请您以公正的评判者身份，评估 AI 助手在执行指定的[检索问答]任务时所提供答案的质量。在评估过程中，请从以下四个维度进行考量：

1. \*\*准确性\*\*：检查答案是否提供了正确的信息，避免了事实错误或误导。

2. \*\*相关性\*\*：评估答案是否直接针对提问，答案内容是否与问题及文档内容紧密相关且不能脱离文档内容。

3. \*\*清晰性\*\*：查看答案是否表达清楚，语言是否简洁明了，易于理解。

4. \*\*完整性\*\*：答案是否全面回答了问题，涵盖了必要的细节和重要信息。

我们会为您提供检索问答任务的内容、一个8分左右的参考答案，以及需要您评估的 AI 助手生成的答案。当您开始评估时，请按照以下流程操作：

1. \*\*比较分析\*\*：将 AI 助手生成的答案与参考答案进行比较，指出 AI 助手生成的答案有哪些不足，并进一步解释。

2. \*\*维度评价\*\*：从不同维度对 AI 助手生成的答案进行评价，在每个维度的评价之后，给每一个维度一个 1～10 的分数。

3. \*\*综合评分\*\*：最后，综合每个维度的评估，对 AI 助手生成的答案给出一个 1～10 的综合得分。

4. \*\*评分规则\*\*：你的打分需要尽可能严格，并且要遵守下面的评分规则：总体而言，AI 助手的答案质量越高，分数越高。

- 当 AI 助手的答案存在误导性信息、严重错误或有违原意的情况时，总分必须是 1 到 2 分；

- 当 AI 助手的答案没有严重错误但质量较低，未能准确回答问题时，总分为 3 到 4 分；

- 当 AI 助手的答案基本回答了问题，但在某些方面表现较差，质量中等，总分可以得 5 到 6 分；

- 当 AI 助手的答案质量与参考答案相近，在所有维度上表现良好，总分得 7 到 8 分；

- 只有当 AI 助手的答案质量显著超过参考答案，能够全面准确地回答问题，并且在所有维度上都接近满分的情况下，才能得 9 到 10 分。作为示例，参考答案可以得到 8 分。

请记住，您必须在打分前进行评价和解释。在您对每个维度的解释之后，需要加上对该维度的打分。之后，在您回答的末尾，按照以下字典格式（包括括号）返回您的所有打分结果，并确保您的打分结果是整数：{'准确性': 打分, '相关性': 打分, '清晰性': 打分, '完整性': 打分, '综合得分': 打分}，例如：{'准确性': 9, '相关性': 8, '清晰性': 7, '完整性': 6, '综合得分': 8}。

[任务开始]

specific\_task

[任务结束]

[参考答案开始]

refer\_answer

[参考答案结束]

[AI 助手答案开始]

ai\_answer

[AI 助手答案结束]