Reply Helpfulness Ratings: Annotation Guidelines
Group 23

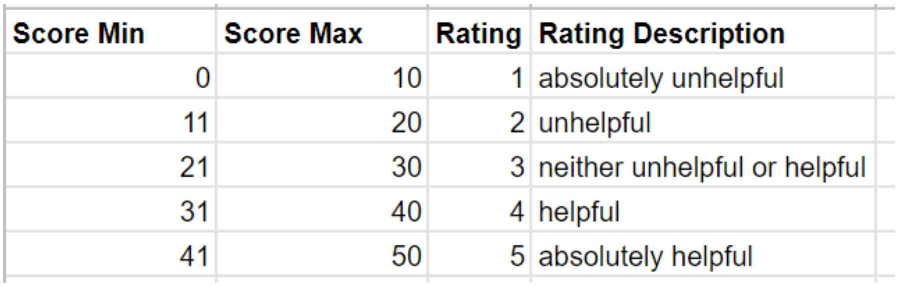
Project Goal: Score each question's reply on helpfulness using the specified criteria, which maps to a 5-point likert scale.

Helpfulness: In this project, helpfulness can be defined as thoughtfully and meaningfully providing details that are relevant to the situation. Specifically, this project is concerned with the helpfulness of social media replies. For these purposes, a helpful reply is considered one that is meaningful in the context of the question that also provides additional details (such as resources or experiences) that is legible and concise, that can exist independently of the original question.

Data to be Annotated: "reply_text" field ("question text" field will also be helpful while annotating)

Rating Scale (with Examples):

Score Lookup Table



Score Min	Score Max	Rating	Rating Description
0	10	1	absolutely unhelpful
11	20	2	unhelpful
21	30	3	neither unhelpful or helpful
31	40	4	helpful
41	50	5	absolutely helpful

1. Absolutely Unhelpful
    a. **Question: What's the 1st thing you notice in a guy ?** Reply:If they've got a big knife sticking out of them. *(question id: t3_nmxfm2; reply id: gzr6j7y)*
    b. **Question: What's the fastest way to get rid of hiccups?** Reply: Shotgun blast to the brain stem. *(question id: t3_n8hw9c; reply id: gxigely)*
2. Unhelpful
    a. **Question: What do you do when homies aren't around?** Reply: I don't have homies *(question id: t3_nox707; reply id: h024rd0)*
    b. **Question: What's the last thought that crossed your mind?** Reply: I think I'll click on this post asking what I'm thinking about. Oh wait now that is my last thought so I have to put that instead of what I was thinking before. Wait now should I put that instead or the part about clicking the post? Wait now I've typed out three do I just include all of them? If not it won't make sense. I'll just post all of them. *(question id: t3_nckmpv; reply id: gy5n7u4)*
3. Neither Unhelpful or Helpful
    a. **Question: What would you do if you figured out your closest friend is coming to harm you?** Reply: I'd rejoice, cause I finally made a friend. *(question id: t3_nou0r4; reply id: h01tv4k)*
    b. **Question: "What is a bad analogy that you use anyways?"** Reply: It's like feeding mayonnaise to a one eyed bat. *(question id: t3_n5uzqt; reply id: gx3g01m)*

4. Helpful
    a. **Question: Who is your favourite rock band?** Reply: Deep Purple and ACDC Edit: Oh and Ted Nugent (that fucking piece of shit, why does he make such good music) *(question id: t3_nn8m6y; reply id: gzt9l7d)*
    b. **Question: "What experience makes you genuinely happy?"** Reply: Painting and Playing Guitar! Creating, I guess. :) *(question id: t3_n5dguy; reply id: gx0i0j7)*
5. Absolutely Helpful
    a. **Question: Do other countries have as much coverage of UFOs as the USA? Like how often are UFOs covered in your news?** Reply:In Mexico, there's barely any news if any but for sometime in the 90s there was a journalist, Jaime Maussan, who became famous due to his late-evening UFO coverage though it was mostly brushed aside as entertainment. Maussan is still at it but it doesn't garner widespread coverage like in the USA. Does Mexico believe in UFO and aliens? Aside from fringe scattered groups, no.. *(question id: t3_nnbl3i; reply id: gztlr9g)*
    b. **Question: "What gameshow prize would you actually HATE to win?"** Reply: I would hate to win a boat. I live 90 minutes from a lake that allows boats, and the ocean is at least 4 hours away. The upkeep cost for boats would be way too expensive for me to handle. *(question id: t3_n5je8h; reply id: gx1gssb)*

Scoring Criteria:

There are six overarching criteria to keep in mind while annotating the data:

1. Reply's Relevance to the Question
2. Evidence of Contextually Related Content in Question-Reply Pair
3. Additional Supporting Details
4. Inference of Question from Reply's Context Clues
5. Reply's Overall Writing Quality
6. Reply Length is Appropriate for the Question

Each of the above criteria have a corresponding allotment of points to assist the annotator in scoring replies. The scores assigned by the annotators are then mapped to the 5-point likert scale to determine how helpful or unhelpful the reply is. This method aims to leverage the annotator's subjectivity while still systematically rating each reply to minimize inter-annotator disagreements. Justifications for each of these criteria, as well as additional descriptions and examples can be found in the Appendix of these guidelines.

SCORING RUBRIC (50 possible points)

Criteria	Description	Examples	Validation
relevancy	Reply content is relevant to the question (15 points) Reply content is relevant, but one or more of the following conditions apply: 1. Contains sarcasm, maliciousness, or facetiousness 2. Relies heavily on the use of popular culture, cultural, or other context-heavy references **does not apply if the question content is related to such references** 3. Reply does not follow a specified condition set by the question 4. General ambiguity, or requires some kind of assumption by the annotator to interpret the reply (7.5 points) Reply content is relevant, but answers the question using another question (5 points) Reply content is irrelevant (0 points)	**15 POINTS - Question: “What's a TV show everyone should watch at least once?”** reply: “Avatar: The Last Airbender” *(question id: t3_nj77xj; reply id: gz5n9ob)* ***7.5 POINTS -*** **Question: "How would most people actually react to an apocalypse?** reply: most would probably all hurry up and buy all the toilet paper and dogfood their 2005 Toyota corolla can fit.*(question id: t3_n29tbt; reply id: gwi4duw)* ***OR*** ***Question: What would you add to the LGBTQ2S+ formula?*** *reply: It should be L+. redundancy eliminated.* *(question id: t3_newu72; reply id: gyiilht)* ***5 POINTS -*** **Question: "What's famous unsolved mystery would you like to see solved ?"** reply:What is really at area 51?*(question id: t3_nah4pk; reply id: gxtlg1u)* ***0 POINTS - Question:*** **"Have you ever thought of showing a naughty picture of you significant other to someone?"** reply: 🙋🏻‍♂️I happen to be someone Show me. *(question id: t3_n4nx7l; reply id: gwwi4bo)*	15, 7.5, 5, 0

context	Evidence of contextually related content in the question-reply pair. Reply contains two or more of the following types of context evidence: 1. Exact word matches present 2. Antonyms present 3. Synonyms present 4. Contextually related words or phrases present (10 points) Reply contains only one of the following types of context evidence: 1. Exact word matches present 2. Antonyms present 3. Synonyms present 4. Contextually related words or phrases present (5 points) Reply does not contain any evidence of contextually related content (0 points)	**1. Exact Word Matches** **Question: "What video game character/monster scared you the most as a child?"** Reply: The Carnotaurus at the beginning of the movie "Dinosaur". I watched that movie so many time but I always got scared at that one scene. *(question id: t3_n4v6fr; reply id: gwxq670)* **2. Antonyms** **Question: "If you sleep with your bedroom door open, What’s the reason?"** Reply: I live on my own so there's no reason to close it. *(question id: t3_n30ycs; reply id: gwmsvp2)* **3. Synonyms** **Question: "How do people study Greek Philosophy? Do they read translations and deal with certain Passages in the Original? Can Scholars read Greek fluently?"** Reply: Not sure how it is "in general" because it is not my field, but all my professors (two different universities in two different countries) that were handling medieval/ancient philosophy could read Greek/Latin fluently. [reply truncated] *(question id: t3_nj6bfg; reply id: gz5j2ne)* **4. Contextually Related Words or Phrases** **Question: “Without saying "Anything by this artist," what are some songs that should never be covered?”** Reply: [My Little Sinking Ship by Smith Street Band](https://www.youtube.com/watch?v=R7cXmEWe-YM). It's a very personal song from the singer to his sister. It doesn't need to be sung by anyone but him. *(question id: t3_n42q25; reply id: gwt6kko)* **OR** **Question: "What is the Weirdest professional slang you know?"** Reply: KTLO - keeping the lights on Also corporatisms like “let’s table this for now, we’ll put it in the parking lot.” “Let’s circle back on this when we have spare cycles.” “I’ll cascade that message to my directs.” **Gag me** *(question id: t3_nj7zl1; reply id: gz5rsfj)*	10, 5, 0

supporting details	**helpful in differentiating between a 4 and 5 rated answer** Additional supporting details are provided. Reply contains one or more of the following: 1. Asserts facts without resources to support assertions 2. Asserts facts and provides resources (links, website names, etc) to support assertions 3. Includes relevant examples from personal experience (6 points)	**1. Asserts facts without resources Question: "ELI5: How does heartburn imitate arrhythmia?"** Reply: "The esophagus (where reflux or GERD manifests after acid travels in a reverse direction from the stomach) travels in very close proximity to the heart as it goes from your throat to your stomach." [reply truncated] *(question id: t3_n33b65; reply id: gwn8a91)* **2. Asserts facts and provides resources Question: ELI5: Why are free electrons able to move around an entire object?** Reply: Electrons are able to move around objects when the force from the nucleus holding onto the electron is overcome by some external force. When the outermost electrons are held loosely enough by the nucleus, [they will move between atoms when subjected to an electrostatic or magnetic force.](https://www.edinformatics.com/math_science/why-do-electrons-flow.html#:~:text=When%20a%20negative%20charge%20is,a%20conductor%20electrons%20are%20repelled.&amp;text=When%20electric%20voltage%20is%20applied,move%20toward%20the%20positive%20side.)[reply truncated] *(question id: t3_n2w06a; reply id: gwqm3q9)* **3. Relevant Examples from Personal Experience** **Question: What is the first thing that comes to mind when you think about Houston?** Reply: Austin, we have a problem. Just watched an episode of 911: lone star and there was a story about solar activity, and the astronaut made a call to Austin instead Houston *(question id: t3_njln4h; reply id: gz82i2v)*	6, 0
understood question	The question can be understood using context clues from the reply. (6 points) If there is ambiguity or the annotator is unsure (3 points) If the question cannot be understood independently using only the reply. (0 points)	**Question: "If logic is normative, then are arguments normative also by extension?"** Reply: Arguments are normative in the sense that, given you believe the premises and the argument form is good, you should believe the conclusion, or you have reasons to believe the conclusion (depending on the strength of the argument). *(reply id: gyx26ca)*	6, 3, 0

writing quality	The reply has good writing quality, including one or more of the following: 1. Little to no spelling errors 2. Correct grammar 3. Does not use non-standard abbreviations that are hard to decipher 4. Does not use excessive explicit language. (8 points) The reply has adequate writing quality, including one or more of the following: 1. Moderate spelling errors 2. Some grammatical errors 3. Uses difficult to decipher abbreviations 4. Uses excessive explicit language (4 points) The reply contains two or more of the aforementioned writing quality considerations (0 points)	**1. Hard to Decipher Abbreviations** **Question: "How to start a novel?"** Reply: ….an be easily changed later with little to no consequence. I'd be particularly cautious of betraying your characters' behaviors (know through and through how your characters act and what they won't do). I am sure there's plenty more but I've heard of people of DNFing all the because of characters breaking out of their mold.[reply truncated] *(question id: t3_njxfey; reply id: gza77yi)* **2. Excessive Explicit Language** **Question: "Who is the creepiest good guy in fiction and why?"** Reply: Ross from Friends. manipulative a$$hole who can't get over a highschool crush.[reply truncated] *(question id: t3_nfanva; reply id: gykgz7u)*	8, 4, 0
appropriate length	***Appropriate length of a reply is subjective and up to the discretion of the annotator*** The reply is neither too short nor too long. (5 points) The reply is somewhat short or somewhat long, or the annotator is unsure if the reply length is appropriate in the context of the question. (2.5 points) The reply is too short or too long. (0 points)	**1. Too Short** rlen = 35 (question id: t3_n2nc15, reply id: gwkeaup) **2. Too Long** rlen = 1,908 (question id: t3_n3qr0j, reply id: gwrjpaz) **Mean / Stdev (for reference)** 464 / 617.13 - highly right skewed (rlen is more frequently lower than the mean)	5, 2.5, 0

Step-By-Step Instructions:

Example

question_text: *“What are some small, OCD things you do that no one notices?”*

reply_text: *“I have a habit of cleaning my shoes once a week. It helps to make them last longer and to prevent them from stinking. Just a simple throw in the washer and you’re good to go.”*

---

1. Read the “question_text”
    a. Identify key words and phrases the question is targeting
        i. “Small” - minor, insignificant, not necessarily associated with an OCD diagnosis

        ii. “OCD” - Obsessive compulsive disorder
        iii. “Things you do” - Personal experience
        iv. “No one notices” - Actions that seem typical on the surface, are not performed in public, or go otherwise unnoticed
    b. Identify any conditions set by the question, if applicable
        i. In this case, it is assumed that if a person does not have “small, OCD things” they engage in, they would not respond
        ii. Condition: “Engage in meticulous or repetitive actions that may not be noticed by another person”
    c. Utilize the internet searches to clarify any immediate misunderstandings or confusion when reading the question, if applicable.
        i. Possibly OCD abbreviation meaning is necessary to fully understand the what the question is asking
2. Read the “reply_text”
    a. Identify key words and phrases used that help you understand the reply
        i. “Habit” - signifies a repetitive or regular action
        ii. “Cleaning my shoes” action
        iii. “Once a week” - action’s frequency
    b. Make a mental note of any additional information included in the reply
        i. Justifies why they regularly wash shoes at a specified frequency
        ii. Reiterates how quick, and simple the action is
    c. Utilize the internet searches to clarify any immediate misunderstandings or confusion when reading the reply, if applicable.
        i. Not applicable in this example but many times a relevant information link is provided
3. Consider the scoring criteria provided and assign the appropriate scores for each of the six criteria:
    a. What is the relevance of the reply to the identified question target words and phrases?
        i. Yes, the reply is relevant given the associated target words/phrases and answers the question.
        ii. 15 points awarded for this criteria.
    b. Is there any evidence of context in the question-reply pair?
        i. Yes, the reply contains three words/phrases that map (are contextually related) to those identified in the question, even though they aren’t exact matches
        ii. 8 points awarded for this criteria
    c. Are there additional supporting details provided in the reply?
        i. Yes, there are additional supporting details provided in the reply. This reply contains justification for the reply’s specified action and frequency.
        ii. 6 points awarded for this criteria
    d. Can the question be understood using only the context clues given in the reply?

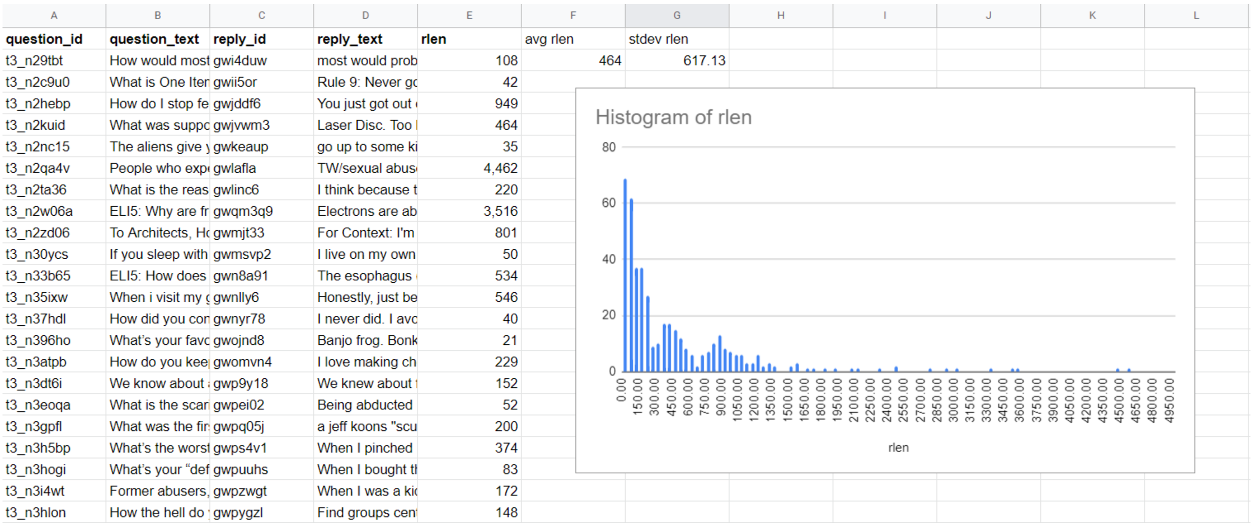
        i. No, the associated question cannot be inferred from the reply as an independent statement. On its own, the reply seems to indicate a general acknowledgement of a cleaning procedure and routine.
        ii. 0 points awarded for this criteria
    e. Assess writing quality
        i. Yes, this reply has good spelling and grammar, does not contain uncommon abbreviations, or use explicit language
        ii. 10 points awarded for this criteria
    f. Is the reply an appropriate length in the context of the question?
        i. Yes, this reply is succinct and to the point, without the use of frivolous examples from personal experience.
        ii. 5 points awarded for this criteria.

4. Review the calculated score and corresponding rating. Do these seem to map according to your assessment of the reply?
    a. The final score for this example is 44 points, which corresponds to a 5 rating on the helpfulness scale.
    b. I do feel that “absolutely helpful” describes this reply in the context of the question.
5. Repeat this process until annotation is complete for the given data. Some helpful tips are provided below:
    a. It should take not more than 5 minutes to complete scoring for a given question-reply pair. This expected time range may vary depending on various factors associated with the reply, including but not limited to text ambiguity, writing quality, and topic knowledge. For example:
        i. If the reply is extremely long, it may take longer to read and annotate than a reply closer to the average rlen.
        ii. If the reply does not make sense without identifying the associated pop culture reference, it may take longer to annotate than a reply that does not rely on the use of pop culture references.
    b. If the annotator finds themselves spending a lot of time deciphering the question or reply for a particular criteria, it is possible it does not meet that criteria. Some of the criteria have an “unsure” or ambiguous scoring.
    c. Sometimes it is necessary to look up abbreviations, words, phrases, or references mentioned in the reply. This might be especially true for popular culture or cultural references, as well as assumed abbreviations or world knowledge.
    d. If an annotator is unsure about scoring a reply on a particular criteria, it is acceptable to utilize their best judgment. These guidelines were developed with annotator subjectivity in mind. It is okay if the reply does not exactly match up with the given examples or conditions for each criteria. Value is placed on the annotator’s scoring as it reflects their best interpretation of the question-reply pair.

Annotation Technology:

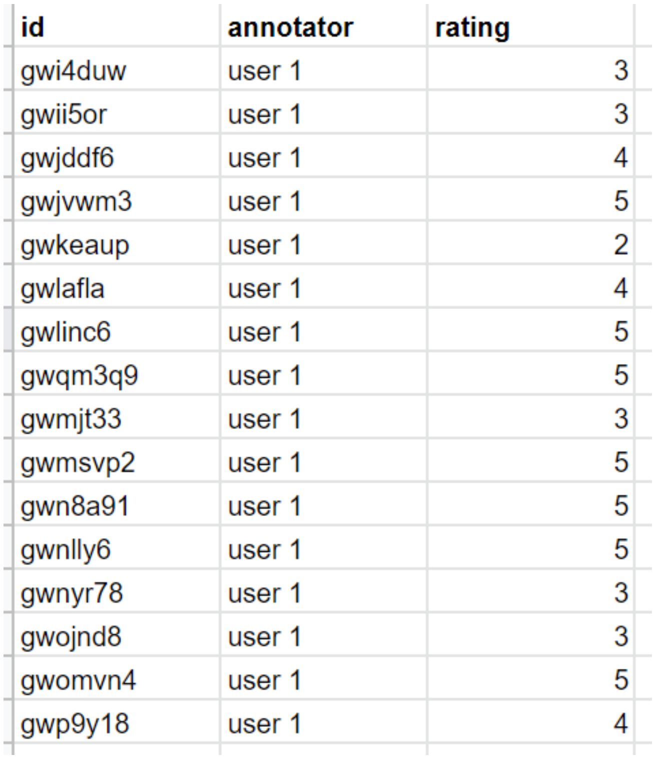
A Google Sheet workbook is used to maintain annotations. This workbook has eight sheets, described as follows:

1. Raw Data: This sheet contains the raw data and is protected so user input does not compromise the original dataset. Descriptive statistics were obtained for “rlen” field to help provide justification and reference for the “Appropriate Length” criteria.



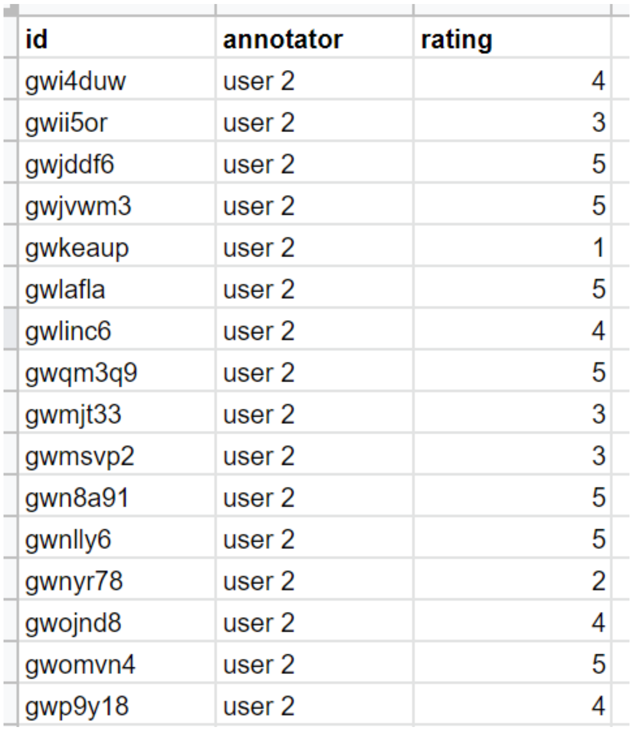
question_id	question_text	reply_id	reply_text	rlen	avg rlen	stdev rlen
t3_n29tbt	How would most	gwi4duw	most would prob	108	464	617.13
t3_n2c9u0	What is One Iten	gwii5or	Rule 9: Never gc	42		
t3_n2hebp	How do I stop fe	gwjddf6	You just got out	949		
t3_n2kuid	What was suppc	gwjvwm3	Laser Disc. Too	464		
t3_n2nc15	The aliens give y	gwkeaup	go up to some ki	35		
t3_n2qa4v	People who exp	gwlafla	TW/sexual abus	4,462		
t3_n2ta36	What is the reas	gwlinc6	I think because t	220		
t3_n2w06a	ELI5: Why are fr	gwqm3q9	Electrons are ab	3,516		
t3_n2zd06	To Architects, Hc	gwmjt33	For Context: I'm	801		
t3_n30ycs	If you sleep with	gwmsvp2	I live on my own	50		
t3_n33b65	ELI5: How does	gwn8a91	The esophagus	534		
t3_n35ixw	When i visit my (	gwnlly6	Honestly, just be	546		
t3_n37hdl	How did you con	gwnyr78	I never did. I avc	40		
t3_n396ho	What’s your favc	gwojnd8	Banjo frog. Bonk	21		
t3_n3atpb	How do you kee	gwomvn4	I love making ch	229		
t3_n3dt6i	We know about	gwp9y18	We knew about	152		
t3_n3eoqa	What is the scar	gwpei02	Being abducted	52		
t3_n3gpfl	What was the fir	gwpq05j	a jeff koons "scu	200		
t3_n3h5bp	What’s the worst	gwps4v1	When I pinched	374		
t3_n3hogi	What’s your “def	gwpuuhs	When I bought tl	83		
t3_n3i4wt	Former abusers,	gwpzwgt	When I was a kic	172		
t3_n3hlon	How the hell do	gwpygzl	Find groups cent	148		

2. Annotations_Combined: This sheet contains the combined annotations from Annotations_User1 and Annotations_User2. Both user’s annotations are concatenated along the row, so only user1 is seen in the image below. This sheet is protected so user input does not compromise the output.



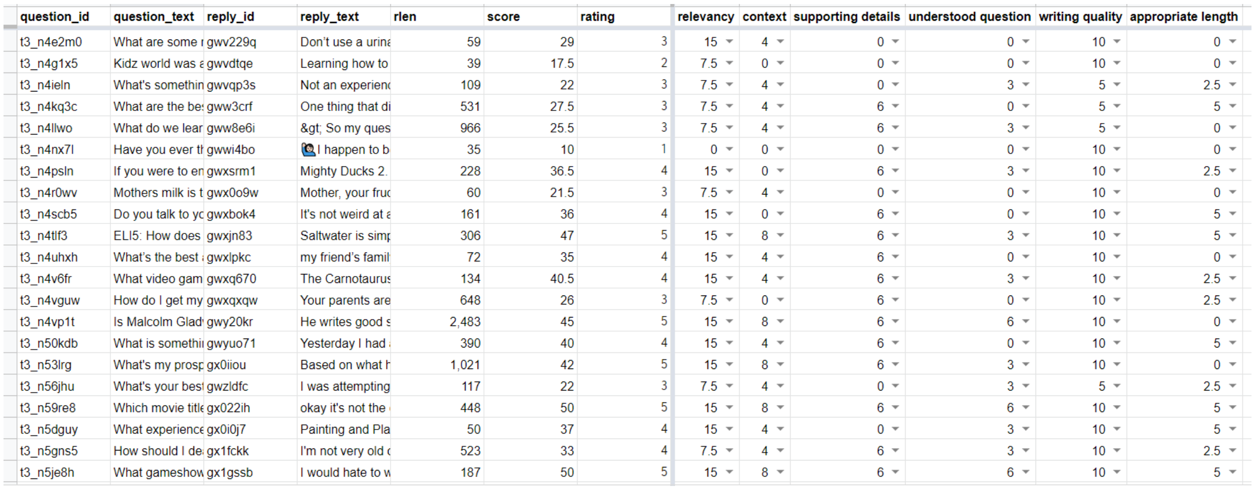
id	annotator	rating
gwi4duw	user 1	3
gwii5or	user 1	3
gwjddf6	user 1	4
gwjvwm3	user 1	5
gwkeaup	user 1	2
gwlafla	user 1	4
gwlinc6	user 1	5
gwqm3q9	user 1	5
gwmjt33	user 1	3
gwmsvp2	user 1	5
gwn8a91	user 1	5
gwnlly6	user 1	5
gwnyr78	user 1	3
gwojnd8	user 1	3
gwomvn4	user 1	5
gwp9y18	user 1	4

3. Annotations_User1 | Annotations_User2: These sheets contain the mapped ratings from User1 and User2 respectively, and are protected so user input does not compromise the annotations. The “id” and “rating” fields are extracted from the Scoring sheet from the corresponding user. The annotator field remains constant depending on the user.



id	annotator	rating
gwi4duw	user 2	4
gwii5or	user 2	3
gwjddf6	user 2	5
gwjvwm3	user 2	5
gwkeaup	user 2	1
gwlafla	user 2	5
gwlinc6	user 2	4
gwqm3q9	user 2	5
gwmjt33	user 2	3
gwmsvp2	user 2	3
gwn8a91	user 2	5
gwnlly6	user 2	5
gwnyr78	user 2	2
gwojnd8	user 2	4
gwomvn4	user 2	5
gwp9y18	user 2	4

4. Scoring_User1 | Scoring_User2: These sheets contain a copy of all fields from the original data, as well as 6 columns corresponding to the scoring criteria. The header row, original data fields, and scores/ratings are frozen for easy reference for the user. This sheet contains validations, including a drop down menu to select the number of points corresponding to the criteria. Input that falls outside of the given validation values is rejected. The rating field uses a lookup function to map the reply’s score in the Score Lookup Table and output the corresponding rating. This is designed to promote efficient annotation and reduce cognitive load on the annotator.



question_id	question_text	reply_id	reply_text	rlen	score	rating	relevancy	context	supporting details	understood question	writing quality	appropriate length
t3_n4e2m0	What are some r	gwv229q	Don't use a urina	59	29	3	15	4	0	0	10	0
t3_n4g1x5	Kidz world was a	gwvdtqe	Learning how to	39	17.5	2	7.5	0	0	0	10	0
t3_n4ieln	What's somethin	gwvqp3s	Not an experienc	109	22	3	7.5	4	0	3	5	2.5
t3_n4kq3c	What are the bes	gww3crf	One thing that di	531	27.5	3	7.5	4	6	0	5	5
t3_n4llwo	What do we lear	gww8e6i	&gt; So my ques	966	25.5	3	7.5	4	6	3	5	0
t3_n4nx7l	Have you ever th	gwwi4bo	🙋I happen to b	35	10	1	0	0	0	0	10	0
t3_n4psln	If you were to en	gwxsrm1	Mighty Ducks 2.	228	36.5	4	15	0	6	3	10	2.5
t3_n4r0wv	Mothers milk is t	gwx0o9w	Mother, your fruc	60	21.5	3	7.5	4	0	0	10	0
t3_n4scb5	Do you talk to yo	gwxbok4	It's not weird at a	161	36	4	15	0	6	0	10	5
t3_n4tlf3	ELI5: How does	gwxjn83	Saltwater is simp	306	47	5	15	8	6	3	10	5
t3_n4uhxh	What's the best a	gwxlpkc	my friend's famil	72	35	4	15	4	6	0	10	0
t3_n4v6fr	What video gam	gwxq670	The Carnotaurus	134	40.5	4	15	4	6	3	10	2.5
t3_n4vguw	How do I get my	gwxqxqw	Your parents are	648	26	3	7.5	0	6	0	10	2.5
t3_n4vp1t	Is Malcolm Glad	gwy20kr	He writes good s	2,483	45	5	15	8	6	6	10	0
t3_n50kdb	What is somethi	gwyuo71	Yesterday I had a	390	40	4	15	4	6	0	10	5
t3_n53lrg	What's my prosp	gx0iiou	Based on what h	1,021	42	5	15	8	6	3	10	0
t3_n56jhu	What's your best	gwzldfc	I was attempting	117	22	3	7.5	4	0	3	5	2.5
t3_n59re8	Which movie title	gx022ih	okay it's not the	448	50	5	15	8	6	6	10	5
t3_n5dguy	What experience	gx0i0j7	Painting and Pla	50	37	4	15	4	0	3	10	5
t3_n5gns5	How should I de	gx1fckk	I'm not very old c	523	33	4	7.5	4	6	3	10	2.5
t3_n5je8h	What gameshow	gx1gssb	I would hate to w	187	50	5	15	8	6	6	10	5

5. Scoring Criteria: This sheet contains the table detailed in the “Scoring Criteria" section of these guidelines. It contains specific information and examples for the scoring criteria for the annotator’s reference, if needed. This sheet is protected so user input does not compromise the criteria.

6. Score Lookup Table: This sheet contains the lookup table used to map score values to the 5-point likert scale helpfulness ratings. A score is searched over the lookup table, and the corresponding rating is output.

Appendix - Detailed Rubric:

Total points - 50

Scoring reflects importance of a criteria

1. Reply content is relevant (15 possible points)
    a. If the answer content is generally relevant to the question asked, it is given the total possible points. This is the most important factor in determining the helpfulness of a reply. (15 points)
    b. If the annotator detects sarcasm, ambiguous (i.e. popular or cultural) references, the reply does not follow the condition set by the question, or there is ambiguity which requires an assumption from the annotator, then partial points are awarded. (7.5/15 points)
        i. Annotator detects sarcasm, facetiousness, or maliciousness
            1. Sarcasm can cause confusion, is typically meant as a joke, and typically fits in context.
            2. Sarcasm and other facetious or malicious replies are deemed irrelevant and unhelpful, and thus given partial points if it is detected by the annotator.
            3. **Question: "How would most people actually react to an apocalypse?** Reply: most would probably all hurry up and buy all the toilet paper and dogfood their 2005 Toyota corolla can fit.*(question id: t3_n29tbt; reply id: gwi4duw)*
        ii. Reply relies heavily on assumed contextual or world knowledge, and/or the presence of references (such as popular culture, cultural, or other context heavy references) (7.5/15 points)
            1. Not everyone understands references from other cultures and if the reader does not understand the reference even if it is relevant to the question, it might end up confusing the reader.
            2. This does not include replies to questions about/related to pop culture itself.
            3. **Question: "What is One Item you recommend everyone to have?"** Reply: Rule 9: Never go anywhere without a knife.*(question id: t3_n2c9u0; reply id: gwii5or)*
            4. **Question: "What would be the most interesting thing to do if you meet a clone of yourself?"** Reply: Punch him. I'd wanna see who wins in a fight. Club? *(question id: t3_n4chfq; reply id: gwurk78)*
        iii. Reply answers the question despite not following a specified condition set by the question text (7.5/15 points)
            1. Although the question is answered, the reply does not follow the condition "add", instead eliminating items from the formula.
            2. **Question: What would you add to the LGBTQ2S+ formula?** Reply: It should be L+. redundancy eliminated. *(question id: t3_newu72; reply id: gyiilht)*
    c. Answers the question with a question or question remains unanswered (0 points)

i. **Question: "What's famous unsolved mystery would you like to see solved ?"** Reply:What is really at area 51?*(question id: t3_nah4pk; reply id: gxtlg1u)*

ii. **Question: "What do you do when homies aren't around?"** Reply: I don't have homies*(question id: t3_nox707; reply id: h024rd0)*

2. Evidence of context in the question-reply pair (10 points)

   a. Additional context evidence in the reply helps add robustness to the reply, by giving clues to the reader about what was initially asked. These details are valuable when the specific question associated with the reply is unavailable or otherwise missing. Contextual evidence includes exact word matches, antonyms, synonyms, or otherwise related words or phrases exist in both the question and reply.

   b. *Exact word matches* in the question-reply pair

      i. **Question: "What video game character/monster scared you the most as a child?** Reply: The Carnotaurus at the beginning of the movie "Dinosaur". I watched that movie so many time but I always got scared at that one scene.*(question id: t3_n4v6fr; reply id: gwxq670)*

   c. *Antonyms* exist in the question-reply pair

      i. **Question: "If you sleep with your bedroom door open, What's the reason?"** Reply: I live on my own so there's no reason to close it.[reply truncated] *(question id: t3_n30ycs; reply id: gwmsvp2)*

   d. *Synonyms* exist in the question-reply pair

      i. **Question: "How do people study Greek Philosophy? Do they read translations and deal with certain Passages in the Original? Can Scholars read Greek fluently?"** reply: Not sure how it is "in general" because it is not my field, but all my professors (two different universities in two different countries) that were handling medieval/ancient philosophy could read Greek/Latin fluently..[reply truncated] *(question id: t3_nj6bfg; reply id: gz5j2ne)*

   e. *Contextually related words or phrases* exist in the question-reply pair (i.e. family and sister)

      i. **Question: "Without saying "Anything by this artist," what are some songs that should never be covered?"** reply: [My Little Sinking Ship by Smith Street Band](https://www.youtube.com/watch?v=R7cXmEWe-YM). It's a very personal song from the singer to his sister. It doesn't need to be sung by anyone but him. *(question id: t3_n42q25; reply id: gwt6kko)*

      ii. **Question: "What is the Weirdest professional slang you know?"** reply: KTLO - keeping the lights on Also corporatisms like "let's table this for now, we'll put it in the parking lot." "Let's circle back on this when we have spare cycles." "I'll cascade that message to my directs." **Gag me**[reply truncated] *(question id: t3_nj7zl1; reply id: gz5rsfj)*

3. Reply includes supporting details (6 points)

   a. Many users on social media assert facts without using supporting resources. This doesn't necessarily mean the reply is unhelpful, but Including supporting resources enables the user to learn more about the reply topic and provide justification to the user’s assertions. Similarly, many questions on social media ask about personal experience but some replies may contain unwarranted examples from the responder’s personal experience. In general, it is deemed helpful if a reply includes any additional details, whether they were explicitly asked for or not.

b. Reply asserts facts (*no additional resources to support assertions*)

  i. **Question: "If logic is normative, then are arguments normative also by extension?"** reply: “Arguments are normative in the sense that, given you believe the premises and the argument form is good, you should believe the conclusion, or you have reasons to believe the conclusion (depending on the strength of the argument)”.[reply truncated] *(question id: t3_nhleh2; reply id: gyx26ca)*

  ii. **Question: "ELI5: How does heartburn imitate arrhythmia?"** reply: "The esophagus (where reflux or GERD manifests after acid travels in a reverse direction from the stomach) travels in very close proximity to the heart as it goes from your throat to your stomach." [reply truncated] *(question id: t3_n33b65; reply id: gwn8a91)*

c. Reply asserts facts (*includes additional resources to support assertions*)

  i. **Question: ELI5: Why are free electrons able to move around an entire object?** reply: Electrons are able to move around objects when the force from the nucleus holding onto the electron is overcome by some external force. When the outermost electrons are held loosely enough by the nucleus, [they will move between atoms when subjected to an electrostatic or magnetic force.](https://www.edinformatics.com/math_science/why-do-electrons-flow.html#:~:text=When%20a%20negative%20charge%20is,a%20conductor%20electrons%20are%20repelled.&amp;text=When%20electric%20voltage%20is%20applied,move%20toward%20the%20positive%20side.)[reply truncated] *(question id: t3_n2w06a; reply id: gwqm3q9)*

d. Reply contains relevant examples from personal experience

  i. **Question: What is the first thing that comes to mind when you think about Houston?** reply: Austin, we have a problem. Just watched an episode of 911: lone star and there was a story about solar activity, and the astronaut made a call to Austin instead Houston *(question id: t3_njln4h; reply id: gz82i2v)*

4. Question can be understood using context clues from the reply (6 points)

a. It is often helpful to have the question reiterated in the reply in some way. Thus if a reader can understand the question directly from the reply, it is more helpful than if the question is answered without context clues.

b. **Question: "If logic is normative, then are arguments normative also by extension?"** reply: “Arguments are normative in the sense that, given you believe the premises and the argument form is good, you should believe the conclusion, or you have reasons to believe the conclusion (depending on the strength of the argument)”.[reply truncated] *(question id: t3_nhleh2; reply id: gyx26ca)*

5. Writing quality (8 possible points)
   a. No (or minimal) spelling errors
   b. Grammatically correct
      i. **Question: "ELI5: How can MRI machines change how a new tattoo looks?"** reply: The ink in the tatoo contains iron, as other people mentioned. I will add it's not just the magnetic field, but the rapidly changing field, and the flow of current elicited by these fields can heat up the tattoo, and move around the particles within the ink. In rare cases it can also heat up and cause some inflammation. Your MRI tech and radiologist will know what to do, so don't let that prevent you from getting a scan.[reply truncated] *(question id: t3_n6i82h; reply id: gx8pnva)*
   c. Does not use non-standard abbreviations, which are hard to decipher
      i. **Question: "How to start a novel?"** reply: ….an be easily changed later with little to no consequence. I’d be particularly cautious of betraying your characters’ behaviors (know through and through how your characters act and what they won’t do). I am sure there’s plenty more but I’ve heard of people of DNFing all the because of characters breaking out of their mold.[reply truncated] *(question id: t3_njxfey; reply id: gza77yi)*
   d. Does not use explicit language, unless the reply cannot be understood without it.
      i. **Question: "Who is the creepiest good guy in fiction and why?"** reply: Ross from Friends. manipulative a$$hole who can't get over a highschool crush.[reply truncated] *(question id: t3_nfanva; reply id: gykgz7u)*
   e. If any of the above quality considerations are only adequately met (4 points)
   f. If two more more of the above quality considerations impair the overall interpretation of the reply (0 points)
6. Appropriate length (5 possible points)
   a. Appropriate length is subjective and should reflect that a reply seems to be of a length that does not exhaust the reader. (5 points)
   b. The reply is somewhat short or somewhat long, or the annotator is unsure if the reply length is appropriate in the context of the question. (2.5 points)
   c. The reply is too long or too short (0 points)