

Confusion (One-vs-Rest): CONTRADICTED (normalize=true)

True (Human)

CONTRADICTED

2

40.0%

3

60.0%

NOT CONTRADICTED

0

0.0%

37

100.0%

CONTRADICTED

NOT CONTRADICTED

Predicted (LLM)

