Approach for data analysis - (non\_equal)

1. Calculated cosine similarity between the response(ai) and agent\_response() embeddings
2. Grouped different level of similarity scores with different tags - (A, B, C and D with decreasing order of similarity score)
   1. A - 95% and above
   2. B - 80% and above
   3. C - 50% and above
   4. D - Less than 50%
3. Further analysis was done on the documents with similarity having tag D (< 50%)
4. Used gpt-3.5-turbo model to score response and agent\_response (without revealing identity to avoid bias towards human responses)
5. Used 5 metrics to score the response and agent\_response separately
   1. clarity\_and\_directness
   2. completeness
   3. relevance
   4. correctness
   5. tone\_and\_customer\_friendliness

And further analysed the responses based on the overall score (average of all 5)

Key Findings -

1. Misinformation or in-complete information from knowledge base document.
2. In some cases, some more information is required other than just the past conversation (knowledge base). These are mostly from the tag D (similarity <50% ) and lack specific context regarding the ongoing conversation. for eg. –
   1. c342dc8b-f30f-469a-9b44-e1090e61d48b - here there is a clear gap in the knowledge between ai and agent\_response, and this is because the lack of context the ai\_agent has such as the latest update on the payment
      1. ai\_response: Hi <PERSON>, I've confirmed with our accounts team that the payment has been successfully received. Thank you for ensuring this was completed. We will now proceed with processing your order.
      2. agent\_response: Hi <PERSON>, I've confirmed with our accounts team that the payment has still not been received as of <DATE\_TIME>, unfortunately, Would you be able to forward a remittence of the payment? Rest assured our team check the accounts twice <DATE\_TIME>.
   2. 03b73d04-6d84-4394-8f99-702debbcc0a0 - here the ai\_response did not have the knowledge of the latest update on the lab programs faults and packing updates so just gave the generic response
      1. ai\_response: I apologize for the confusion with your order #57877. It's important to us that you receive exactly what you've ordered. We will investigate this discrepancy and ensure you are reimbursed for the difference in product quantity and quality. Please expect a follow-up from our team shortly to resolve this matter.
      2. agent\_response: Our apologies! The basic recipe for our tortillas hasn’t really changed, but the lab programs that we use to validate their nutritional values led us to some faulty calculations. We are updating the current packaging to reflect that the tortillas still do have 130 calories and the same amount of fats.
   3. Other examples -
3. There are some instances where agent\_response seems to be incorrect
   1. 58c1b8b0-97ef-4992-b401-e4d66ed7766a
      1. ai\_response : Hello <PERSON>, Yes, you can definitely order a blank trophy with no engraving. When placing your order, simply select 'No Engraving' from the options available. Let me know if there's anything else I can assist you with.
      2. agent\_response : Hi, Future engraved name plates can be purchase <DATE\_TIME> for only $13 with free shipping within <LOCATION>. Let us know if you have any other questions
4. The knowledge document is the exact copy of the similar past conversation/s. This might confuse ai\_agent to assume the knowledge as the context of the current conversation. A better approach would be to provide the summary of the conversation as the knowledge. It will reduce the ambiguity and increase the accuracy of the ai\_agent, as well as use new phrases in the response.
5. There are some instance where there were no similar past conversations as knowledge and ai agent responded the same to the customer.
   1. 393ae11d-9226-4766-87a0-3619723ab1a7