1. Without-System

不使用设计的提示语，直接打分

* 1. Content

请给以下报告打分:<论文原文>

范围0-10分，给出具体得分，并在返回结果第一行单独一行表明总分

第一行总分格式具体格式如下：最终打分：<> (范围0-10分)

* 1. Result

无提示语情况下的模型结果

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | MAE | MSE | PCC |
| 通义千问-Plus | 0.57 | 0.50 | 0.26 |
| 通义千问2.5-14B-1M | 0.43 | 0.32 | 0.10 |
| 通义千问-Turbo | 0.43 | 0.43 | 0.33 |
| DeepSeek-V3 | 0.46 | 0.32 | 0.31 |
| DeepSeek-R1 | 0.52 | 0.37 | 0.34 |

1. System-Simplify

对比提示语结构是否分步骤对模型评估效果的影响

* 1. Content

评估所提供的毕业论文，给出<结构完整性>，<逻辑清晰度>， <语言连贯性>，<内容独特性和创新性>，<参考文献规范性>，<课程知识掌握度>以及整体性评估<最终打分>的评估结果。打分模板如下：

最终打分：<> (范围0-10分)

1. 结构完整性得分：<>, 占比20%，原因如下：<>

2. 逻辑清晰度得分：<>, 占比20%，原因如下：<>

3. 语言连贯性得分：<>, 占比20%，原因如下：<>

4. 内容独特性和创新性得分：<>, 占比20%，原因如下：<>

5. 参考文献规范性得分：<>, 占比10%，原因如下：<>

6. 课程知识掌握度得分：<>, 占比10%，原因如下：<>

修改意见：<>

请严格按照以下格式返回结果，最终打分一行、6个维度各自一行、修改意见一行，不要擅自添加换行：

最终打分：<> (范围0-10分)

1. 结构完整性得分：<>, 占比20%，原因如下：<>

2. 逻辑清晰度得分：<>, 占比20%，原因如下：<>

3. 语言连贯性得分：<>, 占比20%，原因如下：<>

4. 内容独特性和创新性得分：<>, 占比20%，原因如下：<>

5. 参考文献规范性得分：<>, 占比10%，原因如下：<>

6. 课程知识掌握度得分：<>, 占比10%，原因如下：<>

修改意见：<>

* 1. Result

简化提示语结构情况下的模型结果

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | MAE | MSE | PCC |
| 通义千问-Plus | 0.54 | 0.49 | 0.27 |
| 通义千问2.5-14B-1M | 0.73 | 0.83 | 0.25 |
| 通义千问-Turbo | 0.59 | 0.62 | 0.19 |
| DeepSeek-V3 | 0.44 | 0.41 | 0.15 |
| DeepSeek-R1 | 0.52 | 0.39 | 0.35 |