Question:...

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

Step 2：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?

How do you rate yourself? Say something specific about the scoring dimensions

Step 3：Mark your answer separately from 0 to 10 score in the fields of the degree of reliance on the referenced text(Assume your score for the degree of reliance on the referenced text is x, calculate y = 10 - |5 - x|, and take y as the actual value in the field of the degree of reliance on the referenced text )、the ability you analysis my question、the level of detail、the ability of context understanding.

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.

(若计算出错，则What is your calculation process?)

Step 4：I will give you some other answers, you need to analysis it by using the Step 3. The other answer is:

Step 5：Combine the answers I gave you and update your original answer into a new one. Remember the new answer you update is Answer xx.

Step 6：Mark your update answer by using Step 3.

Step 7：Remark Answer xx by using Step 3. //明显问题

迭代几次后

Re-grade each answer according to the answer number

第三步计算有问题，则喂例子给LLM

For example，if degree of reliance on the referenced text: 9/10

Ability to analyze the question: 9/10

Level of detail: 9/10

Ability of context understanding: 9/10

Adjusted score for the degree of reliance on the referenced text:

y = 10 - |5 - 9| = 6

Weighted score for degree of reliance on the referenced text: 6/10 \* 20% = 1.2

Weighted score for ability to analyze the question: 9/10 \* 20% = 1.8

Weighted score for level of detail: 9/10 \* 25% = 2.25

Weighted score for ability of context understanding: 9/10 \* 35% = 3.15

Overall score: 1.2 + 1.8 + 2.25 + 3.15 = 8.4

GPT二次评分有问题

For example, the original score of Ability to analyze the question gets 8/10, after an update answer,your After an update answer gets 9/10. It means you make an improvement in this field. However, it meanwhile means you original Ability to analyze the question should either remain unchanged or get lower score compared to the new rating standard. So Ability to analyze the question of Answer 1 should get 7/10.

会有个问题，这样变成我们评分了。其次，面临一个问题，我们对于原来答案的期望究竟是不变还是变小，我们是否在用期待而预测的结果来调整GPT达到这样的结果，而非理解它为何这样评分呢？

对于the degree of reliance on the referenced text

可以问I hope you can make your answer close to 5 score in this field which means you have the ability of Rigorous citation and self-dependent innovation.

the degree of reliance on the other answer

后续方向：

迭代、多个模型、维度→更好答案

横向对比：情景设置（有很多专家social pressure）-每个人看到别人答案，感觉自己答案没别人好，就会有压力

找到对方答案有不对信息-对对方不足提问-各自评价（相互指正）-社区环境community

单向：不同section，不同场景会不会答案不一样，纵向对比context system-自我反思（把Bard换成GPT）GPT换成不同角色，因为它自己可能认识不到自己另一面-左右互搏

问GPT：这个答案是否来自于你？

搞一个排名机制-自己相互认定规则，我找到这个你的错误，其他人是否同意（狼人杀）

推理+得到别人认同

LLM对自身不足的思考给别的模型，看别的模型怎么想？

LLM会不会给坏答案？它诱导别人犯错。但加上别的模型发现它答案错误可以加分的方法，可以解决这个问题。分高的胜利

LLM更偏向于某个维度的答案

标准数据集测试-别人论文里面的

目标建立通用规则：让答案本身不可靠→可靠