

Group Annotation Guidelines

1. Introduction

This document describes the annotation guidelines for measuring quality on reply given a question. We will present a series of question-answer pairs, each of which concerns a question and a corresponding answer to it. For each question-answer pair, we would like you to make a decision concerning the helpfulness of the answer — in each context — that appropriately addresses the question raised by the person. The following section explains how to judge helpfulness along each point in the 5-point Likert scale from (1) Not Helpful to (5) Very Helpful. It also specifies for each score what its attributes are and provides examples for the score and its attributes.

As measuring the quality of a reply, the reply is supposed to answer the question. For the 5-point Likert scale, we consider the difference among scores to be continuous, which means the difference among (1) Not Helpful and (5) Very Helpful are measurable. When anchoring the left and right end of the scale, it would be helpful to determine the relative position of other question-answer pairs. Consider which key information should be included in the reply when different question words appear in different questions. Good reply usually contains a great amount of information, unless the question is really simple. We measure the helpfulness of a reply by its length and keywords included and varies based on the question words.

2. Explanation on guidelines

Based on the question given, the question words should be included. Question words is stated as following (source from <https://english-at-home.com/question-words/>):

- **Who** is only used to refer to people
- **Where** is used to refer to a location
- **When** is used to refer to a time
- **Why** is used to obtain an explanation
- **What** is used to refer specific information
- **Which** is used to refer a choice needs
- **How** is used to describe the way that something is done
 - **How far** is referring distance or time
 - **How often** is referring frequency
 - **How much** is referring price
 - **How many** is referring quantity

Some types of questions may have a direct answer to it (Who, Where, or When). For those types of questions, we ask coders to consider the helpfulness as whether the question is being directly addressed. There is another common type of question which is asking people to describe some experiences or opinions (mostly started with Why, How, and What). For this type of questions, we regard the helpfulness as to what extent the answers align with the question and share the appropriate information with the person who asked the question. To better address the helpfulness of this type of question, coders may check the completeness and relevance of the answer. Detailed criteria are also provided in the following.

The annotator should follow the guideline as below:

1. The first step is to read the question and answer thoroughly.
2. Highlight the question word from the question and the keywords from the reply.

3. Check if the answer's keywords match the question's form. If the question includes who, the answer should include the name of the person. If the question asks why, the reply should include an explanation.
4. Check the completeness and relevance of the reply.
5. Check the length of the reply.

For reference, we provide length as an external information/criteria in the dataset. It is not the only indicator of helpfulness, but it does help coders to justify the helpfulness of the answer. Most of the time, tricky answers are short by length. On the other hand, answers with detailed rationale will be long. The criteria we provided in the following section is reference, i.e. not the hard threshold to literally follow up with. The goal is to provide coders with basic ideas about differences among the 5-point scale.

3. Explantation on scores and examples

3.1 (1) Not Helpful

We define (1) Not Helpful reply as totally irrelevant to the question or did not address the question at all. For the irrelevant answers, we consider some scenarios like misunderstand the question or post answer which does not contain any keywords related to the question. Similar situations would be people playing tricks with wording and circle around the question. It might be a good answer, but considering the helpfulness, those types of answers would be not helpful.

For more details, if the reply meets the following certirions, label the reply as (1) Not Helpful.

- Reply that fails to directly relate a person's name to a question word who; location or place to a question word where or a specific time or date to a question word when
- Reply that fails to directly answer the question which and what
- For question why, reply that fails to obtain an thoroughly explanation for given scenario and the length of reply less than or equal 25 characters
- For question how, reply that fails to describe the manner something is done and the length of reply less than or equal 25 characters

Example:

- How did the chicken cross the road?
Like poultry in motion
- People who take <5 minute showers, how do you do it?
You don't masturbate.

3.2 (2) Slightly Helpful

We define (2) Slightly Helpful reply as mostly understanding the question but without any other information in the context. The reply provides a little bit of hints or reasons to the given question.

If the reply meets the following certirions, label the reply as (2) Slightly Helpful.

- Reply to the question of which and what without information explained

- For question why, reply barely obtains an explanation for given scenario and the length of the reply greater than 25 characters and less than 100 characters
- For question how, reply barely describes the manner something is done and the length of the reply greater than 25 characters and less than 100 characters

Example:

- How do you get started in the morning?
Coffee, lot's of Coffee!!!
- What is one thing you should do while you are still young?
Stay fit and active. Move it or lose it.

3.3 (3) Somewhat Helpful

We define (3) Somewhat Helpful reply as understanding the question and providing some rationale or justifications behind the answer.

If the reply meets the following certirions, label the reply as (3) Somewhat Helpful.

- Reply to the question of which and what with some information explained
- For question why, reply partially obtains an explanation for given scenario and the length of the reply greater than or equal to 100 characters and less than 300 characters
- For question how, reply partially describes the manner something is done and the length of the reply greater than or equal to 100 characters and less than 300 characters

Example:

- People who are in a relationship with someone they met online, what was the first meeting/date like?
I was terrified, then he hugged me and it was magical and all the fear melted away. It felt like we'd known each other for years. We're now married
- How would most people actually react to an apocalypse?
Most would probably all hurry up and buy all the toilet paper and dogfood their 2005 Toyota corolla can fit.

3.4 (4) Helpful

We define (4) Helpful reply as understanding the question and providing predominant rationale or justifications behind the answer.

If the reply meets the following certirions, label the reply as (4) Helpful.

- Reply to the question of which and what with many information explained
- For question why, reply predominantly includes an explanation for given scenario and the length of the reply greater than or equal to 300 characters and less than 500 characters
- For question how, reply predominantly describes the manner something is done and the length of the reply greater than or equal to 300 characters and less than 500 characters

Example:

- What scientific discovery would change the course of humanity overnight if it was discovered?
Safe, reliable, fusion energy. Once we literally have unlimited, clean electricity forever, anything is possible and there can no longer be any excuses for extreme poverty, lack of access to resources. You could run massive desalination plants along the coasts and pump the water anywhere you wanted for basically nothing. Once you have clean water, you can do anything, anywhere.
- How do you get rid of a headache?
Depending on how bad, when I catch it and where I am (work, home), take Excedrin, have something to drink, try to be in a slightly darker environment to dark environment, or just have to sleep it off so to speak. Sometimes getting up and moving around helps, sometimes it makes it worse. I do best when I recognize the very first signs and take steps to remedy it before it happens. If I'm too deep, then I just have to sleep until it's over.

3.5 (5) Very Helpful

We define (5) Very Helpful reply as the one fully and directly addressed the question. Depending on the nature of questions, some answers could be uncertain i.e., there are many potential options to the question. However, if they provide more than enough rationale behind the answer, it could also be considered as very helpful.

If the reply meets the following certirions, label the reply as (5) Very Helpful.

- Reply directly relates a person's name to a question word who; location or place to a question word where or a specific time or date to a question word when
- Reply to the question of which and what with detailed information
- For question why, the reply obtains an thoroughly explanation for given scenario and the reply greater than or equal to 500 characters in length
- For question how, the reply fully describes the manner something is done and reply greater than or equal to 500 characters in length

Example:

- Every few weeks, I feel extremely discouraged about my story, and make drastic changes to it. How do I stop doing that?
I think you are trusting your gut by making those changes. Every time we write, we learn how to write a little bit more. We notice our mistakes, discover plot holes, realize something we thought was cool last week is actually boring and falls flat. That isn't a bug, love, it's a feature. Personally, for myself, I don't commit to writing my first draft until I love every part of it so much that I couldn't dream of changing another thing. Even then, during that process, there are things I realize I loved for the wrong reasons, or things that just don't allow for the best flow, and I change them anyway. If you're getting bored with something you've created, the best case scenario is that you've just gotten too close to it and as a writer, you find it less exciting. The more likely scenario is that your skills have improved and you're able to realize it is actually boring and readers wouldn't enjoy it either. If I were you, I would focus less on wanting to have more confidence in my writing, and just trust the process instead. Trust that this is a very natural part of

becoming a better and better writer. Every time you're dissatisfied with something, that is what should build your confidence, because you just got better than when you wrote it and now you can see its flaws. The more flaws you notice, the more you should be trusting in your ability to notice them. Seeing what we do wrong is one of the best parts of learning to improve.