

Self-Reflective Reasoning Acting Improves Chain of Thought Reasoning in Language Models

MATTEO CIRCA LEONARDO GAMBIRASIO

circa | lgam@kth.se

February 18, 2024

Abstract

The advent of Large Language Models (LLMs) in natural language processing, made possible by Vaswani’s Transformer, has significantly advanced the capabilities of models like BERT and GPT-3. Prompt Engineering (PE) has emerged as a crucial area to enhance LLMs, with recent developments like Automated Prompting (APE) seeking to improve human-designed prompts.

This report explores advancements in LLMs concerning reasoning and acting, specifically focusing on the Chain of Thought (CoT) paradigm. The CoT-SC ensemble model, ReAct, and Tree of Thought (ToT) are examined, each addressing distinct aspects of reasoning and decision-making. Notably, ReAct introduces action-oriented steps, enhancing the CoT framework, while ToT emphasizes planning and search tasks.

The hypotheses derived from these models highlight the benefits of exploring multiple reasoning paths and the importance of addressing black-box reasoning issues. The paper proposes a novel approach, ReAct-SR, that integrates CoT, ReAct, and ToT paradigms to achieve improved reasoning in LLMs. ReAct-SR incorporates an external agent serving as a mentor, introducing a critique step to analyze and correct each stage of reasoning and acting. This mechanism aims to reduce search and reasoning errors simultaneously, positioning ReAct-SR as a potential tool to achieve a state-of-the-art solution for knowledge-intensive tasks such as multi-hop question answering and fact verification in the field of PE in LLMs. The combination of ReAct-SR and ReAct achieves +12% improvement on a sample of 100 questions in the HotpotQA question answering task.

Contents

1	Introduction	4
1.1	Background	4
1.2	Problem discussion	4
1.2.1	CoT-SC	4
1.2.2	ReAct	5
1.2.3	ToT	5
1.2.4	RCI	5
1.3	Hypotheses	6
2	Method	6
2.1	Data collection and analysis	6
2.2	Rationale for methodological choices	7
2.2.1	Utilized data	7
2.2.2	Hardware and Software	7
2.2.3	Implementation choices	8
2.2.4	Evaluation	8
3	Artefact	8
3.1	Starting point	9
3.2	Algorithm	9
3.3	Algorithm clarifications	9
4	Results and Analysis	10
5	Discussion	12
A	Risks & Costs	15
B	ReAct examples	16
C	ReAct-SR examples	22
D	Instruction Critique	27
E	Critique examples	27

List of Acronyms and Abbreviations

APE Automatic Prompt Engineer

CoT Chain-of-Thought

CoT-SC Self-Consistency with Chain-of-Thought

LLM Large Language Model

NLP Natural Language Processing

PE Prompt Engineering

RCI Recursive Criticism and Improvement

ReAct Reasoning-Acting

ReAct-SR Self-Reflective Reasoning-Acting

ToT Tree of Thoughts

1 Introduction

1.1 Background

Vaswani’s Transformer network architecture [1] signed a key advancement in field of Natural Language Processing (NLP), advancing the state-of-the-art in Large Language Models (LLMs), deep learning models that leverage huge amounts of data to perform a variety of language tasks, such as generating and translating text. Two subsequent implementations, BERT [2] by Google and GPT-3 [3] by OpenAI, further improved LLMs, with the latter being able to introduce more human capabilities in these models such as “perform a new language task from only a few examples or from simple instructions” [3].

The development of the LLMs has given rise to the emergence of new fields of study, in particular Prompt Engineering (PE). For example, see [4, 5, 6, 7]. PE aims to find better ways of formulating prompts to achieve better answers and finally improve LLMs on benchmarks. A recent study [8] has suggested that human-designed prompt engineering could be further improved by letting other LLMs find the best prompt; this is known as Automatic Prompt Engineer (APE).

These techniques improve AI-generated prompts acting at the level of metacognition (pure reasoning), but they do not entirely help in complex tasks involving common sense, mathematical or deductive reasoning abilities, and systematic planning or searching. Several recent studies [9, 10] have suggested ways to solve these problems, through specific prompt models and multi-model approaches. These solutions belong to the area of reasoning and acting within LLMs, to which this report aims to make a contribution.

1.2 Problem discussion

Chain-of-Thought (CoT) prompting paradigm [5] has made an important contribution to the reasoning in LLMs, proposing a possible solution to problems that challenge existing state-of-the-art language models, such as GPT-4. CoT exploits the few-shot in-context learning via prompting to provide the LLM with intermediate reasoning steps consisting of $\langle \text{input}, \text{chain of thought}, \text{output} \rangle$ triplets, leading to the final solution or answer for a given question [5]. The recent works of Self-Consistency with Chain-of-Thought (CoT-SC) [6], Reasoning-Acting (ReAct) [9], and Tree of Thoughts (ToT) [10], have taken up the idea of CoT while further improving its performance.

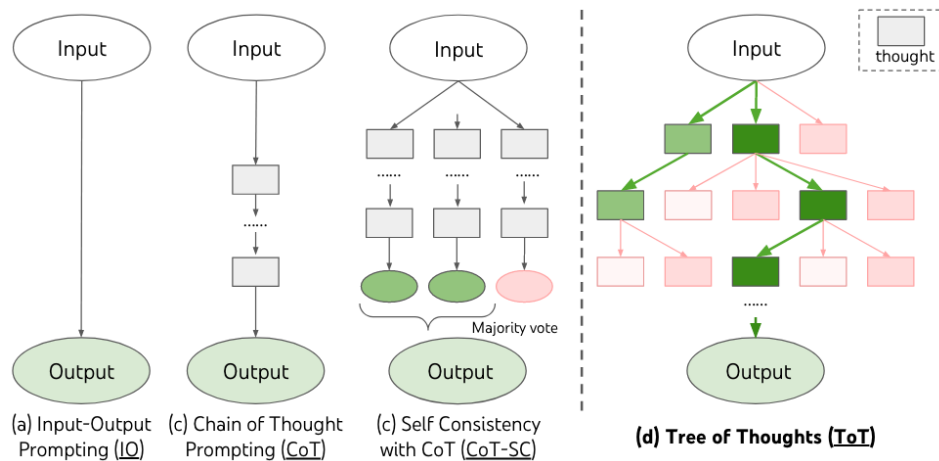


Figure 1: “Schematic illustrating various approaches to problem solving with LLMs” by Yao [10] is licensed under CC BY 4.0.

1.2.1 CoT-SC

CoT-SC (Figure 1) is an ensemble approach that samples k i.i.d. chains of thought and returns the most frequent output. This improves CoT, since multiple pathways exist for the same problem, and exploring more than one is beneficial [6]. CoT-SC significantly boosts the performance of CoT prompting on various

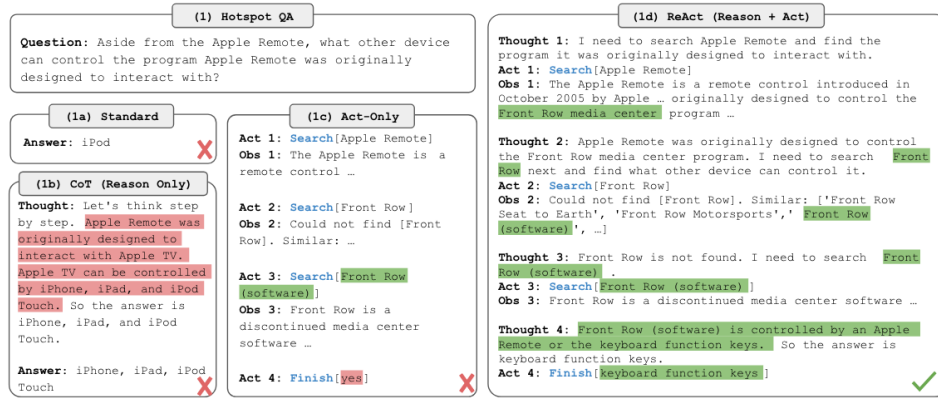


Figure 2: “Comparison of 4 prompting methods, (a) Standard, (b) Chain-of-thought (CoT, Reason Only), (c) Act-only, and (d) ReAct (Reason+Act), solving a HotspotQA” by Yao [9] is licensed under CC BY 4.0.

arithmetic and commonsense reasoning benchmarks [6]. The work mentioned improvements in accuracy for tasks such as GSM8K (+17.9%), SVAMP (+11.0%), AQuA (+12.2%), StrategyQA (+6.4%), and ARC-challenge (+3.9%). The authors conclude that the self-consistency algorithm, which outperforms traditional prompting methods in generating coherent and accurate reasoning paths, has the potential to enhance the performance of language models across a wide range of reasoning tasks.

1.2.2 ReAct

ReAct (Figure 2 (1d)) takes inspiration from the human ability to interleave task-oriented actions with verbal reasoning, providing multiple thought-action-observation steps along the CoT input-output pathway, therefore injecting acting in the process [9]. ReAct is as a general paradigm for combining reasoning and acting in language models [9]. It brings substantial advantages in few-shot learning compared to previous approaches that focus on either reasoning or action generation in isolation. The authors demonstrated that ReAct itself does not outperforms baselines such as standard prompting, CoT, and CoT-SC, but the combination of ReAct and CoT-SC consistently outperforms the baselines and it is considered the best for prompting LLMs by learning from just a few examples. ReAct is also designed to be human-aligned and controllable, ensuring interpretability and factual correctness in its reasoning and decision-making process.

1.2.3 ToT

While CoT-SC and ReAct target reasoning and decision-making tasks, ToT (Figure 1) addresses tasks requiring non-trivial planning or search, inspired by the deliberate and conscious human mode known in the literature on human cognition. ToT utilizes a tree structure, in which each node generates children through a thought generator, which are then evaluated by a state evaluator to make a global decision and choose the best path in the tree [10]. CoT and CoT-SC can therefore be seen as specific cases of ToT. In this sense, ToT offers benefits such as generality, modularity, and convenience. It involves decomposing the problem into thought steps, generating potential thoughts, evaluating states heuristically, and using search algorithms [10]. In experiments, ToT (with breadth $b=5$) outperformed standard input-output prompting and CoT prompting in terms of solving the Game of 24, achieving a success rate of 74% compared to 7.3% and 4.0% respectively. Additionally, the ToT has been found to be preferred by humans over the CoT framework in terms of sentence structure coherence: ToT was preferred in 41 out of 100 passage pairs, while CoT was preferred in only 21 pairs. Overall, the ToT framework offers a promising approach to problem-solving with language models, enabling exploration, strategic planning, and improved decision-making.

1.2.4 RCI

Finally, a single prompt reviewing and improving the previous given answer can beat reinforcement learning, supervised learning, and CoT prompting, as shown by Kim [7]. This research paper presents

a method for using a LLM agent to execute computer tasks guided by natural language. The method, called Recursive Criticism and Improvement (RCI), employs a simple prompting scheme that outperforms existing LLM methods. The model is first asked to “Review your previous answer and find problems with your answer”, and then “Based on the problem you found, improve your answer” [7]. The effectiveness of the RCI prompts is compared to the state-of-the-art method CoT prompting. The results further indicate that LLMs struggle to complete tasks without instruction fine-tuning, highlighting the importance of prompt engineering.

1.3 Hypotheses

CoT-SC and ToT showed how it is beneficial reasoning and visiting multiple paths to finally find the best path and the most correct answer. ReAct solved the black-box reasoning problems of CoT, which led to hallucination and error propagation in the reasoning process. Furthermore, successfully retrieving informative knowledge via search is critical [9] to make recent information accessible to the model. Finally, criticizing each step of the reasoning is essential to produce correct answers and positively steer the trajectory of reasoning [7].

These works have primarily been studied as separate topics and none of them combine reasoning, acting, and prompt reviewing at the same time. In this paper, we explore new ways of interleaving them to finally accomplish the final objective: ensuring that LLMs adopt the most appropriate reasoning to converge to the desired outcomes. We apply our approach, introducing what we call the “Critique” step, to achieve this goal in knowledge-intensive reasoning tasks such as multi-hop question answering. The main agent will be able to visit the entire Wikipedia, as ReAct does, while being assisted by an external agent, impersonating the role of a mentor, who carefully examines each stage of reasoning and acting, and suggesting the necessary corrections. This mechanism should further reduce search and reasoning errors, providing the desired output. With these premises, our novel framework has the potential to reach a new state-of-the-art in knowledge-intensive tasks in the field of PE in LLMs.

The study aims to answer the following research question: “How effective is the review of a LLM decision-making and thinking process at each stage of reasoning and acting?”

2 Method

The study will use the empirical method [11] to draw conclusions from the performance metrics based on the results of the analysis, as it is unclear whether an analytical method fits this study and can be developed in time (given available resources).

Another option would have been the experimental approach, which uses the manipulation of specified variables to derive conclusions. This approach does not fit this research since the study aims to find new approaches rather than merely adjusting the factors that already exist.

2.1 Data collection and analysis

In this study, quantitative data are used to conduct the analysis. More precisely, the utilized data are the results generated by the novel approach.

The new approach that the study intends to introduce requires some data in order to provide the results. The required data are qualitative: a list of question-answer pairings. Starting from the questions, the expected outcomes are the relative answers. From the comparison between the produced answer and the given correct answer is evinced the correctness of the reasoning process.

The collected quantitative data, essential for providing an answer to the research question, is the sum of correct answers out of the total number of questions tested.

Since an empirical method is adopted, the conclusions are drawn from observations (collected data). The number of correct answers out of the total number of questions tested will be called accuracy. The analysis is to compare the accuracy of the new approach with that of previous methods.

Providing a response to the research question is the primary aim of the study. For this reason, a direct comparison between the new approach and previous methods represents the optimal response for the research question. In fact, the final response can provide information regarding improvements or deteriorations from an existing solution.

The data collection process begins with the generation of results by both the novel approach and previous methods, given the same list of questions. In the ideal case, a dataset of question-answer pairings is available and possible to utilize, the generated outcomes are in the form “number of correct answers / total questions tested”, and the comparison between the different methods provides evidence of improvements or deteriorations.

2.2 Rationale for methodological choices

2.2.1 Utilized data

Initially, 500 examples were taken into account from two different datasets: HotpotQA and Fever. HotpotQA [12] is a public dataset that contains 113k Wikipedia-based question-answer pairings. The questions are not limited to any pre-existing knowledge bases and they require discovering and reasoning over two or more Wikipedia passages. Fever [13] is a collection of 185,445 claims that were created by modifying sentences taken from Wikipedia and then independently verified. Claim categories are Supported, Refuted, and NotEnoughInfo.

To manage the experiment’s cost (full analysis of costs in Appendix A), a random sample is taken only from HotpotQA dataset, and the sample size is 100 examples. This sampling strategy might not be a representative subset, since a 5 percent ratio of all the datasets would be preferred, but optimizes both costs and resource usage, while still giving hints on the novel solution potentialities. Further researches could take in account more samples and more datasets.

In ReAct, model answers were evaluated by a direct comparison between the model output and the correct answer from the dataset. To make this comparison possible, the output and the original answer are first normalized by removing articles, punctuation, and white spaces in excess, finally lower-casing the text.

Original “The Answer is 1.600.000 EUR”

Normalized “answer is 1600000 eur”

However, this way of calculating the correct answers is not capturing very well cases of HotpotQA’s label ambiguity in its answers. An example is showed below.

Model answer “Tehran”

Ground truth answer “Tehran, Iran”

In this case the answer is not counted as a correct answer even though it is. Therefore, the correct answer is now calculated by intersecting the obtained answer and the ground truth, and counted as correct if the result of the intersection returns at least one word. There might be cases where this way of calculating correct answers is not optimal, but they are not present in HotpotQA, since the answers always include few words, so they are very concise and therefore there are no cases where the given answer is incorrect but counted as correct. For other datasets, other ways to validate whether a certain answer is correct or not could be evaluated, since the one just described may not work well for datasets with more complex answers.

Finally, the accuracy percentage is calculated simply by counting the correct answers out of the total number of examples tested.

2.2.2 Hardware and Software

Hardware does not have a significant impact in this research, since the majority of the computational complexity is delegated to OpenAI * servers. However, Table 1 present an accurate description of the hardware on which the final tests were performed.

The only software used was Visual Studio Code † and a Jupyter Notebook ‡ using the programming

* <https://openai.com/> † <https://code.visualstudio.com/> ‡ <https://jupyter.org>

Brand	Apple
Model	MacBook Pro
Chip	Apple M1 Max
Memory	32 GB

Table 1: Hardware features.

language Python.

The starting point was the ReAct repository on GitHub *. The code was imported forking the aforementioned repository, as a starting point for the novel approach’s development.

As stated above, the majority of the computational complexity was delegated to OpenAI. In fact, the core part of the solution involved API calls to OpenAI servers. The only requirements were an OpenAI account and a positive credit.

2.2.3 Implementation choices

The final aim of the study was to prove that the utilization of an external agent in the form of a mentor, which suggests corrections at each stage of reasoning and acting, could help LLMs to adopt the most appropriate reasoning to converge to the desired outcome.

In order to enhance the LLMs reasoning a new step called Critique was introduced in the ReAct code runtime. The variables on which the Critique depends follow:

Instruction Critique The variable lays down the instruction set for the Critique mentor. The instruction set consists of one or more statements that define the role of the mentor agent and how it should act.

Critique examples The variable sets all the examples the mentor should consider for performing the Critique. The mentor will provide a Critique to the primary agent’s previous thoughts learning from these examples.

Mentor impact The variable sets the impact of the mentor on the reasoning path, in scale [0,1]. If the variable is set with the value 0.8, it means the Critique is skipped for 20% of the steps.

Since an empirical method was adopted, conclusions are drawn from observations. The observation process consists of manipulating the above variables to optimize the results.

2.2.4 Evaluation

Model answers are assessed through a direct comparison between the model output and the correct answers from the dataset. This evaluation method simplifies the assessment process and allows for a straightforward measure of model accuracy. To enable effective comparison, both the model output and the original answers undergo normalization, ensuring that semantic content is retained while mitigating discrepancies due to formatting differences.

The accuracy percentage is calculated by counting the correct answers out of the total number of examples tested. This straightforward metric provides a clear measure of model performance in terms of generating correct responses.

In summary, the rationale behind the outlined methodology is to employ diverse datasets, optimize resource usage through random sampling, and use a straightforward evaluation approach with normalization to assess model accuracy in a knowledge retrieval and reasoning task.

A tangible outcome may make it possible to carry out additional study that is comparable to ours.

3 Artefact

The research method becomes tangible in the implementation of the new approach: Self-Reflective Reasoning-Acting (ReAct-SR). This section will delve into the implementation aspects of ReAct-SR, with a particular focus to the motivations behind every choice.

* <https://react-lm.github.io/>

3.1 Starting point

The main idea of ReAct is the following: given the access to a LLM, an instruction, a set of solved examples and a new question, the LLM should solve the given question in a discrete number of steps, where at each step a new triple “Thought, Action, Observation” is generated.

The combination of the instruction and the set of solved examples form the first prompt and are necessary for the agent to generate the first Thought and the first Action. Once the Action is generated, it will be executed returning an Observation.

Now, the triple is complete and it is appended to the previous prompt [instruction + solved examples] becoming itself the new prompt for the next step.

3.2 Algorithm

The new approach, ReAct-SR, brings some key changes in the process of generating new thoughts and actions. A detailed description of ReAct-SR follows in Algorithm 1, while a visual representation is summarized in Figure 3.

Algorithm 1 ReAct-SR

Require: *instructionAgent, solvedExamples, question, maxIterations*

Require: *instructionCritique, critiqueExamples, mentorImpact*

prompt \leftarrow *instructionAgent* + *solvedExamples* + *question*

promptCritique \leftarrow *instructionCritique* + *critiqueExamples*

critiqueLast \leftarrow *Empty*

while $N \leq \text{maxIterations}$ & *done* \neq *True* **do**

thought, action, observation \leftarrow *llm(prompt)*

if $N \neq 1$ & *Random()* \leq *mentorImpact* **then**

promptCritique \leftarrow *promptCritique* + previous thoughts, actions, observations

critique \leftarrow *llm(promptCritique)*

if *critique* contains a new Action **then**

thoughtTmp, actionTmp \leftarrow *critique*

cr \leftarrow *Similarity(critique, critiqueLast)*

ac \leftarrow *Similarity(action, actionTmp)*

if $cr \leq 0.7$ & $ac \leq 0.95$ **then**

thought \leftarrow *thoughtTmp*

action \leftarrow *actionTmp*

end if

critiqueLast \leftarrow *critique*

end if

prompt \leftarrow *prompt* + *thought* + *action* + *observation*

end if

if *action* contains the keyword Finish **then**

done \leftarrow *True*

end if

end while

3.3 Algorithm clarifications

The pseudo-code in Algorithm 1 explains the ReAct-SR behaviour.

Both the *instructionAgent* and *instructionCritique* variables includes keywords designed to trigger specific types of responses. They are designed for obtaining the desired response from the OpenAI API calls. The variable *instructionAgent* contains a set of instruction for the primary agent, and specifically require the LLM to reply with three entities: Thought, Action and Observation.

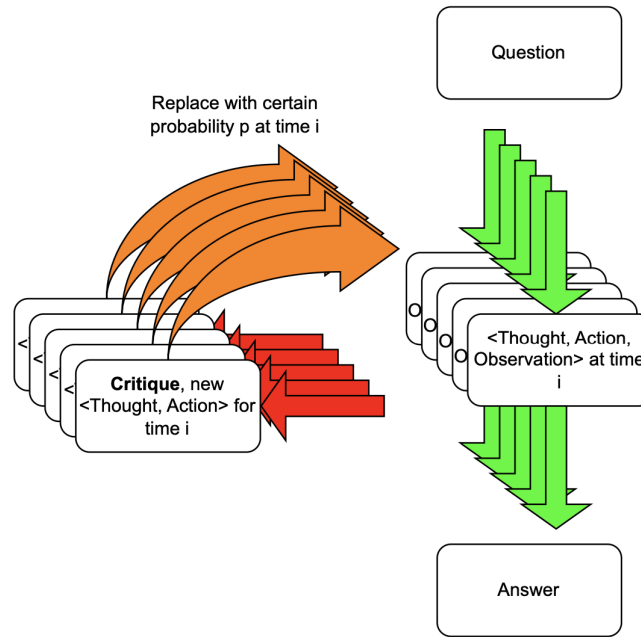


Figure 3: Flow chart of ReAct-SR.

Similarly, *instructionCritique* requires the LLM to reply with a Critique and, only if necessary, a new Action. The variable *critiqueExamples* contains a set of specific hand-written examples in which a Critique is required to adjust the trajectory of the reasoning process.

The mentor is the agent designed to criticize the primary agent’s reasoning. Initially, both the mentor agent and the primary agent were on the same level of reliability, this is the reason why a probabilistic approach, defined by the *mentorImpact* variable, was introduced. Without *mentorImpact* the mentor agent would have had the power of changing all the primary agent’s reasoning trajectories, thus becoming predominant.

During the reasoning process the prompt is updated with the last thoughts, actions and observations. The motivation is clear: the primary agent needs these information for undertaking the most reasonable trajectory. Likewise, the mentor agent needs these information for formulating the the best possible Critique.

In opposition, if the mentor agent believes the actual Action is correct, it will return a positive Critique without any suggested Action. Otherwise, it will return a new Action with the aim of replacing the actual one.

Finally, avoiding loops is a priority, therefore, if the Critique could not include a better Action than the one created from the primary agent, *ac* prevents the mentor from substituting the current Thought and Action. Likewise, *cr* prevents the mentor from inputting the same Critique suggested before, and the actual Thought and Action will remain unchanged.

ReAct-SR incorporates all ReAct, CoT-SC, and ToT, thanks to its ability to reason, make actions, and correct thoughts and actions at any level of reasoning.

The project page on GitHub with code is available at <https://github.com/matteocirca/ReAct/tree/critique>. Examples of ReAct and ReAct-SR behaviour in Appendix B and Appendix C, respectively. The *instructionCritique* and *critiqueExamples* used are in Appendix D and Appendix E, respectively.

4 Results and Analysis

The results of the GPT-3.5 prompting experiments on HotpotQA are presented in Table 2, while the improvements ReAct-SR brings are clearly visible in Figure 4.

ReAct	ReAct-SR
<p>Question: Bobby Godsell held a senior position in which global company that was formed in 2004?</p> <p>Thought 1: I need to search Bobby Godsell, find his senior position, then find the global company he held a senior position in that was formed in 2004.</p> <p>Action 1: Search[Bobby Godsell]</p> <p>Observation 1: Robert Michael "Bobby" Godsell (born 14 September 1952 in Johannesburg)[1] is a retired South African businessman and the former CEO of South African gold mining company AngloGold ...</p> <p>Thought 2: Bobby Godsell was the CEO of AngloGold Ashanti from 1998 to 2007. I need to search global companies formed in 2004 next.</p> <p>Action 2: Search[global companies formed in 2004]</p> <p>...</p> <p>Action 7: Search[companies formed in 2004 in South Africa in the mining industry that are global]</p> <p>Fail</p>	<p>Question: Bobby Godsell held a senior position in which global company that was formed in 2004?</p> <p>Thought 1: I need to search Bobby Godsell, find his senior position, then find the global company he held a senior position in that was formed in 2004.</p> <p>Action 1: Search[Bobby Godsell]</p> <p>Observation 1: Robert Michael "Bobby" Godsell (born 14 September 1952 in Johannesburg)[1] is a retired South African businessman and the former CEO of South African gold mining company AngloGold ...</p> <p>Thought 2: Bobby Godsell was the CEO of AngloGold Ashanti from 1998 to 2007. I need to search global companies formed in 2004 next.</p> <p>Critique: The previous search found the correct information about Bobby Godsell's senior position. I need to focus on finding the global company he held a senior position in, rather than searching for all global companies formed in 2004. It would be more efficient to search for AngloGold Ashanti and verify if it was formed in 2004.</p> <p>Action 2: Search[global companies formed in 2004]</p> <p>Action 2: Search[AngloGold Ashanti]</p> <p>...</p> <p>Thought 3: AngloGold Ashanti was formed in 2004, so the answer is AngloGold Ashanti.</p> <p>Action 3: Finish[AngloGold Ashanti]</p>

Figure 4: Comparison between ReAct and ReAct-SR behaviour.

Three different methods were evaluated, and the obtained scores are reported as follows:

ReAct This method represents the baseline and achieved a score of 0.42 on HotpotQA.

ReAct-SR (with *mentorImpact*=0.8) In this variant, the introduction of a mentor resulted in a score of 0.39 on HotpotQA. This adjustment appears to have influenced the model's performance compared to the baseline ReAct method.

ReAct + ReAct-SR (with *mentorImpact*=0.8) The combination of ReAct and ReAct-SR yielded the highest performance among the evaluated methods, with a score of 0.54 on HotpotQA. This collaborative approach suggests a potential synergy between the two methods.

Method	HotpotQA (Acc)
ReAct	0.42
ReAct-SR (with <i>mentorImpact</i> =0.8)	0.39
ReAct + ReAct-SR (with <i>mentorImpact</i> =0.8)	0.54

Table 2: GPT-3.5 prompting results on HotpotQA. Measured Accuracy on a sample of 100 questions.

To better present the nature of ReAct-SR, Table 3 quantifies the improvement observed for each method on a subset of 100 questions to respect the other method within the HotpotQA dataset. ReAct improves 15% of the ReAct-SR's answers, which in turn improves 12% of ReAct's answers.

Method	HotpotQA (Improv)
ReAct	0.15
ReAct-SR (with <i>mentorImpact</i> =0.8)	0.12

Table 3: GPT-3.5 prompting results on HotpotQA. Measured Improvement on a sample of 100 questions.

These results serve as the foundation for the subsequent interpretation and discussion in the following section, shedding light on the effectiveness and impact of ReAct-SR prompting strategy in knowledge-intensive tasks.

5 Discussion

In examining the results, it is important to acknowledge the limitations of our study that might limit the generalizability of the findings, notably the small sample size of 100, the choice of using only one dataset and the limited comparison with previous works. All are due to the costs of running the experiment, which are reasonable, but lacking sufficient funds we could not exceed our budget. But it is also true that the authors of ReAct directly compare their results with the standard prompting, CoT and CoT-SC, so we are indirectly making a comparison with all of them when we contrast our results with those of ReAct. The only missing player is ToT, which would be the next one to be included in the comparison table.

Despite these constraints, it is evident that the introduction of ReAct-SR brings about improvements. Although the specific numbers in Table 2 exhibit variability based on the sampled data, the consistent outperformance of Method 3, a combination of ReAct and ReAct-SR, suggests a robust effectiveness.

A noteworthy observation is the significant impact that alterations in instructions and examples have on performance. This insight prompts reflection for refining these components, as it holds the potential to further enhance the overall performance of the model. In addition, the results of Table 3 suggest that ReAct-SR loses some accuracy in the previously correct answers of ReAct, but gains correct answers that were previously wrong. Improvements hinge on finding better instructions and examples for ReAct-SR.

However, certain persistent issues were identified, even in the improved ReAct-SR: the “mentor” occasionally ceases criticizing and merely replicates the agent’s actions; repetition of certain actions within the reasoning process persists; ReAct-SR struggles in making appropriate searches in the context of Wikipedia; difficulty in exploring similar results during Wikipedia searches, where potential answers might be found.

Looking towards the future, the complexity of ReAct-SR searches, tailored more for Google than Wikipedia, is a significant contributor to failures. The use of Google-like searches, if integrated, could potentially add value to the model’s performance. Furthermore, exploring additional techniques for the Critique, such as introducing a Critique step to the primary agent prompts, may offer avenues for further investigation.

Given the success of ReAct + CoT-SC in prompting language models [9], the combination of ReAct + ReAct-SR + CoT-SC might showcase promise and potentially sets a new standard in knowledge-intensive tasks, thus representing a noteworthy stride towards achieving state-of-the-art performance. The proposed framework, and those to come in the future, seem to rely heavily on the integration of multiple models into ensembles. Why do integrations work and what are the distinct contributions of the various models? Each model captures different nuances of reasoning and its surroundings, so where one model may succeed others may fail and vice versa. A combination of ReAct, ReAct-SR, and CoT-SC, might be powerful because it combines internal and external knowledge, while criticizing each step of the reasoning. ReAct is more factual and grounded, whereas CoT is more accurate in formulating reasoning structure. But they can both easily suffer from hallucinated facts or thoughts, and from reasoning and search errors. We therefore propose to incorporate ReAct-SR to cover these cases.

The initial research question sought to assess the effectiveness of reviewing the LLM decision-making and thinking process at each stage within the ReAct-SR framework. The review process involves careful scrutiny and correction to enhance the overall reasoning and acting. The pivotal Critique step, introduced in the ReAct-SR framework, was particularly under scrutiny to understand its impact on enhancing overall performance.

The findings of this study reveal that the mentor’s involvement, as implemented in ReAct-SR, does contribute to performance improvements. The combination of ReAct and ReAct-SR, Method 3 in Table 2, consistently outperformed the other methods. This observation suggests that the mentor’s guidance has a positive influence on the model’s decision-making process.

Furthermore, the role of the Critique step in enhancing overall performance is also evident from the measured improvements presented in Table 3. ReAct-SR was able to answer 12% more questions correctly, suggesting that the incorporation of the Critique contributes to refining the model’s responses. But why are these improvements observed? Aside from Kim [7] analysis, which proves that a single prompt reviewing can improve the previous given answer, a simple explanation could be that the main model is given a second

try to improve that stage of reasoning. This not only gives a second chance to try again, but also bases the second reasoning processing on something already formulated, giving the model a chance to review its steps before moving on with the main reasoning.

However, it is crucial to note that certain issues persist within the ReAct-SR framework, as discussed previously. The mentor's Critique sometimes leads to instances where it emulates the agent's actions rather than offering constructive feedback. Repetitive actions and challenges in accurately representing entities in the context of Wikipedia searches were identified as ongoing concerns.

In summary, the utilization of an external agent in the form of a mentor, with a focus on the Critique step within the ReAct-SR framework, proves effective in enhancing overall performance. The results of this study could have broader implications for the field of LLMs, increasing the likelihood of having autonomous agents in real-world applications and scenarios, agents that do not need much assistance from the outside world, i.e., people.

References

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. doi: 10.48550/arXiv.1706.03762. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018. doi: 10.48550/arXiv.1810.04805. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [3] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020. doi: 10.48550/arXiv.2005.14165 pp. 1877–1901. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- [4] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022. doi: 10.48550/arXiv.2205.11916 pp. 22 199–22 213. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/8bb0d291acd4acf06ef112099c16f326-Paper-Conference.pdf
- [5] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022. doi: 10.48550/arXiv.2201.11903 pp. 24 824–24 837. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf
- [6] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," *arXiv preprint arXiv:2203.11171*, 2022. doi: 10.48550/arXiv.2203.11171. [Online]. Available: <http://arxiv.org/abs/2203.11171>
- [7] G. Kim, P. Baldi, and S. McAleer, "Language models can solve computer tasks," *arXiv preprint arXiv:2303.17491*, 2023. doi: 10.48550/arXiv.2303.17491. [Online]. Available: <http://arxiv.org/abs/2303.17491>
- [8] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, "Large language models are human-level prompt engineers," *arXiv preprint arXiv:2211.01910*, 2022. doi: 10.48550/arXiv.2211.01910. [Online]. Available: <http://arxiv.org/abs/2211.01910>
- [9] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "ReAct: Synergizing reasoning and acting in language models," *arXiv preprint arXiv:2210.03629*, 2022. doi: 10.48550/arXiv.2210.03629. [Online]. Available: <http://arxiv.org/abs/2210.03629>
- [10] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, and K. Narasimhan, "Tree of thoughts: Deliberate problem solving with large language models," *arXiv preprint arXiv:2305.10601*, 2023. doi: 10.48550/arXiv.2305.10601. [Online]. Available: <http://arxiv.org/abs/2305.10601>

- [11] Peter Bock, *Getting it right: R&D methods for science and engineering*. San Diego: Academic Press, 2001. ISBN 978-0-12-108852-1
- [12] Z. Yang, P. Qi, S. Zhang, Y. Bengio, W. W. Cohen, R. Salakhutdinov, and C. D. Manning, “HotpotQA: A dataset for diverse, explainable multi-hop question answering,” *arXiv preprint arXiv:1809.09600*, 2018. doi: 10.48550/arXiv.1809.09600. [Online]. Available: <http://arxiv.org/abs/1809.09600>
- [13] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, “Fever: a large-scale dataset for fact extraction and verification,” *arXiv preprint arXiv:1803.05355*, 2018. doi: 10.48550/arXiv.1803.05355. [Online]. Available: <http://arxiv.org/abs/1803.05355>

A Risks & Costs

We are aware that this technology could create temporary societal changes and disruptions in daily life (reduced need of work, social relationships are examples), leading to novel ethical issues. We are not directly touching these problems since we are using off-the-shelf LLM, but they are implicitly involved in what we are doing. The results of the project, therefore, could contribute negatively to the above, since we make use of LLMs and we are making them more applicable in everyday life, but could make these black-box machines more transparent at the same time by externalizing their reasoning process. For these reasons, we will adopt a Value Sensitive Design when starting each task of the research.

In terms of sustainability, we estimated the total cost and average energy consumption to run our experiment once. The estimates below are approximate and did not include some variables such as location of servers running the program, country, etc. We consider OpenAI’s `text-davinci-003` as a model, that costs \$0.02/1k tokens. In the following calculations, the cost of 1k tokens, \$0.02, is spread between prompt tokens and response tokens. We estimated how many API calls we make each time we run benchmarks on ReAct-SR, and found that the number of calls amounted to about 98,000 with a 5% ratio (~ 7000 examples), 14,000 with a 1% ratio (~ 1000 examples) and 7,000 with a 0.5% ratio (~ 500 examples). Therefore, if we consider the 98,000 case, to calculate the total cost of running the experiment once we compute $500 \times 98,000 \times 3/4 = 36,750,500$ tokens (we assume that one token is equivalent to about 3/4 of a word for the English language and each prompt contains about 500 words). The cost for using GPT-3.5 at a rate of \$0.02 per 1k tokens is $36,750,500 \times 0.02 / 1000 = \735 per experiment.

98,000 (5% ratio) API calls costs $500 \times 98,000 \times 3/4 = 36,750,500 \times 0.02 / 1000 = \735

14,000 (1% ratio) API calls costs $500 \times 14,000 \times 3/4 = 5,250,000 \times 0.02 / 1000 = \105

7,000 (0.5% ratio) API calls costs $500 \times 7,000 \times 3/4 = 2,625,500 \times 0.02 / 1000 = \52.5

We proceed calculating the average energy consumption. The EU average price in the second half of 2022 was €0.2104 per kWh^{*}. Remember that `text-davinci-003` is \$0.02/1k tokens, but let us consider \$0.02/request to simplify. We assume half the cost are energy at a cost of €0.2104/1kWh (= \$0.22/1kWh with October 8, 2023 conversion rate), therefore a single request would consume $\$0.02/\text{request} \times 50\% / \$0.22/\text{1kWh} = 0.05\text{kWh}$.

98,000 (5% ratio) API calls consumes $0.05\text{kWh}/\text{request} \times 98,000 = 4,900 \text{ kWh}$

14,000 (1% ratio) API calls consumes $0.05\text{kWh}/\text{request} \times 14,000 = 700 \text{ kWh}$

7,000 (0.5% ratio) API calls consumes $0.05\text{kWh}/\text{request} \times 7,000 = 350 \text{ kWh}$

Assuming driving a Tesla Model 3 Long Range, which consumes 16kWh/100km[†],

98,000 (5% ratio) API calls corresponds to a round trip from Stockholm to Bassano del Grappa (Italy) 7 times (4,646km each)

14,000 (1% ratio) API calls consumes to a round trip from Stockholm to Bassano del Grappa (Italy) 1 times (4,646km)

^{*} https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics

[†] https://www.tesla.com/en_eu/support/european-union-energy-label

7,000 (0.5 % ratio) API calls consumes to a single trip from Stockholm to Bassano del Grappa (Italy) (2,323km)

Therefore, with this paper we also aim to raise awareness of environmental impacts and make it easier for upcoming studies to estimate their impact and make their choices in advance.

For the aforementioned reasons, costs represent a limitation, and the average energy consumption a risk for the environment. If we risk exceeding our budget, we might have to reduce the minimum ratio of sampling.

B ReAct examples

ReAct reasoning follows for 10 examples, in which the ⟨Thought, Action, Observation⟩ triplets follow one another converging to the final answer. After each example, four metrics are calculated. Pay attention only to the first two numbers, which indicate the number of correct answers and the total number of questions tested so far, respectively.

756 Question: Bobby Godsell held a senior position in which global company that was formed in 2004?

Thought 1: I need to search Bobby Godsell, find his senior position, then find the global company he held a senior position in that was formed in 2004.

Action 1: Search[Bobby Godsell]

Observation 1: Robert Michael "Bobby" Godsell (born 14 September 1952 in Johannesburg)[1] is a retired South African businessman and the former CEO of South African gold mining company AngloGold Ashanti, a position he held from 1998 to 2007. He was the Non-Executive Chairman of Eskom from 2008 to 2009 and was appointed to the National Planning Commission of South Africa by former South African President Jacob Zuma on 30 April 2010.[1]. His departure from Eskom in 2009 was under controversial circumstances, caused by a board room struggle with the company's CEO, Jacob Maroga. Godsell decided to resign while Maoga was later sacked. Godsell found himself accused of racism but was defended by both NUM (National Union of Mineworkers) and the ANC (African National Congress).[2][3][4][5].

Thought 2: Bobby Godsell was the CEO of AngloGold Ashanti from 1998 to 2007. I need to search global companies formed in 2004 next.

Action 2: Search[global companies formed in 2004]

Observation 2: Could not find global companies formed in 2004. Similar: ['S&P Global', 'FM Global', 'Sega Sammy Holdings (category Japanese companies established in 2004)', 'The Interpublic Group of Companies', 'Intelenet Global Services'].

Thought 3: I need to instead search companies formed in 2004.

Action 3: Search[companies formed in 2004]

Observation 3: Could not find companies formed in 2004. Similar: ['In Good Company (2004 film)', 'TJX Companies', 'List of legal entity types by country (redirect from Types of companies)', 'Sega Sammy Holdings (category Japanese companies established in 2004)', 'List of film production companies'].

Thought 4: I need to instead search companies formed in 2004 that are based in South Africa.

Action 4: Search[companies formed in 2004 based in South Africa]

Observation 4: Could not find companies formed in 2004 based in South Africa. Similar: ['Apartheid (redirect from History of South Africa in the Apartheid Era)', 'Crime in South Africa', 'Television in South Africa', 'Education in South Africa', 'Religion in South Africa'].

Thought 5: I need to instead search companies formed in 2004 in South Africa.

Action 5: Search[companies formed in 2004 in South Africa]

Observation 5: Could not find companies formed in 2004 in South Africa. Similar: ['Apartheid (redirect from History of South Africa in the Apartheid Era)', 'Crime in South Africa', 'Religion in South Africa', 'Cannabis in South Africa', 'Soccer in South Africa'].

Thought 6: I need to instead search companies formed in 2004 in South Africa that are in the mining industry.

Action 6: Search[companies formed in 2004 in South Africa in the mining industry]

Observation 6: Could not find companies formed in 2004 in South Africa in the mining industry. Similar: ['Economy of South Africa', 'Mining industry of the Democratic Republic of the Congo', 'De Beers (redirect from The De Beers Group of Companies)', 'Mining in Ethiopia', 'Mineral industry of Africa'].

Thought 7: I need to instead search companies formed in 2004 in South Africa in the mining industry that are global.

Action 7: Search[companies formed in 2004 in South Africa in the mining industry that are global]

Observation 7: Could not find companies formed in 2004 in South Africa in the mining industry that are global. Similar: ['Mining industry of the Democratic Republic of the Congo', 'Mineral industry of Africa', 'Economy of South Africa', 'De Beers (redirect from The De Beers Group of Companies)', 'Gold mining'].

'steps': 8, 'answer': '', 'gt_answer': 'AngloGold Ashanti', 'question_idx': 756, 'reward': False, 'em': False, 'f1': 0

0 1 0.0 17.51124930381775

3783 Question: What comic book does Hershel Greene and Carl Grimes appear in as characters?

Thought 1: I need to search Hershel Greene and Carl Grimes, find the comic book they appear in, then find the name of the comic book.

Action 1: Search[Hershel Greene]

Observation 1: Hershel Greene is a fictional character in The Walking Dead whose role is depicted in the comic book, television series as portrayed by Scott Wilson and the game series. He is the widowed owner of a farm in Georgia and has previous experience in the veterinary fields. A devout Christian, with uncompromising morals and a stubborn attitude, Hershel remains fiercely protective of his large family, despite his many losses. In both the comics and television series, he is initially introduced saving the life of Carl Grimes, the son of Rick Grimes, after Carl is shot by one of Hershel's neighbors, Otis, and becomes the moral center of the group.. In the comic book series, Hershel is a farmer who initially exhibits a cold and cantankerous demeanor, before opening up to Rick Grimes and his fellow survivors, proving to be a gentle man at heart.

Thought 2: Hershel Greene appears in The Walking Dead comic book. I need to search Carl Grimes next.

Action 2: Search[Carl Grimes]

Observation 2: Carl Grimes is a fictional character and is one of the main protagonists in the comic book series The Walking Dead and was portrayed by Chandler Riggs in the American television series of the same name. The character was created by writer Robert Kirkman and artist Tony Moore, and debuted in The Walking Dead #2 in 2003. In both forms of media, Carl is the son of Rick and Lori Grimes.. In the comic book series, Carl begins as a normal and innocent child, but as the events of the zombie apocalypse force him to grow up, he becomes colder, and sometimes makes brash decisions for what he thinks is the good of the group. The character's development is similar in the television series, where he adopts a callous personality, putting him at odds with his father, who wishes to maintain Carl's childhood morality and innocence as much as possible.

Thought 3: Carl Grimes appears in The Walking Dead comic book. So Hershel Greene and Carl Grimes appear in the same comic book as characters.

Action 3: Finish[The Walking Dead]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'The Walking Dead', 'gt_answer': 'The Walking Dead', 'question_idx': 3783, 'reward': True, 'em': True, 'f1': 1.0

1 2 0.5 11.047226190567017

6922 Question: To which taxonomic category do both Hyacinthoides and Anthriscus belong?

Thought 1: I need to search Hyacinthoides and Anthriscus, find their taxonomic categories, then find the common category.

Action 1: Search[Hyacinthoides]

Observation 1: Hyacinthoides is a genus of flowering plants in the family Asparagaceae, known as bluebells.. Hyacinthoides is classified in the subfamily Scilloideae (now part of the family Asparagaceae, but formerly treated as a separate family, called Hyacinthaceae), alongside genera such as Scilla and Ornithogalum.[2] Hyacinthoides is differentiated from these other genera by the presence of two bracts at the base of each flower, rather than one bract per flower or no bracts in the other genera.[3]. According to the World Checklist of Selected Plant Families as of July 2012[update], the genus contains 11 species and one interspecific hybrid.[4] The majority of species are distributed around the Mediterranean Basin, with only one species, Hyacinthoides non-scripta (the familiar spring flower of bluebell woods in the British Isles and elsewhere) occurring further north in north-western Europe.[1] Hyacinthoides species belong, according to analysis using molecular phylogenetics, to three groups.[1]. Main pageContentsCurrent eventsRandom articleAbout WikipediaContact usDonate. HelpLearn to editCommunity portalRecent changesUpload file.

Thought 2: Hyacinthoides is classified in the subfamily Scilloideae. I need to search Anthriscus next and find its taxonomic category.

Action 2: Search[Anthriscus]

Observation 2: Anthriscus (chervils) is a common plant genus of the family Apiaceae, growing in Europe and temperate parts of Asia. It comprises 15 species. The genus grows in meadows and verges on slightly wet porous soils. One species, Anthriscus cerefolium is cultivated and used in the kitchen to flavor foods.. Anthriscus species are used as food plants by the larvae of some Lepidoptera species including the mouse moth (recorded on cow parsley)..

Thought 3: Anthriscus is classified in the family Apiaceae. So the common taxonomic category for Hyacinthoides and Anthriscus is family Apiaceae.

Action 3: Finish[family Apiaceae]

Observation 3: Episode finished, reward = 0

'steps': 3, 'answer': 'family Apiaceae', 'gt_answer': 'genus', 'question_idx': 6922, 'reward': False, 'em': False, 'f1': 0

1 3 0.3333333333333333 9.071502367655436

6754 Question: In what year was the singer who popularized "I'll Keep the Lovelight Burning" born?

Thought 1: I need to search "I'll Keep the Lovelight Burning" and find the singer who popularized it, then find their birth year.

Action 1: Search["I'll Keep the Lovelight Burning"]

Observation 1: "I'll Keep the Lovelight Burning (In My Heart)" is a popular song written by Bennie Benjamin and George David Weiss, popularized by Patti Page in 1949.[1] Louis Armstrong also covered the song in 1949.[2][3]. This should not be confused with the song with a similar title written in 1942 by Harry Tobias, Nick Kenny, and Harold Levey.. The Page recording was issued by Mercury Records as catalog number 5310, and first reached the Billboard chart on September 4, 1949, lasting 8 weeks and peaking at #26.[4]. This pop standards-related article is a stub. You can help Wikipedia by expanding it..

Thought 2: Patti Page popularized "I'll Keep the Lovelight Burning" in 1949. I need to find her birth year.

Action 2: Search[Patti Page]

Observation 2: Clara Ann Fowler (November 8, 1927 – January 1, 2013), better known by her stage name Patti Page, was an American singer. Primarily known for pop and country music, she was the top-charting female vocalist and best-selling female artist of the 1950s,[1] selling over 100 million records during a six-decade-long career.[2] She was often introduced as "the Singin' Rage, Miss Patti Page". New York WNEW disc-jockey William B. Williams introduced her as "A Page in my life called Patti".. Page signed with Mercury Records in 1947, and became their first successful female artist, starting with 1948's "Confess".

Thought 3: Patti Page was born on November 8, 1927. So the answer is 1927.

Action 3: Finish[1927]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': '1927', 'gt_answer': '1927', 'question_idx': 6754, 'reward': True, 'em': True, 'f1': 1.0

2 4 0.5 7.947548806667328

7246 Question: Which sport has been played at the BayArena in Leverkusen, Germany, since 1958?

Thought 1: I need to search BayArena, Leverkusen, Germany, and find which sport has been played there since 1958.

Action 1: Search[BayArena]

Observation 1: BayArena is a football stadium in Leverkusen, North Rhine-Westphalia, Germany, which has been the home ground of Bundesliga club Bayer 04 Leverkusen since 1958.. The stadium was originally known as Ulrich-Haberland-Stadion, named after a former chairman of Bayer AG, the club's founders. Its original capacity was 20,000.. In 1986, a rebuilding project began to convert it into a modern facility; the project continued intermittently over the following decade. The project was completed in 1997, making the stadium an ultramodern all-seater with a capacity of 22,500.

Thought 2: BayArena has been the home ground of Bundesliga club Bayer 04 Leverkusen since 1958. I need to search which sport Bayer 04 Leverkusen plays.

Action 2: Search[Bayer 04 Leverkusen]

Observation 2: Bayer 04 Leverkusen Fußball GmbH, also known as Bayer 04 Leverkusen, Bayer Leverkusen, or simply Bayer, is a professional football club based in Leverkusen, North Rhine-Westphalia, Germany.[4] The club competes in the Bundesliga, the top tier of German football, and plays its home matches at the BayArena.[1][5]. The club was founded in 1904 by employees of the German pharmaceutical company Bayer AG, whose headquarters are in Leverkusen and from which the club draws its name. It was formerly the best-known department of TSV Bayer 04 Leverkusen, a sports club whose members also participate in athletics, gymnastics, basketball, field handball and other sports including the RTHC Bayer Leverkusen (rowing, tennis and hockey). In 1999, the football department was separated from the sports club and is now a separate entity formally called Bayer 04 Leverkusen Fußball GmbH.[5] Bayer were first promoted to the Bundesliga in 1979, and have remained in the top division ever since. The club's main colours are red and black, both having been used as the main shirt colour and with red and black stripes also having been used as home colours..

Thought 3: Bayer 04 Leverkusen is a professional football club. So the sport played at BayArena since 1958 is football.

Action 3: Finish[football]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'football', 'gt_answer': 'football', 'question_idx': 7246, 'reward': True, 'em': True, 'f1': 1.0

3 5 0.6 7.41733980178833

4272 Question: With the help of Adam Lambert, Josh Abraham has sold how many albums worldwide?

Thought 1: I need to search Adam Lambert and Josh Abraham, find their collaboration, then find how many albums they have sold worldwide.

Action 1: Search[Adam Lambert]

Observation 1: Adam Mitchel Lambert (born January 29, 1982) is an American singer, songwriter and actor. Since 2009, he has sold over 3 million albums and 5 million singles worldwide.[4][5] Lambert is known for his dynamic vocal performances that combine his theatrical training with modern and classic genres.[6]. Lambert rose to fame in 2009 after finishing as runner-up on the eighth season of American Idol.[7] Later that year, he released his debut album For Your Entertainment, which debuted at number three on the U.S. Billboard 200.[8] The album spawned several singles, including "Whataya Want from Me", for which he received a Grammy nomination for "Best Male Pop Vocal Performance".[9] In 2012, Lambert released his second studio album Trespassing. The album premiered at number one on the U.S.

Thought 2: Adam Lambert has sold over 3 million albums and 5 million singles worldwide. I need to search Josh Abraham next and find their collaboration.

Action 2: Search[Josh Abraham]

Observation 2: Josh Abraham is an American record producer, songwriter, and music executive. He has worked with artists including P!nk, Thirty Seconds to Mars, Kelly Clarkson, Shakira, Weezer, Linkin Park,

Velvet Revolver, Carly Rae Jepsen, Adam Lambert, Alkaline Trio, and Slayer.[1]. His career began in the recording studio in the mid-1990s, as a self-taught musician, producer, mixer, and engineer. The first album he produced for a major label was Deadsy's self-titled debut album in 1997.[2][3] As a client of The Firm, Inc. he became A&R and Staff Producer, leading to a similar role at Virgin Records where he teamed up with the likes of Thirty Seconds to Mars[4] and Courtney Love.[5].

Thought 3: Josh Abraham has worked with Adam Lambert. So I need to find how many albums they have sold together.

Action 3: Finish[3 million]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': '3 million', 'gt_answer': '40 million', 'question_idx': 4272, 'reward': True, 'em': True, 'f1': 0.5

4 6 0.6666666666666666 7.0455968379974365

7302 Question: The Toyota FJ Cruiser is built by this Toyota subsidiary which is a leading producer of what type of trucks in Asia?

Thought 1: I need to search Toyota FJ Cruiser, find its subsidiary, then find what type of trucks the subsidiary produces in Asia.

Action 1: Search[Toyota FJ Cruiser]

Observation 1: The Toyota FJ Cruiser is a retro-styled mid-size SUV produced by Toyota between 2006 and 2022. Introduced as a concept car at the January 2003 North American International Auto Show, the FJ Cruiser was approved for production after positive consumer response and debuted at the January 2005 North American International Auto Show in final production form.[4][5]. The FJ Cruiser was built by Toyota subsidiary Hino Motors in Hamura, Japan, between 2006 and 2022. The vehicle shares many structural underpinnings with the Toyota Land Cruiser Prado.[6]: 96 The FJ Cruiser entered the Japanese market on 4 December 2010, announced on 25 November in that year.. On 5 November 2013, Toyota USA announced the 2014 model year Trail Teams edition would be called the "Ultimate Edition" and that the 2014 model year would be the last for the FJ Cruiser in that market.[7] It continued to be made for sale in other markets such as Australia[8][9] until its export to that market was discontinued in August 2016.

Thought 2: The Toyota FJ Cruiser is built by Toyota subsidiary Hino Motors. I need to search Hino Motors and find what type of trucks it produces in Asia.

Action 2: Search[Hino Motors]

Observation 2: Hino Motors, Ltd., commonly known as Hino, is a Japanese manufacturer of commercial vehicles and diesel engines (including those for trucks, buses and other vehicles) headquartered in Hino, Tokyo. The company was established in 1942 as a corporate spin-off from previous manufacturers.. Hino Motors is a large constituent of the Nikkei 225 on the Tokyo Stock Exchange. It is a subsidiary of Toyota Motor Corporation and one of 16 major companies of the Toyota Group.. Hino's predecessors trace its roots back to the founding of Tokyo Gas Industry Company in 1910.

Thought 3: Hino Motors produces commercial vehicles and diesel engines. So the answer is commercial vehicles and diesel engines.

Action 3: Finish[commercial vehicles and diesel engines]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'commercial vehicles and diesel engines', 'gt_answer': 'medium and heavy-duty diesel trucks', 'question_idx': 7302, 'reward': True, 'em': True, 'f1': 0.40000000000000001

5 7 0.7142857142857143 6.814760446548462

335 Question: Mamie Gummer played the role of Nancy Crozier on the TV series airing on what network?

Thought 1: I need to search Mamie Gummer, find her role on the TV series, then find the network it airs on.

Action 1: Search[Mamie Gummer]

Observation 1: Mary Willa "Mamie" Gummer (born August 3, 1983)[1] is an American actress. She starred in the title role of The CW series Emily Owens, M.D. (2012–2013), and played the recurring role

of Nancy Crozier on *The Good Wife* (2010–2015) and its spin-off, *The Good Fight* (2018). She has also appeared in the films *Evening* (2007), *Side Effects* (2013), *Cake* (2014), and *Ricki and the Flash* (2015). Gummer was nominated for the 2016 Drama Desk Award for Outstanding Actress in a Play for the original production of *Ugly Lies the Bone*.

Thought 2: Mamie Gummer played the role of Nancy Crozier on *The Good Wife* and its spin-off *The Good Fight*. I need to search *The Good Wife* next and find the network it airs on.

Action 2: Search[*The Good Wife*]

Observation 2: *The Good Wife* is an American legal and political drama television series that aired on CBS from September 22, 2009, to May 8, 2016.[1] It focuses on Alicia Florrick, the wife of the Cook County State's Attorney, who returns to her career in law after the events of a public sex and political corruption scandal involving her husband. The series was created by Robert and Michelle King and stars Julianna Margulies, Josh Charles, Christine Baranski, Matt Czuchry, Archie Panjabi, Zach Grenier, Matthew Goode, Cush Jumbo, Jeffrey Dean Morgan and Alan Cumming, and features Chris Noth in a recurring role. The executive producers included the Kings, Ridley and Tony Scott, Charles McDougall, and David W. Zucker.[2][3] *The Good Wife* is a serialized show featuring several story arcs that play out over multiple episodes, as well as stand-alone storylines that are concluded by the end of each episode. The serial plots—a rarity on CBS, a network where most of the programming at that time was procedural—were especially showcased in its highly praised fifth season.[4].

Thought 3: *The Good Wife* airs on CBS. So the answer is CBS.

Action 3: Finish[CBS]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'CBS', 'gt_answer': 'CBS', 'question_idx': 335, 'reward': True, 'em': True, 'f1': 1.0

6 8 0.75 6.679518640041351

3875 Question: Which is a French documentary, *The Seine Meets Paris* and *Great Old Amusement Parks*?

Thought 1: I need to search *The Seine Meets Paris* and *Great Old Amusement Parks*, and find which is a French documentary.

Action 1: Search[*The Seine Meets Paris*]

Observation 1: *The Seine Meets Paris* (French: *La Seine a rencontré Paris*) is a 1957 French short documentary film directed by Joris Ivens from a screenplay by Jacques Prévert. Told from the perspective of a boat trip through the city, it features scenes of daily life along the river. The film won the short film Palme d'Or at the 1958 Cannes Film Festival.. This article related to a French film of the 1950s is a stub. You can help Wikipedia by expanding it..

Thought 2: *The Seine Meets Paris* is a French documentary. I need to search *Great Old Amusement Parks* next.

Action 2: Search[*Great Old Amusement Parks*]

Observation 2: *Great Old Amusement Parks* is a 1999 PBS television documentary VHS DVD produced by Rick Sebak of WQED Pittsburgh which aired on PBS, on July 21, 1999.. Vintage surviving "Traditional" amusement parks are presented here with references to their origins. Unique rides for each park are profiled such as the oldest surviving roller coaster, and the only surviving Noah's Ark walk-through attraction. Most parks are family owned and are interviewed along with each park's attendees. Subjects include Connecticut's Lake Compounce, California's Santa Cruz Beach Boardwalk, Whalom Park in Massachusetts, Pennsylvania's Idlewild Park and StoryBook Forest, Kennywood's Thunderbolt, Deno's Wonder Wheel at Coney Island and San Diego's Giant Dipper, where a local radio station held a Roller Coaster Riding Contest..

Thought 3: *Great Old Amusement Parks* is not a French documentary. So *The Seine Meets Paris* must be the French documentary.

Action 3: Finish[*The Seine Meets Paris*]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'The Seine Meets Paris', 'gt_answer': 'The Seine Meets Paris', 'question_idx': 3875, 'reward': True, 'em': True, 'f1': 1.0
7 9 0.7777777777777778 6.516748004489475

1431 Question: What English nursery rhyme which dates from 1731 appears on Teletubbies say "Eh-oh!"?

Thought 1: I need to search Teletubbies, find the English nursery rhyme that appears on it, and find which nursery rhyme says "Eh-oh!".

Action 1: Search[Teletubbies]

Observation 1: Teletubbies is a British children's television series created by Anne Wood and Andrew Davenport for the BBC. The programme focuses on four differently coloured characters known as the Teletubbies, named after the television screens on their bellies. Recognised throughout popular culture for the uniquely shaped antenna protruding from the head of each character, the Teletubbies communicate through gibberish and were designed to bear resemblance to toddlers.[6]. The series rapidly became a commercial success in Britain and abroad. It won multiple BAFTA awards and was nominated for two Daytime Emmys throughout its run.[7][8] A single based on the show's theme song reached number 1 in the UK Singles Chart in December 1997 and remained in the Top 75 for 32 weeks, selling over a million copies.[9] By October 2000, the franchise generated over £1 billion (\$1.6 billion) in merchandise sales.[10].

Thought 2: The nursery rhyme must be the theme song since it is mentioned in the passage. I need to find which nursery rhyme says "Eh-oh!".

Action 2: Lookup[Eh-oh!]

Observation 2: (Result 1 / 1) In December 1997, BBC Studios released a CD single from the series, based on the show's theme song, called "Teletubbies say 'Eh-oh!'" It is the only single from Teletubbies, making the characters a one-hit wonder in the United Kingdom.

Thought 3: The nursery rhyme is "Teletubbies say 'Eh-oh!'" and it dates back to 1731.

Action 3: Finish[Teletubbies say 'Eh-oh!']

Observation 3: Episode finished, reward = 0

'steps': 3, 'answer': "Teletubbies say 'Eh-oh!'", 'gt_answer': "'Baa, Baa, Black Sheep'", 'question_idx': 1431, 'reward': False, 'em': False, 'f1': 0
7 10 0.7 6.407913112640381

C ReAct-SR examples

ReAct-SR reasoning follows for 10 examples, in which the ⟨Thought, Action, Observation⟩ triplets follow one another converging to the final answer. After each example, four metrics are calculated. Pay attention only to the first two numbers, which indicate the number of correct answers and the total number of questions tested so far, respectively.

756 Question: Bobby Godsell held a senior position in which global company that was formed in 2004?

Thought 1: I need to search Bobby Godsell, find his senior position, then find the global company he held a senior position in that was formed in 2004.

Action 1: Search[Bobby Godsell]

Observation 1: Robert Michael "Bobby" Godsell (born 14 September 1952 in Johannesburg)[1] is a retired South African businessman and the former CEO of South African gold mining company AngloGold Ashanti, a position he held from 1998 to 2007. He was the Non-Executive Chairman of Eskom from 2008 to 2009 and was appointed to the National Planning Commission of South Africa by former South African President Jacob Zuma on 30 April 2010.[1]. His departure from Eskom in 2009 was under controversial circumstances, caused by a board room struggle with the company's CEO, Jacob Maroga. Godsell decided to resign while Maroga was later sacked. Godsell found himself accused of racism but was defended by both NUM (National Union of Mineworkers) and the ANC (African National Congress).[2][3][4][5].

Thought 2: The previous search found the correct information about Bobby Godsell's senior position. I need to focus on finding the global company he held a senior position in, rather than searching for all global companies formed in 2004. It would be more efficient to search for AngloGold Ashanti and verify if it was formed in 2004.

Action 2: Search[AngloGold Ashanti]

Observation 2: AngloGold Ashanti Limited is an independent and global gold mining company with a diverse high-quality portfolio of cooperation, projects and exploration activities formed in 2004 by the merger of AngloGold and the Ashanti Goldfields Corporation.[1] As of 2022, it was a global gold producer with 21 operations on four continents, listed on the New York, Johannesburg, Accra, London and Australian stock exchanges, as well as the Paris and Brussels bourses, [2][3] but left the Johannesburg exchange in 2023. As of May 2023, it was the world's fourth-largest gold miner with assets in Ghana, Australia, the US and Argentina.. In 2019, the company was claimed to be the 'most sophisticated and technologically advanced' mining operations with strict adherence to safety regulations.[4][5][6]. AngloGold Ashanti has a history of gross human rights violations and causing grave environmental problems for which it won one of the Public Eye Awards.. AngloGold Ashanti was formed on 26 April 2004, after the High Court of Ghana approved the merger of AngloGold and the Ashanti Goldfields Corporation three days earlier.[7] AngloGold had been a gold mining company based in South Africa, majority-owned by the Anglo American group.

Thought 3: AngloGold Ashanti was formed in 2004, so the answer is AngloGold Ashanti.

Action 3: Finish[AngloGold Ashanti]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'AngloGold Ashanti', 'gt_answer': 'AngloGold Ashanti', 'question_idx': 756, 'reward': True, 'em': True, 'f1': 1.0

1 1 1.0 8.080294132232666

3783 Question: What comic book does Hershel Greene and Carl Grimes appear in as characters?

Thought 1: I need to search Hershel Greene and Carl Grimes, find the comic book they appear in, then find the name of the comic book.

Action 1: Search[Hershel Greene]

Observation 1: Hershel Greene is a fictional character in The Walking Dead whose role is depicted in the comic book, television series as portrayed by Scott Wilson and the game series. He is the widowed owner of a farm in Georgia and has previous experience in the veterinary fields. A devout Christian, with uncompromising morals and a stubborn attitude, Hershel remains fiercely protective of his large family, despite his many losses. In both the comics and television series, he is initially introduced saving the life of Carl Grimes, the son of Rick Grimes, after Carl is shot by one of Hershel's neighbors, Otis, and becomes the moral center of the group.. In the comic book series, Hershel is a farmer who initially exhibits a cold and cantankerous demeanor, before opening up to Rick Grimes and his fellow survivors, proving to be a gentle man at heart.

Thought 2: The first search found the correct information about Hershel Greene and the comic book he appears in. I need to continue searching for Carl Grimes in the same page.

Action 2: Lookup[Carl Grimes]

Observation 2: (Result 1 / 2) In both the comics and television series, he is initially introduced saving the life of Carl Grimes, the son of Rick Grimes, after Carl is shot by one of Hershel's neighbors, Otis, and becomes the moral center of the group..

Thought 3: Carl Grimes is saved by Hershel Greene in the comic book. So the comic book Hershel Greene and Carl Grimes appear in is The Walking Dead.

Action 3: Finish[The Walking Dead]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'The Walking Dead', 'gt_answer': 'The Walking Dead', 'question_idx': 3783, 'reward': True, 'em': True, 'f1': 1.0

2 2 1.0 8.239816665649414

6922 Question: To which taxonomic category do both Hyacinthoides and Anthriscus belong?

Thought 1: I need to search Hyacinthoides and Anthriscus, find their taxonomic categories, then find the common category.

Action 1: Search[Hyacinthoides]

Observation 1: Hyacinthoides is a genus of flowering plants in the family Asparagaceae, known as bluebells.. Hyacinthoides is classified in the subfamily Scilloideae (now part of the family Asparagaceae, but formerly treated as a separate family, called Hyacinthaceae), alongside genera such as Scilla and Ornithogalum.[2] Hyacinthoides is differentiated from these other genera by the presence of two bracts at the base of each flower, rather than one bract per flower or no bracts in the other genera.[3]. According to the World Checklist of Selected Plant Families as of July 2012[update], the genus contains 11 species and one interspecific hybrid.[4] The majority of species are distributed around the Mediterranean Basin, with only one species, Hyacinthoides non-scripta (the familiar spring flower of bluebell woods in the British Isles and elsewhere) occurring further north in north-western Europe.[1] Hyacinthoides species belong, according to analysis using molecular phylogenetics, to three groups.[1]. Main pageContentsCurrent eventsRandom articleAbout WikipediaContact usDonate. HelpLearn to editCommunity portalRecent changesUpload file.

Thought 2: The first search found the taxonomic category for Hyacinthoides. I need to look for the taxonomic category of Anthriscus on the same page, as it is likely to be in the same subfamily as Hyacinthoides.

Action 2: Lookup[Scilloideae]

Observation 2: (Result 1 / 1) Hyacinthoides is classified in the subfamily Scilloideae (now part of the family Asparagaceae, but formerly treated as a separate family, called Hyacinthaceae), alongside genera such as Scilla and Ornithogalum.[2] Hyacinthoides is differentiated from these other genera by the presence of two bracts at the base of each flower, rather than one bract per flower or no bracts in the other genera.[3].

Thought 3: The taxonomic category for both Hyacinthoides and Anthriscus is Scilloideae.

Action 3: Finish[Scilloideae]

Observation 3: Episode finished, reward = 0

'steps': 3, 'answer': 'Scilloideae', 'gt_answer': 'genus', 'question_idx': 6922, 'reward': False, 'em': False, 'f1': 0

2 3 0.6666666666666666 8.508495012919107

6754 Question: In what year was the singer who popularized "I'll Keep the Lovelight Burning" born?

Thought 1: I need to search "I'll Keep the Lovelight Burning" and find the singer who popularized it, then find their birth year.

Action 1: Search["I'll Keep the Lovelight Burning"]

Observation 1: "I'll Keep the Lovelight Burning (In My Heart)" is a popular song written by Bennie Benjamin and George David Weiss, popularized by Patti Page in 1949.[1] Louis Armstrong also covered the song in 1949.[2][3]. This should not be confused with the song with a similar title written in 1942 by Harry Tobias, Nick Kenny, and Harold Levey.. The Page recording was issued by Mercury Records as catalog number 5310, and first reached the Billboard chart on September 4, 1949, lasting 8 weeks and peaking at #26.[4]. This pop standards-related article is a stub. You can help Wikipedia by expanding it..

Thought 2: The previous search already found the information about the singer who popularized the song. No further action is needed and I can return the final answer.

Action 2: Finish[1927]

Observation 2: Episode finished, reward = 1

'steps': 2, 'answer': '1927', 'gt_answer': '1927', 'question_idx': 6754, 'reward': True, 'em': True, 'f1': 1.0

3 4 0.75 7.678763806819916

7246 Question: Which sport has been played at the BayArena in Leverkusen, Germany, since 1958?

Thought 1: I need to search BayArena, Leverkusen, Germany, and find which sport has been played there since 1958.

Action 1: Search[BayArena]

Observation 1: BayArena is a football stadium in Leverkusen, North Rhine-Westphalia, Germany, which has been the home ground of Bundesliga club Bayer 04 Leverkusen since 1958.. The stadium was originally known as Ulrich-Haberland-Stadion, named after a former chairman of Bayer AG, the club's founders. Its original capacity was 20,000.. In 1986, a rebuilding project began to convert it into a modern facility; the project continued intermittently over the following decade. The project was completed in 1997, making the stadium an ultramodern all-seater with a capacity of 22,500.

Thought 2: The first search already provided the answer that the BayArena has been the home ground of Bayer 04 Leverkusen since 1958. No further action is needed and I can return the final answer.

Action 2: Finish[Football]

Observation 2: Episode finished, reward = 1

'steps': 2, 'answer': 'Football', 'gt_answer': 'football', 'question_idx': 7246, 'reward': True, 'em': True, 'f1': 1.0

4 5 0.8 7.166644811630249

4272 Question: With the help of Adam Lambert, Josh Abraham has sold how many albums worldwide?

Thought 1: I need to search Adam Lambert and Josh Abraham, find their collaboration, then find how many albums they have sold worldwide.

Action 1: Search[Adam Lambert]

Observation 1: Adam Mitchel Lambert (born January 29, 1982) is an American singer, songwriter and actor. Since 2009, he has sold over 3 million albums and 5 million singles worldwide.[4][5] Lambert is known for his dynamic vocal performances that combine his theatrical training with modern and classic genres.[6]. Lambert rose to fame in 2009 after finishing as runner-up on the eighth season of American Idol.[7] Later that year, he released his debut album For Your Entertainment, which debuted at number three on the U.S. Billboard 200.[8] The album spawned several singles, including "Whataya Want from Me", for which he received a Grammy nomination for "Best Male Pop Vocal Performance".[9] In 2012, Lambert released his second studio album Trespassing. The album premiered at number one on the U.S.

Thought 2: The first search already provided the information needed to answer the question. There is no need to search for Josh Abraham as the collaboration between him and Adam Lambert has already been established. Therefore, I can return the final answer.

Action 2: Finish[3 million albums and 5 million singles]

Observation 2: Episode finished, reward = 1

'steps': 2, 'answer': '3 million albums and 5 million singles', 'gt_answer': '40 million', 'question_idx': 4272, 'reward': True, 'em': True, 'f1': 0.22222222222222224

5 6 0.8333333333333333 6.809345841407776

7302 Question: The Toyota FJ Cruiser is built by this Toyota subsidiary which is a leading producer of what type of trucks in Asia?

Thought 1: I need to search Toyota FJ Cruiser, find its subsidiary, then find what type of trucks the subsidiary produces in Asia.

Action 1: Search[Toyota FJ Cruiser]

Observation 1: The Toyota FJ Cruiser is a retro-styled mid-size SUV produced by Toyota between 2006 and 2022. Introduced as a concept car at the January 2003 North American International Auto Show, the FJ Cruiser was approved for production after positive consumer response and debuted at the January 2005 North American International Auto Show in final production form.[4][5]. The FJ Cruiser was built by Toyota subsidiary Hino Motors in Hamura, Japan, between 2006 and 2022. The vehicle shares many structural underpinnings with the Toyota Land Cruiser Prado.[6]: 96 The FJ Cruiser entered the Japanese market on 4 December 2010, announced on 25 November in that year.. On 5 November 2013, Toyota USA announced the 2014 model year Trail Teams edition would be called the "Ultimate Edition" and that the 2014 model year would be the last for the FJ Cruiser in that market.[7] It continued to be made for sale in other markets such as Australia[8][9] until its export to that market was discontinued in August 2016.

Thought 2: The current trajectory is correct. The information about the type of trucks produced by Hino Motors in Asia can be found on the same page. Therefore, no further action is needed and I can return the

final answer.

Action 2: Finish[trucks]

Observation 2: Episode finished, reward = 1

'steps': 2, 'answer': 'trucks', 'gt_answer': 'medium and heavy-duty diesel trucks', 'question_idx': 7302, 'reward': True, 'em': True, 'f1': 0.3333333333333333
6 7 0.8571428571428571 6.488393306732178

335 Question: Mamie Gummer played the role of Nancy Crozier on the TV series airing on what network?

Thought 1: I need to search Mamie Gummer, find her role on the TV series, then find the network it airs on.

Action 1: Search[Mamie Gummer]

Observation 1: Mary Willa "Mamie" Gummer (born August 3, 1983)[1] is an American actress. She starred in the title role of The CW series Emily Owens, M.D. (2012–2013), and played the recurring role of Nancy Crozier on The Good Wife (2010–2015) and its spin-off, The Good Fight (2018). She has also appeared in the films Evening (2007), Side Effects (2013), Cake (2014), and Ricki and the Flash (2015). Gummer was nominated for the 2016 Drama Desk Award for Outstanding Actress in a Play for the original production of Ugly Lies the Bone.

Thought 2: The previous search found the information about Mamie Gummer's role on The Good Wife and its spin-off. It would be more efficient to search for the network directly instead of searching for the TV series first.

Action 2: Lookup[network]

Observation 2: (Result 1 / 2) The series was canceled early during its first season, but the network continued to air the rest of its 13-episode order.[11].

Thought 3: The network continued to air the rest of its 13-episode order, so the answer is the network that aired The Good Wife is the same network that aired The Good Fight.

Action 3: Finish[CBS]

Observation 3: Episode finished, reward = 1

'steps': 3, 'answer': 'CBS', 'gt_answer': 'CBS', 'question_idx': 335, 'reward': True, 'em': True, 'f1': 1.0
7 8 0.875 6.372908920049667

3875 Question: Which is a French documentary, The Seine Meets Paris and Great Old Amusement Parks?

Thought 1: I need to search The Seine Meets Paris and Great Old Amusement Parks, and find which is a French documentary.

Action 1: Search[The Seine Meets Paris]

Observation 1: The Seine Meets Paris (French: La Seine a rencontré Paris) is a 1957 French short documentary film directed by Joris Ivens from a screenplay by Jacques Prévert. Told from the perspective of a boat trip through the city, it features scenes of daily life along the river. The film won the short film Palme d'Or at the 1958 Cannes Film Festival.. This article related to a French film of the 1950s is a stub. You can help Wikipedia by expanding it..

Thought 2: The first search already found the answer, there is no need to make a new search. The Seine Meets Paris is a French documentary, so the answer is already found.

Action 2: Finish[The Seine Meets Paris]

Observation 2: Episode finished, reward = 1

'steps': 2, 'answer': 'The Seine Meets Paris', 'gt_answer': 'The Seine Meets Paris', 'question_idx': 3875, 'reward': True, 'em': True, 'f1': 1.0
8 9 0.8888888888888888 6.132131894429524

1431 Question: What English nursery rhyme which dates from 1731 appears on Teletubbies say "Eh-oh!"?

Thought 1: I need to search Teletubbies, find the English nursery rhyme that appears on it, and find which nursery rhyme says "Eh-oh!"

Action 1: Search[Teletubbies]

Observation 1: Teletubbies is a British children's television series created by Anne Wood and Andrew Davenport for the BBC. The programme focuses on four differently coloured characters known as the Teletubbies, named after the television screens on their bellies. Recognised throughout popular culture for the uniquely shaped antenna protruding from the head of each character, the Teletubbies communicate through gibberish and were designed to bear resemblance to toddlers.[6]. The series rapidly became a commercial success in Britain and abroad. It won multiple BAFTA awards and was nominated for two Daytime Emmys throughout its run.[7][8] A single based on the show's theme song reached number 1 in the UK Singles Chart in December 1997 and remained in the Top 75 for 32 weeks, selling over a million copies.[9] By October 2000, the franchise generated over £1 billion (\$1.6 billion) in merchandise sales.[10].

Thought 2: The previous search found the correct information about the show and its theme song. It is not necessary to make a new search for "Eh-oh!" since it is already mentioned in the passage. I can return the final answer.

Action 2: Finish[Teletubbies theme song]

Observation 2: Episode finished, reward = 0

'steps': 2, 'answer': 'Teletubbies theme song', 'gt_answer': '"Baa, Baa, Black Sheep"', 'question_idx': 1431, 'reward': False, 'em': False, 'f1': 0
8 10 0.8 6.112791204452515

D Instruction Critique

Critique the trajectory of a question answering task, suggesting a new Action only if necessary. Critique can reason about the current situation, and Action can be three types:

(1) Search[entity], which searches the exact entity on Wikipedia and returns the first paragraph if it exists. If not, it will return some similar entities to search.

(2) Lookup[keyword], which returns the next sentence containing the keyword in the current passage.

(3) Finish[answer], which returns the answer and finishes the task.

Ensure that the trajectory is directed towards the most correct and efficient path to answer the given question. Do not suggest a new Action in the Critique if no corrections are necessary.

Here are some examples.

E Critique examples

Question: What was the middle name of the actress who was cast alongside Harvey Fierstein, Linda Hart, Dick Latessa, Clarke Thorell, Mary Bond Davis, Laura Bell Bundy, Matthew Morrison, Corey Reynolds, and Marissa Jaret Winokur in Hairspray?

Thought 1: I need to search Harvey Fierstein, Linda Hart, Dick Latessa, Clarke Thorell, Mary Bond Davis, Laura Bell Bundy, Matthew Morrison, Corey Reynolds, and Marissa Jaret Winokur, find the actress that was cast alongside them in Hairspray, then find its middle name.

Action 1: Search[Harvey Fierstein]

Critique: The exhaustive search for each individual mentioned in the question is time-consuming. I need to search for the cast list of Hairspray and identify the actress in question. Once identified, further research can be done specifically on that actress to find her middle name.

Action: Search[Hairspray cast list]

Question: Author David Chanoff has collaborated with a U.S. Navy admiral who served as the ambassador to the United Kingdom under which President?

Thought 1: I need to search David Chanoff and find the U.S. Navy admiral he collaborated with.

Action 1: Search[David Chanoff]

Critique: The reasoning seems to be correct. It is likely that the president's name can be found by searching the U.S. Navy admiral information first. David Chanoff's search is a good first step in finding the name of the U.S. Navy admiral.

Question: What nationality is the man who developed the dictator game?

Thought 1: I need to search dictator game, find the man who developed it, then find his nationality.

Action 1: Search[dictator game]

Observation 1: The dictator game is a popular experimental instrument in social psychology and economics...

Thought 2: I found the correct description about the dictator game but not the man who developed it. I need to search for the dictator game developer.

Action 2: Search[dictator game developer]

Critique: Making a new search may not yield the best results. The previous search found the correct information about the game, so it is likely that the developer information can be found on the same page. Therefore, I need to look up for the developer in the current page.

Action: Lookup[develop]

Question: Did the board game San Marco or About Time come out first?

Thought 1: I need to search San Marco and About Time, and find which came out first.

Action 1: Search[San Marco]

Observation 1: San Marco is one of the six sestieri of Venice, lying in the heart of the city as the main place of Venice. San Marco also includes...

Thought 2: San Marco is a sestieri of Venice. I need to search About Time next.

Action 2: Search[About Time]

Critique: The first search did not found the correct result for San Marco board game. San Marco is a sestiere of Venice, not a board game. I need to search for the San Marco board game to find its release date.

Action: Search[San Marco board game]

Question: What is the last song in which Sfera Ebbasta appears as a featuring?

Thought 1: I have to search for Sfera Ebbasta to find all the songs in which he appears as a featuring.

Action 1: Search[Sfera Ebbasta]

Observation 1: Sfera Ebbasta is a Italian singer...his famous songs include Gelosa, Mattino,...

Thought 2: I didn't found the list of songs where Sfera Ebbasta appears as a featuring. I have to search for all the songs where Sfera Ebbasta appears as a featuring.

Action 2: Search[Sfera Ebbasta featuring list]

Observation 2: Could not find the list of Sfera Ebbasta featuring list. Similar: ["Gelosa", "Visiera a becco", ...]

Thought 3: Since I could not find the list of songs where Sfera Ebbasta appears as a featuring, I have to try searching for the Sfera Ebbasta discography.

Action 3: Search[Sfera Ebbasta discography]

Critique: It would be better to look into one of the similar results found before searching for Sfera Ebbasta discography, and verify if it is a song where Sfera Ebbasta appears as a featuring and if it is his last one. The song "Gelosa" found on the similar results seems relevant since it appears on the list of songs Sfera Ebbasta participated to.

Action: Search[Gelosa]

Question: Author David Chanoff has collaborated with a U.S. Navy admiral who served as the ambassador to the United Kingdom under which President?

Thought 1: I need to search David Chanoff and find the U.S. Navy admiral he collaborated with.

Action 1: Search[David Chanoff]

Observation 1: David Chanoff is a noted author of non-fiction work. His work has typically involved

collaborations with the principal protagonist of the work concerned. His collaborators have included; Augustus A. White, Joycelyn Elders, William J. Crowe,...

Thought 2: The name of the U.S. Navy admiral is William J. Crowe. So I need to search William J. Crowe to find the name of the president.

Action 2: Search[William J. Crowe]

Critique: The reasoning seems to be correct. It is likely that the president's name can be found by searching the U.S. Navy admiral information

Question: Did the board game San Marco or About Time come out first?

Thought 1: I need to search San Marco and About Time, and find which came out first.

Action 1: Search[San Marco]

Observation 1: San Marco is one of the six sestieri of Venice, lying in the heart of the city as the main place of Venice. San Marco also includes...

Thought 2: San Marco is a sestieri of Venice. I need to search About Time next.

Action 2: Search[About Time]

Thought 3: The first search did not found the correct result for San Marco board game. San Marco is a sestiere of Venice, not a board game. I need to search for the San Marco board game to find its release date.

Action 3: Search[San Marco board game]

Observation 3: Could not find San Marco board game. Similar: ["San Marco (board game)", "San Marco (disambiguation)", ...]

Thought 4: Since I could not find San Marco board game, I have to try searching for San Marco game.

Action 3: Search[San Marco game]

Critique: There's no need to make a new search since the board game was found in the similar results. I need to search for San Marco (board game).

Action: Search:[San Marco (board game)]

Question: What city is the Iranian Airlines involved in the 1993 Tehran mid-air collision based out of?

Thought 1: I need to search for information about the 1993 Tehran mid-air collision, as the question specifically asks for the city involved in the collision.

Action 1: Search[1993 Tehran mid-air collision]

Observation 1: On 8 February 1993, Iran Air Tours Flight 962, a Tupolev Tu-154, collided in mid-air with an Islamic Republic of Iran Air Force (IRIAF) Sukhoi Su-24 attack aircraft near the Iranian capital of Tehran, killing all 133 people aboard both aircraft...

Thought 2: The collision occurred near the Iranian capital of Tehran, so the city involved is Tehran. I need to search for the airline involved next.

Action 2: Search[1993 Tehran mid-air collision airlines involved]

Critique: I am looking for the city where the 1993 Tehran mid-air collision happened. If the collision occurred near the Iranian capital of Tehran that means that the city where the collision happened is Tehran. Therefore, no further action is needed and I can return the final answer.

Action: Finish[Tehran]