|  |  |
| --- | --- |
| Model | Krippendorff's α |
| GPT-4 | 0.78 (0.76, 0.81) |
| GPT-3.5-turbo | 0.62 (0.59, 0.65) |
| Davinci-003 | 0.46 (0.43, 0.49) |
| Flan-T5-XXL | 0.44 (0.41, 0.46) |
| Davinci-002 | 0.42 (0.38, 0.45) |
| Command-XL | 0.31 (0.27, 0.35) |