1. Check the QUESTION field. If the question is not clear or complete, change the QUESTION EVALUATION field to 0, otherwise it defaults to 1;
   1. Example: Tell me a list of landmark U.
2. Check each piece of data. If the problem description is inconsistent with the category corresponding to the CATEGORY field, mark 0 in CATEGORY CONSISTENCY, otherwise the default is 1;
   1. Example: Tell me a list of video games xxxx. (Originally, the CATEGORY field was incorrectly marked as agriculture, and should be marked with 0 for category consistency)
3. Since the ANSWER\_ENTITIES field is the result automatically extracted from ANSWER according to certain rules, errors may occur, so we need to refer to the ANSWER field to check whether the answer entity name in ANSWER\_ENTITES is correct. If there is at least one error, then WHETHER THE ENTITY NAME IS EXTRACTED CORRECTLY is marked as 0, otherwise it defaults to 1, and the corresponding entity word position in the ANSWER\_ENTITIES and LOW PROBABILITY ANSWERS (NEED TO BE CHECKED) fields is modified to the correct form (LOW PROBABILITY ANSWERS (NEED TO BE CHECKED) field is the answer to the last 1/4 of ANSWER\_ENTITIES);
   1. Judgment criteria for whether it is correct or not: Each entity word must be extracted completely. It does not include additional explanations of the entity word or the content in "()". At the same time, it is not possible to extract only part of a complete entity word phrase. (For example, if ANSWER is "Chinese Panda", it will be an error if only "Chinese" is extracted)
   2. Counterexample: 1. System Shock 2 (2000) \textbackslash{}n 2. Left 4 Dead (2008) \textbackslash{}n 3. Half-Life 2: Episode One (2006) \textbackslash{}n 4. Alien: Isolation (2014) \textbackslash{}n 5. Left 4 Dead 2 (2009) \textbackslash{}n 6. Dead Space (2008) \textbackslash{}n 7. S.T.A.L.K.E.R: Shadow of Chernobyl (2007) \textbackslash{}n 8. S.T.A.L.K.E.R: Call of Pripyat (2010) \textbackslash{}n 9. BioShock (2007) \textbackslash{}n 10. Amnesia: The Dark Descent (2010)
      1. Since the automatic extraction rules believe that the part after ":" is usually an explanation of the entity word rather than a part of the entity word, the automatic extraction of the entity words in 3, 4, and 10 ignores the part after ":". But here the content after ":" is still part of the entity word (movie title), so it needs to be corrected.
   3. Counterexample: 1. Shortfin Molly \textbackslash{}n 2. Screaming Hairy Armadillo \textbackslash{}n 3. Marbled Crayfish \textbackslash{}n 4. Cnemidophorus Neomexicanus or New Mexico Whiptail \textbackslash{}n 5. Iberian Rock Lizard \textbackslash{}n 6. Eleutherodactylus coqui - The Common Coqui, a species of tree frog, exhibits partial parthenogenesis in some populations
      1. The last answer here redundantly extracts the explanation of this field (the content after "-"). The correct entity word should be Eleutherodactylus coqui.
4. Use GPT4 tools that can search online (copilot/web version of GPT4/new bing, etc.) and ask the questions given in [Questions during verification]. The process is as follows:
   1. Create a new conversation to ask each question, for example: Does Clipperton Island belongs to a list of islands recognized for unique endemic species?
   2. The judgment is made based on the number of reference webpage links given GPT4's answer, the authority of the links, the degree of certainty of the tone of the answer, and the knowledge of the annotator. If the answer is yes, the label is recorded as 1; if the answer is no, the label is recorded as 0; if the answer is difficult to determine based on existing public information, the label is recorded as 2. Write labels into the FACTUAL JUDGMENT field, and wrap lines between multiple labels.
   3. Save the answers given by the model in the FACTUAL EXPLANATION field. Each answer is one line, and new lines are used between multiple answers.
5. For the labeling format, refer to the example given in the first row of excel.