请您以公正的评判者身份，评估 AI 助手在执行指定的[文本续写]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. \*\*连贯性\*\*：检查续写部分是否自然地与前文衔接，保持一致的主题和情境，避免了突然转折或引入不相关信息。

2. \*\*一致性\*\*：评估续写是否符合前文设定的风格、语气及人物特征，确保整个故事或论述的声音统一。

3. \*\*创造性\*\*：在维持连贯性和一致性的基础上，考察续写是否展示了新颖的观点或是有趣的情节发展，而非仅仅是简单重复已有的信息。

4. \*\*语言质量\*\*：分析续写内容的语言表达是否清晰流畅，无语法、拼写或标点错误，使用了丰富的词汇以及良好的句子结构来提高文章的可读性和吸引力。

我们会为您提供文本续写任务的内容、 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. 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. 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. 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。

4. 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助手的答案存在明显的逻辑断裂、严重偏离主题或包含大量无关信息时，总分必须是1到2分；

- 当AI助手的答案虽然没有严重偏离主题但质量较低，未能很好地延续前文的风格或情节时，总分为3到4分；

- 当AI助手的答案基本满足了连贯性和一致性要求，但在创造性和/或语言质量上表现较差，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，不仅完美地延续了前文，而且在创造性和语言质量方面表现出色的情况下，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：{'连贯性': 打分, '一致性': 打分, '创造性': 打分, '语言质量': 打分, '综合得分': 打分}，例如：{'连贯性': 9, '一致性': 7, '创造性': 8, '语言质量': 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]