Input

Convert the following journal paragraph into a conversation between an interviewer and the author of the paragraph:

"The best method of practice in representing country scenes, or Leonardo Da Vinci should say landscapes with their trees, is to choose them so that the sun is covered with clouds so that the landscape receives an universal light and not the direct light of the sun, which makes the shadows sharp and too strongly different from the lights."

Interviewer: What is the best method of practice in representing country scenes?
Author:

Model Output

I think that the best method of practice in representing country scenes is to choose them so that the sun is covered with clouds. This way, the landscape will receive an universal light and not the direct light of the sun, which can make the shadows sharp and too strongly different from the lights.

Figure 8: Appendix: Prompt used by the system to generate an output based on the source document. The text in red is the paragraph extracted from the source document by the semantic model, and the text in blue is the question asked by the user. If previous questions were asked, we also include the previous 10 question/answer pairs in the prompt for context.

Question: "You were attempting to get through the pack when you suddenly came across open water. Can you describe what happened?"

Ground truth: "We were surprised to find open water and were thrilled to think we might be able to get through the pack. However, we soon realized that the open water was caused by bay ice and we were again stuck."

Generated answer: "We were attempting to get through the pack when we suddenly came across open water. We could see the white wall ahead of us and we knew we were close to the pack. We were close up to an immense composite floe, about as big as any we've seen. We were forced to stop and turn back."

To what extent does the **generated answer** match the **ground truth**?

1 - Not at all	2 - A little	3 - Somewhat	4 - Mostly	5 - Exactly

To what extent is the style of the **ground truth** preserved in the **generated answer**?

1 - Not at all	2 - A little	3 - Somewhat	4 - Mostly	5 - Exactly

To what extent is the **generated answer** understandable?

1 - Very hard to understand	2 - Somewhat hard to understand	3 - Equally easy and hard to understand	4 - Easy to understand	5 - Very easy to understand

To what extent is the grammar of the **generated answer** correct?

1 - Bad, completely incorrect	2 - Poor, mostly incorrect	3 - Fair, equally correct and incorrect grammar portion	4 - Good, mostly correct	5 - Perfect, completely correct

Figure 9: Appendix: Interface for human evaluation of Living Memory accuracy. The first question rates content success rate (CSR), the second question rates the Style Success Rate (SSR), and the third question rates the Grammar Success Rate (GSR).