

# REACT: 在语言模型中协同推理和行动

姚舜宇<sup>\*,1</sup>, 赵杰<sup>2</sup>, 于典<sup>2</sup>, 杜楠<sup>2</sup>, 伊扎克·沙夫兰<sup>2</sup>, 卡尔蒂克·纳拉西曼<sup>1</sup>, 曹远<sup>2</sup>

<sup>1</sup>普林斯顿大学计算机科学系

<sup>2</sup>谷歌研究, Brain团队

<sup>1</sup>{shunyuy, karthikn}@princeton.edu

<sup>2</sup>{jeffreyzhao, dianyu, dunan, izhak, yuancan}@google.com

## 摘要

尽管大型语言模型 (LLMs) 在语言理解和交互式决策方面展示了令人印象深刻的性能, 但它们在推理 (例如思维链提示) 和行动 (例如行动计划生成) 方面的能力主要被研究为独立的主题。在本文中, 我们探讨了使用 LLMs 以交错方式生成推理追踪和任务特定行动的方法, 从而在两者之间实现更大的协同效应: 推理追踪帮助模型诱导、跟踪和更新行动计划, 并处理异常情况, 而行动则使其能够与外部源 (如知识库或环境) 进行接口和收集额外信息。我们将我们的方法命名为 ReAct, 并将其应用于各种语言和决策任务, 并展示其在最先进的基线方法上的有效性, 以及改进的人类可解释性和可信度。具体而言, 在问题回答 (HotpotQA) 和事实验证 (Fever) 方面, ReAct 通过与简单的维基百科 API 进行交互, 并生成更可解释的人类任务解决轨迹, 克服了思维链推理中普遍存在的幻觉和错误传播问题。

此外, 在两个交互式决策制定基准 (ALFWorld 和 WebShop) 上, ReAct 的绝对成功率分别比模仿和强化学习方法高出 34% 和 10%, 而只有一个或两个上下文示例的提示。

## 1 我 NTRODUCTION

人类智能的一个独特特征是能够将面向任务的行动与语言推理 (或内部言语, Alderson-Day & Fernyhough, 2015) 无缝结合, 这被理论化为在人类认知中发挥重要作用, 以实现自我调节或策略化 (Vygotsky, 1987年; Luria, 1965年; Fernyhough, 2010年) 并维持工作记忆 (Baddeley, 1992年)。考虑一下在厨房里做一道菜的例子。在任何两个具体行动之间, 我们可以用语言推理来跟踪进展 (“现在所有的东西都切好了, 我应该把锅里的水烧热”), 处理异常情况或根据情况调整计划 (“我没有盐, 所以我用酱油和胡椒代替”), 并意识到需要外部信息时 (“我如何准备面团? 让我在互联网上搜索一下”)。我们还可以采取行动 (打开烹饪书阅读食谱, 打开冰箱, 检查配料) 来支持推理并回答问题 (“我现在能做什么菜?”)。这种“行动”和“推理”之间的紧密协同使人类能够快速学习新任务, 并在以前未见过的情况下或面对信息不确定性时进行强大的决策或推理。

最近的研究结果暗示了在自主系统中将语言推理与交互决策相结合的可能性。一方面, 适当提示的大型语言模型 (LLMs) 已经展示了进行多步推理过程以得出结论的新能力。

\*在Google实习期间的工作。项目页面和代码: <https://react-lm.github.io/>。

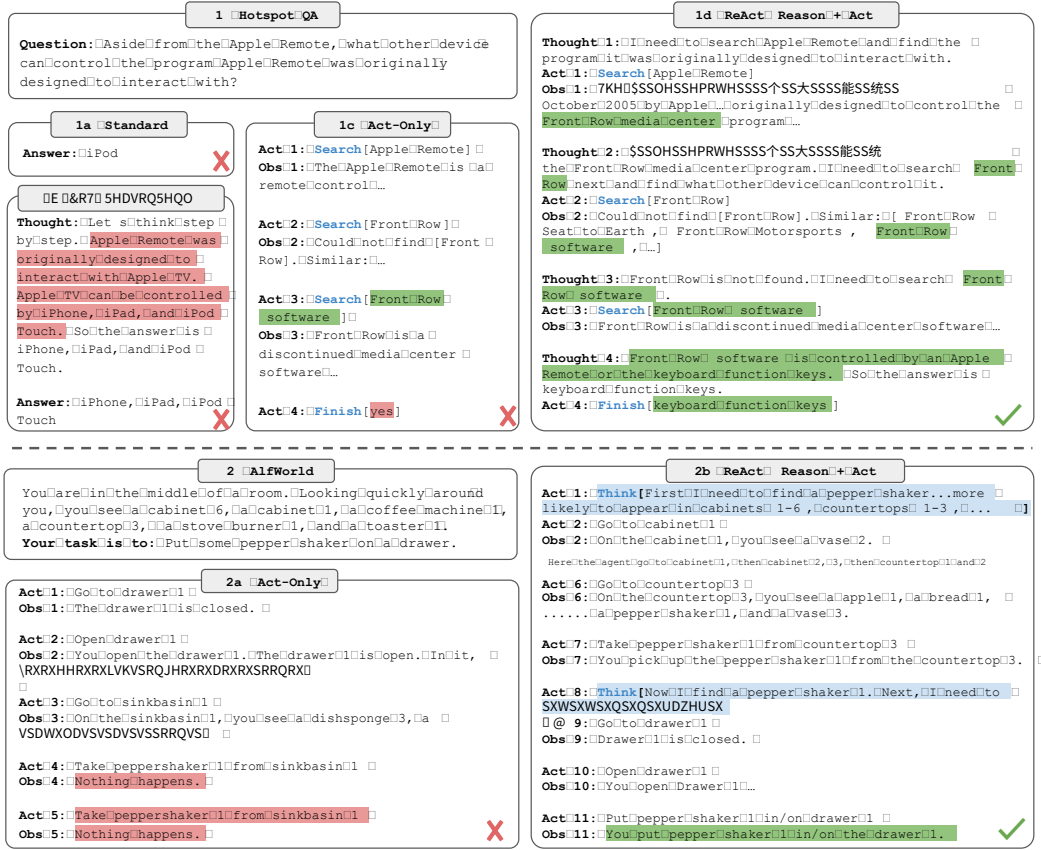


图1: (1) 对比4种提示方法, (a) 标准, (b) 思维链 (CoT, 仅限推理), (c) 仅行动, 和 (d) ReAct (推理+行动), 解决一个HotspotQA (Yang等, 2018)问题。 (2) 对比 (a) 仅行动 和 (b) ReAct 提示来解决一个AlfWorld (Shridhar等, 2020b)游戏。在这两个领域中, 我们省略了提示中的上下文示例, 只展示了模型生成的任务解决轨迹 (行动, 思考) 和环境 (观察)。

算术、常识和符号推理任务中的问题答案 (Wei等, 2022年)。然而, 这种“思维链”推理是一个静态的黑盒子, 模型使用自己的内部表示来生成思维, 而不是基于外部世界, 这限制了它反应性推理或更新知识的能力。这可能导致事实幻觉和错误在推理过程中的传播 (图1 (1b))。另一方面, 最近的研究探索了预训练语言模型在交互环境中规划和行动的使用 (Ahn等, 2022年; Nakano等, 2021年; Yao等, 2020年; Huang等, 2022a年), 重点是通过语言先验预测行动。这些方法通常将多模态观察转化为文本, 使用语言模型生成领域特定的行动或计划, 然后使用控制器选择或执行它们。然而, 它们没有使用语言模型来抽象地思考高层目标或维护工作记忆以支持行动, 除了Huang等人 (2022b年) 在当前状态下执行有限形式的语言推理来重述空间事实。除了与几个块交互的简单体验任务外, 还没有研究如何将推理和行动以协同的方式结合起来进行一般任务解决, 以及这种组合是否与单独的推理或行动相比带来系统性的好处。

在这项工作中, 我们提出了ReAct, 这是一种将推理和行动与语言模型相结合的通用范例, 用于解决各种语言推理和决策任务 (图1)。ReAct提示LLMs以交错的方式生成与任务相关的口头推理追踪和行动, 这使得模型能够进行动态推理, 创建、维护和调整高级行动计划 (推理到行动), 同时与外部环境 (例如维基百科) 进行交互, 将额外信息纳入推理过程 (行动到推理)。

我们对ReAct和最先进的基准模型在四个不同的基准测试上进行了实证评估：问答（HotPotQA, Yang等, 2018），事实验证（Fever, Thorne等, 2018），基于文本的游戏（ALFWorld, Shridhar等, 2020b）和网页导航（WebShop, Yao等, 2022）。对于HotPotQA和Fever，通过与模型可以交互的维基百科API，ReAct在生成行动时优于普通的行动生成模型，同时与思维链推理（CoT）（Wei等, 2022）相竞争。总体而言，最佳方法是将ReAct和CoT结合起来，在推理过程中同时利用内部知识和外部获取的信息。在ALFWorld和WebShop上，仅使用两个甚至一个ReAct提示就能够胜过使用 $10^3 \sim 10^5$ 个任务实例进行训练的模仿学习或强化学习方法，成功率分别提高了34%和10%。我们还通过与仅行动的受控基准模型进行比较，展示了稀疏、多功能推理在决策中的重要性。除了具有广泛的适用性和性能提升外，推理和行动的结还有助于模型的可解释性、可信度和诊断能力，在所有领域中，人们可以轻松区分模型的内部知识和外部环境，并检查推理追踪以理解模型行动的决策依据。

总结一下，我们的主要贡献如下：（1）我们引入了一种新的基于提示的范式ReAct，用于在语言模型中协同推理和行动解决一般任务；（2）我们在各种基准测试中进行了广泛的实验，展示了ReAct在少样本学习设置中相对于之前的方法（在推理或行动生成中单独执行）的优势；（3）我们进行了系统的消融和分析，以了解在推理任务中行动的重要性以及在交互任务中推理的重要性；（4）我们分析了ReAct在提示设置下的局限性（即推理和行动行为的有限支持），并进行了初步的微调实验，展示了ReAct在额外训练数据的情况下改进的潜力。将ReAct扩展到更多任务的训练和操作，并将其与强化学习等互补范式相结合，可以进一步释放大语言模型的潜力。

## 2 REACT：协同推理+行动

考虑一个智能体与环境进行任务解决的一般设置。在时间步骤 $t$ ，智能体从环境中接收到一个观察 $ot \in \mathcal{O}$ ，并根据某个策略 $\pi(at|ct)$ 采取行动 $at$ ，其中 $ct=(o_1, a_1, \dots, o_{t-1}, at_{-1}, ot)$ 是智能体的上下文。

当映射 $c_t \rightarrow a_t$ 高度隐式且需要大量计算时，学习策略是具有挑战性的。例如，图1(1c)中的智能体无法生成正确的最终行动( $Ac_t 4$ )来完成问答任务，因为它需要对轨迹上下文(Question,  $Ac_t 1-3$ , Obs 1-3)进行复杂的推理。同样，图1(2a)中的智能体无法从上下文中理解到sinkbasin 1不包含peppershaker 1，因此会持续产生幻觉行动。

ReAct的想法很简单：我们扩展了代理的行动空间为 $\hat{\mathcal{A}} = \mathcal{A} \cup \mathcal{L}$ ，其中 $\mathcal{L}$ 是语言空间。语言空间中的一个行动 $\hat{a}_t \in \mathcal{L}$ ，我们将其称为思考或推理痕迹，不会影响外部环境，因此不会有观察反馈。

相反，一个思考 $\hat{a}_t$ 旨在通过对当前上下文 $c_t$ 进行推理来组合有用的信息，并更新上下文 $c_{t+1} = (c_t, \hat{a}_t)$ 以支持未来的推理或行动。如图1所示，有各种各样的有用思考，例如分解任务目标并创建行动计划（2b，行动<sub>1</sub>；1d，思考<sub>1</sub>），注入与任务解决相关的常识知识（2b，行动<sub>1</sub>），从观察中提取重要部分（1d，思考<sub>2</sub>，4），跟踪进展并转换行动计划（2b，行动<sub>8</sub>），处理异常情况并调整行动计划（1d，思考<sub>3</sub>）等等。

然而，由于语言空间 $\mathcal{L}$ 是无限的，学习在这个增强的行动空间中是困难的，并且需要强大的语言先验知识。在本文中，我们主要关注的是一个冻结的大型语言模型PaLM-540B（Chowdhery等, 2022）<sup>1</sup>，通过少量上下文示例来生成领域特定的行动和自由形式的语言思考，以解决任务（图1（1d），（2b））。每个上下文示例都是一个人类的行动、思考和环境观察的轨迹，用于解决一个任务实例（见附录C）。对于推理至关重要的任务（图1（1）），我们交替生成思考和行动，使得任务解决轨迹包含多个思考-行动-观察步骤。相反，对于可能涉及大量行动的决策任务（图1（2）），只需要思考即可。

<sup>1</sup>我们在附录A.1中展示了一些GPT-3（Brown等, 2020）的结果，它在PaLM-540B上表现出色。

在轨迹的最相关位置上，思考和行动很少出现，因此我们让语言模型自己决定思维和行动的异步发生。

由于决策和推理能力被整合到一个大型语言模型中，ReAct具有几个独特的特点：A) 直观且易于设计：设计ReAct提示非常简单，人工注释者只需在行动之上用语言记录下他们的思考。本文中我们没有使用特定的格式选择、思考设计或示例选择。我们在第3和第4节详细介绍了每个任务的提示设计。B) 通用且灵活：由于灵活的思考空间和思考-行动发生格式，ReAct适用于具有不同行动空间和推理需求的各种任务，包括但不限于问答、事实验证、文字游戏和网络导航。C) 高性能和稳健：ReAct在仅从一个到六个上下文示例中学习的同时，对新任务实例表现出强大的泛化能力，始终在不同领域中优于仅具有推理或行动的基线模型。我们还在第3节展示了当启用微调时的额外好处，并在第4节展示了ReAct对提示选择的稳健性能。

D) 人类对齐和可控性：ReAct承诺提供一个可解释的顺序决策和推理过程，人类可以轻松检查推理和事实的正确性。此外，人类还可以通过思维编辑来控制或纠正代理行为，如第4节的图5所示。

### 3个知识密集型推理任务

我们从知识密集型的推理任务开始，如多跳问题回答和事实验证。如图1(1d)所示，通过与维基百科API的交互，ReAct能够检索支持推理的信息，同时使用推理来确定下一步要检索的内容，展示了推理和行动的协同效应。

#### 3.1 设置

领域：我们考虑两个具有挑战性的知识检索和推理数据集：(1) Hot-PotQA (Yang等, 2018)，一个需要对两个或更多维基百科段落进行推理的多跳问题回答基准，以及(2) FEVER (Thorne等, 2018)，一个事实验证基准，每个主张都被注释为SUPPORTS、REFUTES或NOT ENOUGH INFO，基于是否存在维基百科段落来验证主张。在这项工作中，我们对两个任务都采用了仅问题的设置，模型只接收问题/主张作为输入，没有访问支持段落的权限，必须依靠其内部知识或通过外部环境的交互来获取知识以支持推理。

行动空间 我们设计了一个简单的维基百科网络API，具有三种类型的操作，以支持交互式信息检索：(1)搜索[实体]，如果存在，则返回相应实体维基页面的前5个句子，否则建议从维基百科搜索引擎中选择前5个相似实体，(2)查找[字符串]，它将返回包含字符串的页面中的下一个句子，模拟浏览器上的Ctrl+F功能，(3)完成[答案]，它将使用答案完成当前任务。我们注意到，这个行动空间通常只能根据确切的段落名称检索到一小部分内容，这明显比最先进的词汇或神经检索器要弱。目的是模拟人类如何与维基百科进行交互，并迫使模型通过明确的语言推理来检索。

#### 3.2 M方法

ReAct：为HotpotQA和Fever提供提示，我们从训练集中随机选择了6个和3个案例，并手动组成了ReAct格式的轨迹，以作为提示中的少量示例。类似于图1(d)，每个轨迹由多个思考-行动-观察步骤（即密集思考）组成，其中自由形式的思考用于各种目的。具体来说，我们使用一种思考的组合，将问题分解为（“我需要搜索x，找到y，然后找到z”），从维基百科观察中提取信息（“x始于1844年”，“这段落没有提到x”），进行常识推理（“x不是y，所以z必须是...”）或算术推理（“1844 < 1989”），引导

<sup>2</sup> 我们发现更多的例子并没有提高性能。

提示方法 <sup>a</sup>	HotpotQA (EM)	Fever (Acc)
标准	28.7	57.1
CoT (Wei等, 2022年)	29.4	56.3
CoT-SC (Wang等, 2022a年)	33.4	60.4
行动	25.7	58.9
反应	27.4	60.9
CoT-SC → 反应	34.2	<b>64.6</b>
反应 → CoT-SC	<b>35.1</b>	62.0
监督式SoTA <sup>b</sup>	67.5	89.5

表1: PaLM-540B在HotpotQA和Fever上的提示结果。

<sup>a</sup> HotpotQA EM分别为27.1、28.9、33.8，对应于标准、CoT和CoT-SC (Wang等, 2022b年)。

<sup>b</sup> (Zhu等, 2021年; Lewis等, 2020年)

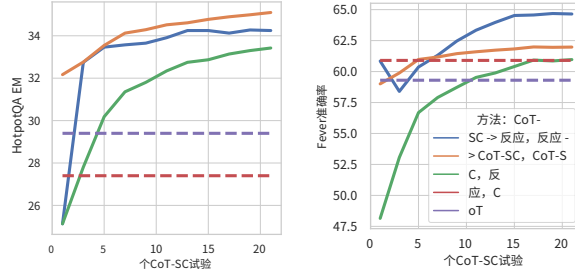


图2: PaLM-540B提示结果与使用的CoT-SC样本数量相关。

搜索重构 (“也许我可以搜索/查找x而不是y”)，并综合最终答案 (“...所以答案是x”)。更多细节请参见附录C。

基线：我们系统地削减ReAct轨迹以构建多个基线的提示（格式如图1 (1a-1c)）：(a) 标准提示 (Standard)，它删除了ReAct轨迹中的所有思考、行动和观察。(b) 思维链提示 (CoT) (Wei等, 2022)，它删除了行动和观察，并作为仅推理的基线。我们还通过在推理过程中使用解码温度0.7对21个CoT轨迹进行采样，并采用多数答案的自洽性基线 (CoT-SC) (Wang等, 2022a;b)，发现它能够持续提升性能。(c) 仅行动提示 (Act)，它删除了ReAct轨迹中的思考，与WebGPT (Nakano等, 2021) 如何与互联网交互以回答问题类似，尽管它在不同的任务和行动空间上操作，并使用模仿和强化学习而不是提示。

将内部和外部知识结合起来，正如在第3.3节中详细介绍的那样，我们观察到ReAct所展示的问题解决过程更加真实和扎实，而CoT在制定推理结构方面更准确，但很容易出现虚构的事实或思想。因此，我们建议将ReAct和CoT-SC结合起来，并让模型根据以下启发式方法决定何时切换到另一种方法：A) ReAct → CoT-SC：当ReAct在给定步骤内无法返回答案时，退回到CoT-SC。我们在HotpotQA和FEVER中分别设置了7和5个步骤，因为我们发现更多的步骤不会提高ReAct的性能。

B) CoT-SC → ReAct：当 $n$ 个CoT-SC样本中的大多数答案出现次数少于 $n/2$ 次时（即内部知识可能不自信地支持任务），退回到ReAct。

微调：由于在大规模上手动注释推理轨迹和行动的挑战，我们考虑了一种类似于Zelikman等人 (2022) 的引导式方法，使用3,000个由ReAct生成的带有正确答案的轨迹（也适用于其他基线模型）来微调较小的语言模型 (PaLM-8/62B)，以便根据输入的问题/主张解码轨迹（所有思考、行动、观察）。更多细节请参见附录B.1。

### 3.3 结果和观察

ReAct在使用PaLM-540B作为基础模型的不同提示方法下，在HotpotQA和Fever结果上表现出色。我们注意到ReAct在两个任务上都优于Act，这证明了推理在指导行动方面的价值，特别是在综合最终答案方面，如图1 (1c-d) 所示。微调结果也证实了推理轨迹对于更明智的行动的好处。

<sup>3</sup>在所有具有正确最终答案的轨迹中，HotpotQA上有7个步骤，FEVER上有5个步骤的轨迹仅占0.84%和1.33%。

	类型	定义	反应	CoT
成功	真正的正例	正确的推理路径和事实	94%	86%
	假正例	虚构的推理路径或事实	6%	14%
失败	推理错误	错误的推理路径（包括无法从重复的步骤中恢复）	47%	16%
	搜索结果错误	搜索结果为空或不包含有用信息	23%	-
	幻觉	虚构的推理路径或事实	0%	56%
	标签模糊	预测正确但与标签不完全匹配	29%	28%

表2: ReAct和CoT在HotpotQA上的成功和失败模式类型，以及人工研究中随机选择的示例中的百分比。

**ReAct**与**CoT**相比，在Fever上表现更好（60.9 vs. 56.3），在HotpotQA上稍逊于**CoT**（27.4 vs. 29.4）。Fever声称的SUPPORTS/REFUTES可能只有微小差异（见附录D.1），因此获取准确和最新知识的行动至关重要。为了更好地理解**ReAct**和**CoT**在HotpotQA上的行为差异，我们从**ReAct**和**CoT**中分别随机抽取了50个具有正确和错误答案的轨迹（由EM判断），总共200个示例，并在表2中手动标记了它们的成功和失败模式。以下是一些关键观察结果：

A) 幻觉是**CoT**的一个严重问题，与**ReAct**相比，在成功模式下，其误报率要高得多（14%对6%），并且构成其主要失败模式（56%）。相比之下，**ReAct**的问题解决轨迹更加扎实、以事实为导向，并且更加可靠，这要归功于对外部知识库的访问。

B) 在交错推理、行动和观察步骤的同时，提高了**ReAct**的扎实性和可靠性，但这种结构约束也降低了其在制定推理步骤方面的灵活性，导致比**CoT**更高的推理错误率。我们注意到，**ReAct**存在一种常见的错误模式，即模型反复生成先前的思考和行动，并将其归类为“推理错误”的一部分，因为模型无法推理出下一步应该采取什么正确的行动并跳出循环。

C) 对于**ReAct**来说，通过搜索成功地检索到信息性知识至关重要。非信息性搜索占错误案例的23%，使模型的推理受阻，并且难以恢复和重新构思思路。这可能是事实性和灵活性之间的预期权衡，这激发了我们提出的结合两种方法的策略。

我们在附录E.1中为每种成功和失败模式提供了示例。我们还发现一些HotpotQA问题可能包含过时的答案标签，例如请参见图4。

**ReAct+CoT-SC**在提示LLMs方面表现最佳。如表1所示，HotpotQA和Fever上最佳的提示方法分别为**ReAct**→**CoT-SC**和**CoT-SC**→**ReAct**。

此外，图2显示了不同方法在使用CoT-SC样本数量方面的表现。虽然两种**ReAct+CoT-SC**方法在各自的任务上具有优势，但它们在不同数量的样本上都明显且一致地优于**CoT-SC**，在仅使用3-5个样本的情况下，即可达到使用21个样本的**CoT-SC**性能。这些结果表明，合理地结合模型内部知识和外部知识对于推理任务非常有价值。

**ReAct**在微调中表现最佳。图3展示了四种方法（标准、CoT、Act、ReAct）在HotpotQA上的扩展效果。使用PaLM-8/62B进行提示，**ReAct**在四种方法中表现最差，因为从上下文示例中学习推理和行动的困难较大。然而，当只使用3,000个示例进行微调时，**ReAct**成为四种方法中最好的方法，使用PaLM-8B微调的**ReAct**优于所有PaLM-62B提示方法，使用PaLM-62B微调的**ReAct**优于所有540B提示方法。相比之下，对于PaLM-8/62B，微调标准或CoT要明显差于微调**ReAct**或**Act**，因为前者实质上教会模型记忆（可能是虚构的）知识事实，而后者教会模型如何（推理和）行动以从维基百科获取信息，这是一种更具普适性的知识推理技能。由于所有提示方法仍然远远落后于特定领域的最先进方法（表1），我们相信使用更多人工编写的数据进行微调可能是释放**ReAct**能力的更好方法。

<sup>4</sup>我们怀疑这可能是由于次优的贪婪解码过程，未来使用更好的解码（例如波束搜索）可能有助于解决这个问题。

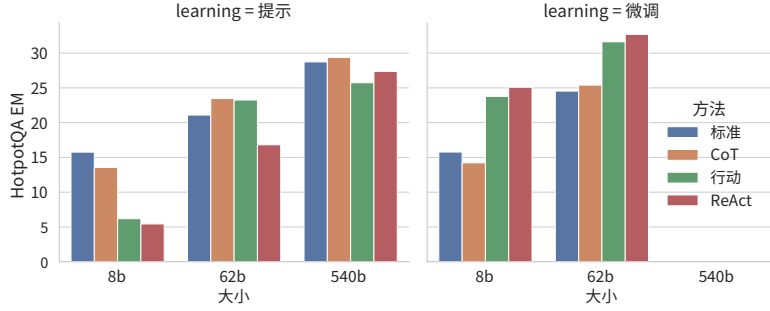


图3：使用ReAct（我们的方法）和基线在HotPotQA上进行提示和微调的扩展结果。

#### 4 DECISION MAKING TASKS

我们还在两个基于语言的交互式决策任务ALFWorld和WebShop上测试了ReAct，这两个任务都涉及到需要在长时间范围内进行行动和有效地进行推理的复杂环境，因此需要推理来行动和探索。

**ALFWorld** ALFWorld (Shridhar等, 2020b) (图1 (2)) 是一个设计用于与具体ALFRED基准 (Shridhar等, 2020a) 对齐的合成文本游戏。它包括6种类型的任务，代理需要通过文本动作（例如，去到咖啡桌1，拿起纸2，使用台灯1）在模拟家庭中导航和交互，以实现高级目标（例如，在台灯下检查纸）。一个任务实例可能有<sup>50</sup>多个位置，并且需要一个专家策略超过<sup>50</sup>个步骤来解决，因此挑战代理人规划和跟踪子目标，以及系统地探索（例如，逐个检查所有桌子上的台灯）。特别是，ALFWorld内置了一个挑战，即确定常见家居物品的可能位置（例如，台灯可能在桌子、架子或梳妆台上），使得这个环境非常适合LLMs利用它们预训练的常识知识。为了提示ReAct，我们随机注释了每种任务类型的三个轨迹，其中每个轨迹包括稀疏的思考，用于（1）分解目标，（2）跟踪子目标完成情况，（3）确定下一个子目标，以及（4）通过常识推理在哪里找到一个物体以及如何处理它。我们在附录C.4中展示了用于ALFWorld的提示。根据Shridhar等人 (2020b) 的方法，我们在134个未见过的评估游戏中进行任务特定的评估。为了稳健性，我们通过从我们注释的3个轨迹的每个排列构建了每种任务类型的6个提示。行动提示使用相同的轨迹构建，但没有思考-由于任务实例是从训练集中随机选择的，它既不偏向ReAct也不偏向Act，并为测试稀疏思考的重要性提供了一个公平和可控的比较。对于基线，我们使用BUTLER (Shridhar等, 2020b)，这是一个在每种任务类型上训练了 $10^5$ 个专家轨迹的模仿学习代理。

**WebShop:** Can ReAct 也能与嘈杂的现实语言环境进行互动，以实现实际应用吗？我们调查了WebShop (Yao等人, 2022年)，这是一个最近提出的在线购物网站环境，拥有1.18M个真实世界的产品和12k个人类指令。与ALFWorld不同，Webshop包含了各种结构化和非结构化的文本（例如产品标题、描述以及从亚马逊爬取的选项），并要求代理根据用户的指令（例如“我正在寻找一个带抽屉的床头柜。它应该有镍饰面，并且价格低于140美元”）通过网络交互（例如搜索“床头柜抽屉”，选择按钮，如“颜色：现代镍白色”或“返回搜索”）购买产品。该任务通过平均分数（选择的产品中所涵盖的所需属性的百分比，对所有剧集进行平均）和成功率（选择的产品满足所有要求的剧集的百分比）在500个测试指令上进行评估。我们制定了带有搜索、选择产品、选择选项和购买的Act提示，而ReAct提示则额外进行推理以确定要探索什么，何时购买以及哪些产品选项与指令相关。请参见附录中的表6以获取示例提示，以及表10以获取模型预测。我们与模仿学习 (IL) 方法进行比较。

<sup>5</sup>Micheli & Fleuret (2021) 在3553个任务实例上对GPT-2模型进行了微调，并取得了比BUTLER更好的性能，但它是在所有任务类型上进行训练的，因此不作为基准。



方法	选择	清洁	加热	冷却	外观	选择2个	全部
行动 (6个中的最佳)	88	42	74	67	72	<b>41</b>	45
ReAct (平均)	65	39	83	76	55	24	57
ReAct (6个中的最佳)	<b>92</b>	58	<b>96</b>	86	<b>78</b>	<b>41</b>	<b>71</b>
ReAct-IM (平均)	55	59	60	55	23	24	48
ReAct-IM (6个中的最佳)	62	<b>68</b>	87	57	39	33	53
BUTLER <sub>g</sub> (6个中的最佳)	33	26	70	76	17	12	22
BUTLER (6个中的最佳)	46	39	74	<b>100</b>	22	24	37

表3: ALFWorld任务特定成功率 (%)。BUTLER和BUTLER<sub>g</sub>的结果来自Shridhar等人的表4 (2020b)。所有方法都使用贪婪解码, 除了BUTLER使用波束搜索。

方法	得分 SR
行动反应	62.3 30.1 <b>66.6 40.0</b>
IL	59.9 29.1
IL+RL	62.4 28.7
人类专家	82.1 59.6

表4: Web-shop上的得分和成功率 (SR)。IL/IL+RL取自Yao等人 (2022)。

使用1,012个人工注释的轨迹进行训练, 并使用10,587个训练指令进行模仿+强化学习 (IL + RL) 方法进行训练。

结果 ReAct在ALFWorld (表3) 和Webshop (表4) 上表现优于Act。在ALFWorld上, 最好的ReAct试验的平均成功率达到71%, 明显优于最好的Act (45%) 和BUTLER (37%) 试验。事实上, 即使是最差的ReAct试验 (48%) 也超过了两种方法的最佳试验。此外, ReAct相对于Act的优势在六个对照试验中保持一致, 相对性能提升范围从33%到90%, 平均为62%。

从定性上看, 我们发现, 没有任何思考, Act无法正确分解目标为较小的子目标, 或者无法跟踪环境的当前状态。附录D.2.1和附录D.2.2中可以找到比较ReAct和Act的示例轨迹。

在Webshop上, 一次性行动提示已经与IL和IL+RL方法表现相当。通过额外的稀疏推理, ReAct的性能显著提高, 成功率绝对提高了10%。通过检查示例, 我们发现ReAct更有可能通过推理来识别与指令相关的产品和选项, 以弥合嘈杂观察和行动之间的差距 (例如, “对于‘客厅节省空间的脚凳’, 该物品有‘39x18x18英寸’和‘蓝色’的选项, 看起来不错。”)。然而, 现有方法在专家人类的表现 (表4) 方面仍然远远落后, 专家人类进行了更多的产品探索和查询重构, 这对于基于提示的方法仍然具有挑战性。

关于内部推理与外部反馈的价值, 据我们所知, ReAct是首个将推理和行动结合应用于交互环境中的封闭循环系统的LLM演示。也许最接近的先前工作是Huang等人 (2022b) 的Inner Monologue (IM), 其中来自具体化代理的行动受到同名“内部独白”的激励。然而, IM的“内部独白”仅限于对环境状态的观察以及代理完成目标所需的内容。相比之下, ReAct用于决策的推理追踪是灵活且稀疏的, 允许针对不同任务引导不同类型的推理 (参见第2节)。

为了展示ReAct和IM之间的差异, 并突出内部推理与简单反应外部反馈的重要性, 我们进行了一项消融实验, 使用了一个由IM类似的密集外部反馈组成的思维模式。如表3所示, ReAct在整体成功率上明显优于IM风格的提示 (ReAct-IM) (71比53), 在六个任务中有五个任务具有一致的优势。从定性上看, 我们观察到ReAct-IM在确定子目标何时完成或下一个子目标应该是什么时经常犯错, 这是由于缺乏高层目标分解。此外, 许多ReAct-IM轨迹在确定ALFWorld环境中物品可能在哪里时遇到困难, 这是由于缺乏常识推理。

这两个缺点可以在ReAct范式中得到解决。有关ReAct-IM的更多细节请参见附录B.2。ReAct-IM的示例提示可以在附录C.4中找到, 示例轨迹可以在附录D.2.3中找到。



## 5个相关工作

用于推理的语言模型也许最著名的使用LLM进行推理的工作是Chain-of-Thought (CoT) (Wei等, 2022年), 它揭示了LLM制定问题解决的“思考过程”的能力。随后进行了几项后续工作, 包括用于解决复杂任务的从最少到最多的提示(Zhou等, 2022年), 零射击-CoT (Kojima等, 2022年) 和自一致性推理(Wang等, 2022a年)。最近, (Madaan & Yazdani, 2022年) 对CoT的制定和结构进行了系统研究, 并观察到符号, 模式和文本的存在对CoT的有效性至关重要。其他工作还扩展到了更复杂的推理架构, 超越了简单的提示。

例如, Selection-Inference (Creswell等, 2022年) 将推理过程分为“选择”和“推理”两个步骤。STaR (Zelikman等, 2022年) 通过对模型自身生成的正确理由进行微调, 引导推理过程。忠实推理 (Creswell & Shanahan, 2022年) 将多步推理分解为三个步骤, 分别由专用的LM执行。类似的方法, 如Scratchpad (Nye等, 2021年), 在中间计算步骤上进行微调, 也展示了在多步计算问题上的改进。

与这些方法相比, ReAct不仅仅执行孤立的、固定的推理, 还将模型的行动和相应的观察整合到一个连贯的输入流中, 以便模型能够更准确地推理并解决超越推理的任务 (例如交互式决策)。

用于决策的语言模型: LLM具有强大的能力, 使其能够执行超越语言生成的任务, 并且越来越流行的是利用LLM作为决策制定的策略模型, 特别是在交互环境中。WebGPT (Nakano等, 2021) 使用语言模型与Web浏览器进行交互, 浏览网页, 并从ELI5 (Fan等, 2019) 中推断出复杂问题的答案。与ReAct相比, WebGPT并没有明确建模思考和推理过程, 而是依赖昂贵的人类反馈进行强化学习。在对话建模中, 像BlenderBot (Shuster等, 2022b) 和Sparrow (Glaese等, 2022) 这样的聊天机器人以及像SimpleTOD (Hosseini-Asl等, 2020) 这样的面向任务的对话系统也训练语言模型来进行API调用的决策。与ReAct不同, 它们也没有明确考虑推理过程, 并且还依赖昂贵的数据集和人类反馈收集进行策略学习。相比之下, ReAct以更便宜的方式学习策略, 因为决策过程只需要推理过程的语言描述。

LLMs在交互和具身环境中的规划和决策中也越来越多地被使用。在这方面, SayCan (Ahn等, 2022年) 和Inner Monologue (Huang等, 2022b年) 可能与ReAct最相关, 它们使用LLMs进行机器人行动规划和决策。在SayCan中, LLMs被提示直接预测机器人可以采取的可能行动, 然后通过基于视觉环境的可行性模型进行最终预测的重新排序。

Inner Monologue通过从环境中注入的反馈进一步改进了这一点。据我们所知, Inner Monologue是第一个展示这种闭环系统的工作, ReAct是基于它构建的。然而, 我们认为Inner Monologue并不真正包含内心的思想, 这在第4节中有详细阐述。我们还注意到, 在交互决策过程中利用语言作为语义丰富的输入已经在其他环境中取得了成功。越来越明显的是, 在LLMs的帮助下, 语言作为一种基本的认知机制将在交互和决策中发挥关键作用。此外, LLMs的进展还激发了像Reed等人 (2022年) 这样多才多艺和通用的代理人的发展。

## 6 C结论

我们提出了ReAct-一种简单而有效的方法, 用于在大型语言模型中协同推理和行动。通过对多跳问题回答、事实核查和交互式决策任务进行多样化的实验, 我们展示了ReAct在可解释的决策轨迹下具有卓越的性能。尽管我们的方法简单, 但是具有大型行动空间的复杂任务需要更多的演示来学习, 而这很容易超出上下文学习的输入长度限制。我们在HotpotQA上探索了微调方法

<sup>6</sup>人类反馈也可以以互补的方式进行整合, 但我们将其留给未来的工作。

初步有希望的结果，但是从更多高质量的人类注释中学习将是进一步提高性能的目标。通过多任务训练扩展ReAct，并将其与强化学习等互补范式相结合，可以产生更强大的代理程序，进一步释放LLM在更多应用中的潜力。

#### 致谢

我们感谢Google Brain团队和普林斯顿NLP团队的支持和反馈。本工作部分得到了国家自然科学基金的资助，资助号为2107048。本材料中表达的任何观点、发现、结论或建议均为作者个人观点，不一定反映国家自然科学基金会的观点。

#### 可复现性声明

我们的主要实验是在PaLM (Chowdhery等, 2022) 上进行的，该模型目前尚不公开可访问。为了增加可复现性，我们在附录C中包含了所有使用的提示，附录A.1中使用了GPT-3 (Brown等, 2020) 进行了额外实验，并在<https://anonymous.4open.science/r/ReAct-2268/>上提供了相关的GPT-3 ReAct提示代码。

#### 伦理声明

ReAct促使大型语言模型生成比以前的方法更具人类可解释性、可诊断性和可控性的任务解决轨迹。然而，将大型语言模型与行动空间连接起来与外部环境（例如网络、物理环境）进行交互存在潜在的危险，例如查找不适当或私人信息，或在环境中采取有害行动。我们的实验通过将交互限制在特定网站（维基百科或WebShop）上来最小化此类风险，这些网站不包含私人信息，并且在行动空间设计中没有任何危险行动（即模型无法在研究基准WebShop上真正购买产品或编辑维基百科）。我们认为研究人员在设计更广泛的实验之前应该意识到这些风险。

#### 参考文献

- Josh Abramson, Arun Ahuja, Iain Barr, Arthur Brussee, Federico Carnevale, Mary Cassin, Rachita Chhaparia, Stephen Clark, Bogdan Damoc, Andrew Dudzik, Petko Georgiev, Aurelia Guy, Tim Harley, Felix Hill, Alden Hung, Zachary Kenton, Jessica Landon, Timothy Lillicrap, Kory Mathewson, Soňa Mokrá, Alistair Muldal, Adam Santoro, Nikolay Savinov, Vikrant Varma, Greg Wayne, Duncan Williams, Nathaniel Wong, Chen Yan和Rui Zhu。模仿交互智能，2020年。网址：<https://arxiv.org/abs/2012.05672>。
- Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil J Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, John Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan和Andy Zeng。不要按我说的做，而是按我能做的做：在机器人能力中落地语言，2022年。网址<https://arxiv.org/abs/2204.01691>。
- Ben Alderson-Day和Charles Fernyhough。内部言语：发展、认知功能、现象学和神经生物学。心理学公告，141(5)：931，2015年。
- Alan Baddeley。工作记忆。科学，255(5044)：556-559，1992年。
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell等。语言模型是少样本学习者。神经信息处理系统的进展，33：1877-1901，2020年。

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann等。Path: 通过路径扩展语言建模。arXiv预印本arXiv:2204.02311, 2022年。

Antonia Creswell和Murray Shanahan。使用大型语言模型进行可靠推理, 2022年。网址<https://arxiv.org/abs/2208.14271>。

Antonia Creswell, Murray Shanahan和Irina Higgins。选择-推理: 利用大型语言模型进行可解释的逻辑推理, 2022年。网址<https://arxiv.org/abs/2205.09712>。

Angela Fan, Yacine Jernite, Ethan Perez, David Grangier, Jason Weston和Michael Auli。ELI5: 长篇问答。在计算语言学协会第57届年会论文集中, 第3558-3567页, 意大利佛罗伦萨, 2019年7月。计算语言学协会。doi: 10.18653/v1/P19-1346。网址<https://aclanthology.org/P19-1346>。

查尔斯·费尔尼霍夫。维果茨基、卢里亚和社会大脑。自我和社会调节: 社会互动与社会理解和执行功能的发展, 2010年, 第56-79页。

Amelia Glaese, Nat McAleese, Maja Trebacz, John Aslanides, Vlad Firoiu, Timo Ewalds, Mari-beth Rauh, Laura Weidinger, Martin Chadwick, Phoebe Thacker, Lucy Campbell-Gillingham, Jonathan Uesato, Po-Sen Huang, Ramona Comanescu, Fan Yang, Abigail See, Sumanth Dathathri, Rory Greig, Charlie Chen, Doug Fritz, Jaume Sanchez Elias, Richard Green, So'na Mokrá, Nicholas Fernando, Boxi Wu, Rachel Foley, Susannah Young, Iason Gabriel, William Isaac, John Mellor, Demis Hassabis, Koray Kavukcuoglu, Lisa Anne Hendricks和Geoffrey Irving。通过有针对性的人类判断改进对话代理的对齐, 2022年。网址: <https://storage.googleapis.com/deepmind-media/DeepMind.com/Authors-Notes/sparrow/sparrow-final.pdf>。

Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz和Richard Socher。用于任务导向对话的简单语言模型。神经信息处理系统的进展, 33:20179-20191, 2020年。

Wenlong Huang, Pieter Abbeel, Deepak Pathak和Igor Mordatch。语言模型作为零射击规划器: 提取行动知识以供具身代理使用。arXiv预印本arXiv:2201.07207, 2022a。

Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar等。内心独白: 通过语言模型进行规划的具身推理。arXiv预印本arXiv:2207.05608, 2022b。

Siddharth Karamcheti, Megha Srivastava, Percy Liang和Dorsa Sadigh。Lila: 基于语言的潜在行动。在CoRL中, 第1379-1390页, 2021年。网址: <https://proceedings.mlr.press/v164/karamcheti22a.html>。

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo和Yusuke Iwasawa。大型语言模型是零-shot推理者。arXiv预印本arXiv:2205.11916, 2022年。

Angeliki Lazaridou, Elena Gribovskaya, Wojciech Stokowiec和Nikolai Grigorev。通过少量提示增强的互联网语言模型, 用于开放域问答。arXiv预印本arXiv:2203.05115, 2022年。

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel等。检索增强的知识密集型NLP任务生成。神经信息处理系统的进展, 33: 9459-9474, 2020年。

Shuang Li, Xavier Puig, Chris Paxton, Yilun Du, Clinton Wang, Linxi Fan, Tao Chen, De-An Huang, Ekin Akyürek, Anima Anandkumar, Jacob Andreas, Igor Mordatch, Antonio Torralba, and Yuke Zhu。用于交互决策的预训练语言模型, 2022年。网址 <https://arxiv.org/abs/2202.01771>。

Aleksandr Romanovich Luria。Ls vygotsky和功能定位问题。神经心理学, 3 (4) : 387–392, 1965年。

Aman Madaan和Amir Yazdanbakhsh。文本和模式: 对于有效的思维链, 需要两个人共同努力, 2022年。网址<https://arxiv.org/abs/2209.07686>。

Vincent Micheli和François Fleuret。语言模型是少样本学习的管家。arXiv预印本arXiv:2104.07972, 2021年。

Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou, Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess和John Schulman。Webgpt: 通过人类反馈进行浏览器辅助问答, 2021年。网址<https://arxiv.org/abs/2112.09332>。

Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, Charles Sutton, and Augustus Odena。展示你的工作: 用于语言模型中间计算的草稿本, 2021年。网址<https://arxiv.org/abs/2112.00114>。

Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley Edwards, Nicolas Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar, and Nando de Freitas。一个通才代理, 2022年。网址 <https://arxiv.org/abs/2205.06175>。

Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer和Dieter Fox。Alfred: 用于解释日常任务的基于地面指令的基准。在计算机视觉和模式识别的IEEE/CVF会议论文集中, 第10740-10749页, 2020a。

Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler和Matthieu Hausknecht。Alfworld: 用于交互式学习的文本和具体环境对齐。arXiv预印本arXiv:2010.03768, 2020b。

Kurt Shuster, Mojtaba Komeili, Leonard Adolphs, Stephen Roller, Arthur Szlam和Jason Weston。寻求知识的语言模型: 对话和提示完成的模块化搜索和生成。arXiv预印本arXiv:2203.13224, 2022a。

Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, Morteza Behrooz, William Ngan, Spencer Poff, Naman Goyal, Arthur Szlam, Y-Lan Boureau, Melanie Kambadur和Jason Weston。Blenderbot 3: 一个部署的对话代理, 不断学习负责任地参与, 2022b。网址 <https://arxiv.org/abs/2208.03188>。

James Thorne, Andreas Vlachos, Christos Christodoulopoulos和Arpit Mittal。Fever: 一个大规模的用于事实提取和验证的数据集。arXiv预印本arXiv:1803.05355, 2018年。

Lev S Vygotsky。思维与言语。LS Vygotsky的著作, 1:39–285, 1987年。

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery和Denny Zhou。自治性改善了语言模型中的思维链推理, 2022a年。网址<https://arxiv.org/abs/2203.11171>。

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi和Denny Zhou。基于理性的语言模型增强集合。arXiv预印本arXiv:2207.00747, 2022b年。

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le和Denny Zhou。思维链提示引发大型语言模型的推理。arXiv预印本arXiv:2201.11903, 2022年。

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W Cohen, Ruslan Salakhutdinov和Christopher D Manning。Hotpotqa: 一个用于多样化、可解释的多跳问题回答的数据集。arXiv预印本arXiv:1809.09600, 2018年。

Shunyu Yao, Rohan Rao, Matthew Hausknecht和Karthik Narasimhan。保持冷静并探索：用于基于文本游戏中的行动生成的语言模型。在2020年经验方法自然语言处理会议论文集中，第8736-8754页，线上，2020年11月。计算语言学协会。doi: 10.18653/v1/2020.emnlp-main.704。URL: <https://aclanthology.org/2020.emnlp-main.704>。

Shunyu Yao, Howard Chen, John Yang和Karthik Narasimhan。Webshop：面向可扩展的真实世界网络交互的基于语言的代理。arXiv预印本arXiv: 2207.01206, 2022年。

Eric Zelikman, Yuhuai Wu, Jesse Mu和Noah D. Goodman。Star：用推理引导推理，2022年。URL: <https://arxiv.org/abs/2203.14465>。

Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Olivier Bousquet, Quoc Le和Ed Chi。最少到最多的提示使得大型语言模型能够进行复杂的推理，2022年。网址: <https://arxiv.org/abs/2205.10625>。

Yunchang Zhu, Liang Pang, Yanyan Lan, Huawei Shen和Xueqi Cheng。自适应信息搜索用于开放领域的问答。arXiv预印本arXiv:2109.06747, 2021年。

## 额外结果

## A.1 GPT-3实验

	PaLM-540B	GPT-3
HotpotQA (完全匹配)	29.4	<b>30.8</b>
ALFWorld (成功率%)	70.9	<b>78.4</b>

表5: 使用PaLM-540B与GPT-3 (text-davinci-002, 贪婪解码) 的ReAct提示结果。在HotpotQA上, 我们随机抽取了500个验证问题的子集。在ALFWorld上, 我们使用了所有134个未见过的验证任务实例, 并根据PaLM-540B选择了最佳提示集。

我们进行了额外的GPT-3 (Brown等人, 2020) 实验, 以确认ReAct提示的性能在不同的大型语言模型中是普遍的。如表5所示, GPT-3 (text-davinci-002, 贪婪解码) 在HotpotQA和ALFWorld上始终优于PaLM-540B, 可能是因为它根据人类指令进行微调的。这表明ReAct提示在不同的大型语言模型和不同的任务上都是有效的。这些实验的代码可以在<https://re-act-lm.github.io/>上找到。

## A.2 REACT在HotpotQA上获得最新知识

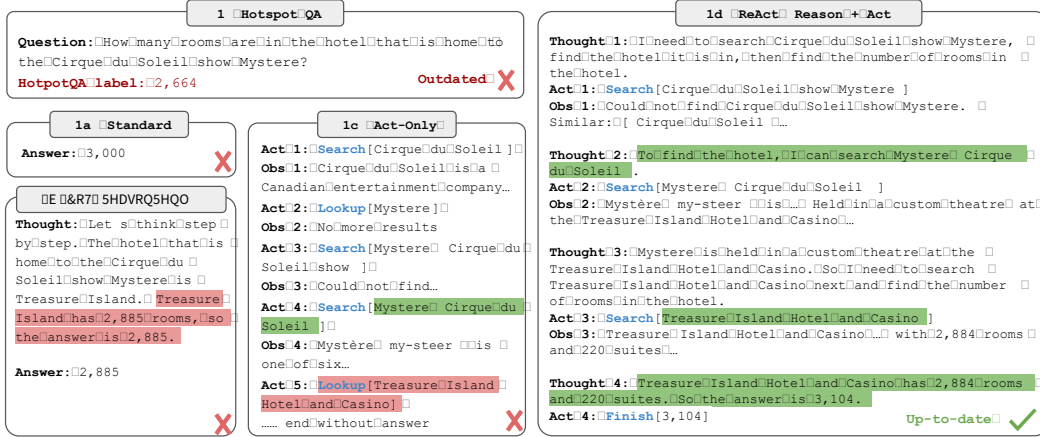


图4: 另一个HotpotQA问题的示例, 原始标签已过时。只有ReAct能够通过与现实世界的网络交互和推理获得最新的答案。

在轨迹检查过程中, 我们还发现有时候ReAct与数据集标签不一致, 因为标签本身可能已过时。例如, 如图4所示, 该问题询问酒店的规模, 这在HotpotQA构建时已经增加。尽管Standard和CoT由于产生幻觉而给出错误答案, Act由于缺乏推理指导如何与互联网进行问答而失败, 只有ReAct能够从互联网中检索到最新信息并提供合理的答案。因此, 更好地融合推理能力可能有助于最近的互联网增强语言模型 (Nakano等, 2021; Lazaridou等, 2022; Shuster等, 2022a) 进行最新任务的解决。

## A.3 HUMAN-IN-THE-LOOP BEHAVIOR CORRECTION ON ALF WORLD

我们还探索了与ReAct的人机交互, 以允许人类检查和编辑ReAct的推理过程。图5显示, 通过简单地删除Act 17中的一句幻想句子, 并在Act 23中添加一些提示, ReAct的行为可以发生drastical变化, 以与这些人类思考编辑对齐并成功完成任务。从人类的角度来看, 解决这样的任务变得更加容易, 从输入数十个动作变为仅仅编辑几个思考, 这使得人机协作的新形式成为可能。我们注意到这样的策略编辑是即时进行的。

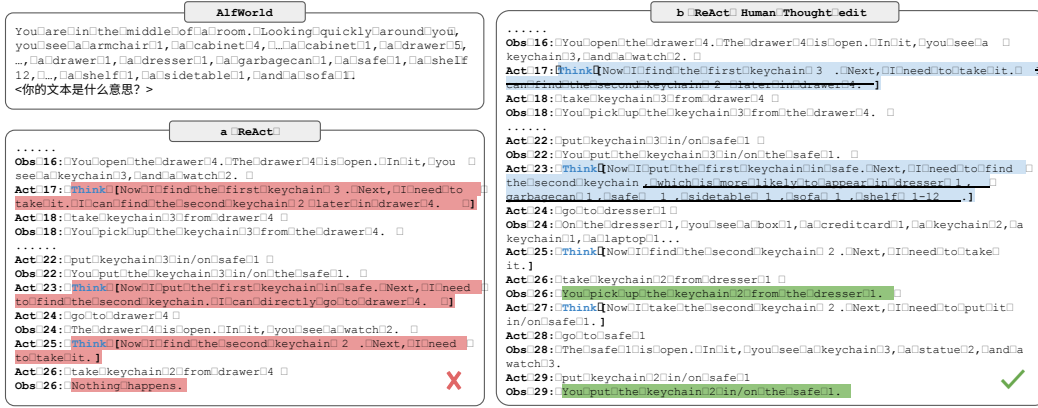


图5：在AlfWorld中使用ReAct进行人机协同行为修正的示例。（a）ReAct轨迹由于产生了幻觉思维（Act 17）而失败。（b）通过人类简单编辑两个思维（Act 17, 23），ReAct轨迹产生了理想的推理轨迹和行动，并成功完成任务。

对于Act和之前的强化学习方法来说，这是困难的，因为人类无法更改模型参数，并且更改少数行动可能无法编辑模型的其余行为。这种范式不仅仅是人类对话以更新目标或子目标，如Huang等人（2022b）所述 - 而编辑ReAct思维可以做到这些，它还可以修改模型的内部信念、推理风格或任何灵活的思维空间支持的内容，以更好地解决任务。我们相信这是人机协同的一个令人兴奋的方向，并将更多系统性研究作为未来的工作留下。

## B EXPERIMENT DETAILS

### B.1 HOTPOT QA FINETUNING DETAILS

对于所有的微调，我们使用批量大小为64。在PaLM-8B上，我们对ReAct和Act方法进行4,000步的微调，对Standard和CoT方法进行2,000步的微调。在PaLM-62B上，我们对ReAct和Act方法进行4,000步的微调，对Standard和CoT方法进行1,000步的微调。我们发现，ReAct和Act方法通常受益于更多的训练步骤（和更多的训练数据），而Standard和CoT方法在微调后很快退化。

### B.2 ALFWORLD IM-STYLE DETAILS

对于IM风格的消融，与ReAct中使用的相同专家轨迹被重新注释为密集型。这些轨迹中的外部反馈思考限制了ReAct-IM仅考虑（1）分解当前目标和（2）需要完成的当前子目标。特别是，ReAct-IM缺乏以下思考：（1）确定何时完成子目标，（2）确定下一个子目标应该是什么，（3）引导LLM参考其内部预训练知识以确定环境中的物品位置。



CP提示

C.1 H热锅QA

热锅QA提示		
原始问题		科罗拉多造山运动的东部区域的海拔范围是多少？
答案		1,800到7,000英尺
问题		音乐家和讽刺作家阿莉·戈茨写了一首关于《辛普森一家》角色米尔豪斯的歌曲，米尔豪斯的名字是由马特·格勒宁以谁的名字命名的？
答案		理查德·尼克松
问题		关于芬兰摇滚乐团的纪录片是关于亚当·克莱顿·鲍威尔还是The Saimaa Gesture？
答案		The Saimaa Gesture
问题		尼古拉斯·雷和埃利亚·卡赞有共同的职业吗？
答案		导演，编剧，演员
问题		Arthur's Magazine和First for Women哪个杂志先创刊？
答案		Arthur's Magazine
问题		帕维尔·乌里索恩和列昂尼德·列文以相同类型的工作而闻名吗？
答案		是的
行动		
问题		科罗拉多造山运动的东部区域的海拔范围是多少？
行动观察1	1	搜索[科罗拉多造山运动] 科罗拉多造山运动是科罗拉多及周边地区的一次山脉形成事件（造山运动）。
行动观察2	2	查找[东部地区] （结果1/1）东部地区延伸到高原地带，被称为中央平原造山运动。
行动观察3	3	搜索[高原地带] 高原地带指的是两个不同的土地区域之一：
行动观察4	4	搜索[高原地带（美国）] 高原是大平原的一个子区域。从东到西，高原的海拔从约1,800到7,000英尺（550到2,130米）升高。[3]行动
	5	完成[1,800到7,000英尺]
问题		音乐家和讽刺作家阿莉·戈茨写了一首关于《辛普森一家》角色米尔豪斯的歌曲，米尔豪斯的名字是由马特·格勒宁以谁的名字命名的？
行动1观察1		搜索[米尔豪斯] 米尔豪斯·穆索里尼·范·豪滕是福克斯动画电视系列《辛普森一家》中的一个经常出现的角色，由帕梅拉·海登配音，由马特·格勒宁创作。
行动2观察2		查找[以...命名] （结果1/1）米尔豪斯是以美国总统理查德·尼克松命名的，他的中间名是米尔豪斯。行动3
		完成[理查德·尼克松]
问题		关于芬兰摇滚乐团的纪录片是关于亚当·克莱顿·鲍威尔还是The Saimaa Gesture？
行动1观察1		搜索[亚当·克莱顿·鲍威尔] 找不到[亚当·克莱顿·鲍威尔]。类似的：['亚当·克莱顿·鲍威尔三世'，'第七大道（曼哈顿）'，'亚当·克莱顿·鲍威尔小学'，'伊莎贝尔·华盛顿·鲍威尔'，'亚当·鲍威尔'，'亚当·克莱顿·鲍威尔（电影）'，'吉安卡洛·埃斯波西托']。行动2
观察2		搜索[亚当·克莱顿·鲍威尔（电影）] 《亚当·克莱顿·鲍威尔》是一部1989年的美国纪录片，由理查德·基尔伯格执导。该片讲述了有影响力的非裔美国政治家亚当·克莱顿·鲍威尔的崛起与衰落。后来，它作为PBS系列节目《美国经历》的一部分播出。行动3
		完成[赛马手势]
问题		尼古拉斯·雷和埃利亚·卡赞有共同的职业吗？
行动1观察1		搜索[尼古拉斯·雷] 尼古拉斯·雷（Raymond Nicholas Kienzle Jr.，1911年8月7日-1979年6月16日）是一位美国电影导演、编剧和演员，以1955年的电影《无因的叛逆》而闻名。行动2
观察2		搜索[埃利亚·卡赞] 埃利亚·卡赞是一位美国电影和戏剧导演、制片人、编剧和演员。

续下一页

火锅QA提示-继上一页

行动	3	完成[导演、编剧、演员]
问题		Arthur's Magazine和First for Women哪个杂志先创刊?
行动1		搜索[亚瑟杂志]
观察	1	《亚瑟杂志》(1844-1846年)是19世纪在费城出版的一本美国文学期刊。
行动2		搜索[First for Women]
观察	2	《First for Women》是由Bauer Media Group在美国出版的女性杂志。该杂志始创于1989年。
行动	3	完成[亚瑟杂志]
问题		帕维尔·乌里索恩和列昂尼德·列文以相同类型的工作而闻名吗?
行动1		搜索[帕维尔·乌里索恩]
观察1		帕维尔·萨穆伊洛维奇·乌里索恩(1898年2月3日-1924年8月17日)是苏联的数学家,以其在维度理论方面的贡献而闻名。
行动2		搜索[列昂尼德·列文]
观察	2	列昂尼德·阿纳托利耶维奇·列文是苏联-美国的数学家和计算机科学家。
行动	3	完成[是]
Cot 问题		科罗拉多造山运动的东部区域的海拔范围是多少?
思考		让我们一步一步地思考。科罗拉多造山运动的东部地区延伸到高原地带。高原地带的海拔从1800到7000英尺,所以答案是1800到7000英尺。答案
		1,800到7,000英尺
问题		音乐家和讽刺作家阿莉·戈茨写了一首关于《辛普森一家》角色米尔豪斯的歌曲,米尔豪斯的名字是由马特·格勒宁以谁的名字命名的?
思考		让我们一步一步地思考。米尔豪斯是以美国总统理查德·尼克松命名的,所以答案是理查德·尼克松。答案
		理查德·尼克松
问题		关于芬兰摇滚乐团的纪录片是关于亚当·克莱顿·鲍威尔还是The Saimaa Gesture?
思考		让我们一步一步地思考。《亚当·克莱顿·鲍威尔(电影)》是一部关于非裔美国政治家的纪录片,而不是关于芬兰摇滚乐团的。所以关于芬兰摇滚乐团的纪录片必须是《赛马手势》。答案
		The Saimaa Gesture
问题		尼古拉斯·雷和埃利亚·卡赞有共同的职业吗?
思考		让我们一步一步地思考。尼古拉斯·雷的职业是导演、编剧和演员。埃利亚·卡赞的职业是导演、制片人、编剧和演员。所以尼古拉斯·雷和埃利亚·卡赞共同拥有的职业是导演、编剧和演员。
答案		导演, 编剧, 演员
问题		Arthur's Magazine和First for Women哪个杂志先创刊?
思考		让我们一步一步地思考。《亚瑟杂志》始创于1844年。《First for Women》始创于1989年。1844年(《亚瑟杂志》)<1989年(《First for Women》),所以《亚瑟杂志》是先创办的。答案
		Arthur's Magazine
问题		帕维尔·乌里索恩和列昂尼德·列文以相同类型的工作而闻名吗?
思考		让我们一步一步地思考。帕维尔·乌里索恩是一位数学家。列昂尼德·列文是一位数学家和计算机科学家。所以帕维尔·乌里索恩和列昂尼德·列文从事相同类型的工作。
答案		是的
ReAct t 问题		科罗拉多造山运动的东部区域的海拔范围是多少?
思考1		我需要搜索科罗拉多造山运动,找到科罗拉多造山运动东部区域延伸的地区,然后找到该地区的海拔范围。
行动1		搜索[科罗拉多造山运动]
观察1		科罗拉多造山运动是科罗拉多及周边地区的一次山脉形成事件(造山运动)。
思考2		它没有提到东部区域。所以我需要查找东部区域。
行动2		查找[东部区域] 观察
	2	(结果1/1) 东部地区延伸到高原地带,被称为中央平原造山运动。
思考	3	科罗拉多造山运动的东部区域延伸到高原。所以我需要搜索高原并找到它的海拔范围。
行动	3	搜索[高原] 观察
	3	高原指的是两个不同的地区之一。
	4	我需要改为搜索高原(美国)。
	4	搜索[高原(美国)]

续下一页

## 火锅QA提示-继上一页

观察4	高原是大平原的一个子区域。从东到西，高原的海拔从约1800到7000英尺（550到2130米）上升。
	高原的海拔从1800到7000英尺不等，所以答案是1800到7000英尺。行动5
	完成[1,800到7,000英尺]
问题	音乐家和讽刺作家阿莉·戈茨写了一首关于《辛普森一家》角色米尔豪斯的歌曲，米尔豪斯的名字是由马特·格勒宁以谁的名字命名的？
思考1	问题简化为“辛普森一家”角色米尔豪斯是以谁的名字命名的。我只需要搜索米尔豪斯并找到它是以谁的名字命名的。
行动1	搜索[米尔豪斯]
观察1	米尔豪斯·穆索里尼·范·豪滕是福克斯动画电视系列《辛普森一家》中的一个经常出现的角色，由帕梅拉·海登配音，由马特·格勒宁创作。
思考2	段落没有告诉米尔豪斯是以谁的名字命名的，也许我可以查一下“以...命名”。
行动2	查找[以...命名]
观察2	(结果1/1) 米尔豪斯是以美国总统理查德·尼克松的名字命名的，他的中间名是米尔豪斯。思考3
	米尔豪斯是以美国总统理查德·尼克松的名字命名的，所以答案是理查德·尼克松。行动3
	完成[理查德·尼克松]
问题	关于芬兰摇滚乐团的纪录片是关于亚当·克莱顿·鲍威尔还是The Saimaa Gesture?
思考1	我需要搜索亚当·克莱顿·鲍威尔和《赛马手势》，找出哪个纪录片是关于芬兰摇滚乐团的。
行动1	搜索[亚当·克莱顿·鲍威尔]
观察1	找不到[亚当·克莱顿·鲍威尔]。类似的：[‘亚当·克莱顿·鲍威尔三世’，‘第七大道（曼哈顿）’，‘亚当·克莱顿·鲍威尔小学办公楼’，‘伊莎贝尔·华盛顿·鲍威尔’，‘亚当·鲍威尔’，‘亚当·克莱顿·鲍威尔（电影）’，‘吉安卡洛·埃斯波西托’]。思考2
行动2	为了找到这部纪录片，我可以搜索亚当·克莱顿·鲍威尔（电影）。
观察2	搜索[亚当·克莱顿·鲍威尔（电影）] 《亚当·克莱顿·鲍威尔》是一部1989年的美国纪录片，由理查德·基尔伯格执导。该片讲述了有影响力的非裔美国政治家亚当·克莱顿·鲍威尔·朱尔回升和垮台的故事。后来，它作为PBS系列节目《美国经历》的一部分播出。思考3
	《亚当·克莱顿·鲍威尔》是一部关于非裔美国政治家的纪录片，而不是关于芬兰摇滚乐团的。所以关于芬兰摇滚乐团的纪录片必须是《赛马手势》。行动3
	完成[赛马手势]
问题	尼古拉斯·雷和埃利亚·卡赞有共同的职业吗？
思考1	我需要搜索尼古拉斯·雷和埃利亚·卡赞，找到他们的职业，然后找到他们共同的职业。
行动1	搜索[尼古拉斯·雷]
观察1	尼古拉斯·雷（原名雷蒙德·尼古拉斯·基恩兹勒二世，1911年8月7日-1979年6月16日）是一位美国电影导演、编剧和演员，以1955年的电影《无因的叛逆》而闻名。思考2
	尼古拉斯·雷的职业是导演、编剧和演员。接下来我需要搜索埃利亚·卡赞，并找出他的职业。
行动2	搜索[埃利亚·卡赞]
观察2	埃利亚·卡赞是一位美国电影和戏剧导演、制片人、编剧和演员。
思考3	埃利亚·卡赞的职业是导演、制片人、编剧和演员。 所以尼古拉斯·雷和埃利亚·卡赞共同的职业是导演、编剧和演员。
行动3	完成[导演、编剧、演员]
问题	Arthur's Magazine和First for Women哪个杂志先创刊？
思考1	我需要搜索《亚瑟杂志》和《First for Women》，并找出哪个先开始。
行动1	搜索[《亚瑟杂志》]
观察1	《亚瑟杂志》（1844-1846年）是19世纪在费城出版的一本美国文学期刊。
思考2	《亚瑟杂志》始于1844年。接下来我需要搜索《First for Women》。
行动2	搜索[《First for Women》]
观察2	《First for Women》是由Bauer Media Group在美国出版的女性杂志。该杂志始创于1989年。
思考3	《First for Women》始于1989年。1844年（《亚瑟杂志》）<1989年（《First for Women》），所以《亚瑟杂志》先开始。行动3
	完成[亚瑟杂志]
问题	帕维尔·乌里索恩和列昂尼德·列文以相同类型的工作而闻名吗？
思考1	我需要搜索帕维尔·乌里索恩和列昂尼德·列文，并找出他们的工作类型，然后判断它们是否相同。
行动1	搜索[Pavel Urysohn]

续下一页

火锅QA提示-继上一页	
观察1	帕维尔·萨穆伊洛维奇·乌里索恩（1898年2月3日-1924年8月17日）是苏联的数学家，以其在维度理论方面的贡献而闻名。
思考2	Pavel Urysohn是一位数学家。 我需要搜索下一个Leonid Levin，并找到他的工作类型。
行动2 察2	搜索[Leonid Levin] 观 列昂尼德·阿纳托利耶维奇·列文是苏联-美国的数学家和计算机科学家。
思考 3	Leonid Levin是一位数学家和计算机科学家。 所以Pavel Urysohn和Leonid Levin有相同的工作类型。
行动 3	完成[是]

C.2F发热

发热提示		
原始		确定是否有支持或反驳主张的观察，或者是否没有足够的信息。
主张		Nikolaj Coster-Waldau曾与福克斯广播公司合作。回答支持
主张		Stranger Things的背景设定在印第安纳州的布卢明顿。回答反驳
主张答案		Beautiful在2003年登上了Billboard Hot 100排行榜的第二名。信息不足
行动		确定是否有支持或反驳主张的观察，或者是否没有足够的信息。
主张		Nikolaj Coster-Waldau曾与福克斯广播公司合作。行动2
观察	1	搜索[Nikolaj Coster-Waldau] 观 尼古拉伊·威廉·科斯特-瓦尔道（出生于1970年7月27日）是一位丹麦演员和制片人。他于1993年毕业于哥本哈根的丹麦国立表演艺术学院，并在丹麦的电影《夜巡》（1994年）中取得了突破性的角色。他在HBO的奇幻剧集《权力的游戏》中饰演詹姆·兰尼斯特，因此获得了两次杰出男配角提名的艾美奖提名。科斯特-瓦尔道在他的祖国丹麦和斯堪的纳维亚地区出演了许多电影，包括《猎头者》（2011年）和《千次晚安》（2013年）。在美国，他的首部电影作品是战争片《黑鹰坠落》（2001年），饰演荣誉勋章获得者加里·戈登。[2]然后他在短命的福克斯电视剧《新阿姆斯特丹》（2008年）中饰演一名侦探，并出演了2009年的福克斯电视电影《虚拟现实》，最初是作为试播集。
行动	2	(2001), playing Medal of Honor recipient Gary Gordon.[2] He then played a detective in the short-lived Fox television series New Amsterdam (2008), and appeared in the 2009 Fox television film Virtuality, originally intended as a pilot. 完成[SUPPORTS]
主张		《怪奇物语》设定在印第安纳州的布卢明顿。
行动1		搜索[怪奇物语]
观察	1	《怪奇物语》是由达弗兄弟创作的美国科幻恐怖剧，主要设定在1980年代的虚构城镇霍金斯，围绕该城镇发生的一系列神秘和超自然事件以及对一群儿童和成年角色的影响展开。
行动	2	完成[REFUTES]
主张		Beautiful在2003年登上了Billboard Hot 100排行榜的第二名。
行动1		搜索[美丽]
观察	1	未找到[美丽]。类似：['美丽', '美丽, 美丽', '美丽心灵（电影）', '美丽（Christina Aguilera的歌曲）', '生活是美丽的']。行动
观察	2	搜索[美丽（Christina Aguilera的歌曲）]
观察	2	《美丽》是美国歌手克里斯蒂娜·阿奎莱拉为她的第四张录音室专辑《剥离》（2002年）录制的一首歌曲。
行动	3	查找[公告牌百强单曲榜]
观察	3	（结果1/3）这首歌在美国公告牌热100榜上排名第二，销量达到50万份，获得金唱片认证。
行动	4	完成[信息不足]
COT		确定是否有支持或反驳主张的观察，或者是否没有足够的信息。
主张		尼古拉·科斯特-瓦尔道曾与福克斯广播公司合作。思考 尼古拉·威廉·科斯特-瓦尔道曾出演2009年的福克斯电视电影《虚拟现实》，所以他曾与福克斯广播公司合作。回答支持
主张		《怪奇物语》设定在印第安纳州布卢明顿。思考 《怪奇物语》发生在虚构的霍金斯镇，而不是布卢明顿，印第安纳州。回答反驳

续下一页

## 发热引发-续前页

主张思考		Beautiful在2003年登上了Billboard Hot 100排行榜的第二名。这首歌在美国公告牌热100榜上排名第二，但不确定是否是在2003年。
答案		信息不足
反应		确定是否有支持或反驳主张的观察，或者是否没有足够的信息。
主张	1	尼古拉·科斯特-瓦尔道曾与福克斯广播公司合作。思考
思考	1	我需要搜索尼古拉·科斯特-瓦尔道并查找他是否曾与福克斯广播公司合作。行动
观察	1	搜索[Nikolaj Coster-Waldau] 观 尼古拉伊·威廉·科斯特-瓦尔道（出生于1970年7月27日）是一位丹麦演员和制片人。他于1993年毕业于哥本哈根的丹麦国立表演艺术学院，并在丹麦的电影《夜巡》（1994年）中取得了突破性的角色。他在HBO的奇幻剧集《权力的游戏》中饰演詹姆·兰尼斯特，因此获得了两次杰出男配角提名的艾美奖提名。科斯特-瓦尔道在他的祖国丹麦和斯堪的纳维亚地区出演了许多电影，包括《猎头者》（2011年）和《千次晚安》（2013年）。在美国，他的首部电影作品是战争片《黑鹰坠落》（2001年），饰演荣誉勋章获得者加里·戈登。[2]然后他在短命的福克斯电视剧《新阿姆斯特丹》（2008年）中饰演一名侦探，并出演了2009年的福克斯电视电影《虚拟现实》，最初是作为试播集。
思考	2	(2001), playing Medal of Honor recipient Gary Gordon.[2] He then played a detective in the short-lived Fox television series New Amsterdam (2008), and appeared in the 2009 Fox television film Virtuality, originally intended as a pilot.
	2	因为他“出演了2009年的福克斯电视电影《虚拟现实》”，所以他应该曾与福克斯广播公司合作。行动
		完成[SUPPORTS]
主张思考1		《怪奇物语》设定在印第安纳州的布卢明顿。我应该搜索《怪奇物语》，看看它是否设定在印第安纳州的布卢明顿。
行动1观察	1	搜索[怪奇物语] 《怪奇物语》是由达弗兄弟创作的美国科幻恐怖剧，主要设定在1980年代的虚构城镇霍金斯，围绕该城镇发生的一系列神秘和超自然事件以及对一群儿童和成年角色的影响展开。
思考	2	观察结果表明它设定在一个名为霍金斯的虚构城镇，所以它并不是设定在布卢明顿。行动
	2	完成[REFUTES]
主张思考1		Beautiful在2003年登上了Billboard Hot 100排行榜的第二名。我需要搜索歌曲《Beautiful》，并查看它是否在2003年登上了公告牌百强单曲榜的第二名。
行动1观察	1	搜索[美丽] 找不到[Beautiful]。类似的：['Beautiful', 'Beautiful, Beautiful', '美丽心灵（电影）', '美丽（Christina Aguilera的歌曲）', '生活是美丽的']。思考
	2	根据建议，我应该搜索《美丽（Christina Aguilera的歌曲）》来找到这首歌。
行动观察	2	搜索[美丽（Christina Aguilera的歌曲）] 《美丽》是美国歌手克里斯蒂娜·阿奎莱拉为她的第四张录音室专辑《剥离》（2002年）录制的一首歌曲。
思考	3	它没有提到公告牌，所以我需要查找“公告牌百强单曲榜”来确定它是否在2003年登上了第二名。
行动观察	3	查找[公告牌百强单曲榜] (结果1/3) 这首歌在美国公告牌热100榜上排名第二，销量达到50万份，获得金唱片认证。
思考	4	它只说这首歌在Billboard Hot 100上排名第二，但没有提到是在2003年。我不确定这个说法是否正确。行动
	4	完成[信息不足]

C.3 WEBSHOP

表6: Webshop实验中使用的提示，见第4节。我们在左边使用提示进行行动，并在右边使用提示（添加推理）进行反应。

指令：我想要一瓶3盎司的明亮柑橘味敏感皮肤除臭剂，价格低于50.00美元	
行动	反应
<p>行动：搜索 [3盎司明亮柑橘味敏感皮肤除臭剂]</p> <p>返回搜索 ]</p> <p>下一页]</p> <p>[B078GWRC1J] 地球妈妈明亮柑橘味除臭剂天然安全，适用于敏感皮肤、怀孕和哺乳期，含有有机金盏花3盎司\$10.99B078GTKVXY</p> <p>地球妈妈姜新鲜除臭剂天然安全，适用于敏感肌肤，孕妇和哺乳期，含有有机金盏花3盎司\$10.99B08KBVJ4XN</p> <p>桶和橡木-无铝除臭剂，男士专用除臭剂，基于精油的香氛，24小时防臭，雪松和广藿香混合，对敏感肌肤温和（山地鼠2.7盎司，2包）\$15.95</p> <p>行动：点击[B078GWRC1J] 观察： 返回搜索 ] 上一页] 香味[混合香味][明亮柑橘][宁静薰衣草][姜[新鲜]简单无味] 尺寸[旅行套装（4包）][3盎司（1包）][3盎司（2包）] 地球妈妈明亮柑橘除臭剂天然安全，适用于敏感肌肤，孕妇和哺乳期，含有有机金盏花3盎司价格：\$10.99评分：N.A.[描述]</p> <p>[特点] [评论] [立即购买]</p> <p>行动：点击[明亮的柑橘] 观察：你点击了明亮的柑橘。</p> <p>行动：点击[3盎司（1包装）] 观察：你点击了3盎司（1包装）。</p> <p>行动：点击[立即购买]</p>	<p>行动：搜索 [3盎司明亮柑橘味敏感皮肤除臭剂]</p> <p>返回搜索 ]</p> <p>下一页]</p> <p>[B078GWRC1J] 地球妈妈明亮柑橘味除臭剂天然安全，适用于敏感皮肤、怀孕和哺乳期，含有有机金盏花3盎司\$10.99B078GTKVXY</p> <p>地球妈妈姜新鲜除臭剂天然安全，适用于敏感肌肤，孕妇和哺乳期，含有有机金盏花3盎司\$10.99B08KBVJ4XN</p> <p>桶和橡木-无铝除臭剂，男士专用除臭剂，基于精油的香氛，24小时防臭，雪松和广藿香混合，对敏感肌肤温和（山地鼠2.7盎司，2包）\$15.95</p> <p>行动：思考 [B078GWRC1J和B078GTKVXY是明亮的柑橘味道的50美元以下的除臭剂。我可以先检查B078GWRC1J。 观察：好的。</p> <p>行动：点击[B078GWRC1J] 观察： 返回搜索 ] 上一页] 香味[混合香味][明亮柑橘][宁静薰衣草][姜[新鲜]简单无味] 尺寸[旅行套装（4包）][3盎司（1包）][3盎司（2包）] 地球妈妈明亮柑橘除臭剂天然安全，适用于敏感肌肤，孕妇和哺乳期，含有有机金盏花3盎司价格：\$10.99评分：N.A.[描述]</p> <p>[特点] [评论] [立即购买]</p> <p>行动：思考[对于敏感肌肤的3盎司瓶装明亮柑橘味道的除臭剂，该商品有“明亮的柑橘”和“3盎司（1包装）”两个选项，看起来不错。] 观察：好的。</p> <p>行动：点击[明亮的柑橘] 观察：你点击了明亮的柑橘。</p> <p>行动：点击[3盎司（1包装）] 观察：你点击了3盎司（1包装）。</p> <p>行动：点击[立即购买]</p>



## C.4 ALFWORLD

表7: ALFWorld清洁任务上的一个行动提示。没有提供任何想法。

---

你正在一个房间的中间。快速环顾四周，你看到一个柜子13，一个柜子12，一个柜子11，一个柜子10，一个柜子9，一个柜子8，一个柜子7，一个柜子6，一个柜子5，一个柜子4，一个柜子3，一个柜子2，一个柜子1，一个咖啡机1，一个台面1，一个餐桌1，一个抽屉1，一个冰箱1，一个垃圾桶1，一个微波炉1，一个架子3，一个架子2，一个架子1，一个水槽1，一个""炉灶炉头4，一个炉灶炉头3，一个炉灶炉头2，一个炉灶炉头1，和一个烤面包机1。

你的任务是：将一个干净的生菜放在餐桌上。

> 去冰箱 1

冰箱1是关闭的。

> 打开冰箱 1

你打开冰箱 1。冰箱 1已经打开。在里面，你看到一个杯子 3，一个鸡蛋 2，一个土豆 3和一个土豆 2。

> 去餐桌 1

在餐桌 1上，你看到一个苹果 1，一个面包 1，一个黄油刀 2，一个杯子 2，一个叉子 2，一个刀子 2，一个刀子 1，一个勺子 1，一个生菜 1，一个杯子 2，一个杯子 1，一个平底锅 2，一个胡椒瓶 1，一个铲子 3，一个番茄 2和一个番茄 1。

> 从餐桌 1上拿起生菜 1

你从餐桌 1上拿起生菜 1。

> 去水槽盆 1

在水槽盆 1上，你看到一个苹果 2，一个勺子 2，一个汤匙 1和一个番茄 3。

> 用水槽 1清洗生菜 1

你用水槽1清洗生菜1。

> 思考：现在我要清洗一个生菜（ 1）。接下来，我需要把它放在餐桌 1上。

好的。

> 去餐桌 1

在餐桌1上，你看到一个苹果1、一个面包1、一个黄油刀2、一个杯子2、一个叉子2、一个刀2、一个刀1、一个勺子1、一个杯子2、一个杯子1、一个锅2、一个胡椒瓶1、一个铲子3、一个番茄2和一个番茄1。

> 把生菜 1放在餐桌 1上

你把生菜1放在餐桌1上。

---

表8: ALFWorld清洁任务上的ReAct提示。

---

你正在一个房间的中间。快速环顾四周，你看到一个柜子13，一个柜子12，一个柜子11，一个柜子10，一个柜子9，一个柜子8，一个柜子7，一个柜子6，一个柜子5，一个柜子4，一个柜子3，一个柜子2，一个柜子1，一个咖啡机1，一个台面1，一个餐桌1，一个抽屉1，一个冰箱1，一个垃圾桶1，一个微波炉1，一个架子3，一个架子2，一个架子1，一个水槽1，一个"炉灶炉头4，一个炉灶炉头3，一个炉灶炉头2，一个炉灶炉头1，和一个烤面包机1。

你的任务是：将一个干净的生菜放在餐桌上。

>

思考：为了完成任务，我需要找到并拿一个生菜，然后用水槽清洗它，然后放在餐桌上。

好的。

> 思考：首先我需要找到一颗生菜。生菜更有可能出现在冰箱（1），餐桌（1），水槽（1），炉灶（1-3），橱柜（1-13）中。我可以逐个检查，从冰箱1开始。

好的。

> 前往冰箱1

冰箱1是关闭的。

> 打开冰箱1

你打开冰箱1。冰箱1已经打开。在里面，你看到一个杯子3，一个鸡蛋2，一个土豆3和一个土豆2。

> 前往餐桌1

在餐桌1上，你看到一个苹果1，一个面包1，一个黄油刀2，一个杯子2，一个叉子2，一个刀子2，一个刀子1，一个勺子1，一个生菜1，一个杯子2，一个杯子1，一个平底锅2，一个胡椒瓶1，一个铲子3，一个番茄2和一个番茄1。

> 思考：现在我找到了一颗生菜（1）。接下来，我需要拿起它。

好的。

> 从餐桌1拿起生菜1

你从餐桌1上拿起生菜1。

> 思考：现在我拿起了一颗生菜（1）。接下来，我需要去水槽（1）清洗它。

好的。

> 前往水槽1

在水槽盆1上，你看到一个苹果2，一个勺子2，一个汤匙1和一个番茄3。

> 使用水槽1清洗生菜1

你使用水槽1清洗了生菜1。

> 思考：现在我清洗了一颗生菜（1）。接下来，我需要将它放在餐桌1上。

好的。

> 前往餐桌1

在餐桌1上，你看到一个苹果1、一个面包1、一个黄油刀2、一个杯子2、一个叉子2、一个刀2、一个刀1、一个勺子1、一个杯子2、一个杯子1、一个锅2、一个胡椒瓶1、一个铲子3、一个番茄2和一个番茄1。

> 把生菜1放在/上餐桌1

你把生菜1放在/上餐桌1。

---

表格9：ALFWorld清洁任务上的ReAct-IM提示。

---

你正在一个房间的中间。快速环顾四周，你看到一个柜子13，一个柜子12，一个柜子11，一个柜子10，一个柜子9，一个柜子8，一个柜子7，一个柜子6，一个柜子5，一个柜子4，一个柜子3，一个柜子2，一个柜子1，一个咖啡机1，一个台面1，一个餐桌1，一个抽屉1，一个冰箱1，一个垃圾桶1，一个微波炉1，一个架子3，一个架子2，一个架子1，一个水槽1，一个""炉灶炉头4，一个炉灶炉头3，一个炉灶炉头2，一个炉灶炉头1，和一个烤面包机1。

你的任务是：将一个干净的生菜放在餐桌上。

>

思考：为了完成任务，我需要找到并拿一个生菜，然后用水槽清洗它，然后放在餐桌上。

好的。

> 思考：首先我需要找到一颗生菜。

好的。

> 前往冰箱1

冰箱1是关闭的。

> 思考：首先我需要找到一颗生菜。

好的。

> 打开冰箱1

你打开冰箱1。冰箱1已经打开。在里面，你看到一个杯子3，一个鸡蛋2，一个土豆3和一个土豆2。

> 思考：首先我需要找到一颗生菜。

好的。

> 前往餐桌1

在餐桌1上，你看到一个苹果1，一个面包1，一个黄油刀2，一个杯子2，一个叉子2，一个刀子2，一个刀子1，一个勺子1，一个生菜1，一个杯子2，一个杯子1，一个平底锅2，一个胡椒瓶1，一个铲子3，一个番茄2和一个番茄1。

> 思考：首先我需要找到一颗生菜。

好的。

> 从餐桌1拿起生菜1

你从餐桌1上拿起生菜1。

> 思考：我需要使用水槽1清洗这颗生菜（1）。

好的。

> 前往水槽1

在水槽盆1上，你看到一个苹果2，一个勺子2，一个汤匙1和一个番茄3。

> 思考：我需要使用水槽1清洗这颗生菜（1）。

好的。

> 使用水槽1清洗生菜1

你使用水槽1清洗了生菜1。

> 思考：我需要把这颗清洗过的生菜（1）放在/上餐桌1。

好的。

> 前往餐桌1

在餐桌1上，你看到一个苹果1、一个面包1、一个黄油刀2、一个杯子2、一个叉子2、一个刀2、一个刀1、一个勺子1、一个杯子2、一个杯子1、一个锅2、一个胡椒瓶1、一个铲子3、一个番茄2和一个番茄1。

> 思考：我需要把这颗清洗过的生菜（1）放在/上餐桌1。

好的。

> 把生菜1放在/上餐桌1

你把生菜1放在/上餐桌1。

---

## D T轨迹

### D.1 FEVER T轨迹

我们在FEVER开发集中随机选择的几个例子中提供了一些由ReAct、CoT和Act生成的轨迹。搜索结果已省略以节省空间。

示例2491 (gt: REFUTES)

<b>ReAct</b>	
主张:	百慕大三角是太平洋中一个定义不明确区域。
思考1:	我需要搜索百慕大三角并找出它是否在太平洋中。
行动1:	搜索[百慕大三角]
观察1:	据报道, 闹鬼的地点...
思考2:	观察表明它位于北大西洋的西部部分 所以它不在太平洋。
行动2:	完成[反驳]
观察2:	剧集结束
<b>行动</b>	
主张:	百慕大三角是太平洋中一个定义模糊的区域。
行动1:	: 搜索[百慕大三角]
观察1:	: 据报道, 闹鬼的地点: ...
行动2:	完成[反驳]
观察2:	剧集结束
<b>Cot</b>	
主张:	百慕大三角是太平洋中一个定义模糊的区域。
思考:	让我们一步一步地思考。百慕大三角是一个定义模糊的区域在大西洋中。太平洋不是大西洋。
答案:	反驳

例子5908 (支持)

<b>ReAct</b>	
主张:	《幽灵公主》是一部电影。
思考1:	我需要搜索《幽灵公主》并找出它是否是一部电影。
行动1:	搜索[《幽灵公主》]
观察1:	《幽灵公主》...
思考2:	从观察中可以得知, 《幽灵公主》是一部电影。
行动2:	完成[支持]
观察2:	剧集结束
<b>行动</b>	
主张:	《幽灵公主》是一部电影。
行动1:	搜索[《幽灵公主》]
观察1:	《幽灵公主》...
行动2:	完成[支持]
观察2:	剧集结束
<b>Cot</b>	
主张:	《幽灵公主》是一部电影。
思考:	让我们一步一步地思考。《幽灵公主》是一部电影, 而电影是电影。
答案:	支持

例子1951 (gt: 反驳)

<b>ReAct</b>	
主张:	联盟号是美国的太空计划的一部分。
思考1:	我需要搜索联盟号并查找它是否是美国太空计划的一部分。
行动1:	搜索[联盟号]
观察1:	联盟号是一个音译...
思考2:	它没有提到美国的太空计划, 所以我需要搜索"美国太空计划"来查找它是否是其中的一部分。
行动2:	搜索[美国太空计划]
观察2:	太空政策...
思考3:	它没有提到联盟号, 所以我不确定这个说法是否正确。
行动3:	完成[信息不足] 观察3:
	剧集结束
<b>行动</b>	
主张:	联盟号是美国的太空计划的一部分。
行动1:	搜索[联盟号]
观察1:	联盟号是一个音译...
行动2:	完成[信息不足] 观察2:
	剧集结束
<b>Cot</b>	
主张:	联盟号是美国的太空计划的一部分。
思考:	让我们一步一步地思考。联盟号是俄罗斯的宇宙飞船。美国的太空计划是NASA。NASA和俄罗斯在国际空间站上合作。答案:
	支持

例子3208 (gt: REFUTES)	
ReAct	
主张:	《重返人间》是一部2010年制作的美国电影。
思考1:	我需要搜索《重返人间》，并查找它是否是一部2010年制作的美国电影。
行动1:	搜索[重返人间] 观察1:
思考2:	《重返人间》...
观察2:	观察结果显示它是一部2007年制作的美国电影，所以不是2010年制作的。
行动2:	完成[反驳]
观察2:	剧集结束
行动	
主张:	《重返人间》是一部2010年制作的美国电影。
行动1:	搜索[重返人间] 行动2:
观察2:	完成[支持]
观察2:	剧集结束
Cot	
主张:	《重返人间》是一部2010年制作的美国电影。
思考:	让我们一步一步地思考。首先，《重返人间》是一部美国电影吗？是的，它是。其次，它是在2010年制作的吗？是的，它是。答案：
	支持

D.2 ALFWORLD轨迹

我们提供了一个关于ReAct、Act和ReAct-IM在同一个ALFWorld游戏中的示例轨迹，pick\_clean\_then\_place\_in\_recep-Knife-None-CounterTop-10/trial\_T20190909\_110445\_675754。

D.2.1 REACT轨迹

ReAct能够正确找到刀子，清洁它，然后将其放在台面上。

## ReAct ALFWorld轨迹

你处在一个房间的中间。快速环顾四周，你看到一个6号柜子，一个5号柜子，一个4号柜子，一个3号柜子，一个2号柜子，一个1号柜子，一个咖啡机1，一个3号台面，一个2号台面，一个1号台面，一个3号抽屉，一个2号抽屉，一个1号抽屉，一个1号冰箱，一个1号垃圾桶，一个1号微波炉，一个3号架子，一个2号架子，一个1号架子，一个1号水槽，一个4号炉灶，一个3号炉灶，一个2号炉灶，一个1号炉灶和一个烤面包机

1.

你的任务是：将一把干净的刀放在台面上。

> 思考：为了解决这个任务，我需要找到并拿起一把刀，然后用水槽清洗干净，最后放在台面上。

好的。

> 思考：首先我需要找到一把刀。刀更有可能出现在橱柜（1-6），抽屉（1-3），台面（1-3），冰箱（1），垃圾桶（1），架子（1-3），水槽（1），炉灶（1-4），烤面包机（1）。我可以逐个检查，从橱柜1开始。

好的。

> 去橱柜 1  
在橱柜1上，你看到一个碗1。

> 去橱柜 2  
橱柜2是关闭的。  
> 去橱柜 3  
在橱柜3上，你看到一个玻璃瓶1。

> 去橱柜 4  
在橱柜4上，你看到一个杯子1。

> 去橱柜 5  
橱柜5是关闭的。

> 去柜子 6  
柜子6号是关闭的。

> 去抽屉1  
抽屉1号是关闭的。

> 去抽屉2  
抽屉2号是关闭的。

> 去抽屉3  
抽屉3号是关闭的。

> 去台面1  
在台面1上，你看到一个生菜2，一个杯子2，一个胡椒瓶1和一个勺子2。

> 去台面2  
在台面2上，你看到一个杯子1，一个洗碗布1，一个玻璃瓶3，一个刀子1，一个盘子2，一个土豆3和一个雕像1。

> 思考：现在我找到了一个刀子（1）。接下来，我需要拿起它。

好的。

> 从台面2拿起刀子1

你从台面2上拿起了刀子1。

> 思考：现在我拿起了一个刀子（1）。接下来，我需要去水槽（1）清洗它。

好的。

> 前往水槽1

在水槽1上，你看到一个叉子3，一个生菜3和一个铲子2。

> 用水槽1清洁刀子1

你用水槽1清洁刀子1。

> 思考：现在我要清洁一把刀（1）。接下来，我需要把它放在/在台面上。

好的。

> 去台面1

在台面1上，你看到一个生菜2，一个杯子2，一个胡椒瓶1和一个勺子2。

> 把刀子1放在/在台面上

你把刀子1放在/在台面上。

## D.2.2 ACT 轨迹

Act能够正确找到刀子，但在去水槽之前尝试使用水槽清洁它。从那时起，模型无法恢复，并陷入重复相同命令序列的困境。使用ReAct，一个现有的思想将明确地确定刀子已经被拿走，下一个子目标是去水槽清洁它。

## 行动ALFWorld轨迹

你处在一个房间的中间。快速环顾四周，你看到一个6号柜子，一个5号柜子，一个4号柜子，一个3号柜子，一个2号柜子，一个1号柜子，一个咖啡机1，一个3号台面，一个2号台面，一个1号台面，一个3号抽屉，一个2号抽屉，一个1号抽屉，一个1号冰箱，一个1号垃圾桶，一个1号微波炉，一个3号架子，一个2号架子，一个1号架子，一个1号水槽，一个4号炉灶，一个3号炉灶，一个2号炉灶，一个1号炉灶和一个烤面包机

1.

你的任务是：将一把干净的刀放在台面上。

> 去柜台<sup>1</sup>

在台面1上，你看到一个生菜<sup>2</sup>，一个杯子<sup>2</sup>，一个胡椒瓶<sup>1</sup>和一个勺子<sup>2</sup>。

> 去柜台<sup>2</sup>

在台面2上，你看到一个杯子<sup>1</sup>，一个洗碗布<sup>1</sup>，一个玻璃瓶<sup>3</sup>，一个刀子<sup>1</sup>，一个盘子<sup>2</sup>，一个土豆<sup>3</sup>和一个雕像<sup>1</sup>。

> 拿刀<sup>1</sup>

你从台面2上拿起了刀子<sup>1</sup>。

> 去台面<sup>1</sup>

在台面1上，你看到一个生菜<sup>2</sup>，一个杯子<sup>2</sup>，一个胡椒瓶<sup>1</sup>和一个勺子<sup>2</sup>。

> 清洁刀<sup>1</sup> 用水槽1清洁

没有任何反应。

> 去柜台<sup>3</sup>

在柜台3上，你看到一个面包<sup>3</sup>，一个黄油刀<sup>2</sup>，一个手机<sup>1</sup>，一个信用卡

1，一个叉子<sup>2</sup>，一个盆栽<sup>1</sup>，一个刀<sup>2</sup>，一个铲子<sup>1</sup>，一个雕像<sup>3</sup>，一个番茄<sup>3</sup>，一个番茄

2，一个番茄<sup>1</sup>和一个花瓶<sup>2</sup>。

> 拿刀<sup>2</sup> 从柜台<sup>3</sup>

没有任何反应。

> 去台面<sup>2</sup>

在柜台2上，你看到一个杯子<sup>1</sup>，一个洗碗海绵<sup>1</sup>，一个玻璃瓶<sup>3</sup>，一个盘子<sup>2</sup>，一个土豆<sup>3</sup>和一个雕像<sup>1</sup>。

> 从柜台2拿刀<sup>1</sup>

没有任何反应。

> 去台面<sup>1</sup>

在台面1上，你看到一个生菜<sup>2</sup>，一个杯子<sup>2</sup>，一个胡椒瓶<sup>1</sup>和一个勺子<sup>2</sup>。

> 用水槽1清洁刀<sup>1</sup>

没有任何反应。

> 去柜台<sup>3</sup>

在柜台3上，你看到一个面包<sup>3</sup>，一个黄油刀<sup>2</sup>，一个手机<sup>1</sup>，一个信用卡

1，一个叉子<sup>2</sup>，一个盆栽<sup>1</sup>，一个刀<sup>2</sup>，一个铲子<sup>1</sup>，一个雕像<sup>3</sup>，一个番茄<sup>3</sup>，一个番茄

2，一个番茄<sup>1</sup>和一个花瓶<sup>2</sup>。

> 拿刀<sup>2</sup> 从柜台<sup>3</sup>

没有任何反应。

> 去台面<sup>2</sup>

在柜台2上，你看到一个杯子<sup>1</sup>，一个洗碗海绵<sup>1</sup>，一个玻璃瓶<sup>3</sup>，一个盘子<sup>2</sup>，一个土豆<sup>3</sup>和一个雕像<sup>1</sup>。

> 从柜台2拿刀<sup>1</sup>

没有任何反应。

> 去台面<sup>1</sup>

在台面1上，你看到一个生菜<sup>2</sup>，一个杯子<sup>2</sup>，一个胡椒瓶<sup>1</sup>和一个勺子<sup>2</sup>。

> 用水槽1清洁刀<sup>1</sup>

没有任何反应。

> 去柜台<sup>3</sup>

在柜台3上，你看到一个面包<sup>3</sup>，一个黄油刀<sup>2</sup>，一个手机<sup>1</sup>，一个信用卡

1，一个叉子<sup>2</sup>，一个盆栽<sup>1</sup>，一个刀<sup>2</sup>，一个铲子<sup>1</sup>，一个雕像<sup>3</sup>，一个番茄<sup>3</sup>，一个番茄

2，一个番茄<sup>1</sup>和一个花瓶<sup>2</sup>。

> 拿刀<sup>2</sup> 从柜台<sup>3</sup>

没有任何反应。

> 去台面<sup>2</sup>

在柜台2上，你看到一个杯子<sup>1</sup>，一个洗碗海绵<sup>1</sup>，一个玻璃瓶<sup>3</sup>，一个盘子<sup>2</sup>，一个土豆<sup>3</sup>和一个雕像<sup>1</sup>。

> 从柜台2拿刀<sup>1</sup>

没有任何反应。

> 去台面<sup>1</sup>

在台面1上，你看到一个生菜<sup>2</sup>，一个杯子<sup>2</sup>，一个胡椒瓶<sup>1</sup>和一个勺子<sup>2</sup>。

> 用水槽1清洁刀<sup>1</sup>

没有任何反应。

## D.2.3 REACT-IM 轨迹

ReAct-IM能够正确找到刀子，但无法清洁刀子。然后，LLM在尝试将这把刀子放在台面上时卡住了。一个可能的问题是错误的思考：我需要找到一把干净的刀子。这似乎让LLM相信刀子已经清洁过了。在ReAct中，这一系列子目标的具体捕捉将由每个思考来完成。



你处在一个房间的中间。快速环顾四周，你看到一个6号柜子，一个5号柜子，一个4号柜子，一个3号柜子，一个2号柜子，一个1号柜子，一个咖啡机1，一个3号台面，一个2号台面，一个1号台面，一个3号抽屉，一个2号抽屉，一个1号抽屉，一个1号冰箱，一个1号垃圾桶，一个1号微波炉，一个3号架子，一个2号架子，一个1号架子，一个1号水槽，一个4号炉灶，一个3号炉灶，一个2号炉灶，一个1号炉灶和一个烤面包机

1.  
你的任务是：将一把干净的刀放在台面上。  
> 思考：为了解决任务，我需要找到并拿起一把干净的刀子，然后放在台面上。

好的。  
> 思考：我需要找到一把干净的刀子。  
好的。  
> 去台面1  
在台面1上，你看到一个生菜2，一个杯子2，一个胡椒瓶1和一个勺子2。  
> 思考：我需要找到一把干净的刀子。  
好的。  
> 去台面2  
在台面2上，你看到一个杯子1，一个洗碗布1，一个玻璃瓶3，一个刀子1，一个盘子2，一个土豆3和一个雕像1。  
  
> 思考：我需要找到一把干净的刀子。  
好的。  
> 拿刀 1 从柜台2  
你从台面2上拿起了刀子1。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 去台面1  
在台面1上，你看到一个生菜2，一个杯子2，一个胡椒瓶1和一个勺子2。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 放置刀子 1 放在/上台面1  
你把刀子1放在/上台面1上。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 放置刀子 1 放在/上台面1  
没有任何反应。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 放置刀子 1 放在/上台面1  
没有任何反应。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 放置刀子 1 放在/上台面1  
没有任何反应。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 放置刀子 1 放在/上台面1  
没有任何反应。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。  
> 放置刀子 1 放在/上台面1  
没有任何反应。  
> 思考：我需要将这把刀子（1）放在/上台面1上。  
好的。

D.3 WEBSHOP TRAJECTORIES

表格10: Act和ReAct预测的Webshop示例轨迹。与Act相比, ReAct使用推理来找到满足所有目标属性的产品。

指令: 给我一个16包装的苹果肉桂冻干香蕉片, 价格低于50.00美元	
行动	反应
<p>行动: 搜索 [16包装苹果肉桂冻干香蕉片]</p> <p>返回搜索 ]</p> <p>下一页]</p> <p>[B00611VFZE]</p> <p>Brothers-ALL-Natural水果脆片, 草莓香蕉口味, 0.42盎司 (100包装) \$85.0 B096H2P6G2</p> <p>[</p> <p>Moon Fruit冻干水果零食。儿童水果零食-多种口味包装\$18.99 B092JLLYK6</p> <p>[</p> <p>Nature's Turn冻干水果零食-香蕉脆片-适合学校午餐或外出零食-无糖, 非转基因, 无麸质, 无人工添加剂 (0.53盎司) 6包装\$12.99</p> <p>行动: 点击[B00611VFZE]</p> <p>观察:</p> <p>返回搜索 ]</p> <p>上一页]</p> <p>口味名称[亚洲梨][香蕉][富士苹果和肉桂][草莓香蕉] Brothers-ALL-Natural水果脆片, 草莓香蕉, 0.42盎司 (100包装), 价格: \$85.0, 评分: N.A., [描述]</p> <p>[特点]</p> <p>[评论]</p> <p>[立即购买]</p> <p>行动: 点击[立即购买]</p>	<p>行动: 搜索 [16包装苹果肉桂冻干香蕉片]</p> <p>返回搜索 ]</p> <p>下一页]</p> <p>[B00611VFZE]</p> <p>Brothers-ALL-Natural水果脆片, 草莓香蕉口味, 0.42盎司 (100包装) \$85.0 B096H2P6G2</p> <p>[</p> <p>Moon Fruit冻干水果零食。儿童水果零食-多种口味包装\$18.99 B092JLLYK6</p> <p>[</p> <p>Nature's Turn冻干水果零食-香蕉脆片-适合学校午餐或外出零食-无糖, 非转基因, 无麸质, 无人工添加剂 (0.53盎司) 6包装\$12.99</p> <p>行动: 思考[B00611VFZE是草莓香蕉, 不是苹果肉桂。B096H2P6G2是水果零食, 不是冷冻香蕉片。B092JLLYK6是香蕉脆片, 不是苹果肉桂。我可以先检查B092JLLYK6。</p> <p>观察: 好的。</p> <p>行动: 点击[B092JLLYK6]</p> <p>观察:</p> <p>返回搜索 ]</p> <p>上一页]</p> <p>口味名称[苹果][苹果肉桂][香蕉][哈密瓜][桃子][梨][草莓][草莓香蕉][多口味样品包][超大口味包][果园口味包]</p> <p>尺寸[0.53盎司 (6包装)][0.53盎司 (8包装)][0.53盎司 (16包装)]</p> <p>大自然的冻干水果零食 - 香蕉脆片 - 完美的学校午餐或外出零食 - 无添加糖, 非转基因, 无麸质, 无人工成分 (0.53盎司) 6包价格: \$12.99评分: N.A.[描述]</p> <p>[特点]</p> <p>[评论]</p> <p>[立即购买]</p> <p>行动: 思考[对于16包苹果肉桂冻干香蕉脆片, 该商品有“苹果肉桂”和“0.53盎司 (16包)”两个选项, 看起来不错, 可以购买。]</p> <p>观察: 好的。</p> <p>行动: 点击[苹果肉桂]</p> <p>观察: 您已点击了苹果肉桂。</p> <p>行动: 点击[0.53盎司 (16包)]</p> <p>观察: 您已点击了0.53盎司 (16包)。</p> <p>行动: 点击[立即购买]</p>
得分: 0.125	得分: 1.0

## E MORE ANALYSIS

## E.1 成功和失败模式分析

我们提供了一些与第3.3节中给出的成功和错误模式分析相对应的示例。为了减少空间，搜索结果和非代表性步骤被省略了。

成功：真阳性	
<b>ReAct</b>	
问题：	作者大卫·查诺夫与一位曾担任美国海军上将并在哪位总统任期内担任驻英国大使的人合作过？
思考1：	我需要搜索大卫·查诺夫并找到他合作过的美国海军上将。
行动1：	搜索[大卫·查诺夫]
观察1：	大卫·查诺夫是一位著名的作者...
行动2：	搜索[威廉·J·克罗]
观察2：	威廉·J·克罗是...
思考3：	威廉·J·克罗曾在比尔·克林顿总统任期内担任驻英国大使。所以答案是比尔·克林顿。行动3：
观察3：	完成[比尔·克林顿] 剧集结束
<b>CoT</b>	
问题：	代表亚利桑那大学参加覆盖西部美国广阔地区（包括亚利桑那州、加利福尼亚州和新墨西哥州）的会议的团队的名称是什么？
思考：	让我们逐步思考。代表亚利桑那大学参加覆盖西部美国广阔地区（包括亚利桑那州、加利福尼亚州和新墨西哥州）的会议的团队的名称是亚利桑那州立大学太阳魔鬼。答案：
亚利桑那州立大学太阳魔鬼	
成功：假阳性	
<b>ReAct</b>	
问题：	桌游圣马可或时间之谜哪个先发布？
思考1：	我需要搜索圣马可和时间之谜，并找出哪个先发布。
行动1：	搜索[圣马可]
观察1：	圣马可可是威尼斯的一个...它是在威尼斯建造的。
思考2：	圣马可可是威尼斯的一个区域。接下来我需要搜索时间之谜。
行动2：	搜索[时间之谜]
观察2：	找不到[时间之谜]。类似的：[...]
思考3：	时间之谜是一部2013年的电影。所以圣马可先发布。
行动3：	完成[圣马可]
观察3：	剧集结束
<b>CoT</b>	
问题：	范德堡大学和埃默里大学哪个先成立？
思考：	让我们一步一步地思考。范德堡大学成立于1873年。埃默里大学成立于1836年。1873年（范德堡大学）>1836年（埃默里大学），所以范德堡大学先成立。答案：
埃默里大学	
失败：推理错误	
<b>ReAct</b>	
问题：	在《喷发》中与哈维·费尔斯坦、琳达·哈特、迪克·拉特萨、克拉克·索雷尔、玛丽·邦德·戴维斯、劳拉·贝尔·邦迪、马修·莫里森、科里·雷诺兹和玛丽莎·贾瑞特·维诺库尔一起出演的女演员的中间名是什么？
思考1：	我需要搜索哈维·费尔斯坦、琳达·哈特、迪克·拉特萨、克拉克·索雷尔、玛丽·邦德·戴维斯、劳拉·贝尔·邦迪、马修·莫里森、科里·雷诺兹和玛丽莎·贾瑞特·维诺库尔，找到与他们一起在《喷发》中出演的女演员，然后找到她的中间名。
行动1：	搜索[哈维·费尔斯坦]
观察1：	哈维·福布斯·费尔斯坦...
思考2：	哈维·费尔斯坦是一位演员、剧作家和编剧。我需要搜索琳达·哈特...
行动2：	搜索[琳达·哈特]...
<b>CoT</b>	
问题：	丹尼·格林和詹姆斯·沃西谁更年长？
思考：	让我们一步一步思考。丹尼·格林出生于1987年。詹姆斯·沃西出生于1961年。1987年（丹尼·格林）>1961年（詹姆斯·沃西），所以丹尼·格林更年长。
答案：	丹尼·格林

失败：搜索错误

<b>ReAct</b>	
问题：	与女神弗里格相关的女神包括哪些故事？
思考1：	我需要搜索女神弗里格并找到其相关的女神。
行动1：	搜索[女神弗里格]
观察1：	找不到女神弗里格。类似：['弗里格'，'弗里格和弗雷娅共同起源假说'，'弗雷娅（部分与弗里格和其他女神和人物的关系）'，'艾尔（重定向自艾尔（女神））'，'巴尔德' ]。思考2：
	我需要搜索Frigg...

失败：幻觉

<b>Cot</b>	
问题：	第一年由位于布朗克斯的组织出版的科学期刊是哪一年？
思考：	让我们一步一步地思考。第一年由位于布朗克斯的组织出版的科学期刊是1916年。
答案：	1916年（标签：1909年）

失败：标签模糊

<b>ReAct</b>	
问题：	开发独裁者游戏的人是哪国人？
思考1：	我需要搜索独裁者游戏，找到开发者，然后找到他的国籍。
行动1：	搜索[独裁者游戏]
观察1：	独裁者游戏是社会心理学和经济学中常用的实验工具...
行动3：	搜索[Amos Tversky]
观察3：	Amos Nathan Tversky...
思考4：	Amos Tversky是以色列的认知和数学心理学家。所以开发独裁者游戏的人是以色列人。行动4：
	完成[以色列]（标签：以色列-美国）
<b>Cot</b>	
问题：	来自加利福尼亚州斯托克顿的Kurt Fuller出演了由Steve Franks创作的哪部美国侦探喜剧剧集？
思考：	让我们一步一步地思考。来自加利福尼亚州斯托克顿的库尔特·富勒出现在《心理追凶》中。《心理追凶》是一部由史蒂夫·弗兰克斯创作的美国侦探喜剧剧情电视剧。答案：
	《心理追凶》（标签：心理追凶是一部美国侦探喜剧剧情）