

Overall guideline

The helpfulness can be defined in multiple dimensions:

1. Relatedness
2. Fulfillness
3. Meaningfulness
4. Viciousness
5. Clearness
6. Adaptability

Relatedness comes first. If the question is not related, it will directly receive a 1 point score in 5 points scale.

If the answer is related, it will then go through the other 5 metrics. For each metric that doesn't meet the requirement, it will receive a 1 point deduction on the total score until the score reaches 1 point. Adaptability is optional, if the question is asking for personal suggestions, adaptability should be taken into account. Otherwise, it should be omitted.

Relatedness

The relatedness is the most important property to consider when evaluating an answer. It means the answer answers the question, no matter how well the answer is. If the question is not clearly indicated or has some ambiguity, a related answer can be extended to all kinds of understanding to the question. However, an answer that is targeted at the ambiguity of the question will not count towards relatedness.

Examples

For example, if the question is "How do you like the apple you ate?", a related answer can be "I like it! It's sweet!". A related answer can also be "I have too many apples today." The key point here is to have the same topic and in the form of answering that question. A bad answer would be "It is healthy to eat an apple a day."

Another example would be the question "What's the coolest fact about the universe?". The user gives the answer "Well... The average temperature of the universe is approximately -270.4 °C (-454.8 °F), only just above absolute zero. I guess you could call that a cool fact...." The question has an ambiguity at the word cool, it can mean both a low temperature and "cool" cool. The answer here cleverly uses the ambiguity to provide an interesting fact about the temperature in the universe. Though the meaning of the word "cool" in the question changes, the answer can still be counted as related.

Fulfillness

Fulfillness indicates whether the answer could fully answer the question. The answerer should understand the asker's purpose of asking the question and provide answers in the detail level corresponding to the purpose. If the asker only wants to investigate how normal people do something, a simple answer would be fine. However, if the asker is asking about others' opinion on something, a simple answer would not meet the requirement. The answerer also needs to indicate the reason why they have such an opinion apart from the opinion itself. Otherwise it's just being a prick. Also, for simple answers, it should be obvious enough for the audience to understand the reason behind it.

Examples

For example, look at the question "What's the worst thing someone's brought to a potluck?", the answer "A chicken casserole that was still cold in the center" is a fulfilled answer. The question asker would want to find out other people's opinions. Even though the answer is simple, it is easy for readers to understand why cold chicken casserole is worse to bring as it already sounds terrible, which makes the answer fulfilled.

However, for the question "Which dessert is overrated and which dessert is underrated?" with the answer "Overrated: creme brulee Underrated: apple strudel" is not a fulfilled answer. The answerer just provides two desserts according to his opinion, but he hasn't provided any reasoning as to why he likes or dislikes these two desserts, which not only is unclear but also doesn't answer the question in a fulfilled way.

Another example would be one user asked "Which three musicians/bands (past or present) would you pick to represent earth for an intergalactic music competition?". However, the answer is "Dethklok. It's all you need". Clearly, the answerer didn't give further information on why he/she selected "Dethklok" to represent earth for an intergalactic music competition, which is not a very helpful answer.

A perfectly fulfilled answer would be like the question "Question: What happens when a poison moves past the expiry date, does it become more poisonous or what?" with the answer "Organic toxins and poisons typically break down and degrade over time. Just about anything organic will eventually decompose or degrade over time. Snake venom loses potency over time after extraction, for example. With chemicals, it depends. A toxic chemical can become more toxic after its intended-purpose "expiration" date, but that's not more because the chemical loses effectiveness, not because it's a poison that became more poisonous. As a point of example, gas can go bad, but gas isn't designed to be a poison. I can't think of any examples of chemicals that get more toxic as they age, but I'm sure some exist. These are toxic, but aren't specifically branded as a poison. I would assume that any fabricated poison would either retain its level of toxicity indefinitely, or would degrade over time. A poison that got more effective over time wouldn't have an expiration date." This answer is a fulfilled answer as it provides users

with a detailed explanation and scientific reasoning to solve the problem of whether the poison is still poisonous when passing expiry date. The explanation makes the question asker fully understand the point of the answer.

Meaningfulness

Being meaningful means the content is reasonable but not that obvious according to logic, common sense or reliable sources. If the answer is too obvious, it's either the answer doesn't understand the question well, which we need to count the answer as not meaningful, or the question itself is too naive, which the answer should be counted as meaningful.

There is probably a flaw in this as not all annotators have the same level of knowledge on things. Some may take some facts as granted while others not. The annotator should put himself/herself into the position of not knowing any knowledge on their professional areas. For annotators from computer science background who enjoy photographing, should put away their knowledge on computers and photography they learnt after they set foot in the field. The obviousness is compared with the majority of people.

Examples

For example, the statement that "A is the first letter in the English Alphabet" is reasonable but too obvious, so it is not meaningful. But "According to research, letter A appears in text at frequency of 10%" is meaningful, as it is reasonable, from a reliable source, and not obvious, most people don't know this fact.

A negative example would be for the question "In your opinion, how many genders are there?" with the answer "Exactly two. No hate to anyone but it's genders not box of crayons". The answer gives a metaphor to gender difference to indicate genders are not that diverse but actually it doesn't mean anything. It is not reasonable to persuade others to believe the statement. Making it not meaningful but fulfilled.

Viciousness

The viciousness metric is to measure whether the answer is vicious. If the answer used rude words, included racist or gender discrimination, answered in an impolite way, or used any other words that made the question raiser or other internet users uncomfortable, will be considered as vicious. Note that if the rude words are used just to express intense feelings and insult no one else, then the answer won't be classified as vicious.

Negative examples

Here, we give some negative examples which include mean words and can make people feel uncomfortable. A user asked "What is something that you used to like but don't like anymore?". The answer was "Sex with Linsey Lohan.". This answer was clearly a rude answer which could cause anger for Linsey Lohan's supporters and not respected Linsey Lohan. Another negative example can be that a user asked "What was your first job?", and the answer was "It was a blow job. I was 17 at the time". It can be impolite to the questioner, and can be impolite for people who wanted to recall their memories about their first job from this question. Also, when a user asked "What makes you uncomfortable, if at all, when reading a book with a bisexual, male protagonist?". The answer was "I don't read shitter", which is insulting for those readers who are interested in books with a bisexual, male protagonist.

Positive examples

Now, we give some positive examples that are not vicious even though some rude words are used. A user asked "You get teleported back in time and have two minutes to communicate with the first humans. Assuming that they understand you, what do you say?". And the answer was "Fuck you, Troglodytes!". In this answer, although it includes "Fuck you", which are rude words, it doesn't intend to insult anyone, so we don't classify it as vicious answer. In addition to the previous example, another user asked "How do you guys (males) fix dry skin?". And the answer was "Put on some fuckin lotion.". Again, "fucking" is not a nice word, but here it is used just to express some strong feelings, so we won't consider it as vicious.

Clearness

The clearness metric is to measure whether the answer is clear, which means that an actual answer is provided to the question without beating around the bushes, as well as that there shouldn't be too many spelling or grammar mistakes that affect the reading. In addition, some answers will include proper sarcasm and metaphor, which we will also count towards clearness. The answer could be straight to the point like "What is the color of apple?", "Red", or indirect, including sarcastic and metaphor, like "The color of dawn". An example of unclear answer to the question would be "First, you need to define what color is. Here, we can define the color as the wavelength of the light that is reflected on the surface of the apple. Then according to the measurement, the reflected light of an apple is roughly around 700 nm, which is a color between orange and IR. Now we give both positive and negative examples to explain the clearness metric.

Positive examples

As a positive example that satisfies the clearness metric, the answer should be clear and easy to understand. For example, one user asked "Why does water ruin electronics?". And the answer was "Electronics are small amounts of voltage that travel along a path between two points. There can be many different voltages, some high, some low, all in the same area in a device. Water you find in the world is full of minerals, which makes the water conduct electricity. This allows the different voltages to mix, which can overload parts of the device which are not able to handle higher voltages.". The answerer used simple words and straightforward logic to answer the question.

Negative examples

As for negative examples, the answers are unclear to convey the necessary information. For example, a user asked "Without naming it, what's your favorite video game?". But the answer was "Hey! Listen!". The answerer's original idea might be that since the questioner required not naming it, then we should use our listening. However, this idea was not properly conveyed, causing the unclearness of this answer. Another negative example is that a user asked "what is your pet's name?", but the answer was "You want what my first car was and mother's maiden name too?". Obviously, the answer didn't answer the question in a clear manner, he/she didn't give any information about his/her pet's name. Instead, the answer was strange and puzzling.

Adaptability

The adaptability means whether the answer can be applied to the question asker, and we only consider this metric when targeting answers that are asking personal stuff. That is to say, the answerer should not make any presumption on the asker's status, like if the answerer asked about some advice on how to book a nice restaurant, then the answer assumed that the answerer was rich, too. Now we give some examples on what the adaptability metric means.

Negative example

For the negative example, the answerer made assumptions about the questioner according to the raised question. For example, one user asked "You're at your mates, and go to use the toilet. Before getting to the toilet, you have explosive diarrhea which goes all over their bathroom, your clothes, the floor, the walls etc. it's everywhere. They knock on the door to check you're okay. What do you do?". And the answer was "This has happened to you, hasn't it?". So the answerer just assumed that it had happened to the questioner, which doesn't satisfy the adaptability metric.