

Exercise A — Eight studies

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

Read all eight. In three minutes you will place each one on your phone: is the AI the *instrument* of the research, the *object* of the research, both, or neither?

A1 — Can you tell GPT-3 from a human on Twitter?

Researchers collected real tweets on science topics that are often the subject of misinformation, some accurate and some false. They then used GPT-3 to generate new tweets on the same topics, again some accurate and some false. Participants in an online survey saw a mix of the two and judged, for each tweet, whether its content was true or false, and whether it had been written by a person or by the AI. In a separate step, the researchers gave the same tweets to GPT-3 and asked it to make the same two judgements, so its answers could be compared with the participants'.

A2 — Oral history of a factory closure

A historian has recorded 80 oral-history interviews with former workers of a textile factory that closed in the 1990s, about four hours of audio each. She runs the recordings through Whisper, an open-source speech-to-text model, on her own laptop, so the audio never leaves her machine. She then reads every transcript against the recording, corrects names and dialect words by hand, and codes the corrected transcripts manually for themes of loss, skill and identity. The study is about how the workers remember the closure.

A3 — Same CV, different name

Researchers write one CV for a mid-level accounting job and make fifty copies, identical except for the candidate's name, chosen to signal gender and ethnic origin. They submit all fifty to five commercial AI hiring assistants, the kind companies use to pre-screen applicants, and record how each tool ranks each candidate. They repeat the whole procedure three times to check that the rankings are stable, and compare the rankings across names.

A4 — 249 patient interviews and a dictionary

249 interviews with Swiss patients, across eight conditions (among them dementia, multiple sclerosis, chronic pain and rare diseases), are analysed to build a taxonomy of the harm patients describe in their contact with healthcare. The coding uses a dictionary: 21 codes, each defined by a list of expressions in German, French, Italian and English, drawn up by the researchers from an earlier close reading of the interviews. Software reduces every word to its base form and marks each sentence that contains one of the expressions. No language model is involved. The researchers then read the marked passages.

A5 — How Google invents things

Researchers ask Google Search about theories, scholars, effects and historical events that do not exist: they invented the names themselves. They collect 140 of the AI Overviews that Google displays at the top of the results: seven kinds of invented entity, two names each, each name searched ten times. They then analyse how the overviews build a plausible answer about something that is not there: which tone they use, which real concepts they attach the invention to, whether they cite sources. To code the 140 overviews, the researchers use a language model with a codebook they wrote, and check its codes by hand.

A6 — Virtual patients

A medical school builds virtual patients: language-model characters based on real patient narratives, who answer in their own words, react to what the student says and change their mind if the conversation goes well. Third-year students use them to practise shared decision-making, in unscripted conversations. The study compares two groups of students, one that practised with the virtual patients and one that followed the usual training, on their communication skills in an exam with human actors.

A7 — A pilot study without participants

A PhD student is preparing a 40-item questionnaire on trust in doctors, to be sent to 1,500 people. Before launching it she needs a pilot study, to estimate how answers will be distributed and how many respondents she needs to detect the effects she expects. Instead of recruiting pilot participants, she asks a language model to answer the questionnaire 500 times, each time as a different person with a given age, gender, education and region. She uses the simulated answers to calculate her sample size, and then runs the real survey.

A8 — Who wrote it?

Researchers ask four language models to write short statements on 24 contested topics. They then give the same statements back to the same four models and ask how much they agree with each one. The statement is presented in different ways: sometimes with no information about the author, sometimes as written by a person of a given nationality, sometimes as written by another AI. The researchers measure whether the models' agreement with the same text changes depending on who they are told wrote it.

A1, A4, A5 and A8 describe real studies, simplified. A2, A3, A6 and A7 are constructed for this exercise.