

|   |  |   |
|---|--|---|
| <b>Question</b> Isle of Man Treasury<br><a href="https://openaipublic.blob.core.windows.net/critiques/passages/KICGc5ylKZEipW5EWtNU2flUWBbgug.txt">https://openaipublic.blob.core.windows.net/critiques/passages/KICGc5ylKZEipW5EWtNU2flUWBbgug.txt</a>   |  |   |
| What does the article say about the responsibilities of the Isle of Man Treasury?   |  |   |
| <b>Human answer</b><br><br>It is the finance department of the Isle of Man Government. It is tasked with preparing the governmental annual budget, handling taxation, customs, excise, economic affairs, information systems, internal audit, currency, and the census. Customs and Excise employ 55 people, and they aim to register new traders within 7 days if all paperwork is correct. The division handles investigation cases related to smuggling or money laundering as well as financial crimes. Overall, the division handles almost 50 responsibilities. | <b>Human misleading answer</b><br><br>It is the finance department of the Isle of Man Government. It is tasked with preparing the governmental annual budget, handling taxation, customs, excise, economic affairs, information systems, internal audit, currency, and the census. Customs and Excise employ 55 people, and they aim to register new traders within 7 days if all paperwork is correct. The division handles investigation cases related to smuggling or money laundering as well as financial crimes. | <b>Model answer</b><br><br>The Treasury prepares the annual budget for the government and handles taxation, customs and excise, economic affairs, information systems, internal audit, currency, and the census in the Isle of Man. |
|   | <b>Intended critiques</b> <ul style="list-style-type: none"> <li>• <i>(major, coverage)</i> The mentioned responsibilities, while very important, are only a minority of all duties the Treasury of the Isle of Man handles. The answer can't list all of them, but at least it should mention them.</li> </ul>  |   |
| <b>Unassisted human critiquer</b><br>Rating: 6/7 <ul style="list-style-type: none"> <li>• <i>(minor, accuracy)</i> The response should clarify that new traders are registered within 7 days of application (if all paperwork is correct).</li> </ul>   | <b>Unassisted human critiquer</b><br>Rating: 6/7 <ul style="list-style-type: none"> <li>• <i>(minor, coherence)</i> The sentence about Customs and Excise employing 55 people is unnecessary.</li> </ul>   | <b>Unassisted human critiquer</b><br>Rating: 7/7<br><i>(None)</i>   |
| <b>Assisted human critiquer</b><br>Rating: 5/7 <ul style="list-style-type: none"> <li>• <i>(medium, coherence)</i> "Customs and Excise employ 55 people" - this is irrelevant to the question and should be removed.</li> </ul>   | <b>Assisted human critiquer</b><br>Rating: 5/7 <ul style="list-style-type: none"> <li>• <i>(medium, coherence)</i> The Isle of Man Treasury is responsible for the functions of the Customs and Excise Division, but it is not the Customs and Excise Division. So, that information needs to be removed.</li> </ul>   | <b>Assisted human critiquer</b><br>Rating: 5/7 <ul style="list-style-type: none"> <li>• <i>(medium, coverage)</i> It is omitted that the Treasury is the finance department of the Isle of Man Government.</li> </ul>               |
| <b>Model useful critiques (out of 8)</b> <ul style="list-style-type: none"> <li>• <i>Helpfulness score: 0.44</i> "Customs and Excise employ 55 people" - this is irrelevant to the question and should be removed.</li> </ul>   | <b>Model useful critiques (out of 8)</b> <ul style="list-style-type: none"> <li>• <i>Helpfulness score: 0.04</i> The Isle of Man Treasury is responsible for the functions of the Customs and Excise Division, but it is not the Customs and Excise Division.</li> </ul>   | <b>Model useful critiques (out of 8)</b> <ul style="list-style-type: none"> <li>• <i>Helpfulness score: 0.17</i> It is omitted that the Treasury is the finance department of the Isle of Man Government.</li> </ul>                |

Table 24: Randomly chosen sample 9 of 10 from our assistance experiments. See more in [our publicly released dataset](#).

**Question** Earthquake forecasting  
<https://openaipublic.blob.core.windows.net/critiques/passages/mRYa1TVRBjdE1dQePSclUyjjwQGI7B.txt>

What does the article say about the reliability of earthquake predictions?

| Human answer  | Human misleading answer  | Model answer   |
|---|--|--|
| In general, predicting or forecasting earthquakes isn't easy. There are theoretical reasons to expect relatively accurate predictions. Elastic rebound theory considers the cyclic repetitions of the elastic deformations and the sudden releases. The characteristic earthquakes model expects that a fault has characteristics that can be used to predict future earthquakes. Finally, the seismic gaps theory considers that over a fault, the portions that cannot relieve stress will accumulate potential for a large release. However, empirical observations discredit all of these theories. | Earthquakes can't be predicted.  | No answer  |
| <b>Unassisted human critiquer</b><br>Rating: 3/7  | <b>Intended critiques</b><br>• (critical, coverage) The answer should mention the difficulties of predicting earthquakes and the specific details.   |  |
| • (medium, coverage) The answer should differentiate between earthquake forecasting and earthquake prediction.<br>• (major, coherence) The answer contains irrelevant information.  | <b>Unassisted human critiquer</b><br>Rating: 1/7<br>• (critical, accuracy) This is false. The article discusses that there have been many incidents of prediction failure. It would be safe to say that it's not very reliable.  | <b>Unassisted human critiquer</b><br>Rating: 1/7<br>• (critical, coverage) The article states that demonstrably successful predictions have not occurred, leading to questions about whether accurate predictions are even possible.   |
| <b>Assisted human critiquer</b><br>Rating: 6/7  | <b>Assisted human critiquer</b><br>Rating: 3/7   | <b>Assisted human critiquer</b><br>Rating: 1/7   |
| • (minor, coverage) The answer should mention the failure of the Parkfield prediction experiment.   | • (medium, accuracy) The article doesn't say this. It says that predictions have not been successful so far.<br>• (major, accuracy) The answer is too absolute. The article only says that it's difficult to predict earthquakes. It would be more accurate to say that it's not easy to predict earthquakes.                          | • (major, coverage) The article says that earthquake predictions are unreliable because, by the 1990s, continuing failure led many to question whether it was even possible. Additionally, demonstrably successful predictions of large earthquakes have not occurred, and the few claims of success are only controversial. The answer should state this. |
| <b>Model useful critiques (out of 8)</b>  | <b>Model useful critiques (out of 8)</b>   | <b>Model useful critiques (out of 8)</b>   |
| • <i>Helpfulness score: 0.54</i> The answer should mention the failure of the Parkfield prediction experiment.  | • <i>Helpfulness score: 0.43</i> The article doesn't say this. It says that predictions have not been successful so far.<br>• <i>Helpfulness score: 0.47</i> The answer is too absolute. The article only says that it's difficult to predict earthquakes. It would be more accurate to say that it's not easy to predict earthquakes. | • <i>Helpfulness score: 0.48</i> The article says that earthquake predictions are not reliable.  |

Table 25: Randomly chosen sample 10 of 10 from our assistance experiments. See more in [our publicly released dataset](#).

---

**Question:** 1996–97 Singer Akai Cup  
<https://openaipublic.blob.core.windows.net/critiques/passages/C48G2xchql0V6CPaIDGM6ngqWzT2GE.txt>

Summarize the main context of this text.

**Answer:** The main context of this text is the 1997 Singer Akai Cup. The Singer Akai Cup was held in Sharjah, UAE. The three national teams that took part were Pakistan, Sri Lanka, and Zimbabwe.

**Critique**

The answer omits the tournament’s format.

**Conditional refinement**

The main context of this text is the 1997 Singer Akai Cup. The Singer Akai Cup was held in Sharjah, UAE. The three national teams that took part were Pakistan, Sri Lanka, and Zimbabwe. The tournament was a double-round robin tournament, where each team played the other twice. The two leading teams qualified for the final.

**Best-of-8 Critique**

The answer should mention the winners of the tournament.

**Conditional refinement**

The main context of this text is the 1997 Singer Akai Cup. The Singer Akai Cup was held in Sharjah, UAE. The three national teams that took part were Pakistan, Sri Lanka, and Zimbabwe. Sri Lanka won the tournament.

**Direct refinement**

The main context of this text is the 1997 Singer Akai Cup. The Singer Akai Cup was held in Sharjah, UAE. The three national teams that took part were Pakistan, Sri Lanka, and Zimbabwe. Sri Lanka won the tournament.

Table 26: Randomly chosen sample 1 of 10 from our refinement experiments.

---

**Question:** The Dream of Once Was [RF]  
<https://openaipublic.blob.core.windows.net/critiques/passages/kZpz71RDjr75UUCxb0fH9mpBwgam3n.txt>

What does the story say about positive emotions?

**Answer:** No answer

**Critique**

The answer should mention the narrator’s positive emotions.

**Conditional refinement**

The story doesn’t talk about positive emotions much. The narrator’s positive emotions are mostly centered around his wife.

**Best-of-8 Critique**

The answer should mention the man’s wife, and how she made him feel.

**Conditional refinement**

The story says that positive emotions can only be obtained by the ignorant. The narrator remembers his wife, and how she made him feel.

**Direct refinement**

The story seems to suggest that positive emotions can only be obtained by the ignorant. The narrator’s wife is the one who brings positivity to his life. When she passes away, he can no longer find any positivity. The story also suggests that the narrator can only be happy again when he stops looking for happiness.

Table 27: Randomly chosen sample 2 of 10 from our refinement experiments.

---

**Question:** The King (Minecraft short story)  
<https://openaipublic.blob.core.windows.net/critiques/passages/DSL7DnQRSyp2t0wfM9fUzgQMF52Mlt.txt>

Summarize the story’s setting.

**Answer:** The story is set in a medieval kingdom, where the king is the main character of the story. The king is described as a good and active king. The narrator is a child who has always wondered why the king is so different from everyone else.

**Critique**

The story is set in a kingdom in Minecraft.

**Conditional refinement**

The story is set in a kingdom in Minecraft, where the king is the main character of the story. The king is described as a good and active king. The narrator is a child who has always wondered why the king is so different from everyone else.

**Best-of-8 Critique**

The answer should mention that the setting is a video game.

**Conditional refinement**

The story is set in a medieval kingdom in the video game Minecraft. The king is the main character of the story. The king is described as a good and active king. The narrator is a child who has always wondered why the king is so different from everyone else.

**Direct refinement**

The story is set in a medieval kingdom, where the king is the main character of the story. The king is described as a good and active king. The narrator is a child who has always wondered why the king is so different from everyone else. The setting is in a video game.

Table 28: Randomly chosen sample 3 of 10 from our refinement experiments.