# Performance Comparison of LLMs in Fact and Relationship Extraction

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metric** | **Paper** | **Manual** | **DeepSeek** | **ChatGPT** | **Mistral 24B** |
| Total facts identified | P1  P2 | 32  52 | 15  15 | 25  10 | 62  27 |
| Total unique nodes | P1  P2 | 90  208 | 27  78 | 78  36 | 108  28 |
| Total relationships extracted | P1  P2 | 30  52 | 15  29 | 22  10 | 62  27 |
| Matching facts with manual | P1  P2 |  | 2 (6.25%)  3 (5.80%) | 11 (34.38%)  1 (1.92%) | 11 (34.38%)  4 (7.69%) |
| Missed facts compared to manual | P1  P2 |  | 30 (93.75%)  49 (94.23%) | 21 (65.63%)  51 (98.08%) | 22 (68.75%)  48 (92.31%) |
| Hallucinated /incorrect | P1  P2 |  | 11 (34.38%)  2 (3.85%) | 3 (9.375%)  9 (17.31%) | 0 (0%)  0 (0%) |