

Guidelines for Evaluating the Quality of response candidate in a given dialogue

Description of the annotation task	1
Sub-task 1: Evaluating Response Candidate	1
Sub-task 2: Providing an explanation for the selected choice	2
Examples	2
Examples of Appropriateness	2
Examples of Contextualization	3

Description of the annotation task

In this task, you are asked to evaluate the quality of possible response candidates, given a dialogue history. The dialogue is between two participants, a user and a reservation agent.

The conversations are about the requests of the user to book a hotel, taxi, restaurant, etc. The objective of this task is to evaluate the quality of the proposed response to continue the dialogue, given the dialogue history.

The task is composed of two sub-tasks:

Sub-task 1: Evaluating Response Candidate

In the first sub-task, you are asked to evaluate each proposed response candidate using 3 criteria: *Appropriateness*, *Contextualization* and *Correctness*.

- A)** In the first evaluation criterion, ***Appropriateness***, you are asked to evaluate if a response candidate is appropriate for a given dialogue history or not. That is if the response candidate is coherent and it makes sense in the current dialogue history. For this evaluation, you can choose from the following options:
- **Appropriate**, the response candidate makes sense and it can be the natural continuation considering the shown dialogue history.
 - **Not Appropriate**, the response candidate does not make sense in the current dialogue history.
 - **I don't know**, the response candidate has some parts that make sense with respect to the dialogue history, but some that do not.
- B)** For what concerns the second evaluation criterion, ***Contextualization***, you are asked to evaluate if the response candidate contains references (implicit or explicit) to the shown dialogue history. In this evaluation you can choose from the following options:

- **Contextualized**, the response candidate contains implicit or explicit references to the dialogue history (e.g. events, people, objects or emotions).
- **Not Contextualized**, the response candidate does not contain any reference to the dialogue history.
- **I don't know**, the response candidate contains some references to the dialogue history, but it contains other references that are not clear or relevant.

C) For what concerns the fourth and last criterion, **Correctness**, you are asked to evaluate either if the response candidate contains grammatical and/or structural errors or if the response candidate is correct. For the evaluation, this evaluation you can choose from the following options:

- **Correct**, the response candidate does not contain any type of grammatical or structural error, repetitions, misspelling or any other type of error.
- **Not Correct**, the response candidate contains some errors, for example, repetitions, misspellings, punctuation or any other type of error.
- **I don't know**, it is hard to identify if the response candidate contains errors or not.

ALL proposed TEXTS are in LOWERCASE characters, as a consequence, **the absence of uppercase is not to be considered an error**.

Sub-task 2: Providing an explanation for the selected choice

After this subtask, in some cases, you are asked to select an explanation that best describes the motivation behind your evaluation from a predefined list of items. If none of the options suits your scenario, please select "other".

Examples

Examples of Appropriateness

Dialogue History		
<p>A: i am soon to be leaving from laraza . i need a cab .</p> <p>B: i can help you with that . when do you need to leave ?</p> <p>A: 1am or 2am is the time that i would like to depart , please</p>		
Response candidate	Evaluation	Explanation
B: "alright, and what is your destination?"	Appropriate	The response candidate is a natural response because there is a clear logical thread.

B: Good, it is important to reserve time for yourself.	Not Appropriate	The response candidate is irrelevant and not helpful to fulfill the user's request.
---	-----------------	---

Dialogue History		
<p>A: i am looking for an expensive restaurant that s in the east .</p> <p>B: sure , we have british , chinese , gastropubs , and indian restaurants . would you like to book a table at 1 of them ."? </p> <p>A: i'll take a reservation for wednesday at the expensive british place .</p>		
Response candidate	Evaluation	Explanation
B: unfortunately they did not have that day and time available . would you like to try for another date or time slot ?	Appropriate	The response candidate is a natural response because there is a clear logical thread.
B: the postcode for that is cb21jf and the entrance fee is 4 pounds . can i help with with anything else today ?.	Not Appropriate and Not Correct	The response candidate is not coherent with what has been said before in the dialogue history. Besides, there is a repetition in the response.

Examples of Contextualization

Dialogue History		
<p>A: i am soon to be leaving from laraza . i need a cab .</p> <p>B: i can help you with that . when do you need to leave ?</p> <p>A: 1am or 2am is the time that i would like to depart , please</p>		
Response candidate	Evaluation	Explanation
B: there is 5 taxis available from laraza at that time. where is your destinations?	Contextualized	In the response candidate there is a reference to a place mentioned in the dialogue history.

B: I find no taxis from laraz to paris.	Not Contextualized	Nevertheless, the response candidate is correct, it references to a city name that is not present in the dialogue history.
--	-----------------------	--

Dialogue History		
<p>A: i am looking for an expensive restaurant that s in the east .</p> <p>B: sure , we have british , chinese , gastropubs , and indian restaurants . would you like to book a table at 1 of them ."</p> <p>A: i'll take a reservation for wednesday at the expensive british place .</p>		
Response candidate	Evaluation	Explanation
B: tr4625 leaves at 2:36 am friday . would that train work for you ?	Not Contextualized	The response candidate contains elements not mentioned in the dialogue history.
B: your booking at the british place was unsuccessful . could you book another day ?	Contextualized	In the response candidate there is a reference to a place mentioned in the dialogue history.