Group Annotation Guidelines

Overview:
The following guidelines aim to provide an annotator a better sense of how to determine the helpfulness of an answer to a question on Reddit. Because annotators will be judging helpfulness using a 5-point Likert scale, these guidelines will offer concepts to look for and examples that would satisfy all five levels. This is the best approach for accommodating the wide range of topics and response approaches from Reddit users.

Group Definition of Helpful:
An answer is helpful when it answers the question with relevant, well-organized information in a manner that someone with no or little background knowledge could easily understand. Additionally, the tone of the response is supportive, encouraging, and sincere.

1 - Strongly Disagree (i.e. least helpful)
- If the answer is not at all related to question this score should automatically be given
- Generally one sentence or less
- Insults the original poster for asking the question
- Cheap, lazy grab for attention
- When asking for an experience, the answer makes a joke to where it's obvious that the given experience has not happened.
- If the question is asking for advice, the answer will cause harm to the question asker if they carry out the instructions in the answer.
- Incorrect answer that makes no attempt at being correct
- Writing is unclear and advice, if present, is useless
- ELI5: Answers the question incorrectly
- Asking for personal experience: Posting something unrelated to the question
- The vast majority of the answer is rambling off topic
- Asking a follow-up question that is insulting to the question asker.
- Answer is not in the spirit of what the question asker intended. Focuses on loopholes rather.
- Leaves an outside reader very confused (unless the question is about a niche subject)

2 - Disagree
- Contains inside jokes or references that may not be obvious to an outside reader
- Makes many assumptions as to what the question asker knows.
- Leaves an outside reader asking more questions than they came in with (i.e. counter-productive)
- If the question is asking for advice, the answer could potentially cause harm to the question asker if they carry out the instructions in the answer.
- Incorrect answer that makes an attempt at being correct
- Vague, ambiguous answers with little or no utility

- ELI5: Answers it with the complexity that only an expert could understand
- When related to artistic and visual media (images, music, videos), the answer does not provide a link to said media (unless it's very commonly known such as the Mona Lisa).
- Asking for personal experience: Sentence fragment related statement with no further explanation.
- Moderate amounts of rambling off topic
- Asking a follow-up question that could be insulting to the question asker.

3 - Neither Agree nor Disagree
- Does not answer the question directly, but makes a sincere attempt to do so.
- ELI5: Answers it with the complexity that an 18 year old could understand, but not a 5 or 12 year old.
- If the question is asking for advice, the answer would neither help nor harm the question asker if they carried out the instructions in the answer.
- Doesn't add or detract from the value of the discourse
- Somewhat organized ideas; isn't decidedly difficult or easy to follow
- Neutral in tone; not sincere or insincere, warm or cold, encouraging or discouraging, flippant or supportive
- Might be considered off-topic, aimless, unnecessary, reactive, or other ways of describing an ultimately useless response
- Asking for personal experience: 1-2 sentences with no or little further explanation.
- Mild amounts of rambling off topic
- Genuinely answers the question correctly/well, but makes little to no effort

4 - Agree
- ELI5: Answers it with complexity that a 12 year old could understand, but a 5 year old could not.
- If the answerer finds the question unclear, they ask for clarification from the question asker. Additionally, if they find the question unclear, they might also offer a strongman argument (i.e. interpreting one or several variations of the question being asked and responding to those)
- Makes few assumptions as to what the question asker knows
- If the question is asking for advice, the answer may help the question asker if they carried out the instructions in the answer
- Largely useful, thoughtful, and addresses the question thoroughly enough to qualify as a sufficient (but not exceptional) response
- Mild vagueness in response. May leave an outside reader asking one or two questions.
- Asking for personal experience: ~1 paragraph with some context
- Minimal amounts of rambling off topic
- Puts a significant amount of thought and effort into a long response, but somewhat difficult to follow or doesn't include a TL;DR (too long didn't read) summary.
-

5 - Strongly Agree

- Answered in such a manner that allows someone who knows nothing about the topic to fully understand
- Answers 'why' even when one is not asked
- ELI5: Answers it to where a kindergartener could understand
- Makes no assumptions as to what the question asker knows
- Gives direct links to outside sources
- Shows empathy towards the question asker when necessary
- If the question is asking for advice, the answer will very likely help the question asker if they carried out the instructions in the answer
- Sincere and encouraging in tone
- Thoughtful and even meticulous in addressing the material; well-organized ideas that are clear and easy to understand
- Provides all necessary context for an outside reader to understand, particularly when describing a personal experience
- Asking for personal experience: 2+ paragraphs with fully sufficient context
- No rambling off topic
- If the answer is very long, it includes a TL;DR (too long didn't read) summary.

Scenarios to think about

- If the question is really a commentary statement in context (that is just formatted like a question)
  - Ex: How is twitch allowing tub streamers and no bans?
- Spelling/grammar
- Asking for personal experiences, not necessarily for help
- Questions that are difficult to understand without context
  - Ex) Could you reason with a T-1000?