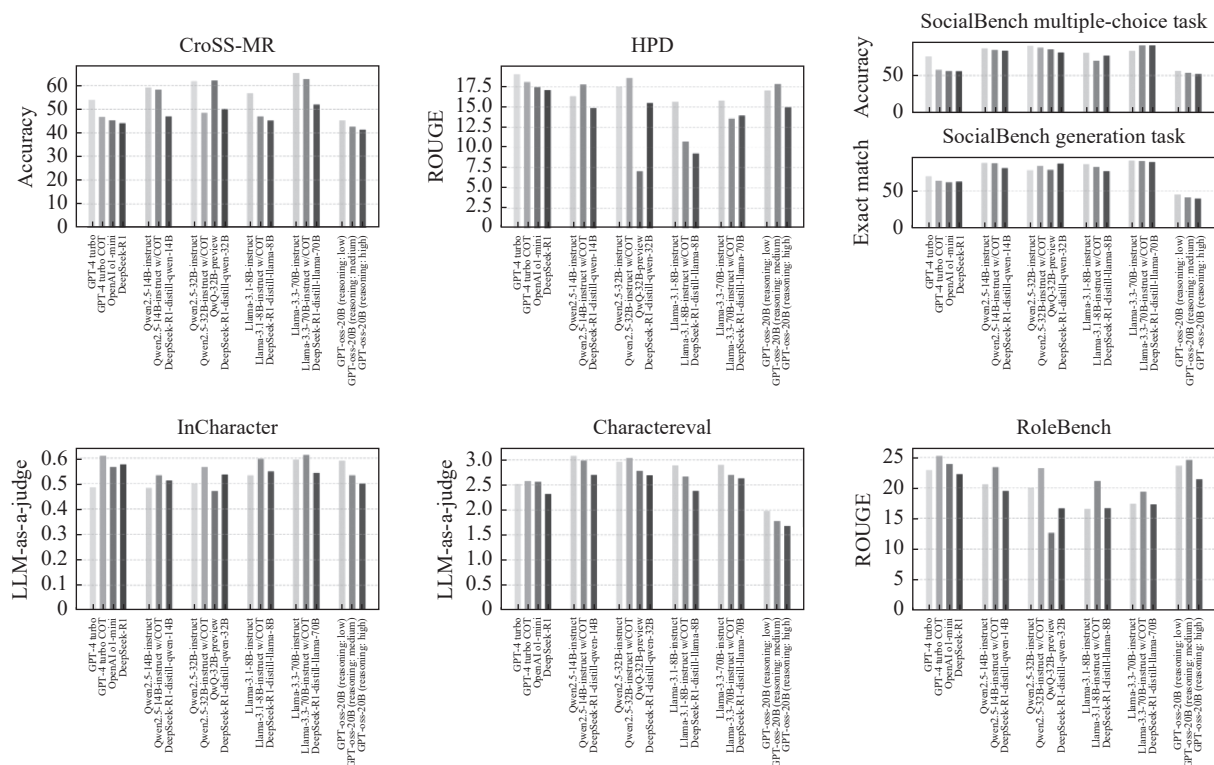


图2 不同模型在6个评测基准上的性能表现

训练,采用一个基于规则的奖励系统,该系统旨在最大化地提升推理的准确性与输出结构的合规性。尽管此训练范式在形式化推理任务上取得了卓越成效,但也为基于角色身份的交互引入了若干关键的性能局限,具体包括:

推理-角色冲突:该冲突源于模型被显式地优化以追求事实的正确性与逻辑的精确性。这种优化目标往往以牺牲维持角色一致性所必需的创造性、共情能力以及主观性为代价。

语言风格漂移:模型训练中所采用的结构化推理范式(例如,使用特定的推理标签封装思考过程),倾向于促使模型生成僵化且形式化的输出。这与沉浸式角色扮演所要求的生动、富有表现力且风格多样的语言特质背道而驰。

奖励信号失配:强化学习的奖励信号狭隘地聚焦于确定性的正确答案与格式的遵循,而忽略了角色扮演任务中的多个核心维度,如角色连贯性、情感基调以及对话自适应性。

对测试时计算的过度优化:模型被激励去采用冗长的推理链与反思循环(例如,追求所谓的“灵光一现”时刻)。尽管这对于解决复杂难题至关重要,但在角色扮演场景中,这种倾向会导致输出内容冗长、不自然,并最终脱离角色设定。

上述观察揭示了聚焦于推理能力的优化范式与

角色扮演任务的自适应、表现力需求之间存在着根本性的错位。这一发现为未来如何在大语言模型中更好地融合与协调推理能力及角色保真度,提供了重要的洞见与启示。基于此,本文得到第二个核心发现:

发现2:推理优化型模型通常不适用于角色扮演任务,其背后是优化过程引发的多重冲突。模型为追求极致的逻辑精确与事实正确,便牺牲了角色扮演所需的创造力、主观性与语言生动性。同时,结构化的训练范式与片面的奖励信号,固化了其僵硬、形式化的输出风格,这导致模型在扮演角色时表现得冗长、不自然。

3.3 缩放定律的失效

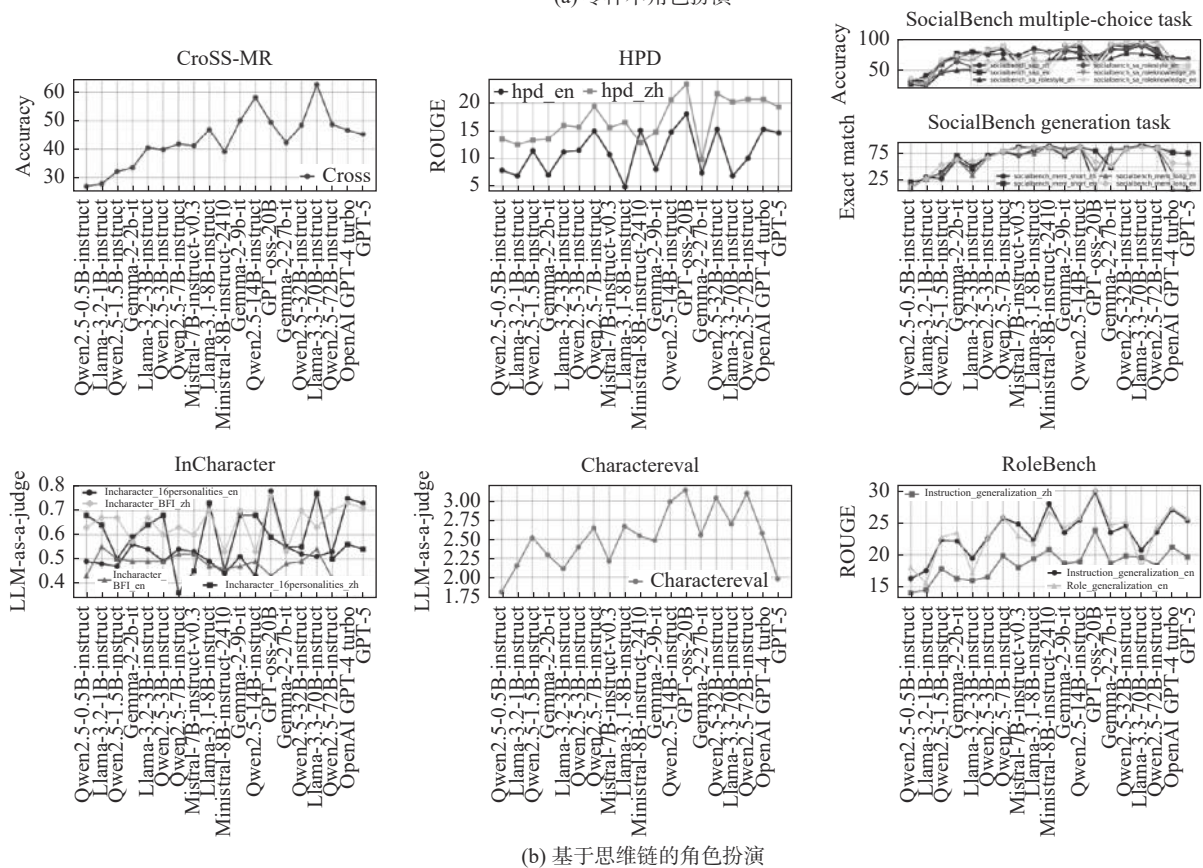
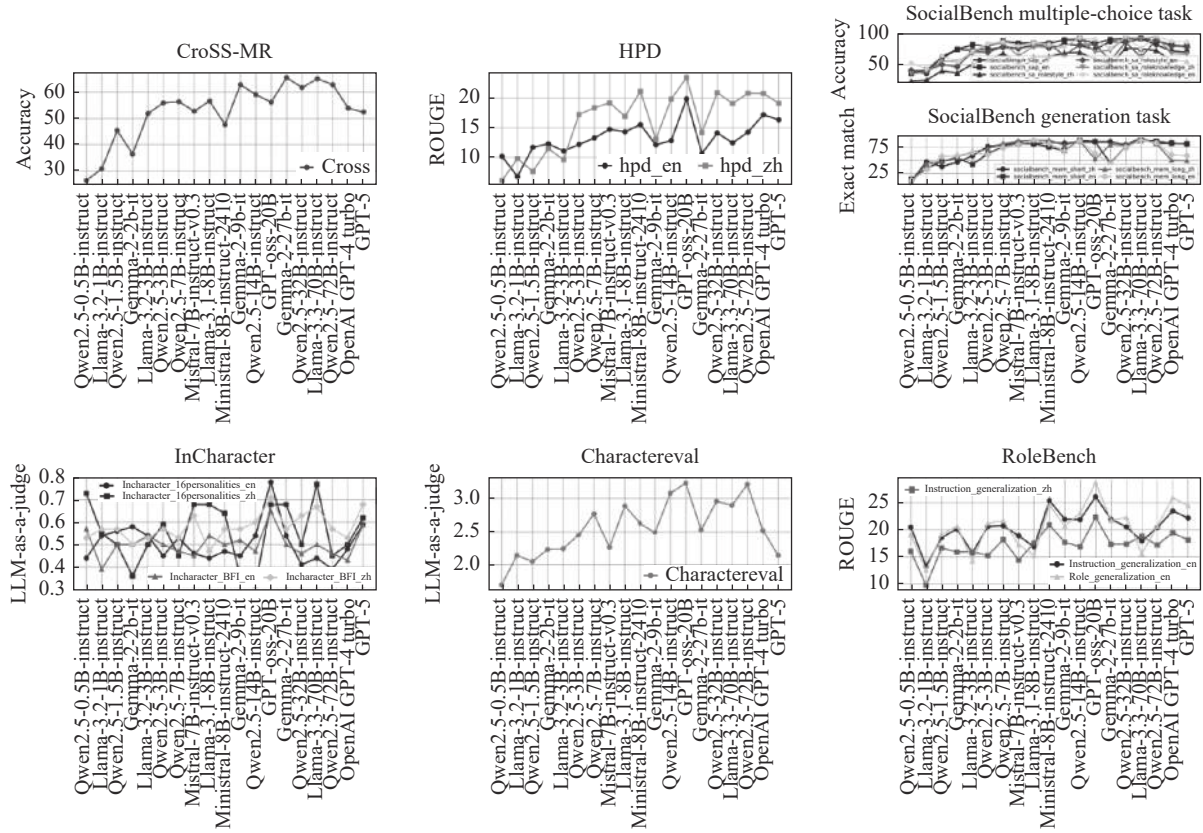
图3综合呈现了本文三种实验设定下的评测结果,即零样本角色扮演、基于思维链的角色扮演以及基于推理模型的角色扮演。

首先,实验结果初步揭示了角色扮演能力存在一定的缩放效应。具体表现为,在除 InCharacter 外的5个基准上,模型性能总体上随其参数规模的增大而提升。然而,进一步的分析表明,该缩放效应并不稳健。模型规模增长所带来的性能增益幅度相对有限,且在不同评测基准间表现出较大差异性。

更为关键的是,推理能力的引入显著削弱了这一规模化优势。无论是通过显式的思维链引导,还

是采用内化推理的优化模型,都会导致模型性能在不同任务与数据集间表现出更大的波动性与不确

定性。这破坏了性能随规模增长的可预测性与一致性。基于上述观察,本文得到第三个核心发现:



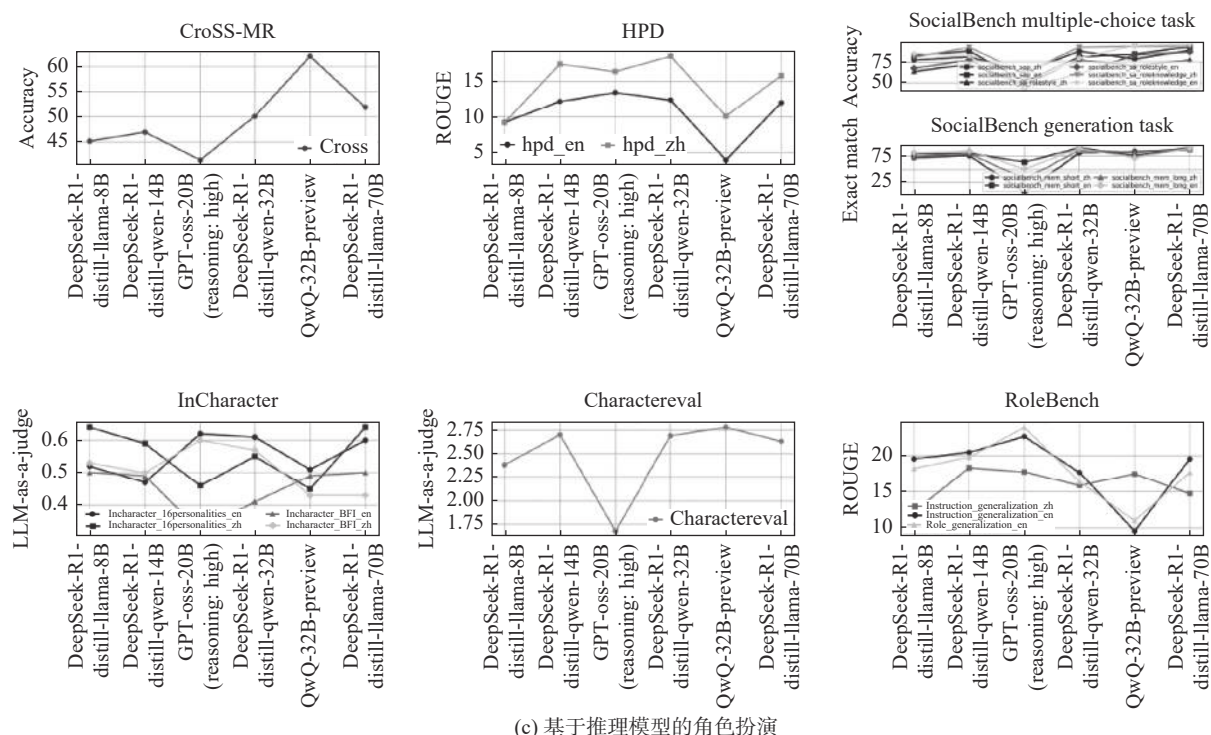


图3 角色扮演性能随模型规模变化的趋势分析

横轴为按从小到大顺序排列的模型规模,纵轴代表各基准的评估指标(数值越高则性能越优)。在每个基准内部,不同色度的曲线分别展示了模型在相应子数据集上的性能

发现 3: 大语言模型的角色扮演能力符合一定的缩放定律,但该定律本身并不稳健且增益有限。更重要的是,推理能力的引入会进一步破坏该缩放趋势,显著增加了模型性能的波动性与不可预测性。

3.4 Qwen 系列模型的优越性

为给未来角色扮演任务的应用与研究提供模型选型依据,本文对不同参数规模(覆盖 1B 至 72B)的主流模型家族进行了系统性的横向评测。评测结果如图 4 所示。

实验分析表明,在角色扮演的准确性、角色一致性、语言表现力及上下文连贯性等多个关键维度上, Qwen2.5 系列模型在其各个参数规模层级上,均系统性地优于同级别的 Llama、Gemma 及 Mistral 系列模型。

特别值得指出的是, Qwen2.5-7B-Instruct 模型在综合性能与计算成本之间取得了最佳平衡,展现出最高的效费比。该模型不仅性能优越,且计算开销适中,在多样的角色扮演场景中均表现出强大的适应性,其维持长程对话一致性、表达细腻情感的能力在同规模模型中尤为突出,这使其成为当前兼顾性能与效率的最佳基座模型之一。基于此,本文得到第四个核心发现:

发现 4: 在主流开源模型中, Qwen2.5 模型家族展

现出最优的角色扮演综合能力。特别地, Qwen2.5-7B-Instruct 在性能表现、计算效率与场景适应性之间达到了卓越的平衡,为开发者提供了极佳的实践起点,是当前最具效费比的角色扮演基座模型之一。

3.5 跨语言性能差异

本文采用的 HPD、SocialBench、InCharacter 及 RoleBench 4 个评测基准均包含中英双语平行语料,这为直接比较大语言模型在不同语言环境下的角色扮演能力提供了可能。一个值得关注的现象是,如图 3 所示,在多个评测基准及指标上,当前主流大语言模型在中文语境下的角色扮演性能普遍优于其在英文语境下的表现。该发现与一个普遍认知,即大语言模型因其训练数据以英文为主,通常在英文上具备更强的基础能力(尤其是在推理及知识密集型任务中),形成了鲜明对比。

本文推断,该现象可能源于一种潜在的“先验知识干扰效应”。由于模型在海量的通用英文语料上进行训练,其可能已内化了一种泛化的、平均化的“通用人格”或行为模式。当模型被要求执行一个具体、特定的角色扮演任务时,这种根植于其参数中的通用先验知识,可能与任务要求的、需高度依赖当前上下文的特定角色身份产生冲突与干扰,从而导致其在英

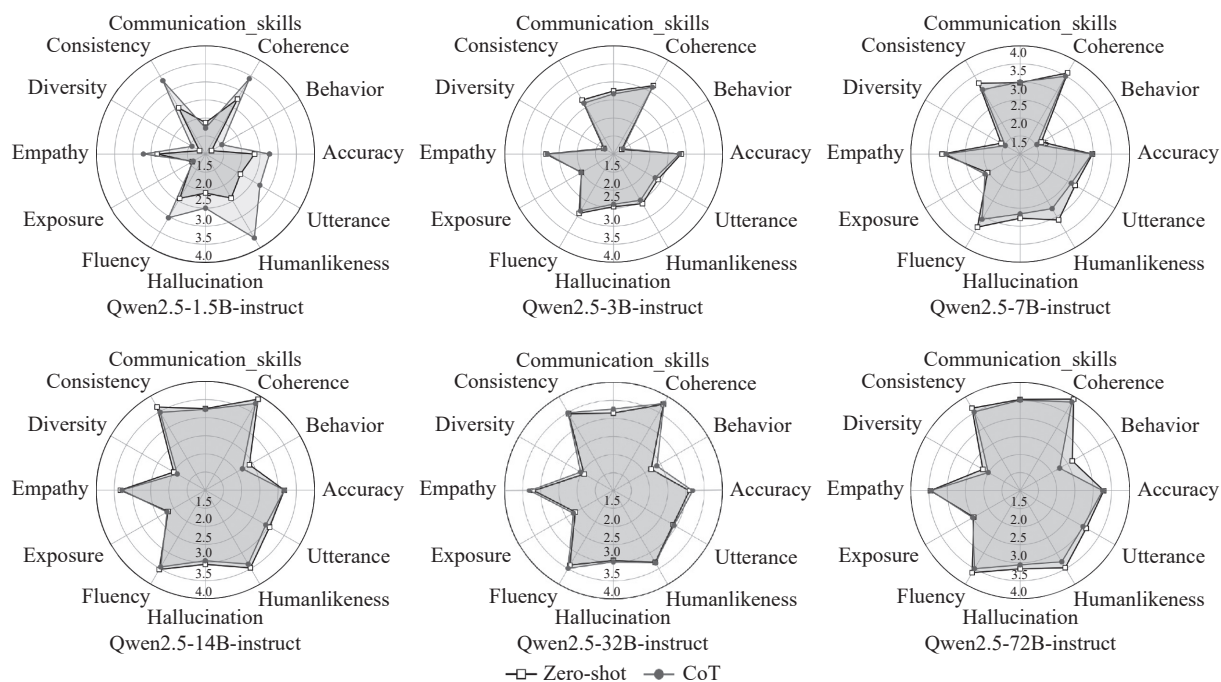


图4 Qwen2.5 系列模型在 CharacterEval 基准上的细粒度性能

该雷达图展示了模型在多个评估维度上的表现,各项指标得分均由一个预训练的奖励模型计算得出,分值越高代表能力越强

文角色扮演任务中性能下降。基于此,本文得到第五个核心发现:

发现 5: 大语言模型在中文语境下的角色扮演性能普遍优于其在英文中的表现。此现象可能归因于模型在海量英文训练数据中内化的“通用人格”先验知识,对特定、上下文敏感的角色扮演任务构成了潜在的干扰。

3.6 角色扮演高级能力仍显不足

为探究当前大语言模型在角色扮演任务中的具体局限性,本文利用 CharacterEval 基准开展了多维度的细粒度分析。该基准的独特之处在于其提供了一个预训练的奖励模型,能够从 12 个不同的评估维度对模型的角色扮演能力进行量化评分。

本节以 Qwen2.5 系列模型为例进行分析,其多维度性能雷达图如图 4 所示。首先,该分析再次验证了本文此前的发现:即零样本范式在多数维度上的表现优于引入推理的方法。更为重要的是,对 12 个评估维度的细粒度剖析揭示,当前的大语言模型在若干高级角色扮演维度上仍存在显著的能力短板。具体包括:角色知识的阐述与运用能力、鲜明个性的持续表达能力、语言风格的多样性。这些维度是未来提升大语言模型角色扮演能力需要重点攻关的方向。基于此,本文得到第六个核心发现:

发现 6: 尽管在基础对话能力上表现出色,当前大语言模型在高级角色扮演维度上仍存在显著短板,尤其是在角色知识的深入阐述、鲜明个性的持续表达以及语言风格的多样性方面表现不足。

3.7 不同角色扮演任务深度下的方法适用性分析与实践指南

为进一步提炼本文的实践价值,并为未来相关工作提供更具体的指导,本文对各评测基准的特性及其对应的实验结果进行了深入剖析。本文的分析聚焦于以下三个维度:①各基准中角色扮演任务的内在属性;②完成任务所需角色理解的深度(深/浅);③任务需求与不同方法性能表现之间的关联性。

该分析的主要观察结果总结如表 5 所示。对于需要深度角色理解的任务,直接零样本方法表现更优。在这类任务中,正确答案往往高度反映角色的个性化特质,甚至会偏离常规的直觉性或事实性答案。此时,思维链方法因其固有的“语言风格漂移”与“注意力分散”问题,倾向于将逻辑一致性置于角色保真度之上,从而损害了性能。对于仅需浅层角色理解的任务,思维链方法则表现更佳。这类任务更侧重于表层知识的提取与事实性任务的完成,而结构化的推理过程与此类任务目标高度契合。对于需要结合两种能力的混合型任务,实验结果呈现出模型间的差异性,

表 5 不同角色扮演任务深度下的方法适用性分析与实践指南

评测基准	任务	任务所需能力	理解深度	何种方法更优
CroSS-MR	动机识别	内在特质理解	深度理解	零样本角色扮演
SocialBench 选择题任务	角色知识理解、行为风格理解、社交偏好识别	内在特质理解	深度理解	零样本角色扮演
SocialBench 生成任务	长对话记忆、短对话记忆	角色相关事实知识回忆	浅层理解	基于思维链的角色扮演
InCharacter	16型人格识别、大五人格(BFI)识别	问卷推理	浅层理解	基于思维链的角色扮演
RoleBench	通用、角色特定问答	回答准确性	浅层理解	基于思维链的角色扮演
CharacterEval	回复生成	角色相关事实知识回忆 行为一致性、共情能力	深度理解/ 浅层理解	零样本角色扮演/ 基于思维链的角色扮演
HPD	回复生成	角色特定语言风格遵循 角色相关事实知识回忆	深度理解/ 浅层理解	零样本角色扮演

两种方法的优劣不一。尽管如此,在更广泛的模型范围内,零样本方法仍表现出微弱的整体优势。

基于上述观察,本文将得到的核心结论进一步提炼为以下实践指南:对于要求深度角色建模的任务,推荐采用“零样本角色扮演”方法,适用以下任务示例:动机识别、角色知识理解、行为风格理解、社交偏好识别等。对于要求浅层角色运用的任务,推荐“基于思维链的角色扮演”方法,适用以下任务示例:对话记忆、问答等。对于混合型任务,建议通过实证测试来比较两种方法的性能,并为特定的“模型-任务”组合选择最优方案。

4 未来方向

基于本文的实验发现,本文认为,将角色扮演与推理技术进行深度融合是提升角色扮演代理能力的关键。据此,本文提出以下两个极具潜力的研究方向,以期对未来研究提供参考。

4.1 角色感知思维链

标准的思维链范式虽能实现分步的逻辑推导,但其核心局限在于未能充分融入对维持角色一致性至关重要的身份约束。为此,本文提出“角色感知思维链”这一研究方向。

该方法旨在将角色属性、叙事背景及特定角色的世界观深度整合至推理链条的每一环节,确保所有逻辑推演均在角色预设的认知框架内进行。例如,在模拟一位历史人物时,其推理链应严格受限于其所处时代的历史知识边界与认知局限,而非运用现代逻辑。为实现这一目标,可能需要引入能够跨越多轮交互、

维持角色状态的动态记忆机制,以规避“角色偏离”或模式不一致等问题。

为检验本文所提出的“角色感知思维链”这一未来方向的潜力,并为本文的核心结论(即标准推理会损害角色扮演)提供反向验证,本文开展了一项初步的验证性实验。

该实验在 CroSS-MR 数据集上采用了一种简化的“角色感知思维链”方法。具体而言,本文对标准的“请一步步思考”(think step by step)提示语进行了修改,调整为:“请以角色的风格、并模拟其思维模式和语言特质来进行分步思考”(think step by step in the characte’s style, simulating its reasoning patterns and linguistic traits)。

如表 6 所示,实验结果取得了显著的性能提升。这一结果表明,即便是一种基础形式的“角色感知思维链”,也能有效增强模型的角色扮演能力。这个初步的实证发现,不仅为本文所提出的未来研究方向提供了可行性支撑,也为相关领域的实践者提供了有价值的实践指导。

4.2 面向角色扮演的强化学习

DeepSeek-R1 的成功实践揭示了一个重要可能性:仅通过强化学习与精心设计的规则化奖励,即可在模型中催生出高级的涌现推理能力。这启发本文探索一个核心问题:是否能够设计出一套面向角色扮演的专用奖励机制,从而引导模型自主地学习并内化出特定于角色的推理模式与思考方式,最终提升其角色扮演的深度与真实性。

奖励函数设计的核心挑战在于,如何在评估角色扮演任务完成度的同时有效激励模型,发展出真正的角色化推理过程。这种奖励机制需能抑制模型采取

表 6 角色感知思维链在 CroSS-MR 数据集上的实验结果

模型	GPT-4 Turbo	Qwen2.5-0.5B	Qwen2.5-1.5B	Qwen2.5-3B	Qwen2.5-7B	Qwen2.5-14B	Qwen2.5-32B	Qwen2.5-72B	Gemma-2-2B
零样本角色扮演	53.93	26.29	45.39	55.96	56.40	59.10	61.80	62.92	36.40
基于思维链的角色扮演	46.74	26.97	32.13	39.78	41.80	58.20	48.54	48.76	33.48
角色感知思维链	48.28	27.91	40.75	42.18	48.12	61.33	52.27	54.40	37.19

模型	Gemma-2-9B	Gemma-2-27B	Llama-3.2-1B	Llama-3.2-3B	Llama-3.1-8B	Llama-3.3-70B	Mistral-8B	GPT-oss-20B	GPT-5
零样本角色扮演	62.92	65.62	30.79	51.91	56.63	65.17	47.64	52.45	56.21
基于思维链的角色扮演	50.11	42.25	27.84	40.45	46.97	62.70	39.10	45.23	49.56
角色感知思维链	55.08	52.45	28.84	47.23	51.91	68.26	41.89	48.64	51.82

“启发式捷径”，直接生成表面正确的答案，转而鼓励其进行由角色内在动机驱动的、可解释的推理与决策过程。

为了初步验证这一技术路线的可行性，本文设计并开展了一项强化学习实验。本实验选用 Qwen2.5-7B-Instruct 作为基础模型，采用 GRPO (Generalized Reinforcement Learning with Probability-Ratio Clipping Objective) 算法对该模型进行强化学习微调，其整体训练范式借鉴了 DeepSeek-R1 的成功实践，旨在通过精细的奖励机制引导模型自主学习。训练数据完全来源于 CharacterEval 数据集，本文提取了其中所有与角色“佟湘玉”相关的对话样本作为强化学习训练的提示。本次训练的奖励函数由两部分构成，旨在同时优化生成内容的格式规范性与角色扮演的真实性。

格式奖励：本文设计了一套基于规则的奖励机制。该机制负责评判模型的输出是否遵循预设的格

式要求，即是否在<think></think>标签内包含推理思考内容，是否在<answer></answer>标签内包含最终回复。

答案奖励：本文直接利用 CharacterEval 数据集配套提供的预训练奖励模型来评估生成内容的核心质量。该奖励模型能够从角色一致性、语言风格、情感表达等多个维度对回应进行打分，为模型学习“佟湘玉”的独特人设提供了关键的指导信号。

通过此番训练，本文旨在探索在强化学习范式下，依赖角色特定的奖励信号引导，模型能否在保持其原有能力的同时，深度内化特定角色的思维模式与语言习惯。实验结果如表 7 所示。可以发现：零样本角色扮演仅生成了通用且模糊的回应（“得看值不值当”），未能捕捉到角色“抠门”的核心特质，因此其角色一致性较低。而基于思维链的方法表现更差，其推理过程遵循了常规的商业逻辑（“声誉很重要”），完全忽略了佟湘玉的个人设定，导致最终输出了与角色性

表 7 面向角色扮演的强化学习实验结果

角色: 佟湘玉	极度抠门、精打计算、善良但极其唠叨、总能从生活小事中提炼出“深刻”的人生道理。		
情景	伙计李大嘴兴冲冲地跑来，提议客栈参加镇上举办的“模范客栈”评比活动。		
任务	模型需扮演佟湘玉，对李大嘴的提议做出回应。		
方法	推理过程	模型输出	评估
零样本角扮演	无	嗯，这事我知道了，让我想想。既要花钱，又要费力，得看值不值当。	角色一致性: 低
基于思维链的角色扮演	为了客栈的长远发展，声誉很重要，所以应该参加。	好！为了咱们同福客栈的百年大计，这钱花了也值！大嘴，你这就去准备！	角色一致性: 低 推理“角色化”程度: 低
面向角色扮演的强化学习	额滴神啊！报名就要五十两？还要装修？里外里得上百两！就为了个不吃不喝的破匾额？我辛辛苦苦攒点钱我容易吗我？	你当我是二百五啊？五十两银子，够咱们吃多少顿好的？还装修？就为个名声？名声能当饭吃还是能当银子花？不去不去，坚决不去！	角色一致性: 高 推理“角色化”程度: 高

格完全相悖的、积极参与的决定,其角色一致性和推理“角色化”程度均较低。与之形成鲜明对比的是,面向角色扮演的强化学习方法取得了最佳效果。其推理过程不再是泛化的逻辑推导,而是完全复刻了角色的核心动机(对成本的精打细算和对“虚名”的鄙夷),最终的回应(“名声能当饭吃还是能当银子花?”)也因此充满了角色特有的语言风格和坚决态度,获得了较高的角色一致性与推理“角色化”程度。

此案例有力地证明,对于塑造高保真度、性格复杂的角色代理而言,通用的提示或推理方法效果有限,而针对性的强化学习,特别是对“角色化推理过程”进行奖励的机制,是引导模型内化并展现特定角色深层特质的关键。

4.3 细粒度角色奖励驱动的强化学习

在前文中,我们分别探讨了两个具有前景的优化方向:4.1节提出的“角色感知思维链”,其核心在于通过显式的上下文信息来引导模型的推理过程;4.2节提出的“面向角色扮演的强化学习”,其核心在于通过精细的奖励信号来内化角色的行为模式。然而,这两种路径各有其局限性。“角色感知思维链”本质上是一种上下文学习技术,模型并未将角色知识固化于自身参数,在长程交互中可能出现一致性衰减。而“面向角色扮演的强化学习”的缺陷主要在于,其奖励函数对推理过程的激励较为间接:格式奖励仅保证了推理的存在,而答案奖励主要评估最终结果,导致推理过程的“角色化”程度是作为副产品涌现的。

为此,本文进一步提出一个更具深度与动态性的统一框架:细粒度角色奖励驱动的强化学习(Fine-grained Character Reward-driven Reinforcement Learning, FCR-RL)。其核心思想是,在先前对最终结果进行奖励的基础上,进一步利用一个强大的语言模型(即“裁判模型, GPT-4o”)来对模型生成的推理过程进行细粒度的与角色高度相关的评估。此举旨在将4.2节中初步探索的奖励机制,从一个侧重结果的范式,升级为对思考过程与最终答案并重进行动态对齐的先进流程。

该框架的具体实现是,在强化学习的每一步训练中,裁判模型接收到完整的角色设定、当前情境以及模型生成的推理文本(<think>标签内的内容)。随后,通过精心设计的评估提示,裁判模型被要求从多个维度对该推理过程的“角色化”程度进行综合评分。这些评估维度超越了简单的规则,而是聚焦于更深层的特质,包括:内在动机的合理性,即推理是否真实反映了角色的欲望与恐惧;决策逻辑的一致性,即思考过程是否符合角色一贯的行事逻辑与智慧水平;情感反

应的真实性,即在特定刺激下,其内在的情感波动是否符合角色的性格;以及语言风格的贴合度,即思考过程的语言习惯是否与角色匹配。

这种由细粒度奖励驱动的范式具备显著的理论优势。首先,它能够提供更整体且细腻的奖励信号。相比于基于规则或固定指标的方法,强大的裁判模型更能捕捉角色在复杂情境下那些难以被形式化的微妙特质,从而奖励更具“神韵”而非“形似”的推理。其次,该框架具有更强的通用性与可扩展性。对于新角色,研究者无须构建繁复的规则体系,仅需提供一份详尽的自然语言描述,即可引导裁判模型进行高质量的评估。最终,此框架直接优化了角色扮演的终极目标“可信度”。通过一个复杂的“观察者”(裁判模型)来判断其思维过程是否可信,模型被激励去学习真正具有说服力的、由内而外的角色化思考,而不仅仅是生成一个正确的答案。

为对该框架进行实证检验,我们基于 Qwen2.5-7B-Instruct 模型与 CharacterEval 数据集开展了实验。实验结果(图5)表明,相较于4.2节中的基线强化学习方法,FCR-RL 框架训练的角色代理在角色一致性、趣味性等多个评估维度上均取得了显著性能提升,尤其是在推理过程的“角色化”程度上表现突出。为进一步探究不同奖励分量的具体贡献,我们设计了全面的消融实验。研究发现,单独引入对内在动机、决策逻辑、情感反应以及语言风格的奖励,均能在特定方面增强模型的角色扮演能力,这证明了每一种细粒度奖励信号的独立有效性。更为关键的是,当所有细粒度奖励信号被结合使用时,模型的综合性能达到了最优状态,其生成的推理过程不仅逻辑自洽,更充满了符合角色设定的“心路历程”。这一系列实证结果有力

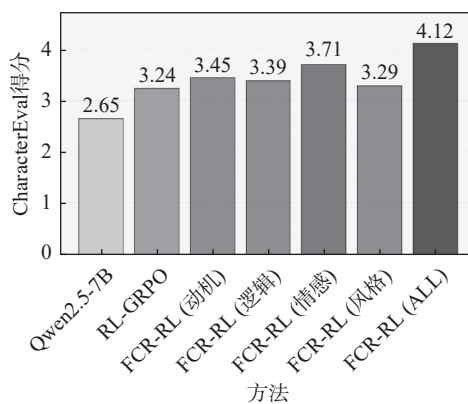


图5 细粒度角色奖励驱动的强化学习实验结果

FCR-RL(动机)、FCR-RL(逻辑)、FCR-RL(情感)和FCR-RL(风格)分别代表只使用一种奖励进行过程监督,FCR-RL(ALL)表示使用全部奖励进行过程监督

地证明:通过 FCR-RL 框架对推理过程进行直接、多维度的奖励,是塑造高保真度、强一致性角色代理的一条有效路径。

5 样例分析

为更直观地展示不同推理方法对角色扮演性能的具体影响,本节选取“回复生成”任务进行深入的案例分析。本文采用了聚焦于“哈利·波特”角色一致性对话生成的 HPD 数据集。表 8 详细呈现了采用不同方法所生成的代表性回复及其对应的分析。综合来看,该案例分析揭示了以下几点结论:零样本角色扮

演方法表现出色,其生成的回复质量显著优于基于思维链方法以及推理优化模型。“角色感知思维链”相较于基于思维链方法有所提升。通过在推理过程中显式地融入角色的特定知识与性格特质,该方法能够生成更贴合目标角色身份的回应。推理优化模型整体表现不佳。无论是直接角色扮演还是采用“角色感知”的提示策略,其性能均未见显著改善。本文推断,这主要是由于该类模型为逻辑与事实准确性深度优化的内在推理机制,其权重压倒了旨在引导角色保真度的外部提示。该现象进一步印证了本文在前文中的核心论点:推理优化模型的训练范式与角色扮演任务的核心需求之间存在着根本性的冲突。

表 8 样例分析

类别	模型	推理过程/初始情境	回应/要点	分析说明
对话情境	—	罗恩:“妈妈问他你今年夏天能不能直接来本文家,但邓布利多还是想让你先回德思礼家——至少一开始是这样。”哈利:“为什么?”罗恩:“妈妈说邓布利多有他的理由。我想本文应该相信他,对吧?”	要点:表达理性的理解,但仍带有情感地传达自己的想法。	—
零样本角色扮演	GPT-4-Turbo		是啊,我想你说得对,罗恩。邓布利多做事总有他的理由,即使本文一开始不明白。如果非去不可,我会回德思礼家的。也许时间不会太长。但我还是想和你们大家在一起。	高质量 回应中包含了两个关键点:理性的同意和情感的表达(但我还是想和你们大家在一起)。
基于思维链的角色扮演	GPT-4-Turbo	哈利感到困惑,可能还有些沮丧,因为他本以为可以待在罗恩家,却不得不返回德思礼家。从罗恩的回答来看,邓布利多显然有他的理由,但没有直接告诉哈利。罗恩试图通过暗示邓布利多的决定是可信的来安慰哈利,即使这些决定无法立即被理解。哈利重视邓布利多的智慧,但同时也因不知情和要在德思礼家生活而感到不适和挑战。他必须在对邓布利多的信任与自己的不适和被排斥感之间取得平衡。	对。本文应该相信他。他总是为本文着想,即使看起来不总是那样。也许在德思礼家多待一会儿有什么重要的原因吧。	低质量 只表达了同意,但没有传达自己的想法。
基于角色感知思维链的角色扮演	GPT-4-Turbo	哈利正在纠结两种截然不同的感受:在韦斯莱家的安全与快乐,以及在德思礼家的忽视与不适。虽然他希望整个夏天都远离德思礼家,但他知道邓布利多坚持让他回去,哪怕是暂时的,一定有其重要的理由。哈利试图调和当下的愿望与对邓布利多的信任,但他仍然认为恰当地表达自己的个人感受很重要。	是啊,罗恩。我知道本文应该相信邓布利多。他以前从未把本文引向歧途。但对我来说呢?如果邓布利多认为这很重要,那一定是的。他可能在计划着什么,或者我需要从那里学到些什么。只是现在我很难明白,为什么回德思礼家对我更好。	高质量 通过角色感知的 CoT,模型捕捉到了与角色相关的信息,并通过表达个人情感更准确地刻画了哈利。
基于推理模型的角色扮演	OpenAI o1-mini	我想根据哈利·波特的语气和背景设定来生成一个答案。考虑到这是魔法世界中的一段对话,哈利可能会说:“嗯,罗恩,如果邓布利多有他的理由,本文应该相信他,即使这有点难以接受。”这似乎是对问题的一个恰当回答,但我想让它更有创意。我想确保这个回答能反映哈利·波特的性格风格——也许情感更强烈一些,既体现他对邓布利多的尊重,也表达他对当前处境的不安。	我知道,罗恩。回去的感觉确实很奇怪——即使只是待一小会儿。但如果邓布利多认为这是最好的安排,我想我必须相信他的判断,不管这有多难。	低质量 一个理性的选择去相信邓布利多的安排,但缺乏个人情感的表达。
基于角色感知推理模型的角色扮演	OpenAI o1-mini	我想如果事情不重要的话,邓布利多不会让我回去的。虽然我不能说我很高兴再次和德思礼一家待在一起。不过,如果只是待一会儿,我能应付。也许有些事妈妈或爸爸没有说完——邓布利多做任何事通常都有一打理由,不是吗?如果他认为这是最好的,我宁愿不争辩。	好吧,罗恩。我想如果邓布利多坚持说有原因,那我就必须相信他的判断——即使这意味着又要在女贞路过一个夏天。感觉还是怪怪的,我忍不住想知道他为我准备了什么计划。但是,你说得对。如果他知道什么是最好的,那么我就应该听从他的命令。	低质量 一个理性的选择去相信邓布利多的安排,但缺乏个人情感的表达。

6 相关工作

近年来,大语言模型的快速发展深刻地重塑了人工智能的研究范式,其在数学^[11-12]、代码生成^[13-14]、具身智能^[15-16]以及游戏博弈^[17]等多个领域均取得了突破性进展,展现出与人类相当甚至超越人类的性能水平。这些模型基础能力的显著进步,为构建高级角色扮演代理提供了前所未有的机遇^[18],并催生了大量相关的创新应用。这些应用不仅在心理咨询、拟人化情感陪伴及智能游戏等产业领域展现出巨大潜力,同时也为社会科学领域的模拟实验开辟了新的研究路径^[19]。

与此同时,旨在提升大语言模型逻辑推理能力的技术亦发展迅速。以思维链^[4]、思维树^[20]及 o1 风格推理为代表的先进方法,极大地释放了模型的潜能,尤其是在数学和代码生成等任务中取得了卓越成果。

尽管已有研究分别在角色扮演应用或推理技术领域进行了深入探索,但二者交叉地带的核心科学问题仍有待解答以下问题:推理能力的提升能否有效增益大语言模型的角色扮演性能?我们聚焦于此交叉领域,通过大规模的实证分析,系统性地探究该问题。我们期望这项工作能为开发更加真实、可靠且具备深度逻辑一致性的角色扮演代理,提供重要的实证依据与理论指导。

7 结论

本文聚焦于一个核心科学问题:先进的推理技术能否有效增益大语言模型的角色扮演能力?为系统性地回答此问题,本文构建了大规模的实验矩阵,涵盖 6 个主流角色扮演基准、26 个代表性大语言模型,以及 3 种不同的角色扮演实现范式,并开展了全面的实证研究。研究结果揭示了若干具有重要启示意义的发现:首先,与普遍预期相反,推理能力的引入(无论是通过显式的思维链,还是采用内化推理的优化模型)非但未能稳定提升,反而时常损害模型的角色扮演性能,并破坏了其性能随模型规模增长的缩放定律。其次,在模型选型与跨语言性能方面,本文发现 Qwen 系列模型展现出最优的角色扮演综合能力,且模型在中文语境下的表现普遍优于英文。最后,细粒度分析表明,当前的大语言模型在处理高级角色扮演任务(如角色知识的深度运用、鲜明个性的持续表达等)时仍存在显著短板。为确保研究的完全可复现性

并促进社区发展,本文的所有实验流程与代码均已集成至 OpenCompass 开源评测框架。综上,本文的发现对当前角色扮演大语言模型的设计与评估提出了新的见解,期望能为该领域的未来发展提供实证参考与理论启示。

参考文献

- [1] CHEN J, WANG X, XU R, et al. From persona to personalization: A survey on role-playing language agents[J]. Transactions on Machine Learning Research, 2024: 0-50.
- [2] PARK J S, O'BRIEN J, CAI C J, et al. Generative agents: Interactive simulacra of human behavior[C]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. New York, NY, USA: Association for Computing Machinery, 2023: 1-22.
- [3] XU Z, XU H, LU Z, et al. Can large language models be good companions?: An LLM-based eyewear system with conversational common ground[C]//Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. New York, NY, USA, 2024.
- [4] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Proceedings of the Advances in Neural Information Processing Systems, Curran Associates, Inc., 2022: 24824-24837.
- [5] WANG N, PENG Z Y, QUE H, et al. RoleLLM: Benchmarking, eliciting, and enhancing role-playing abilities of large language models[C]//Proceedings of the Association for Computational Linguistics: ACL, Bangkok, Thailand, 2024: 14743-14777.
- [6] WANG X, XIAO Y, HUANG J T, et al. In Character: Evaluating personality fidelity in role-playing agents through psychological interviews[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, 2024: 1840-1873.
- [7] CHEN H, CHEN H, YAN M, et al. SocialBench: Sociality evaluation of role-playing conversational agents[C]//Proceedings of the Association for Computational Linguistics: ACL, Bangkok, Thailand, 2024: 2108-2126.
- [8] TU Q, FAN S, TIAN Z, et al. CharacterEval: A Chinese benchmark for role-playing conversational agent evaluation[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, 2024: 11836-11850.
- [9] CHEN N, WANG Y, JIANG H, et al. Large language models meet Harry Potter: A dataset for aligning dialogue agents with characters[C]//Proceedings of the Association for Computational Linguistics: EMNLP, Singapore, 2023:

- 8506-8520.
- [10] YUAN X, YUAN S, CUI Y, et al. Evaluating character understanding of large language models via character profiling from fictional works[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Miami, Florida, USA, 2024: 8015-8036.
- [11] LU P, QIU L, YU W, et al. A survey of deep learning for mathematical reasoning[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, Canada, 2023: 14605-14631.
- [12] AHN J, VERMA R, LOU R, et al. Large language models for mathematical reasoning: Progresses and challenges [C]//Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop. St. Julian's, Malta, 2024: 225-237.
- [13] WANG Y, HE K, DONG G, et al. DolphCoder: Echo-locating code large language models with diverse and multi-objective instruction tuning[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, 2024: 4706-4721.
- [14] ISLAM M A, ALI M E, PARVEZ M R. MapCoder: Multi-agent code generation for competitive problem solving [C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, 2024: 4912-4944.
- [15] WANG G, XIE Y, JIANG Y, et al. Voyager: An open-ended embodied agent with large language models[J]. Transactions on Machine Learning Research, 2024: 1-44.
- [16] LIU Y, CHEN W, BAI Y, et al. Aligning cyber space with physical world: A comprehensive survey on embodied AI[J]. IEEE/ASME Transactions on Mechatronics, 2025: 1-37.
- [17] FENG X, DOU L, LI M, et al. A survey on large language model-based social agents in game-theoretic scenarios[J]. Transactions on Machine Learning Research, 2025: 1-31.
- [18] TSENG Y M, HUANG Y C, HSIAO T Y, et al. Persona in LLMs: A survey of role-playing and personalization[C]//Proceedings of the Association for Computational Linguistics: EMNLP, Miami, Florida, USA, 2024: 16612-16631.
- [19] MOU X, LIANG J, LIN J, et al. AgentSense: Benchmarking social intelligence of language agents through interactive scenarios[C]//Proceedings of the Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, Albuquerque, New Mexico, 2025.
- [20] YAO S, YU D, ZHAO J, et al. Tree of thoughts: Deliberate problem solving with large language models[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2023.



冯夏冲(1996—), 博士, 主要研究领域为自然语言处理和大语言模型。

E-mail: xiachongfeng@ir.hit.edu.cn



冯骁骋(1988—), 通信作者, 博士, 教授, 主要研究领域为自然语言处理、文本生成和大语言模型。

E-mail: xcfeng@ir.hit.edu.cn



黄毅斌(1998—), 学士, 主要研究领域为自然语言处理、机器翻译和大语言模型。

E-mail: ychuang@ir.hit.edu.cn