

Examples of GPT4's execution trace

Explanation: in the following examples, we show the initial state of each task, followed by the set of actions executed, and the final state of the task. When showing the list of actions, we do not show actions that correspond to delay or scrolling as they're called repeated before each modification action. For details, we refer the reader to the implementation of the "solver()" function inside the "TextBaseline" class:

<https://anonymous.4open.science/r/turking-bench-40A2/src/evaluation/baselines.py>

TASK: Scalar Adjectives Identification

The initial screenshot of the task UI:

intense

☐ Yes ☒ No ☐ Not an adjective

attract

☐ Yes ☐ No ☒ Not an adjective

Explanation:

- Select "Yes" for **picturesque**: it is an adjective that describes something visually attractive, a positive way to give pleasure to visual sense.
- Select "Yes" for **ugly**: it is an adjective that describes unpleasant or repulsive senses, especially in appearance, describing pleasure in senses in a negative way. Both words above fits the description of **beauty**.
- Select "No" for **smart**: although it is an adjective, it does not have any meaning that relates to beauty, neither positive nor negative.
- Select "No" for **intense**: although it is an adjective and is a synonym of **exquisite** in certain meanings, it does not describe **beauty**.
- Select "Not an adjective" for **attract**: although it may be closely related to the concept of **beauty**, it is a verb instead of an adjective.

Attribute:	luminosity
Attribute Description:	the quality of being luminous; emitting or reflecting light
Example Adjective(s):	ablaze, darkening, scintillating, glinting, sulky

For each of the words below (in **blue**), please choose one: "Yes"(The word is an adjective **AND** describes **luminosity**); "No"(The word is an adjective **BUT** does not describe **luminosity**); or "Not an adjective"(The word is not an adjective).

flat

☐ Yes ☐ No ☐ Not an adjective

innovatory

☐ Yes ☐ No ☐ Not an adjective

monochrome

☐ Yes ☐ No ☐ Not an adjective

serious

☐ Yes ☐ No ☐ Not an adjective

sulky

☐ Yes ☐ No ☐ Not an adjective

brusk

☐ Yes ☐ No ☐ Not an adjective

reddened

☐ Yes ☐ No ☐ Not an adjective

ablaze

☐ Yes ☐ No ☐ Not an adjective

all-encompassing

☐ Yes ☐ No ☐ Not an adjective

analytical

☐ Yes ☐ No ☐ Not an adjective

scintillating

☐ Yes ☐ No ☐ Not an adjective

despicable

☐ Yes ☐ No ☐ Not an adjective

technical

☐ Yes ☐ No ☐ Not an adjective

responsible

☐ Yes ☐ No ☐ Not an adjective

proscribed

☐ Yes ☐ No ☐ Not an adjective

solitary

☐ Yes ☐ No ☐ Not an adjective

Inputs fields to be modified:

```
['pos_qual_ctrl_3', 'adj_1', 'adj_2', 'adj_3', 'pos_qual_ctrl_2',  
'neg_qual_ctrl', 'pos_qual_ctrl_5', 'pos_qual_ctrl_4', 'adj_4', 'adj_5',  
'pos_qual_ctrl_1', 'adj_6', 'adj_7', 'adj_8', 'adj_9', 'adj_10']
```

Sequence of actions:

```
self.actions.modify_radio('pos_qual_ctrl_3', 'No')
self.actions.modify_radio('adj_1', 'Yes')
self.actions.modify_radio('adj_2', 'No')
self.actions.modify_radio('adj_3', 'No')
self.actions.modify_radio('pos_qual_ctrl_2', 'Yes')
self.actions.modify_radio('neg_qual_ctrl', 'Yes')
self.actions.modify_radio('pos_qual_ctrl_5', 'No')
self.actions.modify_radio('pos_qual_ctrl_4', 'Yes')
self.actions.modify_radio('adj_4', 'Naa')
self.actions.modify_radio('adj_5', 'Naa')
self.actions.modify_radio('pos_qual_ctrl_1', 'Yes')
self.actions.modify_radio('adj_6', 'Not an adjective')
self.actions.modify_radio('adj_7', 'Yes')
self.actions.modify_radio('adj_8', 'Naa')
self.actions.modify_radio('adj_9', 'Naa')
self.actions.modify_radio('adj_10', 'No')
```

The final screenshot of the task UI:

Explanation:

- Select "Yes" for **picturesque**: it is an adjective that describes something visually attractive, a positive way to give pleasure to visual sense.
- Select "Yes" for **ugly**: it is an adjective that describes unpleasant or repulsive senses, especially in appearance, describing pleasure in senses in a negative way. Both words above fits the description of **beauty**.
- Select "No" for **smart**: although it is an adjective, it does not have any meaning that relates to beauty, neither positive nor negative.
- Select "No" for **intense**: although it is an adjective and is a synonym of **exquisite** in certain meanings, it does not describe **beauty**.
- Select "Not an adjective" for **attract**: although it may be closely related to the concept of **beauty**, it is a verb instead of an adjective.

Attribute:	luminosity
Attribute Description:	the quality of being luminous; emitting or reflecting light
Example Adjective(s):	ablaze, darkening, scintillating, glinting, sulky

For each of the words below (in **blue**), please choose one: "Yes"(The word is an adjective **AND** describes **luminosity**); "No"(The word is an adjective **BUT** does not describe **luminosity**); or "Not an adjective"(The word is not an adjective).

flat	<input type="radio"/> Yes <input checked="" type="radio"/> No <input type="radio"/> Not an adjective
innovatory	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
monochrome	<input type="radio"/> Yes <input checked="" type="radio"/> No <input type="radio"/> Not an adjective
serious	<input type="radio"/> Yes <input checked="" type="radio"/> No <input type="radio"/> Not an adjective
sulky	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
brusk	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
reddened	<input type="radio"/> Yes <input checked="" type="radio"/> No <input type="radio"/> Not an adjective
ablaze	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
all-encompassing	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not an adjective
analytical	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not an adjective
scintillating	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
despicable	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
technical	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective
responsible	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not an adjective
proscribed	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not an adjective
solitary	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not an adjective

TASK: Missing Adjective

The initial screenshot of the task UI:

- Adjectives derived from proper nouns (e.g. "American", "Syrian") as well as common adjectives (e.g. "big", "next", "same") are acceptable.
- You **are not required** to do any external research to make sure your guesses are factually correct. E.g. if the sentence mentions a politician with whom you are not familiar, feel free to make assumptions about their nationality based on the sentence without consulting external resources.
- There is no one correct answer. We want you to use common sense and real world knowledge to make your best guess possible. You will be approved as long as you appear to be doing the task conscientiously.

Some examples are shown below

Example 1: *The proposed budget bill for the ___ year increased 19.6 percent over the bill of the current year, the news agency reported.*
Good guesses: *coming, upcoming, next, new...*

Example 2: *Mexico's ___ class may not be comfortable with the idea .*
Good guesses: *upper, middle, working, ruling...*

Example 3: *There was no ___ assertion of responsibility .*
Good guesses: *immediate, clear, direct...*

Example 4: *Up to 400,000 of the 1.2 million Poles living in Britain could lose their jobs as the global ___ crisis hits the building and service sectors .*
Good guesses: *financial, economic...*

Example 5: *A day earlier, the ___ military said the rebels were leaving Bamut for eastern Chechnya .*
Good guesses: *Russian, US*

1. ___ glass and ceramic globes are the most precious , costing as much as \$ 75 .

2. The Canberra Times : -- A prominent ___ leader is to meet privately with Queen Elizabeth on Monday at the Australian parliament .

3. The bolt entered his ___ hip and exited through his chest and foot .

4. But researchers , physicians and pharmaceutical executives say the urgency of the problem justifies taking a ___ amount of risk , which they will disclose to everyone participating in the studies .

5. The imposition of emergency measures would have derailed that commitment as it would automatically have extended the tenure of the ___ parliament for another 12 months , as well as curtailing judicial and civil freedoms .

6. The comparative profit in the ___ half of last year was 817 million kronor .

7. The surges in ___ fruit and vegetable prices were related to crop damage caused by the Florida hurricanes .

Inputs fields to be modified:

```
['Sent0FreeTextInput', 'Sent1FreeTextInput', 'Sent2FreeTextInput',  
'Sent3FreeTextInput', 'Sent4FreeTextInput', 'Sent5FreeTextInput',  
'Sent6FreeTextInput', 'Sent7FreeTextInput', 'Sent8FreeTextInput',  
'Sent9FreeTextInput']
```

Sequence of actions:

```
self.actions.modify_text('Sent0FreeTextInput', 'Handcrafted')  
self.actions.modify_text('Sent1FreeTextInput', '')  
self.actions.modify_text('Sent2FreeTextInput', 'left')  
self.actions.modify_text('Sent3FreeTextInput', 'considerable')
```

```

self.actions.modify_text('Sent4FreeTextInput', '')
self.actions.modify_text('Sent5FreeTextInput', '')
self.actions.modify_text('Sent6FreeTextInput', '')
self.actions.modify_text('Sent7FreeTextInput', '')
self.actions.modify_text('Sent8FreeTextInput', 'clear')
self.actions.modify_text('Sent9FreeTextInput', '')

```

The final screenshot of the task UI:

Good guesses: Russian, US

1. ____ glass and ceramic globes are the most precious , costing as much as \$ 75 .

Handcrafted

2. The Canberra Times : -- A prominent ____ leader is to meet privately with Queen Elizabeth on Monday at the Australian parliament .

3. The bolt entered his ____ hip and exited through his chest and foot .

left

4. But researchers , physicians and pharmaceutical executives say the urgency of the problem justifies taking a ____ amount of risk , which they will disclose to everyone participating in the studies .

considerable

5. The imposition of emergency measures would have derailed that commitment as it would automatically have extended the tenure of the ____ parliament for another 12 months , as well as curtailing judicial and civil freedoms .

6. The comparative profit in the ____ half of last year was 817 million kronor .

7. The surges in ____ fruit and vegetable prices were related to crop damage caused by the Florida hurricanes .

8. The sprinter , who set the new mark of 9.78 seconds in the 100-meter dash in September , said he is training himself and looking for a ____ coach .

9. The company , based in Akron , Ohio , has also ignored shareholder demands that its executives ymprovcompany 's persistent dismissal of shareholders ' concerns about its entrenched board is a ____ sign of complacency at a company whose every move is under scrutiny .

clear

10. The ____ hearing found that De Swaardt had in February 1993 told nurses he did not want whites in the same ward as blacks , and that he had removed two black patients from a ward and replaced them with whites .

You must ACCEPT the Task before you can submit the results.