Confusion Matrix for Model gpt-4-turbo-preview and Prompt P4 1 600: Remaining 0 -1: Non-systematic-review 2: Human-non-RCT-non-drug-intervention 3: Human-non-RCT-drug-intervention 4: Human-case-report - 505: Animal-other 2 -6: Animal-drug-intervention 7: Human-systematic-review 3 - 1 8: In-vitro-study 9: Human-RCT-non-drug-intervention 4 -_ 40 10: Animal-non-drug-intervention 11: Human-RCT-drug-intervention 5 - 1 12: Clinical-study-protocol **True Labels** 13: Human-RCT-non-intervention 6 -8 -- 20 9 -10 - 0 11 - 0 - 10

12 - 0

- 0

13 - 1

Predicted Labels