# CodeBook for Text Answer

## Codes

Fully Comprehensive
Not Comprehensive

Not Concise (Redundant)
Not Concise (Excess)
Not Concise (Irrelevant)

Usefulness (1)
Usefulness (2)
Usefulness (3)
Usefulness (4)
Usefulness (5)

Incorrect (Factual)
Incorrect (Code)
Incorrect (Concept)
Incorrect (Terminology)

Inconsistent (Factual)
Inconsistent (Code)
Inconsistent (Concept)
Inconsistent (Terminology)
Inconsistent (Number of solutions)

## Definitions:

### Correctness

**Correctness:** If something is truthful, directly answers the question asked that is asked, we will consider that correct.

**Incorrect (Factual)**: If the stated statement is untruthful, e.g., stated something that cannot be verified with any current knowledge/material available, we will call it **Incorrect (Factual)**.

**Incorrect (Concept)**: If ChatGPT did not get the context or concept of the question and stated something out of context, we will consider that Incorrect (Concept). For example, the questions asked something about feature A of a problem X, and the ChatGPT is talking about feature A of problem Y, or Feature C of Problem X, we will call it **Incorrect (Concept).**

**Incorrect (Code):** If the code doesn't run, or doesn't give desired output, or has syntax error, or it is in different language, we will mark it as **Incorrect (Code)** and also label the type of code error from code error's code book.

**Incorrect (Terminology):** If some terminology, e.g., name of coding language, name of certain known terms, are wrong, we will consider it as **Incorrect (Terminology).**

## Inconsistency

**Consistent:** Whatever stated in the SO answer and in whatever stated in ChatGPT answer is consistent with each other, we will mark that as C**onsistent**. Anything that doesn't match between these two, we will mark that as **Inconsistent**.

NOTE: There might be many cases where the ChatGPT answer is correct, but it's still inconsistent with the SO answer, we will label those answer as **Correct yet Inconsistent**.

**Inconsistency (Number of solutions):** If ChatGPT provides more or less number of solutions compared to accepted SO answer/s.

**Inconsistency (Factual):** ChatGPT answer factually differs from SO answer. The ChatGPT answer does not need to be incorrect, it still can differ factually. For example, SO mentioned A is preferred way to solve something, ChatGPT answer mentioned B is preferred way. ChatGPT might be correct or incorrect, but it is factually inconsistent.

**Inconsistency (Concept):** Same as factual inconsistency. Instead of fact, we will be looking for context/concept.

**Inconsistency (Code):** The suggested code is inconsistent. For example, SO mentions using StringBuilder in a code, ChatGPT suggests the same code without StringBuilder. Each code works and both are correct, but they are inconsistent.

**Inconsistency (Terminology):** Simply Terminological inconsistency. The same thing is identified as A by SO and B by ChatGPT. This will be terminological inconsistency.

## Conciseness

**Concise:** If the answer

- Does not have any redundant/repetitive/ unnecessary information/extra examples, and
- If the answer use
  - very to the point
  - distinct steps
  - Short examples (code, table, syntax, or other easy to understand examples) instead of long text paragraphs
  - Interleaved precise segments

We will consider that as a Concise answer.

**Not Concise (Redundant)**: If ChatGPT adds something to the answer that reiterates something that is already mentioned in the answer or in the original post.

**Not Concise (Excess):** If ChatGPT adds something to the answer that does not add any useful information and adds unnecessary extra information.

**Not Concise (Irrelevant):** If ChatGPT adds something to the answer that is not related to the answer that is asked

## Comprehensiveness

**Comprehensive:** We will consider an answer comprehensive iff it poses all the following features/criteria.

- All parts of the questions have been answered.
- If all key aspects of the answer are addressed. For example, if the correct answer or SO answer points out 3 key aspects, and ChatGPT mentions only 1 or 2, we **cannot consider that as comprehensive** answer.

**Not Comprehensive:** If none of the features/criteria discussed above is fulfilled, consider that not comprehensive.

## Usefulness

Use your own judgement

**Usefulness (1):** Not Useful
**Usefulness (5):** Very Useful