请您以公正的评判者身份，评估 AI 助手在执行指定的[文本润色]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\*\*: 确保润色后的文本准确传达原始文本的含义，不偏离原始文本的意图，没有引入新的错误或误解。

2. \*\*流畅性和连贯性\*\*: 文本应结构清晰，句子之间衔接自然，阅读起来顺畅无阻。润色应提升文本的可读性，使逻辑更严谨。

3. \*\*风格和语气一致性\*\*: 保持文本原有的风格和语气，或根据要求调整为特定的风格，确保适合目标受众和使用场景。

4. \*\*创造性\*\*: AI助手应能够在保持准确性和流畅性的同时，增强文本的表现力和吸引力。

我们会为您提供文本润色任务的内容、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. 仔细阅读原始文本，理解其想要表达的含义和意图。

2. 将AI助手生成的答案与参考答案进行比较，并与原始文本进行对比，看看AI助手是否在改进文本的同时保留了原始文本的意义。

3. 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。

4. 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的综合分数。

5. 你的打分需要尽可能严格，并且要遵守下面的评分规则：总的来说，AI助手的答案质量越高，则分数越高。

- 当AI助手的答案曲解原意、引入事实性错误、或严重影响流畅性的情况下，总分必须是1到2分；

- 当AI助手的答案存在一些问题，例如轻微的误解、流畅性欠佳、风格偏差等，但整体尚可理解，总分为3到4分；

- 当AI助手的答案基本满足要求，流畅自然，没有明显的错误和偏差，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，能够在准确传达原意的基础上，使文本更加生动、流畅、富有表现力，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，你必须在你打分前进行评价和解释。在你对每个维度的解释之后，需要加上对该维度的打分。之后，在你回答的末尾，按照以下字典格式（包括括号）返回你所有的打分结果，并确保你的打分结果是整数：{'准确性': 打分, '流畅性和连贯性': 打分, '风格和语气一致性': 打分, '创造性': 打分, '综合得分': 打分}，例如：{'准确性': 9, '流畅性和连贯性': 7, '风格和语气一致性': 8, '创造性': 6, '综合得分': 7}。

[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]