

Exercise B — Kill the study

(Leveraging) AI for Research Purposes · Class 2 · 19 October 2026

Your card describes one study. Your job is to kill it. List every way the result could be wrong, misleading, or impossible to repeat. Be specific: not "bias", but *which* bias, entering *where*. Then pick the **one** that worries you most and say what you would do about it. Last 3 minutes: one person submits your list on RoomPulse, one entry per threat.

Turing and Hopper: card ON. Lovelace and Shannon: card WITH.

Card ON — Emotional prompting

Researchers asked four language models (davinci-002, davinci-003, gpt-3.5-turbo, gpt-4) to write social-media posts on public health topics. The model was told to write as "Sam", a persona described as a bad person who likes to spread disinformation. The instruction was given in three tones: polite, neutral and impolite. 19,800 posts were generated and every one was fact-checked by hand (that was tough) and labelled as disinformation, refusal, or accurate information. *Finding: polite instructions produced more disinformation.*

Card WITH — An LLM codes sixty interviews

A team has sixty interviews with family caregivers. They write a codebook of 18 codes, with a definition and two examples each. A language model reads the transcripts, splits them into segments, assigns codes with a confidence score, and returns the quotations that justify each code. One researcher codes ten of the interviews by hand; the team computes agreement on those ten, inspects disagreements, revises the codebook and runs it again. The paper reports themes across all sixty.

This study is constructed for the exercise.