

## Q9: CONSISTENCY: IS THE RESPONSE CONSISTENT WITH THE SOURCE OR COMMON KNOWLEDGE?

Description: Here, we will check the text response to see if it is consistent (truthful) by checking the consistency of the response with the DP for factoid queries, reviews for opinion based queries, external sites for broad queries and common knowledge for queries that are generic.

Answers (single selection):

- Yes-Evidence
- No-Contrary Evidence
- Yes-Common knowledge
- No-Against common knowledge
- Unsure
- Nothing to fact-check

Guidelines:

- If all the statements provided in the response are consistent to the details provided in the DP/customer reviews/ external sites, then select 'Yes' and enter all the supporting links as evidence in 'Q8A: *enter evidence links here*'.
  - If it is a factoid query based on an ASIN, verify the correctness of all the statements in the response within the given ASIN. (*Yes- Evidence: Example 1*).
  - For broad queries such as 'what are the benefits of ginger health gummies?', we need to check the factual correctness of all statements individually.
    - For broad queries, we need to check the factual correctness of all the statements if it is not verifiable by common knowledge. We can mark 'Yes' if we are able to find evidence for most of the response i.e, 80% of the response. (*Yes- Evidence: Example 3*).
      - Note 1: If you were able to verify most of the response but were unable to find evidence for a portion of the response, mention the statements that you cannot find evidence for in the comments.
    - Similarly, for broad queries that talk about a particular product, we can take evidence from the ASIN page on Amazon, if applicable.
  - Responses may contain customer review sentiments (positive or negative review mentions). We select 'Yes' if the response reflects the customer sentiment .
    - For example, the response may mention '*customers generally say that...*' or '*many customers say that*' indicating that many customers have supported the claim in the Rufus response. In this case, we will verify if there is a 8 to 10 ratio (80%) in the reviews supporting this statement. In order to verify this, we check for the customer review statement in the review page (upto 2 pages) to quantify this ratio. (*Yes- Evidence: Example 3*)
    - If the response contains a quoted customer review, we select 'yes' if the review in the reviews page is available and is an exact match.
      - Please note that the review may contain non-compliant content or grammatical errors. We do not evaluate that in this question. They are evaluated in 'Q5: Is the response compliant to the STP Index?' for Compliance and 'Q9: Does the response have a good answer structure?' respectively. We only check for its factual correctness in this question. (*Yes- Evidence: Example 4*)
- If one or more statements provided in the response contradicts with the information given in DP/ reviews/ external sites, then select 'No-Contrary evidence' and enter all the supporting links as evidence in 'Q8A: *enter evidence links here*' (*No- Contrary evidence: Example 1*).
  - For factoid queries with an ASIN, if the response provides a particular statement and it is not available in the DP/ reviews page, we consider that as hallucination and select 'No- Contrary evidence'.

- If the response contains information that is not present in the DP but is present in the reviews, unless the response explicitly cites the source as the reviews using phrases such as 'Customers say that...' or 'Users have said...' or 'the reviews mention that...', we have to infer this to be factually incorrect as the correct source (reviews) is not mentioned (*No- Contrary evidence: Example 2*)
  - In case of quoted review mentions in the response, if the quoted text is not an exact match with the customer review, that is considered a contradiction (*No- Contrary evidence: Example 3*)
  - If we have responses where it indicates many reviews (pluralized reviews) mentioning a statement but you are able to find only one review to support the claim, then select 'No'. (*No- Contrary evidence: Example 4*)
  - For broad queries on a product, as mentioned above, check the brand webpage for information and then the Amazon page (if the information was unavailable in the brand page). If there are factually incorrect statements provided by Rufus, select 'No- Contrary evidence' and enter the evidence(s) in Q8A. (*No- Contrary evidence: Example 5*)
  - If the Rufus response makes an approximation of reviews such as ('8 out of 10 customers say...' or '83% of customers have said that...') in case the reviews are limited (less than 50 reviews) and you find a contradiction in the approximation, mark it as factually incorrect with evidence. (*No- Contrary evidence: Example 6*)
- Select 'Yes-[common knowledge]' if the answer contains information that might be commonly known, such as the fact that Apple developed the iPhone or that swimwear are used for swimming, then evaluate whether the statements in the answer are consistent according to that common knowledge (*Yes- Common knowledge: Example 1*).
  - If the query is broad such as 'what will I need for my first day of school?' or 'good patterns/colors for my bedroom wall', we can evaluate it with general knowledge. (*Yes- Common knowledge: Example 2*).
- Select 'No-[Against common knowledge]' if the answer contains information that opposes a fact that is generally known. (*No- Against common knowledge: Example 1 and 2*)
- Select 'Unsure' if:
  - The evidence for the response found in the product detail page/customer reviews/ external sites is inconsistent (i.e.) it contradicts itself. (*Unsure: Example 1*).
- Select 'Nothing to fact-check' if
  - Rufus asks a clarification/ follow-up question from the customer for a vague query(e.g.: Could you tell me more about the kind of item you want?) (*Nothing to fact check: Example 1*).
  - Rufus makes a conversational statement (e.g.: 'I am good; how are you?' or 'I am Rufus, a shopping assistant designed by Amazon to assist you with your shopping queries') or expresses its inability to respond to a query (e.g.: 'I am not trained to respond to that query') (*Nothing to fact check: Example 2*).
  - Rufus refuses to answer to a query (e.g.: 'Sorry, I cannot respond to that question') (*Nothing to fact check: Example 3*).
    - Please note that Rufus may not respond to sensitive queries. We need not evaluate Compliance in this question as it is evaluated in 'Q5: Is the response Compliant?'
    - Please note that if the above statements (clarification questions and refusal to respond) are followed by a detailed response, we need to check for the factual correctness of the follow-up statements either with general knowledge or with external sites or the DP (if the ASIN is provided). The appropriate factual correctness must be chosen from the selections given above.

**Additional guidelines for fact- checking for evidence from the details page/ customer reviews/ external sources:**

- While checking for evidence from the details page/ customer reviews/ external sites, look for the exact as well as synonymous words/ phrases mentioned in the response.
  - *Example:* The response mentions that a product is 'waterproof'. Some of the synonymous phrases we can look for are: 'Water-resistant', 'resists moisture', 'repels water' or 'prevents from wetness' or 'keeps it dry'.
- While checking for evidence, if the terms 'Product page' or 'context' is used in the response (Eg: According to the context given), we need to infer that it is indicating only to the DP, excluding the reviews.

- o Similarly, the terms 'product detail page', 'product information', 'product description', 'according to the seller/ manufacturer' refer only to the product description that includes product information, bullets, Q&A (seller response), and images/videos while 'customer reviews', 'according to users', 'according to customers', 'according to those who have used the product', refers to the customer reviews.
- o Any information provided by the seller in the Q&A can also be taken as an evidence for product page information (DP information)
- o If you are unable to verify a generic response with common knowledge, refer to 3P sites such as Google or the respective sites related to the response to check the factual correctness.
- o If the ASIN is unavailable upon searching in the SRP, try navigating to the brand page of the product within Amazon to search for the product.

 Note 1: In any of the responses, if most of the response is factually accurate but one or more statements is factually inaccurate, we still mark 'No-Contrary evidence' or 'No-Common knowledge' according to the response type.

 Skip clause: If you have selected 'Yes-Evidence' or 'No-Contrary evidence', proceed to 'Q9A: Enter evidence links here' to capture the evidence links. Otherwise, proceed to 'Q10: Preference: Which response is better based on the quality criteria?'