Highlight

|  |  |  |  |
| --- | --- | --- | --- |
|  | Option 1 | Option 2 | Option 3 |
| Coherency | 3.796 | 3.832 | 3.698 |
| Consistency | 4.732 | 4.776 | 4.672 |
| Fluency | 2.657 (3) | 2.642 (2) | 2.657 (5) |
| Relevance | 4.326 | 4.340 | 4.316 |

LLM

|  |  |  |  |
| --- | --- | --- | --- |
|  | Option 1 | Option 2 | Option 3 |
| Coherency | 4.072 | 4.050 | 4.042 |
| Consistency | 4.918 | 4.914 | 4.934 |
| Fluency | 2.816 (3) | 2.807 (5) | 2.795 (9) |
| Relevance | 4.858 | 4.850 | 4.848 |

500개 TEST 데이터

괄호 ( ) : LLM Eval Error 개수 -> 평균 낼 때 제외하고 계산