Confusion Matrix for Model gpt-3.5-turbo and Prompt P2 HIERARCHY based on binary prompt P3 800: Remaining 1: Non-systematic-review 2: Human-non-RCT-non-drug-intervention 3: Human-non-RCT-drug-intervention - 70<mark>4: Human-case-report</mark> 5: Animal-other 2 -6: Animal-drug-intervention 7: Human-systematic-review 3 -608: In-vitro-study 9: Human-RCT-non-drug-intervention 4 - 1 10: Animal-non-drug-intervention 11: Human-RCT-drug-intervention 5012: Clinical-study-protocol - 2 **True Labels** 13: Human-RCT-non-intervention 6 - 0 - 40 7 -8 - 2 - 30 9 -

10 - 0 - 20

- 0

12 - 0

13 - 1

Predicted Labels

- 10

- 0