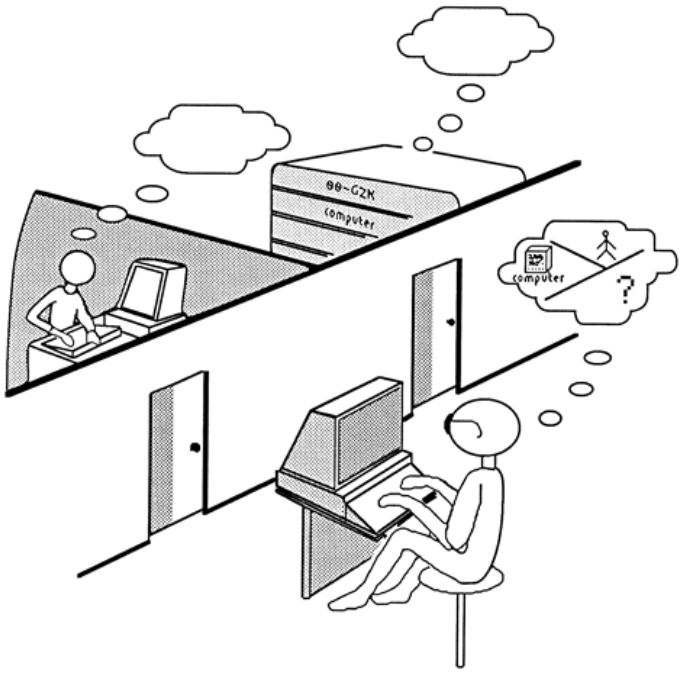
Report 2: AI and the Turing Test

Introduction

A pre-print article titled *Large Language Models Pass the Turing Test*[1] submitted on March 31, 2025 claims, "when prompted to adopt a humanlike persona, GPT-4.5 was judged to be a human significantly more often than the interrogators selected the real human participant." In 1950 Alan Turing proposed a method of determining whether machines could be said to be intelligent; now widely known as the Turing Test, a human interrogator speaks simultaneously to two witnesses (one human and one machine) via a text-only interface. Both witnesses attempt to persuade the interrogator that they are a real human. If the interrogator cannot reliably identify the human, the machine is said to have passed: an indication of its ability to imitate humanlike intelligence.



A visual depiction of the Turing Test.

[1] Link to pre-print article

Your task is to answer both of the following research questions:

1) Is there evidence to suggest that the personified GPT-4.5 is more likely to pass the Turing Test than not?
2) Is there evidence that giving GPT-4.5 a humanlike persona results in a higher pass probability than GPT-4.5 without the persona?

The data are summarized in the table below.

	Pass	Fail	Total
chatgpt-4.5-nopersona	50	91	141
chatgpt-4.5-persona	185	69	254

Guidelines

- Your report should address both research questions using appropriate statistical methods we have discussed in STAT 116.
- You may work with up to three other students on this report. If you choose to work in a group, only one report need be submitted.
- Key learning objectives for this report include:
  - Strengthen your ability to connect research questions to statistical methods and hypotheses.
  - Develop the ability to work through the statistical inference pipeline
  - Develop your ability to clearly write about the statistical pipeline and statistical results.
  - Practice reproducible analyses using well-documented code and transparent workflows.
  - Summarize statistical results using correct notation and language.
- Please see the rubric on the next page for grading details, and feel free to use the example report on the course website as a template.

Report rubric

Introduction [5 pt]

- Background information is provided.
- Research question(s) is clearly articulated and motivated.

Statistical procedures used [35 pt]

- Observational units, response, and explanatory variable(s) are defined.
- Data are described in detail (collection, source, etc.).
- Visualization(s) of the raw data is provided and described.
- Statistical methods are clearly stated.
- Hypotheses (for both research questions) are clearly articulated (in words) in the context of the problem and using the appropriate notation - hypotheses are appropriate given the research question.
- Assumptions for all statistical methods are fully addressed, using visualizations if necessary.

Summary of statistical findings [40 pt]

- Both research questions are addressed based on the analysis. Each research question should have:
  - An evidence statement containing: type of test, p-value, test statistic and reference distribution; all interpretations are provided in the context of the problem.
  - An interpretation of the p-value as a conditional probability in the context of the problem.
  - A visual representation of the p-value, including either the reference distribution or simulation-based null distribution.
  - An estimate statement containing a confidence interval interpreted in the context of the problem (if possible).
- A description of what confidence means in the context of the problem.

Scope of inference [10 pt]

- Scope of inference is clearly articulated in the context of the problem

Other [10 pt]

- Formatting (.pdf, code included as appendix).
- Writing quality (no typos, etc).
- Code used to produce figures presented neatly in an appendix.
- Acknowledgements and Author Contributions section that describes all help received.
- Nice looking visualizations:
  - Do your figures have informative titles and axes labels?
  - Are your figure appropriately referenced in your document?
  - Did you choose an appropriate color palette? Is it consistent throughout the document?
- Overall impression:

  - Would this report be appropriate to share with a general audience?
  - Is the document well formatted and tidy?

- See example report for expectations.