Detecting AI Content

Artificial intelligence (AI) can be used to write, create art, and compose music. As AI becomes more common, determining whether humans can distinguish between AI-created content and human created content becomes increasingly important.

Step 1: Ask a research question

Can people reliably tell the difference between AI created content and human recreated content?

Step 2: Design a study and Collect Data

1. If you were a researcher trying to answer this research question, how would you design a study?

Some tools available on the Internet allow a person to test how well they can distinguish AI-created content from human-created content. These tools are set up like a quiz. In each question, the person looks at an object, such as photographs, art, or a piece of writing. They then guess whether the object was created by AI or not. Here are some examples of questions.

Example 1: Images

5 → Is this a real photo of a cat or an image generated by AI?*

A Real photo

B Generated by AI

Example 2: Text

11 → Was this text written by a human or AI?*

One tends to think that the best way to lose weight is through a diet program of some sort. A diet program is definitely the way to go if you want to shed some pounds in order to fit yourself into that dress or jeans that you have been eyeing on buying. A weight loss program would consist of taking in less refined and processed carbs, and more of the complex carbohydrates. It would also mean that you would have to avoid processed foods, and instead eat all the healthy and fresh fruits and vegetables that you can.

A Human

B AI

Imagine we are investigating whether one person, say a professor at your college, can detect AI content with a quiz made of questions like those above.

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

2. Explain this data collection method in your own words. Apply vocabulary by explaining:
    a. What are the observational units?
    b. What is the variable? Is it quantitative or qualitative?
    c. What is the parameter of interest?

3. **Null Hypothesis.** One possibility is that the professor cannot tell the difference between AI created content and human generated content. They are just guessing whether the image or block of text was created by AI.
    a. In this case, what proportion of the questions do you think the professor would answer correctly?
    b. Restate your idea from part (a) with symbols and numbers by completing the phrase below.

$$H_o: \pi = \_\_\_\_\_$$

4. **Alternative Hypothesis.** Another possibility is that the professor *can* tell the difference between AI and human created content at a rate better than random chance.
    a. In this case, which of the following statements seem to make the most sense?
        - The professor will choose the right answer on 50% of the questions.
        - The professor will choose the right answer on more than 50% of the questions.
        - The professor will choose the right answer on less than 50% of the questions.

    b. Restate your idea from part (a) with symbols and numbers by completing the phrase below. Circle the correct symbol. (< , > or =).

$$H_a: \pi \ (>, <, =) \ \_\_\_\_\_$$

One professor tried a quiz with four of these types of questions. The professor correctly answered all 4 of the questions.

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

Step 3: Explore the data.

One professor tried a quiz with four of these types of questions. The professor correctly answered all 4 of the questions.

5. Calculate the relevant statistic. What proportion of the questions did the professor get correct?

$$\hat{p} = \_\_\_\_\_$$

6. What is the sample size in this study?

$$n = \_\_\_\_\_$$

7. Looking at the data, make a conjecture. do you think that you have enough evidence to say that the professor can distinguish between AI and human created content?

Step 4: Draw an Inference

If could be that the professor can distinguish between AI and human generated content. Afterall, they got all 4 questions right. But it could be that they were just lucky.

8. Explain how we can use a coin flip to simulate this scenario.

   Heads means the professor _____

   Tails means the professor _____

   One coin flip represents _____ so we should flip the coin _____times

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

9. Use the Rossman Chance applets to simulate the null hypothesis.
    a. If a professor was just randomly guessing, how common would it be that they got 4 out of 4 correct?

    b. Report the p-value.

    c. Considering this p-value, would you accept or reject the null hypothesis? What does this mean in terms of the strength of evidence about the professor's ability to distinguish between AI and human generated content?

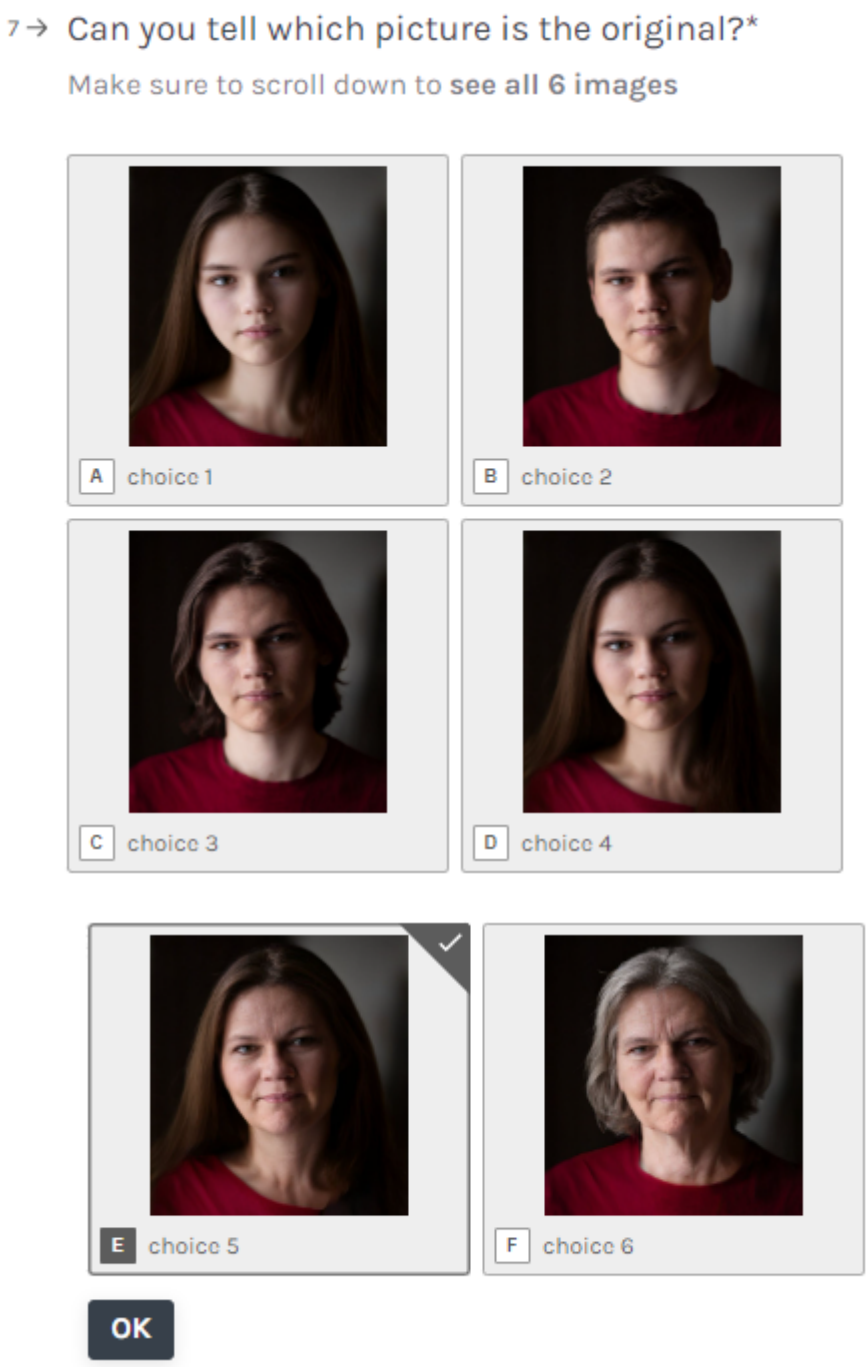
10. Suppose the professor had gotten 8 out of 8 possible questions correct. How would that change your strength of evidence or your conclusion?
    a. First use your intuition to predict whether you would have stronger or weaker evidence about the strength of evidence for the professor's ability to distinguish between AI and human generated content.

       Would you predict the p-value to be bigger or smaller?

    b. Then, use the Rossman Chance applet. What is the p-value? Did this align with your prediction?

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

11. Other types of quizzes have questions where the person has to choose the real image from a selection of images that includes AI editing, like the question below.



12. In your opinion, how is this situation different from the earlier situation?

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

13. Online, people can take a series of questions like those in #11, where a person tries to choose which of the six images is the original. Imagine we are trying to detect whether a person can identify whether an image has been edited by AI using a set of questions like these. In your own words, what do you think the null and alternative hypotheses would be?

    One of the online quizzes has 4 (four) questions like the ones shown in #11. On these questions, the professor got 3 out of 4 (three out of four) correct.

14. How would you simulate this situation? You could use coins, spinners, dice, or other methods.

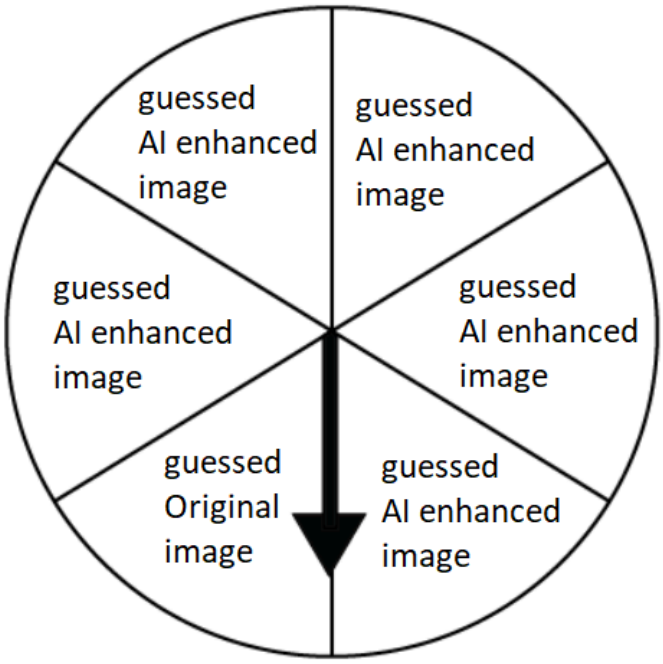
Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

15. How do you think the p-value would compare to the p-value you found earlier? How would the p-value between these two situations compare?

First Situation Determine whether an image or piece of writing is created by a human or AI. 4 Questions. Professor got 4 out of 4 correct.	
Second Situation Choose which of 6 provided images is an original image. The other 5 were edited by AI. 4 Questions. Professor got 3 out of 4 correct.	

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

16. One student decided to spin a spinner with six sections four times to simulate this situation. One section of the spinner was labeled as guessing the original image correctly correct image, as shown below.

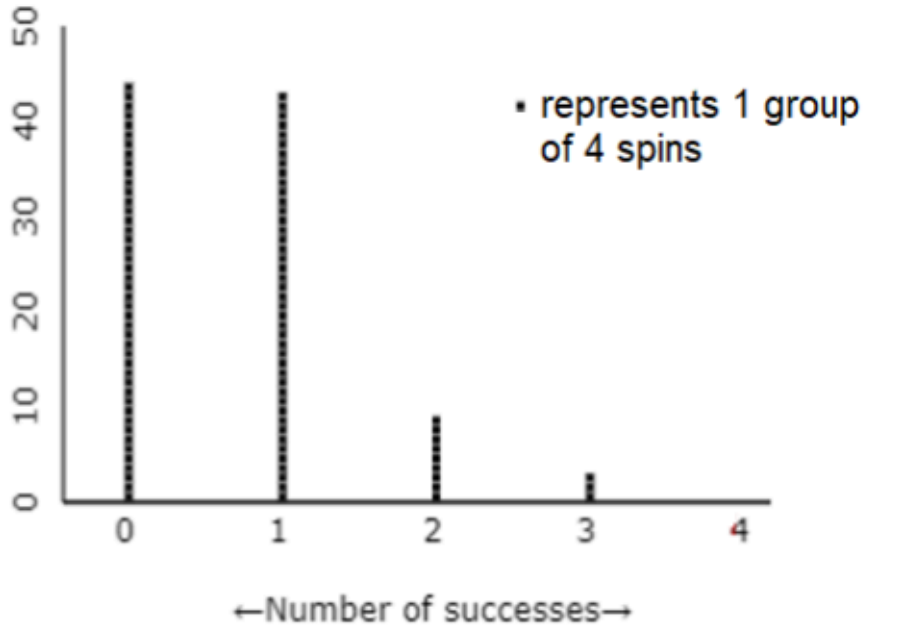


How would you explain to someone else why the student is doing this? The table below might help.

Why spin a spinner?	
Why are there 6 (six) sections?	
Why spin it 4 (four) times?	

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

Within each group of 4 (four) spins, the student counted the number of times the spinner landed on the section representing correctly guessing the original image. The graph of the first 100 simulations is shown below. One dot represents 1 group of 4 spins.



Number of Times Spinner Landed on "Guessed Original Image" in 4 Spins

17. How could you use this graph to think about the strength of evidence about whether the professor was guessing?

18. What is the p-value?

19. Do you accept or reject the null hypothesis? What does this mean in terms of the professor being able to detect AI-edited images?

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.

Step 5: Formulate conclusions

20. Consider how the data were collected, the results, and the p-values.
    a. Do you think these results are statistically significant?
    b. How broadly are you willing to generalize your conclusions? Do you think these results would be true for all professors? All adults? Why or why not?

Step 6: Look forward and back

21. Suggest a new research question that you could investigate that builds on these results.

Source: https://www.tidio.com/blog/ai-test/

Developed by Laura Callis, Curry College, for the DISCUS-IS Project. Funded by the National Science Foundation IUSE Engaged Student Learning # 2314358.