1. Reflection on Ours

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model |  | 1 | 2 | 3 |  |  |
| Llama3.1-8b | R@1 | 44.89% | 48.98% | 36.73% |  |  |
|  | R@2 | 59.18% | 51.02% | 57.14% |  |  |
|  | R@3 | 63.26% | 61.22% | 68.09% |  |  |
|  | R@4 | 72.92% | 61.22% | 69.39% |  |  |
|  | R@5 | 75.51% | 75.51% | 73.47% |  |  |
|  | R@0 | 12.0% | 6.0% | 4.0% |  |  |
|  | Trained | 62.0% | 60.0% | 56.0% |  |  |
| Llama3.1-70b | R@1 | 74.0% | 72.0% | 74.0% |  |  |
|  | R@2 | 76.0% | 78.0% | 78.0% |  |  |
|  | R@3 | 80.0% | 84.0% | 78.0% |  |  |
|  | R@4 | 82.0% | 84.0% | 80.0% |  |  |
|  | R@5 | 86.0% | 86.0% | 82.0% |  |  |
|  | R@0 | 16.32% | 12.24% | 10.20% |  |  |
|  | Trained | 64.0% | 60.0% | 58.0% |  |  |
| Qwen2.5-7b | R@1 | 63.3% | 67.3% | 67.3% |  |  |
|  | R@2 | 73.5% | 77.6% | 81.6% |  |  |
|  | R@3 | 77.6% | 69.4% | 75.5% |  |  |
|  | R@4 | 81.6% | 85.7% | 81.6% |  |  |
|  | R@5 | 83.7% | 79.6% | 83.3% |  |  |
|  | R@0 | 20.0% | 14.0% | 12.0% |  |  |
|  | Trained | 68.0% | 64.0% | 62.0% |  |  |
| Glm4-9b | R@1 | 68.0% | 76.0% | 74.0% |  |  |
|  | R@2 | 80.0% | 84.0% | 84.0% |  |  |
|  | R@3 | 86.0% | 88.0% | 88.0% |  |  |
|  | R@4 | 84.0% | 90.0% | 88.0% |  |  |
|  | R@5 | 88.0% | 86.0% | 86.0% |  |  |
|  | R@0 | 26.53% | 20.41% | 12.24% |  |  |
|  | Trained | 76.0% | 72.0% | 70.0% |  |  |
|  |  |  |  |  |  |  |

2. Reflection on Public

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model |  | 1 | 2 | 3 |  |
| Llama3.1-8b | R@1 | 36.0% | 48.0% | 44.0% |  |
|  | R@2 | 42.0% | 56.0% | 62.0% |  |
|  | R@3 | 44.0% | 64.0% | 62.0% |  |
|  | R@4 | 48.0% | 68.0% | 70.0% |  |
|  | R@5 | 50.0% | 70.0% | 74.0% |  |
|  | R@0 | 20.0% | 14.0% | 12.0% |  |
|  | Trained | 58.0% | 56.0% | 54.0% |  |
| Llama3.1-70b | R@1 | 70.0% | 68.0% | 66.0% |  |
|  | R@2 | 78.0% | 70.0% | 72.0% |  |
|  | R@3 | 82.0% | 78.0% | 80.0% |  |
|  | R@4 | 88.0% | 80.0% | 86.0% |  |
|  | R@5 | 90.0% | 82.0% | 90.0% |  |
|  | R@0 | 20.0% | 18.0% | 18.0% |  |
|  | Trained | 64.0% | 64.0% | 58.0% |  |
| Qwen2.5-7b | R@1 | 62.0% | 58.0% | 63.0% |  |
|  | R@2 | 66.0% | 59.2% | 63.4% |  |
|  | R@3 | 72.0% | 74.0% | 75.5% |  |
|  | R@4 | 76.0% | 75.5% | 69.4% |  |
|  | R@5 | 66.0% | 74.0% | 72.0% |  |
|  | R@0 | 32.0% | 26.0% | 24.0% |  |
|  | Trained | 74.0% | 72.0% | 68.0% |  |
| Glm4-9b | R@1 | 76.0% | 78.0% | 70.0% |  |
|  | R@2 | 72.0% | 88.0% | 84.0% |  |
|  | R@3 | 90.0% | 92.0% | 86.0% |  |
|  | R@4 | 96.0% | 90.0% | 90.0% |  |
|  | R@5 | 88.0% | 94.0% | 92.0% |  |
|  | R@0 | 36.0% | 34.0% | 30.0% |  |
|  | Trained | 78.0% | 76.0% | 76.0% |  |
|  |  |  |  |  |  |