**请您以公正的评判者身份，评估 AI 助手在执行指定的[消息总结]任务时所提供答案的质量。 在评估过程中，请从以下五个维度进行考量：**

**1. \*\*一致性\*\*：检查消息总结内容是否与待总结的原始文本内容保持一致，是否有误导性的信息或错误的细节。**

**2. \*\*完整性\*\*：参照参考答案，评估消息总结内容是否涵盖了原始内容中的关键点和重要信息，确保没有遗漏核心内容。**

**3. \*\*简洁性\*\*：参照参考答案，查看消息总结内容是否尽量简洁，去除了不必要的细节，同时避免过于简化导致信息丢失。**

**4. \*\*流畅性\*\*：消息总结内容应该逻辑清晰，语言流畅，易于理解，确保信息的自然过渡。**

**5. \*\*指令遵循\*\*：输出结果的格式应与任务要求的输出格式保持一致。**

**我们会为您提供消息总结任务的内容、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，未能准确反映原文的主要内容时，总分为3到4分；**

**- 当AI助手的答案基本反映了原文的核心内容，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，能够全面准确地概括原文的所有要点，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：{'一致性': 打分, '完整性': 打分, '简洁性': 打分, '流畅性': 打分, '指令遵循': 打分, '综合得分': 打分}，例如：{'一致性': 9, '完整性': 7, '简洁性': 8, '流畅性': 6, '指令遵循': 8, '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]**