

Self-Guide:一种基于自我规划的大语言模型推理增强方法

刘艺彬¹, 刘正皓^{1*}, 闫宇坤², 于是², 王硕², 杨麟儿³, 陈慧敏⁴, 谷峪¹, 于戈¹

¹东北大学计算机科学与工程学院, 辽宁省沈阳市 110169

²清华大学计算机科学与技术系, 北京市 100084

³北京语言大学信息科学学院, 北京市 100083

⁴清华大学新闻与传播学院, 北京市 100084

liuyibin@stumail.neu.edu.cn, {liuzhenghao, guyu, yuge}@mail.neu.edu.cn
{yanyk.thu, lineryang, wangshuo.thu}@gmail.com
{yus21, huimchen}@mails.tsinghua.edu.cn

摘要

尽管大语言模型在自然语言处理任务中取得显著进展, 但其在复杂问题推理等领域还面临着认知负荷问题, 即大语言模型在推理过程需要记忆并处理大量信息。因此, 如何有效地减少语言模型推理过程中的认知负荷, 缓解推理过程中可能出现的认知过载是一个亟待解决的问题。对此本文提出了Self-Guide方法, 用于增强语言模型的推理能力。该方法通过指引大语言模型生成常识知识和推理指导, 让语言模型基于自我规划来增强其推理能力, 并通过与推理链结合的方式对模型的推理过程进行校准。与现有方法不同的是, 本文在不对大语言模型进行微调或使用外部工具的情况下, 显著提升了语言模型的推理性能。实验结果表明, Self-Guide方法在四种常见推理任务上性能显著优于基线方法, 同时相比传统的推理链模型, Self-Guide方法在推理能力较弱的模型上也具有良好的泛化性能。通过结合大语言模型的自我规划和推理能力, Self-Guide方法为提升语言模型的推理能力提供了一种新的有效途径。

关键词: 大语言模型; 自我引导规划; 推理增强; 认知负荷

Self-Guide: Enhancing LLM Reasoning Ability via Self-Plan

Yibin Liu¹, Zhenghao Liu^{1*}, Yukun Yan², Shi Yu², Shuo Wang²,
Liner Yang³, Huimin Chen⁴, Yu Gu¹, Ge Yu¹

¹Department of Computer Science and Technology, Northeastern University, Shenyang, Liaoning

²Department of Computer Science and Technology, Tsinghua University, Beijing

³School of Information Science, Beijing Language and Culture University, Beijing

⁴School of Journalism and Communication, Tsinghua University, Beijing

liuyibin@stumail.neu.edu.cn, {liuzhenghao, guyu, yuge}@mail.neu.edu.cn
{yanyk.thu, lineryang, wangshuo.thu}@gmail.com
{yus21, huimchen}@mails.tsinghua.edu.cn

Abstract

Despite significant advancements of Large language Models (LLMs) in NLP tasks, they still face the cognitive overload problem, especially in domains requiring complex reasoning. Specifically, during reasoning, LLMs need to process and memorize vast amounts of information during the reasoning process. Thus, it is a pressing issue to effectively reduce the cognitive load during the reasoning process of LLM to alleviate potential cognitive overload. We introduce the Self-Guide method to alleviate the problem, which boosts LLMs' reasoning abilities by leveraging self-generated common sense knowledge and reasoning instructions. Experimental results demonstrate that our Self-Guide method outperforms baseline methods significantly on four common reasoning tasks, which effectively enhances the reasoning ability of LLMs.

Keywords: LLMs, Self-Guide, Reasoning Enhancement, Cognitive Load

* 通讯作者

1 引言

在自然语言处理及相关领域，大语言模型例如Chat-GPT、ChatGLM、MiniCPM、文心一言、通义千问(Brown et al., 2020; Du et al., 2021; Hu et al., 2024; Sun et al., 2021; Bai et al., 2023)等已经取得了重大进展。研究表明，随着模型参数规模的快速增长，语言模型展现出了一种涌现现象，使其在许多自然语言任务上的性能都显示了其有效性，同时在很多推理、规划任务中也展现了更为强大的能力(Wei et al., 2022a)。然而，尽管大语言模型在很多任务上表现出色，在解决复杂问题时候，其多步推理能力仍然面临着重要挑战(Valmeekam et al., 2022; Huang and Chang, 2022)。

传统的基于思维链 (Chain-of-Thought, CoT) (Wei et al., 2022b)的推理方法在许多复杂的推理任务，如：数学计算、关系推理等任务中展现了良好的效果。其通过指令指引大语言模型对于给定问题进行一步步思考。在思维链的形成过程中，需要大语言模型在能够对问题的解决思路进行规划，同时又能针对分解后的子问题进行回答，形成一个子问题-答案交融的推理链结果。澳大利亚新南威尔士大学的认知心理学家约翰·斯威勒 (John Sweller) 于1988年提出认知负荷理论 (Cognitive Load Theory) (Sweller, 1988)。他认为人类的认知能力在处理新信息的过程中，其工作记忆的容量是相当有限的，因此我们需要考虑如何高效地呈现信息，来降低工作记忆的负荷量。考虑到语言模型的上下文学习能力和有限的上下文学习窗口(Dong et al., 2023)，在思维链生成过程中常常会面临的认知过载的问题，即产生幻觉或无效推理过程。因此，如何能够显著地降低大语言模型在思维链推理过程中所面临的认知负荷问题，缓解可能出现的认知过载是一个亟待解决的问题。

本文提出了Self-Guide方法——一种基于自我规划的大语言模型推理增强方法，用于增强语言模型的推理能力。该方法通过提示帮助大语言模型生成针对复杂问题的常识知识和推理指导 (Guideline)，让语言模型通过自我规划来预演其推理过程。并通过与推理链结合的方式来降低大语言模型在推理过程中所展现出的认知负荷。值得注意的是，我们在没有对大语言模型进行微调或使用外部工具的情况下，显著提升了语言模型在多种推理等任务上的性能表现。具体来说，本文的贡献主要包括以下三个方面：

- 本文提出了一种新的提示学习方法，通过大模型自身对给定问题的解决过程进行推演进而对给定问题产生相应的问题推理指导 (Guideline)。
- 本文通过生成的指导在一定程度上缓解了传统基于思维链的推理模型在解决复杂问题中所具有的认知负荷问题，提高了语言模型的推理的性能和可靠性。
- 本文通过实验验证了较大规模模型通过规划所产生的问题推理指导能够有效地指导较小规模模型产生更加精准的结果，从而弥补了后者在推理任务上的不足。

我们的实验利用gpt-3.5-turbo-1106来实现Self-Guide方法，并在四种常见推理任务——语言理解、多跳问答、时间推理、关系推理上性能显著优于传统的基于思维链的推理模型 (实现了平均约2%的提升)。同时我们通过实验证明了Self-Guide方法在推理能力较弱的模型 (如LLaMA-7B-Chat和LLaMA-13B-Chat) 上的泛化性，使其在自身推理能力较弱的条件下通过自我规划实现超过10%的提升。本文所有的数据和代码已在GitHub上开源。

2 相关工作

复杂推理为基于语言模型构建大量应用程序提供了机会，从而使语言模型有机会成为下一代“操作系统”(OpenAI et al., 2024)。因此大语言模型应当在复杂推理任务上具有较好的逻辑推理能力，并可以通过与工具、用户和外部环境的所有元素互动来完成复杂的指令。尽管大语言模型在很多任务上表现出色，多步推理能力仍然是其短板之一(Valmeekam et al., 2022; Huang and Chang, 2022)。为了增强大语言模型的推理能力，先前的研究提出了多种方法：Nye等人提出在生成最终答案之前，通过生成“Scratchpads” (即中间步骤) 来微调语言模型，让语言模型具备合成/执行多步推理的能力(Nye et al., 2021)。推理链提示 (Chain-of-Thought, CoT)

©2024 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

项目资助：国家自然科学基金 (No. 62106126, No. 62272093)，辽宁省科技厅联合计划项目 (No. 2023-MSBA-081) 以及中央高校基本科研业务费项目 (No. N2416012)。

以“Let’s think step by step”之类的指令来引导大语言模型先给出推理步骤以及中间的结果，通过推理链增强最终帮助大语言模型生成更加精准的结果。推理链模型在较大规模的语言模型上表现出优越的性能，然而在较小模型上其效果表现差强人意(Wei et al., 2022b)。此外，诱导语言模型生成推理原因需要构建大规模的数据集，或者仅使用少样本提示来牺牲准确性，因此Zelikman等人提出Self-Taught Reasoner(STaR)，利用大语言模型的现有推理能力，在给出答案之前迭代地生成高质量的推理原因(Zelikman et al., 2022)。Wang和Huang等人认为大语言模型有自我改进答案的能力，提出Self-Consistency解码策略，利用CoT生成多个推理路径和答案，最终选择出现次数最多的作为最终答案输出(Wang et al., 2022a; Huang et al., 2022)。

尽管使用高质量数据微调或少样本提示的方法显著增强了大语言模型的复杂推理能力，但在实际业务或问题中，常面临数据量不足或质量较低的情况，且微调和少样本提示的方法泛化性较低。为应对这一问题，Kojima等人证明大语言模型在Zero-Shot-CoT提示下仍然具有优秀的推理能力(Kojima et al., 2022)。而推理链提示在较难的问题上通常会出现推理与答案不一致的现象，为了克服从易到难的泛化挑战，Zhou等人提出Least-to-Most的提示方法，即将复杂问题分解为一系列更简单的子问题，然后按顺序解决它们(Zhou et al., 2022)。考虑到充分利用大语言模型海量的参数知识，Liu等人提出可以通过让语言模型生成知识作为额外输入来增强语言模型的常识推理能力(Liu et al., 2021)。

在大语言模型解决复杂问题时，语言模型的幻觉问题是另一个亟待解决的问题。由于预训练阶段记忆的带有噪音或错误的知识以及模型在特定任务上缺乏先验知识，大语言模型可能会出现生成文本不符合事实及不遵循原文的现象，即大语言模型的幻觉问题(Ji et al., 2023)。为缓解这一问题，研究者提出了多种方法：检索增强被证明可以显著减少幻觉问题。进一步地，Peng等人提出可以通过利用外部知识和语言模型自我反馈缓解大语言模型的幻觉(Peng et al., 2023)。同时，许多学者认为提示大语言模型在给出答案之前给出分步推理的过程能够缓解大语言模型的幻觉问题(Wei et al., 2023; Yao et al., 2022; Wang et al., 2022b; Li et al., 2022)，但这种假设只有在生成的推理符合模型真实推理的情况下才成立(Jacovi and Goldberg, 2020)。

然而语言模型生成的推理与其真实的推理过程并不总是保持一致(Turpin et al., 2024)。对此，Zhao等人提出Verify-and-Edit模型，其是一种知识增强的CoT提示框架，能够让语言模型利用外部知识编辑推理链来提高预测的真实性(Zhao et al., 2023)。Lyu等人提出Faithful CoT，通过将自然语言表示的推理过程转换为符号语言求解从而保证推理具有更高的忠实性(Lyu et al., 2023)。虽然，上述方法取得了较好的效果，但是在在大语言模型进行推理时，传统方法往往需要语言模型同时处理问题规划和答案生成，可能导致认知过载的问题出现。为此，本文提出了Self-Guide模型，通过将语言模型的问题规划和答案生成进行解耦以期望提升大模型对复杂问题的推理效果。

3 Self-Guide: 融合自身指导的大语言模型推理增强技术

在本节中，我们介绍了Self-Guide方法。传统的思维链增强推理方法(Chain-of-Thought, CoT)在具有较好的泛化性的前提下保持了良好的推理能力，但此方法在推理过程中需要同时对问题进行规划和回答，模型在整个推理过程需要处理并记忆大量信息，受到语言模型推理时间和自身能力的限制，语言模型通常会面临幻觉问题或无效推理，如何有效地减少思维链中的认知负荷，缓解大语言模型在推理过程中出现的认知过载问题是一个亟待解决的问题。

如图1所示，受到认知负荷理论(Sweller, 1988)的启发，本文提出了Self-Guide模型来缓解在传统推理链模型当中所面临的认知负荷问题。该方法融合了语言模型的自身指导对语言模型的推理过程进行增强，将语言模型的推理过程解耦为问题规划理解和答案生成。总的来说，我们首先介绍了基于大语言模型自适应问题规划方法(3.1节)，然后我们利用语言模型的自适应性问题规划对其推理过程进行增强(3.2节)。

3.1 基于大语言模型自适应问题规划方法

对于给定的输入 Q ，Self-Guide模型通过问题规划生成以及基于问题规划的大语言模型增强两步方法提升大模型对于复杂问题的推理效果推理。该方法通过对传统推理链模型的问题规划以及答案生成进行解耦，以缓解大模型进行复杂推理过程中所具有的认知负荷问题，最终提升大模型的复杂推理效果。具体的提示模板在表5中展示。

根据认知负荷理论，我们采用Zero-Shot的设定，使大语言模型可以自适应地针对给定问题



Figure 1: Self-Guide模型概述。模型分为思维链模块（Step A），基于大语言模型的自适应问题规划方法（Step B）和基于问题规划的大语言模型推理增强方法（Step C）。其中背景知识（Knowledge）、问题理解（Understanding）和问题规划（Planning）对应内容已高亮显示。

让大语言模型为问题的解决方案形成规划。对于给定的问题 Q ，语言模型生成规划Guideline。具体来说，它提供了基于大语言模型自身经验的指导方针，帮助语言模型减少推理过程中无关的认知负荷，从而将其注意力重新集中在推理过程上。对于给定问题 Q ，我们通过大语言模型生成任务解决规划：

$$\text{Guideline} = \text{Generate}(\text{Instruction_Guide}, Q), \quad (1)$$

其中，Instruction_Guide表示语言模型结合问题 Q 生成自适应问题规划Guideline的过程，具体提示词模板见表5。通过这种方法利用大语言模型自适应地生成问题规划，以降低在复杂推理过程中可能出现的认知负荷，并提高大模型在复杂推理任务上的性能。

如图1中步骤B的输出部分所示，生成的Guideline包含了背景知识（Knowledge）部分作为对问题进行合理推理的基础，根据问题的特定领域提供相关的先验知识。对于大语言模型而言，它们已经通过海量的文本训练获取了广泛的知识，因此可以根据问题内容提供相关的背景知识，以帮助模型更好地理解 and 解决问题。问题理解（Understanding）部分需要模型准确把握问题的隐含信息、逻辑关系等，从而为后续的推理和解决提供基础。其关键在于从问题描述中提取有效信息，并将其与第一部分“背景知识”相结合，形成对问题的全面把握，从而有针对性地展开后续的推理和解决过程。问题规划（Planning）部分是Self-Guide中的关键内容，模型需要根据前面对问题的知识和理解，制定出一套合理的解决方案。通过合理的规划，模型可以避免盲目地进行推理，而是有条不紊地按照既定的步骤逐步推进，从而提高解决问题的准确性。

3.2 基于问题规划的大语言模型推理增强方法

根据Wei和Lanham等人的实验结果，规模较小的模型往往不具备根据链式推理方法提供推理过程的能力，如LLaMA-7B, LLaMA-13B等(Wei et al., 2022b)。，即在推理过程中表现出较高的内在认知负荷，当模型参数增大到大约10B时，大语言模型才会逐渐表现出推理能力(Lanham et al., 2023)。

由于GPT3.5模型有着优秀的基础推理能力，因此我们采取了一种自适应的学习流程（见图1）。首先，对于问题 Q ，我们让语言模型在独立思考 and 探索问题解决方案的过程中运用自身的认知资源，自主解决问题，并生成初步答案推理链：

$$\text{CoT} = \text{Generate}(\text{Instruction_CoT}, Q), \quad (2)$$

其中推理链生成指令 Instruction_CoT 为“Let’s think step by step”。接下来，我们让大语言模型利用公式1所产生的 Guideline 对语言模型的原始思维链进行增强。具体地，我们让大语言模型根据 Guideline 重新审视和反思初始答案 CoT ，修正推理过程中的错误，并根据 Guideline 完善推理过程和最终答案：

$$\text{GuideCoT} = \text{Template}(\text{CoT}, \text{Guideline}, Q), \quad (3)$$

其中模板如图1中步骤C的输入部分所示。最终，语言模型通过结合独立思考和自我规划的学习过程，生成信息密度更高且推理过程更缜密的答案。我们通过Self-Guide的自我规划在语言模型推理过程中减少了其在推理过程中冗余的推理步骤，从而有效降低了推理过程中的无关的认知负荷。这种方法旨在最大程度地发挥大语言模型的自主性和推理能力，并通过结合自我规划和自我反馈来提高针对于复杂推理问题的学习效果和质量。

4 实验设计

在本节中，我们描述了实验中使用的数据集、基线、评估指标以及实现细节。

4.1 数据集：

在实验中，我们采用了知识理解、多跳问答、时间推理以及关系推理四个任务来评估Self-Guide模型在不同任务上的性能。由于ChatGPT的推理成本，对于数据集MMLU和CLUTRR，我们在实验中随机抽取了500个样本进行测试，这与之前的工作保持一致 (Yoran et al., 2023; Zhang et al., 2024; Xu et al., 2024)。由于StrategyQA和Date数据集的样本数量均小于500，因此我们在实验过程中在这两个数据集的全集上进行了实验。

知识理解 知识理解任务旨在衡量模型在预训练期间获得的知识。我们应用MMLU（大规模多任务语言理解）数据集(Hendrycks et al., 2021)，旨在通过对模型进行零样本或者少样本场景下的评估，来衡量预训练期间获得的知识。这使得该基准更具挑战性，也更类似于我们评估人类的方式。该基准涵盖57个学科，涉及科学、技术、工程和数学、人文科学、社会科学等。其难度从初级水平到高级专业水平不等，既测试世界知识，也测试解决问题的能力。这些科目的细化和广泛性使该基准成为识别模型知识薄弱领域的理想工具。

多跳问答 我们考虑了数据集StrategyQA(Geva et al., 2021)。这是一个开放域问题数据集，需要隐含的多步骤策略来回答，而回答问题所需的推理步骤隐含在问题中，应使用策略来推断。比如对于问题“亚里士多德使用笔记本电脑吗？”，回答时隐含着需要分别回答以下三个子问题“1.亚里士多德生活于何时？”“2.笔记本电脑是什么时候发明的？”“3.子问题2答案对应时间在子问题1答案对应时间的前面吗？”。

时间推理 我们采用来自BIG-bench基准测试(bench authors, 2023)的Date Understanding数据集。这是一个日期推理的数据集，旨在通过询问日期来测试语言模型理解日期的能力。它要求模型通过对相对时间段的计算，根据上下文推断日期。给定上下文，模型应该回答以下问题“*What is the date [day] in MM/DD/YYYY?*”，其中[day]可以是today、tomorrow等。虽然这项任务对人类来说很容易，但对于许多语言模型却很困难，因此用来评估各种优化模型推理的方法较有价值。

关系推理 我们使用CLUTRR基准(Sinha et al., 2019)，这是一个人物关系推理的数据集，要求自然语言理解(NLU)系统推断短篇小说中人物之间的亲属关系。要成功完成这项任务，既要提取实体之间的关系，又要推断出支配这些关系的逻辑规则。CLUTRR允许通过评估逻辑规则的保留组合来精确测量模型的系统泛化能力，并允许通过添加经过策划的噪声事实来评估模型的鲁棒性。

4.2 基线

基于Zero-Shot的实验设定，我们将Self-Guide模型与不同的提示工程方法进行了对比。我们的基线主要由标准零样本提示(Zero-Shot)和链式推理方法(Chain-of-Thought, CoT)组成：

参数	checkpoint	temperature	top_p	max_seq_len	max_gen_len	max_batch_size
设置值	meta_llama	0.2	0.9	2048	128	8

Table 1: LLaMA大语言模型推理过程中的参数设置。

标准零样本提示学习 (Zero-Shot)：在不提供示例的情况下，对于给定的问题，通过给定的指令让大语言模型直接输出预测答案。

链式推理方法 (Chain of Thought, CoT)：在标准提示方法(Brown et al., 2020)的基础上，研究人员要求大语言模型在输出最终答案之前，显式输出逐步的中间推理步骤来增强大模型的常识和推理等能力。而这一系列逐步的中间推理步骤的内容就被称为链式推理方法(Wei et al., 2022b)。链式推理方法区别于传统的提示方法，实现了从输入到思维链再到输出的映射，即<Input-Reasoning Chain-Output>，而不是从输入直接到输出的映射<Input-Output>的方式。另外，链式推理方法分为Zero-Shot-CoT与Few-Shot-CoT。本实验采用Zero-Shot-CoT作为基线，不添加示例（即一些针对如何进行中间推理的示范内容），而仅在指令中添加“Let’s think step by step”(Wei et al., 2022b)，就可以使大模型表现出较强的零样本推理能力。

4.3 评估指标

准确率 (Accuracy)：我们使用准确率 (Accuracy, Acc) 来评估不同任务中模型的性能，这与先前的工作(Xu et al., 2024)类似。我们首先将大语言模型的输出和标准答案都转换为小写，然后执行字符串匹配 (StringEM) 方法来计算每个模型预测结果和标准答案之间的准确率。具体匹配规则见附录B。

困惑度 (Perplexity)：对于Self-Guide方法在不同规模语言模型上的泛化能力实验，我们使用困惑度 (Perplexity, PPL) 来评估语言模型在推理过程中的生成文本质量。对于词元化的序列 $X = (x_0, x_1, \dots, x_t)$ ，序列 X 的困惑度为：

$$\text{PPL}(X) = \exp \left\{ -\frac{1}{t} \sum_i^t \log p_{\theta}(x_i | x_{<i}) \right\}, \quad (4)$$

其中 $\log p_{\theta}(x_i | x_{<i})$ 是根据我们的模型，第 i 个标记在前面的标记 $x_{<i}$ 的条件下的对数似然。通常而言，PPL 越小，表示文本的上下连贯度越高，文本越通顺。

4.4 实验细节

对于Self-Guide方法，即表2，表3以及表6，我们通过OpenAI API访问ChatGPT模型来评估其性能。具体地，我们使用gpt-3.5-turbo-1106作为推理的基础模型，并将temperature设置为0.2进行生成。对于Self-Guide方法在LLaMA模型上的泛化性实验，即表4和图2，我们首先使用gpt-3.5-turbo-1106作为自我规划的生成模型，然后分别在开源大语言模型LLaMA2-7B-Chat和LLaMA2-13B-Chat上评估生成的自我规划结果 (Guideline) 的性能，LLaMA模型的参数设定见表1。

5 实验结果

在本节中，我们展示了Self-Guide的整体性能。随后，我们通过消融实验证明了Self-Guide方法的有效性。接着，我们深入分析了该方法在较小规模模型上的泛化能力，Self-Guide方法的作用机制以及其特征。最后，我们进行了案例分析。

5.1 Self-Guide模型整体性能效果

我们在表2中呈现了Self-Guide方法在不同领域数据集上的综合表现。在Zero-Shot的实验设置下，我们将Self-Guide方法与不同类型的基线进行了比较，包括标准Zero-Shot提示方法 (Zero-Shot) 和Zero-Shot设定下的链式思维提示方法 (Chain of Thought, CoT)。

为了深入探究大语言模型如何通过自我规划来增强推理能力，我们采用了三种方法对大语言模型的推理进行增强：让模型直接通过自我规划进行推理 (Self-Guide (No Instruct))，在

方法	多任务语言理解 MMLU	多跳问答 StrategyQA	时间推理 Date	关系推理 CLUTRR
Zero-Shot	56.0	68.2	54.0	36.6
CoT	66.2	72.7	64.9	55.6
Self-Guide (No Instruct)	65.8	69.8	61.6	47.4
Self-Guide	67.4	70.2	61.3	48.8
CoT w. Self-Guide	68.8	74.3	66.9	56.4

Table 2: 各模型实验结果。其中Self-Guide为我们的方法。在该实验中，所有方法均使用gpt-3.5-turbo-1106作为推理的基础模型。我们在附录展示了不同模型在MMLU数据集中不同领域上的实验结果（表6）。

方法	实验设置	多任务语言理解 MMLU	多跳问答 StrategyQA	时间推理 Date	关系推理 CLUTRR
Zero-Shot	Vanilla	56.0	68.2	54.0	36.6
	w. Reflect	56.6	47.1	51.0	24.0
	w. Debate	60.8	67.8	55.7	38.0
	w. Self-Guide	67.3	70.2	61.3	48.8
CoT	Vanilla	66.2	72.7	64.9	55.6
	w. Reflect	36.2	51.0	60.2	25.4
	w. Debate	63.2	65.3	61.3	36.4
	w. Self-Guide	68.8	74.3	66.9	56.4

Table 3: 消融实验结果。在该实验中，所有方法均使用gpt-3.5-turbo-1106作为推理的基础模型。其中Self-Reflect为在大语言模型生成答案后自我反思更新答案，Self-Debate为让大语言模型检查其他模型的答案后更新答案。

标准提示方法的基础上利用自我规划来增强推理过程并更新答案（Self-Guide），以及在链式推理方法下利用自我规划来增强推理过程并更新答案（CoT w. Self-Guide）。

如表2中实验结果所示，三种采用Self-Guide方法进行推理增强的方法相比基线模型（Zero-Shot）在所有数据集上都表现出较为显著的提升。特别值得注意的是，我们的模型CoT w. Self-Guide在所有数据集上都表现出最佳的性能。相较于Zero-Shot模型，我们的模型在所有数据集上的平均准确率都有显著提升（+6.1至+19.8），而相较于基于推理链的推理增强模型，我们的模型在所有数据集上的平均准确率也有一定程度的提升（+0.8至+2.6），证明了我们模型在解决复杂推理问题时的有效性，同时说明Self-Guide方法能够通过自然语言表达语言模型的所形成的自动规划，来缓解传统推理链中的认知负荷问题。

鉴于ChatGPT-3.5模型本身具备较强的推理能力，因此，仅仅使用Self-Guide方法进行自我规划（Self-Guide和Self-Guide (No Instruct)）的效果相较于推理链模型，其并没有表现出显著提升。换言之，虽然Self-Guide模型能够通过生成Guideline的方式缓解大模型的认知负荷问题，但是诸如推理链模型当中针对于每一步规划生成相对应的中间结果亦是至关重要的。

5.2 消融实验

为了进一步验证Self-Guide方法的有效性，即Self-Guide所带来的性能提升是否是由于模型通过自我规划提高了推理能力，而不是由于模型的二次推理调用所产生更好地结果，我们进行了消融实验。

如表3中所示，我们将我们的模型与不同的推理增强方法进行比较。我们以标准提示方法（Zero-Shot）和链式思维推理方法（CoT）为基础答案（Vanilla），并使用不同的方法对基础答案进行增强，包括自我反思（Reflect）和对其他语言模型的推理过程进行检查（Debate）。

如实验结果所示，相比基线方法，让大语言模型对自身推理进行自我反思（Zero-Shot w. Reflect）时，相比较Zero-Shot模型，模型的推理能力略有下降（-21.1到+0.6）。而相比较推理链模型（CoT）时，自我反思（CoT w. Reflect）模型的推理能力下降的更为明显（-30.2到-

方法	实验设置	多任务语言理解 MMLU	多跳问答 StrategyQA	时间推理 Date	关系推理 CLUTRR
LLaMA2-7B-Chat					
LLaMA-Guide	Zero-Shot	44.8	59.6	28.4	27.2
	CoT	40.8	60.4	32.0	34.6
GPT-Guide	CoT	52.0	64.7	52.9	37.6
	Guideline	48.8	66.1	44.3	45.0
	CoT w. Guideline	53.0	65.5	54.9	45.6
LLaMA2-13B-Chat					
LLaMA-Guide	Zero-Shot	46.8	62.2	39.0	34.4
	CoT	45.6	51.2	45.7	39.4
GPT-Guide	CoT	54.0	62.7	59.9	43.0
	Guideline	54.2	65.9	57.4	46.0
	CoT w. Guideline	56.8	64.9	61.6	47.2

Table 4: Self-Guide在不同规模语言模型上的泛化能力实验结果。其中LLaMA-Guide方法由LLaMA模型独立完成推理问题，GPT-Guide方法由gpt-3.5-turbo-1106生成CoT和Guideline，在去除最终答案后指导LLaMA模型解决复杂的推理问题。

4.7)。而让语言模型对其他语言模型的推理过程进行检查并给出自身答案 (Debate) 时，相比较Zero-Shot模型，语言模型的答案准确率有所提升 (-0.4到+4.8)，但相比较CoT模型，CoT w. Debate模型的性能反而下降 (-19.2到-3.0)。

实验结果表明，相比Reflect和Debate的推理增强方法，Self-Guide方法能够通过缓解大语言模型的认知负荷问题来提升大语言模型的推理能力，而不是仅仅通过自我反思 (Reflect) 以及自我检查 (Debate) 以期实现更加精准的推理过程。相比较Zero-Shot模型，Self-Guide方法使得语言模型在所有数据集上的推理能力提升+2.0到+12.2，甚至在部分数据集中达到了和推理链模型相匹敌的效果。这证明了类似于推理链等诸多模型，其自我规划的结果能够在处理逻辑推理问题时起到至关重要的作用。

此外，Self-Guide也是一种有效方法能够通过推理链结合的方式来进一步提升模型效果，Self-Guide方法使得语言模型在所有数据集上的推理能力提升+0.8到+2.6。这表明，使用Self-Guide方法所取得的提升并不是由于模型的二次推理或调用所产生的，而是由于模型通过自我规划来对思维链进行校准，进而提升了大语言模型针对复杂任务的推理能力。

5.3 Self-Guide在不同规模语言模型上的泛化能力

该实验展示了Self-Guide在不同模型上的泛化能力，如表4所示，我们利用Self-Guide方法增强推理能力较弱的大语言模型，如LLaMA2-7B-Chat和LLaMA2-13B-Chat。使这些模型展现出更强的推理能力。由于较小规模的LLaMA模型基础推理能力较弱，无法很好地生成推理链并解决复杂的推理任务。在本实验中，我们尝试将ChatGPT-3.5视为教师模型，通过Self-Guide方法中生成的CoT和Guideline来指导小语言模型解决复杂的推理问题 (GPT-Guide)。具体地，在GPT-Guide的方法中，我们首先利用ChatGPT-3.5生成推理链和Guideline，在去除ChatGPT-3.5生成的推理链和Guideline中的最终答案后，分别采用ChatGPT-3.5生成的推理链引导LLaMA模型推理 (CoT)，采用ChatGPT-3.5生成的Guideline引导语言模型进行推理 (Guideline) 和同时使用ChatGPT-3.5生成的CoT和Guideline来引导LLaMA模型进行推理 (CoT w. Guideline)。

如表4所示，CoT w. Guideline方法在所有任务和数据集上都表现出了显著的性能，特别是在多跳问答、时间推理和关系推理任务上。具体来说，与由LLaMA模型独立完成推理问题的方法比较 (LLaMA-Guide)，在LLaMA2-7B-Chat模型上，我们的方法相比于标准Zero-Shot提示方法，性能提升范围为+5.9到+26.5；相比于CoT提示方法，性能提升范围为+5.1到+22.9。而在LLaMA2-13B-Chat模型上，相比于标准Zero-Shot提示方法，性能提升范围为+2.7到+22.6；相比于CoT提示方法，性能提升范围为+7.8到+15.9。这些结果表明，通过使用较大规模模型基于Self-Guide方法生成的Guideline来引导较小规模模型，较大规模模型所获得的丰富语义表示和推理能力可以有效地传递给原本推理能力较弱的较小规模模型，从而弥补了后者在推理任务上的不足。

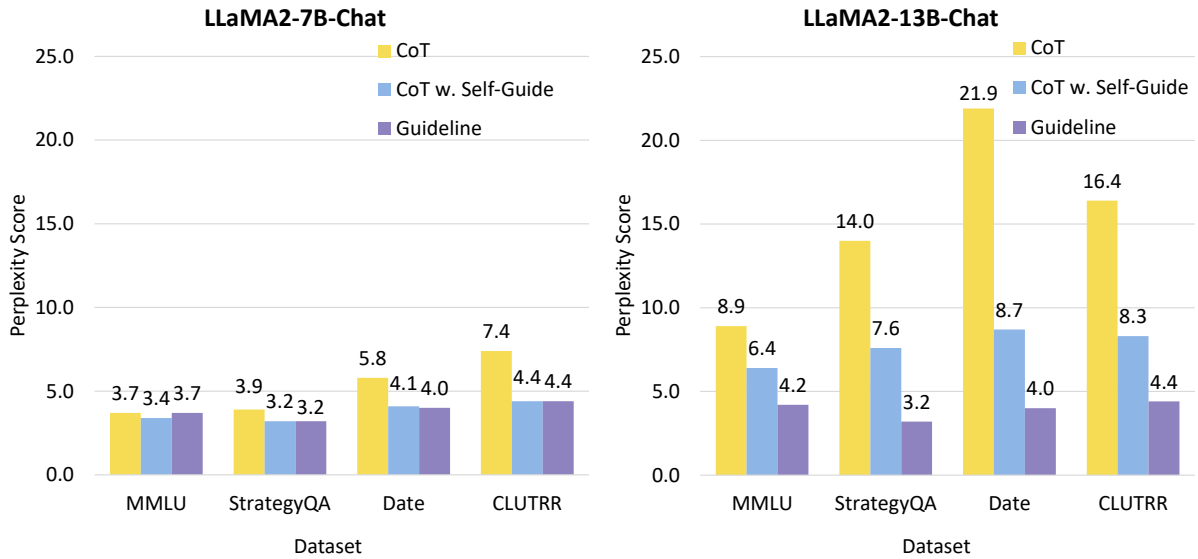


Figure 2: LLaMA模型在使用不同提示方法时困惑度（Perplexity）的差异。

除此之外，在使用ChatGPT-3.5模型的推理过程引导LLaMA完成推理问题的方法中（GPT-Guide），基于CoT w. Guideline引导的学习方法相比于基于CoT引导的学习方法，其推理性能都有一定程度的提升。在LLaMA2-7B-Chat模型上，CoT w. Guideline引导的模型相比于CoT引导的模型，性能提升范围为+0.8到+8.0。而在LLaMA2-13B-Chat模型上，CoT w. Guideline引导的模型相比于CoT引导的模型，性能提升范围为+1.7到+4.2。同时，随着模型参数的不断减小，CoT和CoT w. Guideline方法之间的性能差距也在不断扩大，从ChatGPT-3.5上的CoT和CoT w. Self-Guide之间的0.8%至2.6%，到LLaMA2-13B-Chat模型上CoT和CoT w. Guideline方法之间的1.7%至4.2%，到LLaMA2-7B-Chat模型上的CoT和CoT w. Guideline之间的0.8%至8.0%。这些结果表明由于模型的参数规模不同，其具有的认知容量也不同，推理链通过一边规划一边进行解答的过程所带来的认知负荷问题在小规模的语言模型中得到了放大。推理链中的问题解决思路难以适配到参数较小的模型当中。而Self-Guide模型所产生的Guideline更类似于一种解题思路，其并未规定题目的具体解题方法，而是尝试引导模型考虑到解题当中的必要知识、步骤等，能够更好地适配较小规模语言模型的能力，以此帮助模型产生更好的生成结果。

为了更好地量化大语言模型在推理过程中的内在认知负荷，如图2所示，我们采用困惑度（Perplexity）来衡量语言模型在推理过程中生成答案的置信程度。如实验结果所示，相比起其他的提示方法，使用Guideline引导的模型在推理过程中困惑度显著降低，且使用LLaMA2-7B-Chat和LLaMA2-13B-Chat作为语言模型在所有数据集上PPL值稳定始终在4.0左右（-0.8到+0.4）。该现象证明了Self-Guide方法所产生的Guideline能够更加容易地帮助不同规模的语言模型更加精确地产生正确答案，减小模型的不确定性。除此之外，Self-Guide方法能够为大模型指导小模型学习提供一种可能。不同于通过推理链进行蒸馏的框架(Hsieh et al., 2023)，通过生成Guideline进行蒸馏能够更加适配小模型的学习过程，帮助匹配较小规模语言模型的能力，进而能够更好地帮助其理解大模型问题解决思路、必要知识与规划，而不是一味地模仿更大规模模型的较难的解题思路。

5.4 案例分析

最后，我们在MMLU数据集中随机抽取了一个案例进行分析，以显示Self-Guide方法的有效性。其他数据集的案例分分析详见附录D。

如表5所示，经过Self-Guide方法增强后的模型的推理过程相比较推理链模型，其更加简洁且信息密度更高，同时推理步骤之间的逻辑关系也更加缜密。具体来说，在我们选择的案例中，Self-Guide首先定位问题领域（医学临床诊断），结合问题所述症状提供了与超声检查、尿液分析等相关的先验知识；之后提取问题内容（症状）的关键信息（血尿、右侧肾积水等），

<p>问题: 一名56岁的男子来到急诊科就诊, 原因是他有4天的右侧腹部绞痛病史, 疼痛还放射到腹股沟并伴有血尿。肾脏超声波检查显示右侧肾积水和输尿管扩张。尿液分析最有可能发现以下哪项?</p> <p>选项: A: 葡萄糖 B: 白细胞管型 C: 椭圆形脂肪体 D: 尿酸结晶</p>
<p>原始推理链 (CoT) 输入提示</p> <p>这里有一个问题: {问题} 为了解决这个问题, 请逐步分析和推理, 并给出你解决问题时的思考和推理过程。</p>
<p>原始推理链 (CoT) 输出回答</p> <p>1. 首先, 让我们分析一下这名56岁男子的症状.....根据这些症状, 患者很可能患有肾结石或某种形式的尿路梗阻, 导致肾积水和输尿管扩张。</p> <p>2. 现在, 让我们考虑一下尿液分析的选项: A.....B.....C.....D.....</p> <p>3. 根据症状和检查结果, 最有可能的答案是B.白细胞管型, 因为.....</p>
<p>步骤1: 基于Self-Guide的自我规划</p> <p>获取Self-Guide的输入提示</p> <p>你是一位知识渊博、耐心细致的教授, 其职责是指导学生正确解决问题。以下是一个问题: {问题} 注意: 因为你的责任是引导学生回答问题, 所以你的分析应该循序渐进, 请注意你的角色是引导他们一步一步地地解决问题, 所以请不要给他们最终的结果。</p>
<p>得到的Guideline</p> <p>1. 好吧, 让我们一步一步解决问题。首先, 我们来看看病人的症状...超声波检查显示右侧肾积水和输尿管扩张进一步证实了这种可能性。</p> <p>2. 现在, 让我们来想想在这种情况下, 尿液分析最有可能发现什么。尿液分析是一种常见的检查方法, 用于评估尿液中的各种成分, 包括血液、蛋白质、葡萄糖以及不同类型的细胞和结晶。根据症状和超声波检查结果..... 血尿的出现表明尿液中含有血液, 这与腹部绞痛以及肾结石或梗阻的可能性是一致的。</p> <p>3. 考虑到这一点, 让我们对所提供的选项进行评估: A.....B.....C.....D..... 而尿酸结晶是肾结石患者尿液中常见的结晶, 尤其是那些由尿酸组成的结晶。考虑到腹部绞痛的症状以及肾结石或梗阻的可能性, 尿酸结晶是本病例尿液分析中最有可能发现的。</p> <p>4. 根据以上分析, 您认为哪个选项最有可能是正确的?</p>
<p>步骤2: 最终答案生成</p> <p>获取最终答案的输入提示</p> <p>{CoT输出}</p> <p>这是一个对这个问题的可靠的分析: {Self-Guide输出}。</p> <p>请根据这个分析验证您的推理过程是否存在错误, 然后完善你的推理过程和答案。</p>
<p>得到的最终答案</p> <p>推理过程: 首先, 考虑到症状和超声波检查结果, 患者很可能患有肾结石或某种形式的尿路梗阻, 导致肾积水和输尿管扩张。右侧腹部绞痛、向腹股沟放射的疼痛和血尿也支持这一推断。接下来, 我们对尿液分析的选项进行评估: A. 尿液中的葡萄糖通常与糖尿病有关, 在这种情况下是不可能出现的。B. 白细胞管型-虽然表明有炎症或感染, 但在本病例中可能不是最有可能的发现。C. 椭圆形脂肪体-与患者的症状和超声检查结果没有直接关系。D. 尿酸盐结晶-考虑到患者有腹部绞痛的症状, 以及肾结石或梗阻的可能性, 尿酸盐结晶是本病例最有可能的尿液检查结果。</p> <p>最终答案: 根据综合推理过程, 本病例尿检最有可能发现的是尿酸结晶 (选项D)。这是因为症状和超声波检查结果强烈提示存在肾结石或梗阻, 而肾结石或梗阻与尿液中存在尿酸结晶有关。</p>

Table 5: 案例分析。我们随机选择了一个MMLU数据集的例子来展示Self-Guide的效果, 并突出显示了答案和相关内容。这是一个经过翻译的中文版本。原始内容在表7中展示。

将这些和先验知识结合, 形成对整个临床病情的深入理解; 最后根据深入理解, 提供了从各个选项开始的分析方案。在通过Self-Guide得到Guideline之后, 模型能够首先清晰地给出针对整个问题的推理过程, 其结合了Guideline所提供的理解内容并按照Self-Guide从选项逐步剖析, 并结合初始的推理链进行反思与分析, 最后直接给出了正确答案。这表明Self-Guide充分利用了模型自身的强大生成能力, 使其能够根据具体问题灵活地调整推理策略, 从而更好地应对各种复杂推理场景。Self-Guide能够帮助大语言模型对思维链结果进行反思, 引导大语言模型对问题进行细致的解析, 在处理复杂推理问题时能够通过形成知识密集的Guideline的相关内容, 以在处理复杂推理任务时降低大语言模型所面临的认知负荷问题。

6 结论

本文提出了一种结合了大语言模型生成的常识知识和推理指导, 让语言模型基于自我规划来增强其推理能力的方法Self-Guide。在不对语言模型进行微调的同时显著增强了语言模型的推理能力。我们证明了我们的方法在四种常见推理任务——语言理解、多跳问答、时间推理、关系推理上性能显著优于基线。相关实验结果分析进一步说明了Self-Guide模型可以通过Guideline将较大规模语言模型的语言表示和推理能力有效地传递给原来推理能力较弱的较小规模模型, 同时显著地降低较小规模模型生成答案时候的不确信度。

参考文献

- Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. 2023. Qwen technical report.
- BIG bench authors. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*.
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.
- Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, Lei Li, and Zhifang Sui. 2023. A survey on in-context learning.
- Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2021. Glm: General language model pretraining with autoregressive blank infilling. *arXiv preprint arXiv:2103.10360*.
- Mor Geva, Daniel Khashabi, Elad Segal, Tushar Khot, Dan Roth, and Jonathan Berant. 2021. Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies.
- Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring massive multitask language understanding.
- Cheng-Yu Hsieh, Chun-Liang Li, CHIH-KUAN YEH, Hootan Nakhost, Yasuhisa Fujii, Alex Jason Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.
- Shengding Hu, Yuge Tu, Xu Han, Chaoqun He, Ganqu Cui, Xiang Long, Zhi Zheng, Yewei Fang, Yuxiang Huang, Weilin Zhao, Xinrong Zhang, Zheng Leng Thai, Kaihuo Zhang, Chongyi Wang, Yuan Yao, Chenyang Zhao, Jie Zhou, Jie Cai, Zhongwu Zhai, Ning Ding, Chao Jia, Guoyang Zeng, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2024. Minicpm: Unveiling the potential of small language models with scalable training strategies.
- Jie Huang and Kevin Chen-Chuan Chang. 2022. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*.
- Jiaxin Huang, Shixiang Shane Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2022. Large language models can self-improve. *arXiv preprint arXiv:2210.11610*.
- Alon Jacovi and Yoav Goldberg. 2020. Towards faithfully interpretable nlp systems: How should we define and evaluate faithfulness? *arXiv preprint arXiv:2004.03685*.
- Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Comput. Surv.*, 55(12), mar.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.
- Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, Danny Hernandez, Dustin Li, Esin Durmus, Evan Hubinger, Jackson Kernion, et al. 2023. Measuring faithfulness in chain-of-thought reasoning. *arXiv preprint arXiv:2307.13702*.
- Shiyang Li, Jianshu Chen, Yelong Shen, Zhiyu Chen, Xinlu Zhang, Zekun Li, Hong Wang, Jing Qian, Baolin Peng, Yi Mao, Wenhui Chen, and Xifeng Yan. 2022. Explanations from large language models make small reasoners better.

- Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2021. Generated knowledge prompting for commonsense reasoning. [arXiv preprint arXiv:2110.08387](#).
- Qing Lyu, Shreya Havaldar, Adam Stein, Li Zhang, Delip Rao, Eric Wong, Marianna Apidianaki, and Chris Callison-Burch. 2023. Faithful chain-of-thought reasoning. [arXiv preprint arXiv:2301.13379](#).
- Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, et al. 2021. Show your work: Scratchpads for intermediate computation with language models. [arXiv preprint arXiv:2112.00114](#).
- OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kafan, Lukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Lukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rameez Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O’Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayarvija, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. Gpt-4 technical report.
- Baolin Peng, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang, Lars Liden, Zhou Yu, Weizhu Chen, and Jianfeng Gao. 2023. Check your facts and try again: Improving large language models with external knowledge and automated feedback.
- Koustuv Sinha, Shagun Sodhani, Jin Dong, Joelle Pineau, and William L. Hamilton. 2019. CLUTRR: A diagnostic benchmark for inductive reasoning from text. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, [Proceedings of the 2019 Conference on Empirical Methods in Natural](#)

- Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4506–4515, Hong Kong, China, November. Association for Computational Linguistics.
- Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiaxiang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. 2021. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. arXiv preprint arXiv:2107.02137.
- John Sweller. 1988. Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2):257–285.
- Miles Turpin, Julian Michael, Ethan Perez, and Samuel Bowman. 2024. Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting. Advances in Neural Information Processing Systems, 36.
- Karthik Valmeekam, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. 2022. Large language models still can't plan (a benchmark for llms on planning and reasoning about change). arXiv preprint arXiv:2206.10498.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022a. Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, and Denny Zhou. 2022b. Rationale-augmented ensembles in language models.
- Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022a. Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022b. Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35:24824–24837.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models.
- Zhipeng Xu, Zhenghao Liu, Yibin Liu, Chenyan Xiong, Yukun Yan, Shuo Wang, Shi Yu, Zhiyuan Liu, and Ge Yu. 2024. Activerag: Revealing the treasures of knowledge via active learning.
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629.
- Ori Yoran, Tomer Wolfson, Ben Bogin, Uri Katz, Daniel Deutch, and Jonathan Berant. 2023. Answering questions by meta-reasoning over multiple chains of thought.
- Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. 2022. Star: Bootstrapping reasoning with reasoning. Advances in Neural Information Processing Systems, 35:15476–15488.
- Jintian Zhang, Xin Xu, Ningyu Zhang, Ruibo Liu, Bryan Hooi, and Shumin Deng. 2024. Exploring collaboration mechanisms for llm agents: A social psychology view.
- Ruochen Zhao, Xingxuan Li, Shafiq Joty, Chengwei Qin, and Lidong Bing. 2023. Verify-and-edit: A knowledge-enhanced chain-of-thought framework. arXiv preprint arXiv:2305.03268.
- Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc Le, et al. 2022. Least-to-most prompting enables complex reasoning in large language models. arXiv preprint arXiv:2205.10625.

方法	多任务语言理解(MMLU)				
	全部	人文伦理	社会科学	理工科	其他
Zero-Shot	56.0	47.2	67.8	46.3	68.3
CoT	66.2	59.0	74.8	63.4	71.3
Self-Guide	65.8	57.4	73.9	55.4	82.7
Zero-Shot w. Self-Guide	67.4	62.6	72.2	61.2	77.6
CoT w. Self-Guide	68.8	58.4	76.5	65.9	80.2

Table 6: 不同模型在MMLU数据集中不同领域上的实验结果。其中Self-Guide为我们的方法。在该实验中，所有方法均使用gpt-3.5-turbo-1106作为推理的基础模型。

附录

A 模型在MMLU数据集中不同领域上的性能

MMLU基准涵盖57个学科，涉及科学、技术、工程和数学、人文科学、社会科学等。其难度从初级水平到高级专业水平不等，既测试世界知识，也测试解决问题的能力。这些科目的细化和广泛性使该基准成为识别模型知识薄弱领域的理想工具。我们在所有学科（All），人文伦理（Hum.），社会科学（Soc. Sci.），理工科（STEM），和其他问题（Other）等多个领域进行了实验（表6）。

B 评估指标细节

在执行字符串匹配（StringEM）方法来计算每个模型预测结果和标准答案之间的准确率的过程中，具体地，我们首先将大语言模型的输出和标准答案都转换为小写。对于MMLU和StrategyQA数据集，我们提取大预言模型生成结果中的第一个单词作为模型预测结果与标准答案进行匹配。对于Date数据集，我们提取出所有由数字、斜杠、数字、斜杠、数字组成的模式串，在转换为标准日期格式（MM/DD/YYYY）后与标准答案进行匹配。对于CLUTRR数据集，我们直接执行字符串匹配方法来计算每个模型预测结果和标准答案之间的准确率。

C 英文原版案例分析

我们在MMLU数据集中各随机选择了一个例子来展示Self-Guide的效果，并突出显示了答案和相关内容。我们在表5中展示了案例分析的原始输入和输出。为了方便阅读我们在正文中展示了简略版的中文案例分析。

D 其他数据集案例分析

我们在CLUTRR和Date数据集中各随机选择了一个例子来展示Self-Guide的效果，并突出显示了答案和相关内容。由于StrategyQA数据集需提交至Leaderboard进行评测，我们无法知道单个例子的对错，因此不对StrategyQA进行案例分析。

<p>问题: A 56-year-old man comes to the emergency department because of a 4-day history of colicky right flank pain that radiates to the groin and hematuria. Ultrasound examination of the kidneys shows right-sided hydronephrosis and a dilated ureter. Which of the following is most likely to be found on urinalysis?</p> <p>选项: A: Glucose B: Leukocyte casts C: Oval fat bodies D: Uric acid crystals</p>
<p>原始推理链 (CoT) 输入提示 Here is a question:{Question} In order to solve this problem, analyze and reason step by step, and give your thinking and reasoning process while solving the problem.</p>
<p>原始推理链 (CoT) 输出回答 1. First, let's analyze the symptoms ... Based on these symptoms, it is likely that the patient is experiencing a kidney stone or obstruction in the urinary tract, leading to hydronephrosis and dilated ureter. 2. Now, let's consider the options for urinalysis: A. Glucose in the urine is typically associated with diabetes and would not be expected in this case. B. Leukocyte casts in the urine indicate inflammation or infection in the kidneys or urinary tract... C. Oval fat bodies are associated with nephrotic syndrome and would not be expected in this case. D. Uric acid crystals are associated with gout, but they are not associated with the symptoms ... 3. Based on the symptoms and findings, the most likely answer would be B.leukocyte casts, Because...</p>
<p>步骤1: 基于Self-Guide的自我规划 获取Self-Guide的输入提示 You are a knowledgeable and patient professor whose role is to guide students in solving problems correctly. Here is a question: {Question} Note: Since your responsibility is to guide students in answering the question, your analysis should think step by step, Please note that your role is to guide them step by step through the problem, so please don't give them the final result.</p>
<p>得到的Guideline 1. Alright, let's break this down step by step. First, let's consider the symptoms presented by the patient. ... The ultrasound examination showing right-sided hydronephrosis and a dilated ureter further supports this possibility. 2. Now, let's think about what would be most likely to be found on urinalysis in this case. Urinalysis is a common test that evaluates various components of urine, including the presence of blood, protein, glucose, and different types of cells and crystals. Given the symptoms and the ultrasound findings, ... The presence of hematuria suggests that there is blood in the urine, which is consistent with the colicky flank pain and the possibility of a kidney stone or obstruction. 3. Considering this, let's evaluate the options provided:... Therefore, we can eliminate option D. uric acid crystals are a common finding in the urine of patients with kidney stones, especially those composed of uric acid. Given the symptoms of colicky flank pain and the possibility of a kidney stone or obstruction, uric acid crystals are the most likely finding on urinalysis in this case. 4. Based on this analysis, which option do you think is most likely correct?</p>
<p>步骤2: 最终答案生成 获取最终答案的输入提示 {CoT-log} This is another analysis of this question:{Self-Guide}. Please combine this reasoning process with your reasoning processes, then give me a better reasoning process. Generation Format: 1. Inference process: 2. Answer:</p>
<p>得到的最终答案 Inference process: 1. First, considering the symptoms and ultrasound findings, it is likely that the patient is experiencing a kidney stone or some form of obstruction in the urinary tract, leading to hydronephrosis and dilated ureter. The presence of colicky right flank pain, radiating to the groin, and hematuria supports this inference. 2. Next, let's evaluate the options provided for urinalysis: A. Glucose in the urine is typically associated with diabetes and would not be expected in this case. B. Leukocyte casts - While indicative of inflammation or infection, it may not be the most likely finding in this case. C. Oval fat bodies - Not directly related to the symptoms and ultrasound findings presented by the patient. D. Uric acid crystals. Given the symptoms of colicky flank pain and the possibility of a kidney stone or obstruction, uric acid crystals are the most likely finding on urinalysis in this case. Answer: Based on the combined reasoning processes, the most likely finding on urinalysis in this case is uric acid crystals (option D). This is because the symptoms and ultrasound findings strongly suggest the presence of a kidney stone or obstruction, which would be associated with the presence of uric acid crystals in the urine.</p>

Table 7: 案例分析。我们随机选择了一个MMLU数据集中的例子来展示Self-Guide的效果，并突出显示了**答案和相关内容**。这是英文原始输入和输出内容，经过翻译的中文版本在正文表5中展示。

<p>信息: [Davis] played chess with his brother [Rick]. [Davis] was disappointed that his father, [Kent], wouldn't be at the play to see him perform. [Sharon] likes taking her son [Oliver] to baseball practice. She also likes taking her son [Rick] to soccer practice. [Dorothy] made cookies with her grandson [Davis]. [Davis]'s father [Kent] is not allowed to eat sugar.</p> <p>问题: How is [Dorothy] related to [Oliver]?</p> <p>答案: grandmother</p>
<p>原始推理链 (CoT) 输入提示</p> <p>To solve the problem, Please think and reason step by step, then answer. Question: {Question}</p> <p>Generation Format: 1. Reasoning process: 2. Answer: -----</p>
<p>原始推理链 (CoT) 输出回答</p> <p>Reasoning process:</p> <ol style="list-style-type: none"> [Davis] is the grandson of [Dorothy]. [Oliver] is the son of [Sharon]. There is no direct relationship mentioned between [Dorothy] and [Oliver]. <p>Answer: There is no direct relationship between [Dorothy] and [Oliver] mentioned in the information.</p>
<p>步骤1: 基于Self-Guide的自我规划</p> <p>获取Self-Guide的输入提示</p> <p>You are a knowledgeable and patient professor whose role is to guide students in solving problems correctly. Here is a question:{Question} please provide a detailed analysis. Note: Since your responsibility is to guide students in answering the question, your analysis should think step by step, Please note that your role is to guide them step by step through the problem, so please don't give them the final result.</p> <p>-----</p> <p>得到的Guideline</p> <p>Step 1: Identify the family members mentioned in the story: - Davis - Rick - Kent - Sharon - Oliver - Dorothy</p> <p>Step 2: Determine the relationships between the family members that are explicitly mentioned in the story: - Davis is the grandson of Dorothy - Rick is the brother of Davis - Kent is the father of Davis - Sharon is the mother of Oliver and Rick</p> <p>Step 3: Identify any indirect relationships that may help us determine the relationship between Dorothy and Oliver: - Davis is the grandson of Dorothy - Kent is the father of Davis - Sharon is the mother of Oliver - Sharon is also mentioned as the mother of Rick</p> <p>Step 4: Consider the relationships between the family members and how they might connect Dorothy and Oliver: - Since Davis is the grandson of Dorothy, and Kent is the father of Davis, it means that Kent is the son of Dorothy. This makes Dorothy the grandmother of Kent's children, including Oliver.</p> <p>Step 5: Based on the relationships identified, we can infer that Dorothy is the grandmother of Oliver. Now, based on the analysis, we can determine the relationship between Dorothy and Oliver.</p>
<p>步骤2: 最终答案生成</p> <p>获取最终答案的输入提示</p> <p>{CoT-log}</p> <p>This is a credible analysis of this question: {Self-Guide}. Please verify your reasoning process for errors based on this analysis, then refine your reasoning process and answer.</p> <p>Generation Format: 1. Inference process: 2. Answer: -----</p>
<p>得到的最终答案</p> <p>Inference process:</p> <ol style="list-style-type: none"> Identify the family members mentioned in the story. Determine the relationships between the family members that are explicitly mentioned in the story. Identify any indirect relationships that may help determine the relationship between the two family members in question. Consider the relationships between the family members and how they might connect the two family members in question. Verify the reasoning process for errors based on the analysis and refine if necessary. <p>Answer: Based on the analysis, Dorothy is the grandmother of Oliver.</p>

Table 8: 案例分析。我们随机选择了一个CLUTRR数据集中的例子来展示Self-Guide的效果, 并突出显示了**答案**和**相关内容**。这是英文原始输入和输出内容。

<p>问题: Jane was born on the last day of February in 2000. Today is her 16-year-old birthday. What is the date yesterday in MM/DD/YYYY?"</p> <p>答案: 02/28/2016</p>
<p>原始推理链 (CoT) 输入提示</p> <p>To solve the problem, Please think and reason step by step, then answer. Question: {Question} Please format the date in the following manner: MM/DD/YYYY.</p> <p>Generation Format: 1. Reasoning process: 2. Answer:</p>
<p>原始推理链 (CoT) 输出回答</p> <p>Reasoning process:</p> <ol style="list-style-type: none"> Jane was born on February 29, 2000, which was a leap year. Since today is her 16th birthday, we need to calculate the date 16 years ago. 16 years ago from 2021 is 2005. Since 2005 was not a leap year, February 29th did not exist. Therefore, her birthday would have been on February 28th. <p>Answer: 02/28/2005</p>
<p>步骤1: 基于Self-Guide的自我规划</p> <p>获取Self-Guide的输入提示</p> <p>You are a knowledgeable and patient professor whose role is to guide students in solving problems correctly. Here is a question:{Question} please provide a detailed analysis. Note: Since your responsibility is to guide students in answering the question, your analysis should think step by step, Please note that your role is to guide them step by step through the problem, so please don't give them the final result.</p> <p>得到的Guideline</p> <p>Step1: Determine the current date. Today is Jane's 16th birthday, so we need to calculate the current date. Since she was born on the last day of February in 2000, her birthday would be on February 29th, 2016 (since 2000 was a leap year).</p> <p>Step2: Calculate yesterday's date. To find yesterday's date, we need to subtract 1 day from Jane's birthday. Since February 2016 has 29 days, we need to subtract 1 from 29 to get yesterday's date.</p> <p>Step3: Consider the month and year after subtracting 1 from 29.</p> <p>Step4: Finalize the answer now that we have calculated yesterday's date. By following these steps, we can determine the date yesterday in MM/DD/YYYY format.</p>
<p>步骤2: 最终答案生成</p> <p>获取最终答案的输入提示</p> <p>{CoT-log}</p> <p>This is a credible analysis of this question: {Self-Guide}. Please verify your reasoning process for errors based on this analysis, then refine your reasoning process and answer.</p> <p>Generation Format: 1. Inference process: 2. Answer: MM/DD/YYYY.</p> <p>得到的最终答案</p> <p>Inference process:</p> <ol style="list-style-type: none"> The current date is February 29, 2016, as it is Jane's 16th birthday and she was born on the last day of February in 2000, which was a leap year. To find yesterday's date, we subtract 1 day from February 29, 2016, which gives us February 28, 2016. Considering the month and year, yesterday's date is February 28, 2016. <p>Answer: 02/28/2016</p>

Table 9: 案例分析。我们随机选择了一个Date数据集中的例子来展示Self-Guide的效果，并突出显示了**答案**和**相关内容**。这是英文原始输入和输出内容。