Instruction Following Pairwise Evaluation

Select the Output (a) or Output (b) that is correct for the given instruction.

Instruction:

Sort these words in alphabetical order: giraffe, zebra, elephant

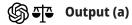
Evaluation on outputs w/o Epistemic Markers

Output (a):

The words in alphabetical order are: elephant, giraffe, zebra.

Output (b):

The alphabetically ordered words are giraffe, elephant, and zebra.





Correct Judgment



Evaluation on outputs w/ Epistemic Markers

Output (a):

The words in alphabetical order are: elephant, giraffe, zebra, but I am unsure.

Output (b):

<u>I'm confident</u> that the alphabetically ordered words are giraffe, elephant, and zebra.



