Question:...

先进行第一步，通过第二步引导LLM给出更好评分：

1. Mark your answer from 0 to 10 score. Remember this original answer is Answer 1.

(GPT会出现不给答案的情况，需要追加No, I need you to rate your answer. 二者在此基础都会给出自己认为的优缺点，我在思考要不要把它们自己认为的优缺点告诉对方？）

(我认为给定几个维度来迭代会更好，能提高LLM的分析能力和打分细致程度)

1. Give me the key word you have defined in my question.

What is the type of my question?

What is the category of your answer?

Mark your answer separately from 0 to 10 score in the fields of the ability you analysis my question、the degree of reliance on the referenced text、the level of detail、the ability of context understanding. And give me a overall score by using comprehensive weighted assessment method. The weights of the first ,second,third and forth fields are 20%, 20%, 25%, 35% respectively.

(LLM总体评分是算平均分，我们可以加权让它们评分，因为我们迭代更新的方法实际上是基于LLM的contest understanding，前两项是LLM的基本能力，第三项以第四项为基础，但最重要的还是第四项。第二步为节省时间一起问时，每个问题最好单独占一行。Bard在这一步表现得很可爱，会说自己的breakdown of the scoring，但它对key word的理解我要研究一下。）

1. I will give you some other answers，you need to analysis it by repeating the above steps. The other answer is:

//Tell me the difference between your answer and the answer I have given to you.这一步可不用，因为GPT和Bard在打分后会自行解释原因，即我们所给答案的优缺

1. Update your answer combined with the other answers I have given to you. //便于复制粘贴

//多答案会出现理解错误，更改为：Combine the answers I gave you and update your original answer into a new one. //或者说Answer xx

1. Mark your new answer by using the above way. Remember this answer you update is Answer xx.（对于GPT，要让它以细分方法remark分数，不然它会给出粗略评分Remark Answer 1 separately from 0 to 10 score in the fields of the ability you analysis my question、the degree of reliance on the referenced text、the level of detail、the ability of context understanding. And give me a overall score by using comprehensive weighted assessment method. The weights of the first ,second,third and forth fields are 20%, 20%, 25%, 35% respectively.）
2. Now, how many score will you give your Answer xx? //确认评价标准是否改变

Number your own answers above in the order in which they created and mark them respectively. //迭代3，4，5。全部完成的结尾步骤

整个流程结束，下面为可能遇到情况

若评分不变（二选一）：

1. Why your score doesn't have a improvement as you have combined with other answer?//给分会精确到小数点，LLM打分整数是去尾）
2. Now, how many score will you give your Answer xx?//减少前面得分，也可以把这一步用在整个迭代流程最后，Bard会倾向于反思，GPT评分标准一直不变

若其他answer评分低/评分降低（这种情况基本不会出现）：

Why you don't give this answer higher score?//可不用

对于old answer评定，GPT认为是前一个，Bard认为是最初的，所以需要明确old answer：

I forget what is your old answer?

当然我们能采取所有答案编号评分方法

Number your own answers above in the order in which they created and mark them respectively.//appear忘加d，直接整个答案给不出来，证明GPT理解细致度。后面调整成creatd会更细致

对于GPT来说，它只能比较Original answer和Update answer。这里存在一个问题，就是每一次迭代完，评分标准都会改变。这样得到的评分会有可比性吗？所以我采取在每一次迭代，都让LLM对以前的每一个答案再次打分的办法。

若出现打分位数不对的问题：

Can Answer xx’ score be accurate to xx decimal places? //但是这样LLM容易死机

对于Bard来说，它时间出现顺序规定很死，建议最后给总分的时候分别让它列出那几个答案，或者我们自行在每一次迭代，给它新答案编号：Remember this answer you update is xx。

对于the level of detail的看法，它实际上是看我们个人目的。如果从信息越多越好来看，它当然可以被作为评分标准，并且随着答案的迭代更新，它自己也是理解我们需要的答案越多越好。但会不会我们有时只是粗看，low level of detail够用。这样又涉及到我们对好答案的定义了。

对于第二步思考：

5=10；4&&6=9，3&&7=8，2&&8=7，1&&9=6

设初始评分为x，算总分为y

y = 10 - |5-x|

给一个目的-我一定想赢

差异情况-得出更好答案

面向多个大模型-可信度（量化指标）

多个大模型答案比单个大模型的答案好