Reasoning-trace guided fine-tuning for autonomous web agents

Aadithya Bodke
abodke2@illinois.edu

Hanxi Fang
hanxif2@illinois.edu

Sreenivas Tadikonda
sst11@illinois.edu

Abstract

For a large language model (LLM) to act correctly and effectively in a complex interactive environment, it is crucial for the model to generate coherent reasoning and make decisions aligned with that reasoning. However, traditional LLMs are not inherently trained to operate in this manner. Reasoning traces refer to the step-by-step thought processes generated by LLMs to explain or justify their decisions and actions in completing complex tasks. Aligning an LLM with such reasoning traces can help it develop an understanding of the reasoning processes required to act appropriately in dynamic environments. Our goal is to fine-tune an open-source LLM using reasoning traces generated from actions performed in the WebShop environment. This fine-tuning aims to improve the model's ability to make decisions and act effectively within interactive settings. The code can be found at `https://github.com/Sreenivas-root/WebShop_ReasoningTraces`

1 Introduction

Large language models (LLMs) have demonstrated impressive capabilities across a wide range of tasks, from question answering to complex reasoning [12, 14, 15, 25]. However, their ability to perform multi-step logical reasoning and make decisions in interactive environments remains a significant challenge [4]. This limitation becomes particularly evident when LLMs are required to generate coherent reasoning traces and align their actions with these thought processes in complex, dynamic settings.

Reasoning traces, which represent the step-by-step thought processes generated by LLMs, play a crucial role in enhancing the model's ability to explain and justify its decisions and actions [12, 16, 18, 21, 23]. These traces not only improve the interpretability of the model's outputs but also contribute to the development of more trustworthy and reliable AI systems. By aligning an LLM with well-formed reasoning traces, we can potentially enhance its capacity to understand and navigate complex interactive environments.

Recent research has explored various approaches to improve the reasoning capabilities of LLMs. One such method is the Chain-of-Thought prompting (Wei et al. [21]), which has shown promising results in enhancing the performance of large language models on arithmetic, commonsense, and symbolic reasoning tasks. This technique demonstrates the potential of incorporating explicit reasoning steps into the model's decision-making process. However, this chain-of-thought reasoning is a static black box, in that the model uses its own internal representations to generate thoughts and is not grounded in the external world, which limits its ability to reason reactively or update its knowledge.

Another innovative approach is the ReAct (Yao et al. [23]) paradigm, which combines reasoning traces with task-specific actions in an interleaved manner. This method has shown significant improvements in both language reasoning and decision-making tasks, outperforming traditional imitation and reinforcement learning methods in interactive environments [6, 17, 22].

Preprint. Under review.

Despite these advancements, LLMs still struggle with generating valid reasoning traces and making correct decisions in complex interactive environments [5]. As such, there is a growing interest in researching grounded language agents.

Grounded language agents are AI systems designed to interpret and process language in relation to the physical world or specific contexts. These agents aim to bridge the gap between natural language understanding and decision-making in dynamic, real-world environments. They usually use methods like taking actions to generate extra observations from the real world or search for extra context in complementary to the output of generative LLM . This approach is especially valuable in applications like autonomous web browsing, robotics, game-playing, and virtual agents operating in interactive domains. [8, 9, 10]

In this context, our research aims to address this challenge by fine-tuning an open-source LLM using refined reasoning traces generated from actions performed in the WebShop (Yao et al. [22]) environment. By doing so, we seek to enhance the model's ability to generate coherent reasoning processes and make decisions that align with these thought patterns, ultimately improving its performance in complex, interactive settings.

1.1 Background

1.1.1 Simulated interactive environments

In order to build grounded language agents, we require interactive environments that contain: (1) language elements that reflect real-world usage (2) task feedback that is well-defined and automatically computable to facilitate interactive learning. Several approaches have been proposed to enable large language models (LLMs) to make decisions and execute actions within an environment. Here are two examples:

ReAct combines reasoning and acting with LLMs to solve tasks involving language reasoning and decision-making. It prompts LLMs to interleave reasoning traces with corresponding actions, allowing them to dynamically plan and adjust high-level strategies (*reason to act*) while interacting with external environments to gather new information (*act to reason*).

Reflexion (Shinn et al. [16]) proposes an alternative approach where agents learn from past mistakes through verbal reinforcement. It converts binary or scalar feedback from the environment into textual summaries, which are then provided as additional context for the agent in future interactions. This self-reflective feedback serves as a "semantic" gradient, guiding the agent toward improvements and better task performance in subsequent episodes.

2 Related Works

Early works in web navigation with large language models (LLMs) include WebGPT Nakano et al. [13] that used GPT-3 (Brown et al. [3]) as a browsing assistant, but did not use web elements.

There were qttempts to pre-train LLMs on HTML content but struggled with benchmarking on classical NLP tasks, limiting their use in web navigation [2]. However, breakthroughs through works like Gur et al. [7] demonstrated the potential of LLM in web content comprehension, surpassing previous supervised learning (SL) approaches with significantly less data.

Kim et al. [9] further showcased the potential of larger LLMs such as GPT4 [1] in web navigation tasks through iterative prompting, achieving state-of-the-art baselines in a few shots manner.

Despite these advancements, it remains challenging to develop efficient and adaptable models for Web navigation. Existing methods often require extensive training data and may struggle with fast inference times, highlighting the need for smaller, more capable models.

The paper closest to what we have done is Thil et al. [19]. The authors here describe a new approach that involves Supervised Learning and Reinforcement Learning to improve performance over web navigation tasks on the MiniWoB++ benchmark [11]. This work points out the limitations of former models in understanding HTML contents and proposes ways to enhance their comprehension ability to achieve better performance with much less training data.

Another paper in the space is Lai et al. [10]. AutoWebGLM is designed as a web navigating agent which employs HTML simplification, curriculum training, and reinforcement learning for handling diverse tasks across the open web.

Zheng et al. [24] proposes a system that leverages GPT-4V for visual understanding and action on websites. However, the study identifies grounding as a major challenge, with a 20-25% performance gap between the best grounding strategies and oracle grounding.

3 Limitations of Reflexion and ReAct

Large language models (LLMs) such as GPT-4 have revolutionized natural language processing by demonstrating impressive capabilities in understanding and generating human-like text. To enhance their functionality in complex interactive environments, techniques like **Reflexion** and **ReAct** have been developed. These methods aim to ground LLMs by integrating reasoning processes that guide decision-making and actions. However, despite their innovative approach, Reflexion and ReAct suffer from significant limitations in accuracy, which undermines their effectiveness in real-world applications.

Reflexion and **ReAct** are established methodologies designed to imbue LLMs with structured reasoning capabilities. They operate by generating reasoning traces that inform the model's actions and decisions within interactive settings. While these techniques represent substantial advancements in grounding LLMs, they exhibit several critical shortcomings:

- **Low Accuracy in Reasoning Traces**: Both Reflexion and ReAct often produce reasoning traces that lack precision and consistency. Inaccurate reasoning can lead to flawed decision-making processes, resulting in actions that are misaligned with the intended objectives of the interactive environment.
- **Inconsistent Decision-Making**: The reliance on generated reasoning traces means that any errors or inconsistencies within these traces directly impact the model's ability to make reliable decisions. This inconsistency can manifest as erratic behavior, reducing the trustworthiness and dependability of the LLM in critical applications.
- **Limited Adaptability**: Reflexion and ReAct struggle to adapt to diverse and dynamic scenarios due to their inherent inaccuracies. In environments that require nuanced understanding and flexible responses, such as the *WebShop* platform, these techniques may fail to provide the necessary support for effective task execution.
- **Challenges in Debugging and Improvement**: The low accuracy of reasoning traces complicates the process of identifying and rectifying errors. Without precise and reliable reasoning data, developers face significant difficulties in debugging the model and implementing targeted improvements.

4 Methodology

Our methodology for improving the performance of a large language model (LLM) in interactive environments like WebShop consists of three main phases: data collection, data refinement, model fine-tuning, and evaluation.

4.1 Data Collection and Refinement

We have collected reasoning traces for the WebShop dataset using two advanced prompting techniques - ReAct and Reflexion. Each trace includes the model's thought process, decisions, and actions taken to complete the shopping task.

To obtain the traces, we have run 3 trails with 10,000 environments where each environment is an instruction to complete a shopping task. Trail 0 is running ReAct and consequent trails prompt the model to analyze its previous responses, identify potential errors or areas for improvement, and generate refined reasoning traces (Reflexion technique).

Successful traces get a score of 1 and failure traces get a score of less than 1 based on how close they were to accurately finding the product and buying it. But more often than not the environment task fails as WebShop demands long contexts and traces fail due to various reasons. Failure criteria may include incorrect item selection, invalid action, or inability to navigate the shopping interface effectively, causing long, unfinished traces (e.g. hallucination). We have corrected some traces manually that were close to finishing the task but couldn't and we have collated a total of 1448 successful traces (see example A.2) ready for fine-tuning.

4.2 Model Fine-tuning

4.2.1 Model Architecture and Loading

The base model used is a 4-bit quantized version of the Llama 3.1 8B instruct model [20]. We have utilized the Unsloth library, which provides optimizations for faster fine-tuning.

4.2.2 Dataset Preparation

The trajectory is processed to fit the Llama 3.1 chat template format, with system prompts and user/assistant exchanges properly formatted. Each generated trace is split into a json with 5 keys

- instruction (shopping task)
- score (between 0 and 1)
- success/failure
- trajectory (whole reasoning trace with instructions)
- conversation (reasoning trace as a conversation between user and assistant)

4.2.3 Training Configuration and Hyperparameters

- A maximum sequence length of 2048 tokens is set.
- LoRA (Low-Rank Adaptation) is applied with a rank of 16
- The SFTTrainer (Supervised Fine-Tuning Trainer) from the TRL (Transformer Reinforcement Learning) library is used with the following key settings:
  - Batch size of 16 per device with gradient accumulation steps of 2
  - AdamW 8-bit optimizer
  - 5 training epochs

The trainer is configured to only train on the assistant's responses, ignoring the user inputs and system prompts during backpropagation.

5 Result analysis

5.1 Accuracy improvement

To assess the effectiveness of our fine-tuning approach, we compare the accuracy of different models before and after the fine-tuning process. The accuracy is measured as the success rate of correctly completing the shopping task based on the reasoning traces collected during our experiments. We ran three distinct trails, using the ReAct prompting technique in trail 1, and the Reflexion technique in Trails 2 and 3. The accuracy scores are summarized in fig. 1 and fig. 2. The results clearly demonstrate an improvement in the accuracy of the fine-tuned Llama model, especially when compared to the baseline models, GPT-4 and GPT-mini. Notably, the performance of the "fine-tuned-llamma" model remains consistent across all trails, with no major deviations observed in Trails 2 and 3 using Reflexion.

Prior Fine-Tuning. Prior to fine-tuning, the llama 3.1 model performed at a baseline accuracy of 0% for all trails. This highlights that the model had not yet been optimized for the reasoning and decision-making tasks within the WebShop environment.

During the ReAct and Reflexion trails (trails 1–3), we observed that the GPT-4 model consistently achieved a modest accuracy of 0.4 to 0.5. However, the GPT-mini model showed lower performance with an accuracy of 0.3, underperforming relative to GPT-4. This suggests that larger models, like GPT-4, have more effective reasoning capabilities in complex environments like WebShop.

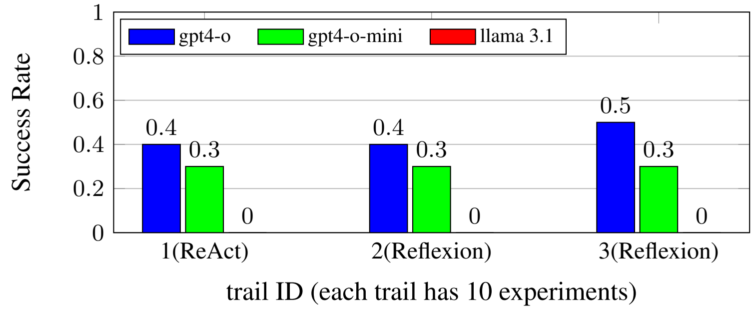

Figure 1: Originally, the open source model (lamma 3.1) has 0 success rate when trying to order merchandise from webshop environment.

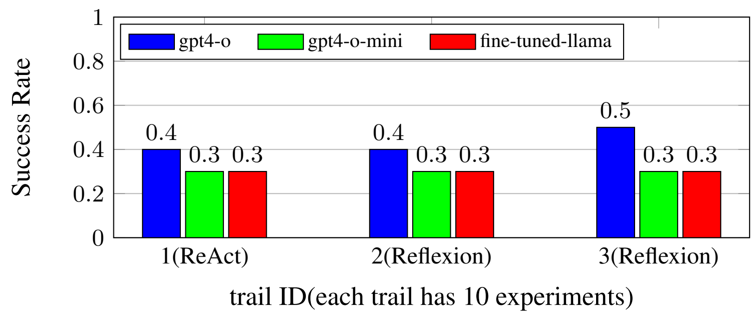

Figure 2: After finetuninig, open-source model (llamma) has comparable success rate with some close-source model (GPT-4o-mini).

Post Fine-Tuning. The final fine-tuned Llama model (after applying LoRA and optimization techniques) achieved an accuracy of 0.3 across all trails, suggesting that our fine-tuning approach using reasoning traces has made substantial progress in improving the model's decision-making abilities. However, the results also indicate that there is still room for improvement, especially in further optimizing the model to reach or exceed the performance of larger models like GPT-4.

5.2 Failure Analysis

For the fine-tuned Llama 3.1 model, we performed an in-depth analysis to identify the causes of failure and categorized the failure cases into different types. The results are shown in fig. 3

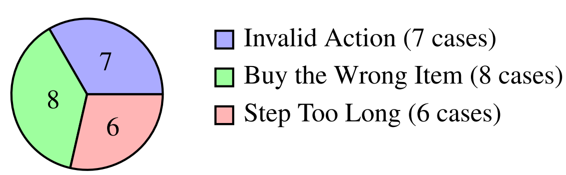

Figure 3: Distribution of Failure Types for Fine-tuned Llama Model

The following are the types of failures observed, along with their corresponding case counts:

- **Invalid Action:** The model occasionally takes an action that is not relevant or valid within the shopping task. For example, selecting a non-existent item or interacting with an incorrect interface element. This category accounts for 7 failure cases.
- **Buy the Wrong Item:** The model sometimes identifies an item but selects the wrong one, such as picking an item that does not match the given criteria (e.g., wrong size, color, or product type). This accounts for 8 failure cases.

- **Step Too Long:** In some cases, the model takes an excessive number of steps, causing it to exceed the allowed time or step limit. The model either becomes too indecisive or cannot find the correct next step promptly. This category includes 6 failure cases.

These failure types point to some common challenges faced by the fine-tuned model, particularly in understanding the nuances of the task and effectively completing it within the given constraints.

6 Conclusion and Future Work

Our study shows us that a small fine-tuned model can perform as well as a very large model on WebShop. Future directions include

- Evaluate transferability of our model against real websites like Amazon
- Related works in the field seem to opt Reinforcement Learning, we would like to see if that will improve the model accuracy
- Increasing the model size might show comparable results to GPT-4

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[2] Armen Aghajanyan, Anchit Gupta, Akshat Shrivastava, Xilun Chen, Luke Zettlemoyer, and Sonal Gupta. Muppet: Massive multi-task representations with pre-finetuning. *arXiv preprint arXiv:2101.11038*, 2021.

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020. URL `https://arxiv.org/abs/2005.14165`.

[4] Antonia Creswell and Murray Shanahan. Faithful reasoning using large language models. *arXiv preprint arXiv:2208.14271*, 2022.

[5] Antonia Creswell, Murray Shanahan, and Irina Higgins. Selection-inference: Exploiting large language models for interpretable logical reasoning. *arXiv preprint arXiv:2205.09712*, 2022.

[6] Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun, and Yu Su. Mind2web: Towards a generalist agent for the web, 2023.

[7] Izzeddin Gur, Ofir Nachum, Yingjie Miao, Mustafa Safdari, Austin Huang, Aakanksha Chowdhery, Sharan Narang, Noah Fiedel, and Aleksandra Faust. Understanding html with large language models. *arXiv preprint arXiv:2210.03945*, 2022.

[8] Karl Moritz Hermann, Felix Hill, Simon Green, Fumin Wang, Ryan Faulkner, Hubert Soyer, David Szepesvari, Wojciech Marian Czarnecki, Max Jaderberg, Denis Teplyashin, et al. Grounded language learning in a simulated 3d world. *arXiv preprint arXiv:1706.06551*, 2017.

[9] Geunwoo Kim, Pierre Baldi, and Stephen McAleer. Language models can solve computer tasks. *Advances in Neural Information Processing Systems*, 36, 2024.

[10] Hanyu Lai, Xiao Liu, Iat Long Iong, Shuntian Yao, Yuxuan Chen, Pengbo Shen, Hao Yu, Hanchen Zhang, Xiaohan Zhang, Yuxiao Dong, et al. Autowebglm: Bootstrap and reinforce a large language model-based web navigating agent. *arXiv preprint arXiv:2404.03648*, 2024.

[11] Evan Zheran Liu, Kelvin Guu, Panupong Pasupat, Tianlin Shi, and Percy Liang. Reinforcement learning on web interfaces using workflow-guided exploration. In *International Conference on Learning Representations (ICLR)*, 2018. URL `https://arxiv.org/abs/1802.08802`.

[12] Grégoire Mialon, Roberto Dessì, Maria Lomeli, Christoforos Nalmpantis, Ram Pasunuru, Roberta Raileanu, Baptiste Rozière, Timo Schick, Jane Dwivedi-Yu, Asli Celikyilmaz, et al. Augmented language models: a survey. *arXiv preprint arXiv:2302.07842*, 2023.

[13] Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. Webgpt: Browser-assisted question-answering with human feedback. *arXiv preprint arXiv:2112.09332*, 2021.

[14] Bhargavi Paranjape, Scott Lundberg, Sameer Singh, Hannaneh Hajishirzi, Luke Zettlemoyer, and Marco Tulio Ribeiro. Art: Automatic multi-step reasoning and tool-use for large language models. *arXiv preprint arXiv:2303.09014*, 2023.

[15] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36, 2024.

[16] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.

[17] Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. Alfworld: Aligning text and embodied environments for interactive learning. *arXiv preprint arXiv:2010.03768*, 2020.

[18] Haotian Sun, Yuchen Zhuang, Lingkai Kong, Bo Dai, and Chao Zhang. Adaplanner: Adaptive planning from feedback with language models. *Advances in Neural Information Processing Systems*, 36, 2024.

[19] Lucas-Andrei Thil, Mirela Popa, and Gerasimos Spanakis. Navigating webai: Training agents to complete web tasks with large language models and reinforcement learning. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, pages 866–874, 2024.

[20] unsloth. unsloth/meta-llama-3.1-8b-instruct-bnb-4bit. URL `https://huggingface.co/unsloth/Meta-Llama-3.1-8B-Instruct-bnb-4bit`.

[21] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

[22] Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. Webshop: Towards scalable real-world web interaction with grounded language agents. *Advances in Neural Information Processing Systems*, 35:20744–20757, 2022.

[23] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

[24] Boyuan Zheng, Boyu Gou, Jihyung Kil, Huan Sun, and Yu Su. Gpt-4v (ision) is a generalist web agent, if grounded. *arXiv preprint arXiv:2401.01614*, 2024.

[25] Yuchen Zhuang, Yue Yu, Kuan Wang, Haotian Sun, and Chao Zhang. Toolqa: A dataset for llm question answering with external tools. *Advances in Neural Information Processing Systems*, 36:50117–50143, 2023.

A Appendix

A.1 Base Prompt for getting Traces

```
You are a shopping agent shopping in Webshop. You need to buy the item that
↪ matches the requirements.

The actions available to you are :
```

```
- reset
- think[Thought]
- search[Search query]
- click[Button to click] (Example : click[< Prev])

Rules:
- IMPORTANT : You can ONLY reply with the action you want to take and
↪  nothing else!
- You can only click buttons available on the page described in the current
↪  observation. Buttons are defined between square brackets []. Ensure you
↪  click the correct buttons.
- You can reset from any page, think on any page.
- You can only search from a page with [Search], so click on the back
↪  buttons to reach such a page before you search again.
- You must end after a few tries by attempting to buy something.

Tips:
- Carefully surf the webshop to fullfil requirements.
- If any items match some of the requirements, click on them to see a
↪  detailed description and check if they match all the requirements.
↪  Quantity requirements can be met by buying multiple items.
- Don't just give up on a search at the 1st page of results. Move through
↪  the result pages by pressing the [Next >] button. You may decide to
↪  give up at a reasonable point such as when the results are empty or too
↪  different from the requirements (usually 3-4 pages).

Examples are given below.

Webshop
Instruction:
i would like a 3 ounce bottle of bright citrus deodorant for sensitive skin,
↪  and price lower than 50.00 dollars
[Search]

Action: search[3 ounce bright citrus deodorant sensitive skin]
Observation:
[Back to Search]
Page 1 (Total results: 50)
[Next >]
[B078GWRC1J]
Bright Citrus Deodorant by Earth Mama | Natural and Safe for Sensitive Skin,
↪  Pregnancy and Breastfeeding, Contains Organic Calendula 3-Ounce
$10.99
[B078GTKVXY]
Ginger Fresh Deodorant by Earth Mama | Natural and Safe for Sensitive Skin,
↪  Pregnancy and Breastfeeding, Contains Organic Calendula 3-Ounce
$10.99
[B08KBVJ4XN]
Barrel and Oak - Aluminum-Free Deodorant, Deodorant for Men, Essential
↪  Oil-Based Scent, 24-Hour Odor Protection, Cedar & Patchouli Blend,
↪  Gentle on Sensitive Skin (Mountain Sage, 2.7 oz, 2-Pack)
$15.95

Action: think[B078GWRC1J and B078GTKVXY are bright citrus deodorant less
↪  then 50 dollars. I can check B078GWRC1J first.]
Observation: OK.

Action: click[B078GWRC1J]
Observation:
```

```
[Back to Search]
[< Prev]
scent [assorted scents][bright citrus][calming lavender][ginger
↪  fresh][simply non-scents]
size [travel set (4-pack)][3 ounce (pack of 1)][3-ounce (2-pack)]
Bright Citrus Deodorant by Earth Mama | Natural and Safe for Sensitive Skin,
↪  Pregnancy and Breastfeeding, Contains Organic Calendula 3-Ounce
Price: $10.99
Rating: N.A.
[Description]
[Features]
[Reviews]
[Buy Now]

Action: think[For 3 ounce bottle of bright citrus deodorant for sensitive
↪  skin, the item has options 'bright citrus' and '3 ounce (pack of 1)'
↪  and seems good to buy.]
Observation: OK.

Action: click[bright citrus]
Observation: You have clicked bright citrus.

Action: click[3 ounce (pack of 1)]
Observation: You have clicked 3 ounce (pack of 1).

Action: click[Buy Now]
```

A.2 Example Success Environment

```
You are a shopping agent shopping in Webshop. You need to buy the item that
↪  matches the requirements.

The actions available to you are :
- reset
- think[Thought]
- search[Search query]
- click[Button to click] (Example : click[< Prev])

Rules:
- IMPORTANT : You can ONLY reply with the action you want to take and
↪  nothing else!
- You can only click buttons available on the page described in the current
↪  observation. Buttons are defined between square brackets []. Ensure you
↪  click the correct buttons.
- You can reset from any page, think on any page.
- You can only search from a page with [Search], so click on the back
↪  buttons to reach such a page before you search again.
- You must end after a few tries by attempting to buy something.

Tips:
- Carefully surf the webshop to fullfil requirements.
- If any items match some of the requirements, click on them to see a
↪  detailed description and check if they match all the requirements.
↪  Quantity requirements can be met by buying multiple items.
- Don't just give up on a search at the 1st page of results. Move through
↪  the result pages by pressing the [Next >] button. You may decide to
↪  give up at a reasonable point such as when the results are empty or too
↪  different from the requirements (usually 3-4 pages).
```

```
Here is the task:

WebShop
Instruction:
i need a long clip-in hair extension which is natural looking, and price
↪  lower than 30.00 dollars
[Search]
> reset

WebShop
Instruction:
i need a long clip-in hair extension which is natural looking, and price
↪  lower than 30.00 dollars
[Search]
> search[long clip-in hair extension]

[Back to Search]
Page 1 (Total results: 50)
[Next >]
[B09C337K8S]
Hair Clip Extensions for Women Natural Looking Hair Extensions Clip in
↪  Hairpieces Black Synthetic Hair Extensions
$29.99
[B093BKWHFK]
MJMOJO Hair Extensions Clip in Human Hair Extensions 120g 7pcs Straight
↪  Natural Black Hair Clip in Remy Human Hair Extensions Double Weft (16
↪  inches #NST)
$63.99
[B099K9Z9L2]
Clip in Hair Extensions 20" Double Weft 100% Remy Human Hair Grade 8A
↪  Quality Thick Long Soft Silky Straight 7pcs 16clips for Women 70grams
↪  Chocolate Brown/Light Blonde #4/27 Color
$43.99
> click[B09C337K8S]

[Back to Search]
[< Prev]
Hair Clip Extensions for Women Natural Looking Hair Extensions Clip in
↪  Hairpieces Black Synthetic Hair Extensions
Price: $29.99
Rating: N.A.
[Description]
[Features]
[Reviews]
[Attributes]
[Buy Now]
> click[Buy Now]
Your score (min 0.0, max 1.0): 1.0

STATUS: OK
```

A.3 Example Failed Reasoning Trace

```
Here is the task:

WebShop
Instruction:
```

```
find me a high speed dual style package with  12" power amplifier car
↪  subwoofer, and price lower than 180.00 dollars
[Search]
> reset

WebShop
Instruction:
find me a high speed dual style package with  12" power amplifier car
↪  subwoofer, and price lower than 180.00 dollars
[Search]
> search[high speed dual style package 12" power amplifier car subwoofer]

[Back to Search]
Page 1 (Total results: 50)
[Next >]
[B01BGW8GOM]
MTX TNP212D2 12" 1200 Watt 4 Ohm Dual Loaded Car Audio Subwoofer Package
↪  with Sub Enclosure, Planet 1500W Monoblock A/B Amplifier & 8 Gauge
↪  Wiring Kit
$277.99
[B08Q9SCL49]
Rockville SK510 Package 10" 1200w Loaded K5 Car Subwoofer Enclosure+DB11
↪  Amp
$164.95
[B06XW8PL68]
Planet Audio AC12D 12" 1800W 4 Ohm DVC Car Subwoofer Package with Boss
↪  Audio 1500W Mono A/B Amplifier, Wiring Amp Kit and QPower Vented Sub
↪  Enclosure
$100.0
> click[B06XW8PL68]

[Back to Search]
[< Prev]
Planet Audio AC12D 12" 1800W 4 Ohm DVC Car Subwoofer Package with Boss
↪  Audio 1500W Mono A/B Amplifier, Wiring Amp Kit and QPower Vented Sub
↪  Enclosure
Price: $100.0
Rating: N.A.
[Description]
[Features]
[Reviews]
[Attributes]
[Buy Now]
> click[Buy Now]
Your score (min 0.0, max 1.0): 0.5

STATUS: FAIL
```