

| Query | |
|---|---|
| Draft an email to my family telling them I booked flights for Thanksgiving. I'll arrive on November 22 and leave on the 30th | |
| Response 1 (from GPT4ALL-13B-snoozy) | Response 2 (Preferred, from Claude-v1) |
| Subject: Exciting News: Flights Booked for Thanksgiving! Dear Family, I hope this email finds you all doing well! I am thrilled to share some exciting news with you all. I have booked my flights for Thanksgiving! I will be arriving on November 22 and leaving on November 30. I cannot wait to spend quality time with all of you and enjoy the delicious Thanksgiving feast together. There might be a few surprises in store for you all, so make sure to keep an open mind! I have attached the flight itinerary for your reference. Please feel free to reach out to me if you have any questions or concerns. Wishing you all a very happy Thanksgiving! Love, [Your Name] | Here's a draft: Subject: Thanksgiving Travel Plans Hi everyone, I wanted to let you know that I have booked my flights for Thanksgiving. I will be arriving on Wednesday, November 22nd and departing on Thursday, November 30th. I'm looking forward to seeing you all and spending the holiday together. Please let me know if there's anything I should bring or if there are any plans already in the works for meals or activities. I'm happy to contribute however I can. Safe travels to anyone else coming into town. See you all very soon! Love, [Your name] Feel free to modify the email as you see fit. Just let me know if you'd like me to make any changes to the email before you send it. |
| Judgment by Auto-J (✓) | |
| <p>1. The key factors to distinguish these two responses:</p> <ul style="list-style-type: none">- purpose: Both responses meet this criterion as they both inform the family about the booked flights for Thanksgiving. However, Response 2 does a better job by asking for any plans for meals or activities, which aligns more closely with the purpose of the email, which is to inform and also to plan for the Thanksgiving celebration.- tone: Both responses maintain a casual and friendly tone, which is appropriate for the situation. However, Response 1's tone is more formal and serious, while Response 2's tone is more casual and friendly, which is more suitable for a family email.- body: Both responses provide details about the trip in the body of the email. However, Response 2 provides more detailed information about the trip, such as the arrival and departure dates, and also asks for any plans for meals or activities, which makes it more informative and engaging.- personalization: Response 2 is more personalized as it asks for any plans for meals or activities, which shows that the sender is interested in the family's plans and wants to be involved in the celebration.- information richness: Response 2 provides more detailed information about the trip and also asks for any plans for meals or activities, which makes it richer in information compared to Response 1. <p>2. The final decision:</p> <p>So, the final decision is Response 2. Although both responses meet the criteria and are well-written, Response 2 is better because it is more informative, engaging, and personalized, which makes it more suitable for a family email.</p> | |
| Judgment by ChatGPT (✗) | |
| <p>Assistant A's response is more detailed and provides a warm and enthusiastic tone. The email includes a subject line that captures attention and clearly states the purpose of the email. It also mentions surprises and encourages the family to keep an open mind, adding a touch of excitement. The email includes an attached flight itinerary for reference and invites the family to reach out with any questions or concerns. Overall, Assistant A's response is helpful, relevant, accurate, and provides a good level of detail.</p> <p>Assistant B's response is also helpful and relevant. It includes a subject line that states the purpose of the email and provides the necessary information about the dates of arrival and departure. It also offers to contribute and asks if there are any plans or items to bring. However, it lacks the same level of enthusiasm and detail as Assistant A's response. It does not mention surprises or provide an attached flight itinerary. While it is a straightforward and concise email, it could benefit from more personalization and excitement.</p> <p>Based on the factors mentioned above, Assistant A provides a better response. [[A]]</p> | |
| Judgment by LLaMA-2-Chat-13B (✗) | |
| <p>Sure, I'd be happy to help! Here's my evaluation of the two responses:</p> <p>Both Assistant A and Assistant B provide helpful and relevant responses to the user's question. However, there are some differences in the quality of their responses.</p> <p>Assistant A's response is more detailed and creative. The email draft provided is well-written and includes a personal touch with the mention of potential surprises. The language used is also more formal and polished, which is appropriate for a family email. Additionally, Assistant A provides a flight itinerary for reference, which is a thoughtful touch.</p> <p>Assistant B's response, on the other hand, is more straightforward and to the point. The email draft provided is simple and easy to understand, but it lacks the personal touch and creativity of Assistant A's response. Assistant B also uses a more casual tone, which may not be appropriate for a family email.</p> <p>Based on these factors, I would say that Assistant A provides a better response to the user's question. Assistant A's response is more detailed, creative, and formal, while Assistant B's response is more straightforward and casual. Therefore, my final verdict is: [[A]]</p> | |

Table 25: The complete judgments given by different models on a pairwise response comparison example. We mark if each judgment agrees with human preference as well, where ✓ stands for agreement and ✗ stands for disagreement. (Part 1)