

Confusion Matrix for Model gpt-4-turbo-preview and Prompt P2_HIERARCHY based on binary prompt P3

True Labels

0	74	1	65	3	24	2	0	1	2	0	0	0	0	0
1	21	43	1	1	6	2	0	0	0	0	1	0	0	0
2	2	0	31	2	2	0	0	0	0	5	0	0	0	0
3	2	0	5	22	3	0	0	0	0	0	0	4	0	0
4	0	0	0	0	33	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	26	5	0	0	0	0	0	0	0
6	0	0	0	0	0	3	27	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	21	0	0	0	0	0	0
8	0	0	1	0	0	5	1	0	13	0	1	0	0	0
9	0	0	0	0	0	0	0	0	0	17	0	1	0	0
10	0	0	0	0	0	1	0	0	0	0	15	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	15	0	0
12	0	0	0	0	0	0	0	0	0	5	0	0	7	0
13	1	0	5	1	0	0	0	0	0	4	0	0	0	0
	0	1	2	3	4	5	6	7	8	9	10	11	12	13

Predicted Labels

- 0: Remaining
- 1: Non-systematic-review
- 2: Human-non-RCT-non-drug-intervention
- 3: Human-non-RCT-drug-intervention
- 4: Human-case-report
- 5: Animal-other
- 6: Animal-drug-intervention
- 7: Human-systematic-review
- 8: In-vitro-study
- 9: Human-RCT-non-drug-intervention
- 10: Animal-non-drug-intervention
- 11: Human-RCT-drug-intervention
- 12: Clinical-study-protocol
- 13: Human-RCT-non-intervention

0

10

20

30

40

50

60

70