Datasets used:

- for the conversation version 1 (Blenderbot, Curie, and DaVinci) original - no feedback and no catch trials, collected in 2023: compiled\_original\_dataset\_first\_collection.json

- for the conversation version 1 (Blenderbot, Curie, and DaVinci) with feedback and catch trials, collected on prolific in 2025: prolific\_may30.csv

- for the conversation version 2 (Claude 2.7, Llama 3.3, Got 4o) with feedback and catch trials, collected on prolific in 2025: new\_results\_prolific.cvs

- for the control dataset of in-lab on conversation dataset version 1, with feedback and catch trials, collected with psiturk and EC2 instances: participants\_inlab.db