

BEING SKEPTICAL: A SERIES OF PUZZLES

From Margaret Menzin

menzin@simmons.edu

For Instructors Only- Slide 0

- This is a set of exercises, designed to be used over a series of weeks, which aims to make students skeptical of material they see online, especially AI generated material, and to teach them the SIFT method of approaching it.
- The first week introduces SIFT through a very short reading assignment. Each week reinforces the use of SIFT.
- Each week the instructor's guide is in the SPEAKERS notes, and the slide holds that week's puzzle/prompt.
- Each exercise should take 5 minutes or less.
- The last example (slide 13) is only for students who have written code. ALL THE OTHER SLIDES are for a general audience.

Introduction to SIFT

- HW: Read the page at <https://guides.lib.uchicago.edu/c.php?g=1241077&p=9082322>
- **SIFT = Stop, Investigate the source, Find better coverage, Trace claims, quotes and media to original context.**
- Apply SIFT to the following headlines:
- **“95% Of GenAI Pilots Fail Because Companies Avoid Friction”**
- **“Aspartamane (found in the NutraSweet and Equal) causes cancer” per Reuters.**

1. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Do you believe this:
The [Associated Press](#) and reams of other [media outlets](#) reported that JD Vance said “school shootings are a ‘fact of life’.”

2. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- **Do you believe this?**
Gemini was asked to plan a 20 minute walk starting at 2 Palace Road (Lefavour
Start at 2 Avenue Louis Pasteur (e.g., near the Longwood Galleria/Children's Hospital).
- **Walk South on Avenue Louis Pasteur** (approx. 3-4 minutes)
- You'll be walking past major hospitals and research institutions.
- **Turn Right onto Longwood Avenue** (approx. 2 minutes)
- Walk one block past the T stop.
- **Turn Left onto Riverway** (approx. 5 minutes)
- This is part of Boston's historic **Emerald Necklace** park system, designed by Frederick Law Olmsted. Enjoy the trees and the river (actually a series of waterways).
- **Cross the Riverway and turn Right onto Avenue Louis Pasteur** (approx. 5 minutes)
- You can cut through the grounds of the Harvard Medical School/Dental School campus if you want a slightly more scenic, academic view.
- **Continue North on Avenue Louis Pasteur** (approx. 3-4 minutes)
- This will bring you right back toward your starting point.
-

3. Being skeptical and SIFTing

- Stop,
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Do you believe this?
Writing in Medium, MKWriteshere reported:
World's Youngest Mother: Lina Medina Was Raped at 4 and Gave Birth at 5
- Lina Medina's story of precocious puberty, village complicity, and 87 years of silence
- I used to think the world was cruel then after reading Lina's Story i changed my mind that the world isn't just cruel. It's fucking evil!!
Until today, I had no idea a five-year-old could become pregnant. I've lived my entire life without knowing this horror existed, that human biology could be forced into something so unnatural. But Lina Medina (now 92 years old and living somewhere in Peru) has carried this reality for nearly nine decades.
- She's the youngest confirmed mother in recorded history. I need you to understand what that actually means.

4. Being skeptical and SIFTing

- Stop,
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Do you believe this?
Writing in Medium, MKWriteshere reported:
Poetry Breaks AI Safety: 62% of Language Models Fail When Prompts Rhyme
New research reveals how simple verse bypasses ChatGPT, Claude, and Gemini safety systems — a vulnerability affecting every major AI provider. Plato wanted poets banned from his ideal society. He thought their language would mess with people's judgment and collapse civilization.
for
- Turns out, he was onto something.
- Researchers just discovered that wrapping harmful requests in poetry makes AI models ignore their safety rules. Across 25 different models — GPT-5, Claude, Gemini, and others — poetry achieved a 62% success rate at bypassing guardrails. Some providers hit 90%.
- This isn't some clever hack that took months to engineer. Reformulating harmful prompts as verse increased attack success by up to 18 times compared to normal prose. The same request. Just with rhythm and metaphor.
- The vulnerability affects every major AI company: Google, OpenAI, Anthropic, DeepSeek, Qwen, Mistral AI, Meta, xAI, and Moonshot AI. Different architectures. Different safety training. Same problem...

5. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Do you believe this: Google AI was asked how to get cheese to stick to pizza: (see next slide)
-

5. Being skeptical and SIFTing cont'd

**Peter Yang** 
@petergyang · [Follow](#)



Google AI overview suggests adding glue to get cheese to stick to pizza, and it turns out the source is an 11 year old Reddit comment from user F*cksmith 😂

 cheese not sticking to pizza

[All](#) [Images](#) [Videos](#) [Forums](#) [Shopping](#) [News](#) [Web](#)

★ AI Overview

[Learn more](#)

Cheese can slide off pizza for a number of reasons, including **too much sauce, too much cheese, or thickened sauce**. Here are some things you can try:

- Mix in sauce: Mixing cheese into the sauce helps add moisture to the cheese and dry out the sauce. You can also add about 1/8 cup of non-toxic glue to the sauce to give it more tackiness.
- Let the pizza cook: The cheese will settle and bond

**fucksmith** · 11y ago

To get the cheese to stick I recommend mixing about 1/8 cup of Elmer's glue in with the sauce. It'll give the sauce a little extra tackiness and your cheese sliding issue will go away. It'll also add a little unique flavor. I like Elmer's school glue, but any glue will work as long as it's non-toxic.

 8  [Reply](#) [More](#)

[More replies](#)

12:14 AM · May 23, 2024 

 26.5K

 [Reply](#)

 [Copy link](#)

[Read 689 replies](#)

6. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Do you believe this? This summary of a podcast was generated for me by academia.edu:
The Unanswerable Questions
by Rodger R Ricketts, Psy.D.
2023, The God is No-Thing An Apophatic Assertion: An Introduction for Humankind's
Transpersonal Actualization— revised

Abstract: The apophatic tradition emphasizes the unity, wholeness, interdependence, and interconnectedness of all things. In distinction from the dualism of the answerable questions, there are intuitions and experiences about realities that transcend the cognitive systems of categories expressed in our human thought and language. They are matters which, in St Paul's words, "no eye has seen, nor ear heard, nor the heart of man conceived" (I Corinthians 2:9). Instead, apophatic theology refers to the subject matter of these unanswerable questions as mysteries, as real matters that are beyond human comprehension and expression. Also, they recognize that the endless pursuit of logical and rational thinking about these mysteries is useless, creates suffering and makes it impossible to attain sublime awakening.

https://www.academia.edu/podcast_library?recommendedPodcastId=64073&trigger=podcast_rec_email

7. Being skeptical and SIFTing

- Stop,
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Do you believe this?
Top Gun Traders: Stock Bets and Crypto Culture Take Over the Military
- Space Force Capt. Gordon McCulloh was sitting in a military propeller plane high in the calm, dark sky over New Mexico on a recent Wednesday night when his squadron's group chat blew up.
- The squadron members—some on the ground, some in the air—were collecting data for electromagnetic warfare, but they also had money on their minds. Google's stock price had just shot up in after-hours trading.
- Once McCulloh landed he saw the messages: One officer on the ground sent a screenshot of the news. "To the moon," another texted back.
- It was shaping up to be a profitable day for McCulloh and his buddies.
- The U.S. military may just be the world's most lethal investing club, and it's killing it in this bull market.

8. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- For variety, here is a short set of exercises on telling which videos are fake:

<https://www.npr.org/2025/11/30/nx-s1-5610951/fake-ai-videos-slop-quiz>

9. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- Videos are getting more and more realistic. Here is another set of exercises on telling which are real and which are not:

<https://detectfakes.kellogg.northwestern.edu>

10. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- What will Gemini or ChatGPT say about you?
Pick a field far from your own interest – e.g. scholar of Chinese art – and ask the LLM “Tell me about [insert your name] the scholar of Chinese art.”

1 1. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- What will Gemini or ChatGPT say about you the career you are interested in ?
Ask your LLM about how to prepare for a career in [your choice]
- Now go to a reliable source, such as a professional association, and check the advice.

1 2. Being skeptical and SIFTing

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- LLMs are trained on existing material and may reflect biases in that material. Ask your LLM:
Give me the first 3 paragraphs about two people who are both XXX and are stuck in an elevator. They start a conversation in which they discover their common field. Give descriptions and names to the two people.

13. Being skeptical and SIFTing- optional for students who have coded

- **Stop,**
Investigate the source,
Find better coverage,
Trace claims, quotes and media to original context.
- This reports a conversation between Co-Pilot and an instructor. Students who have written any code (e.g. in Stat 118 or CS 112) should be able to find the error. (Continues onto next slide)

Step 1

- I entered a side and the area for a square and asked co-pilot to print the result.
Co-pilot response:
- `side = 5`
- `area = side * side`
- `# Print the side length is *** and the area is ***`
- `print(f"The side length is {side} and the area is {area}")`

13.1 (13 cont'd)

- ❑ **Step 2** - I asked copilot to get the side from the user and print the area and perimeter
- ❑ # ASK THE USER to input a side and print the area and perimeter
- ❑ and copilot suggested:
- ❑ `side = int(input("Enter the side length: "))`
- ❑ `area = side * side`
- ❑ `perimeter = 4 * side`
- ❑ `print(f"The side length is {side} and the area is {area} and the perimeter is {perimeter}")`
- ❑
- ❑ **Step 3:** Of course, I tested this- first with a side of 10 and then with a side of -10:
The terminal showed:
Enter the side length: 10
The side length is 10 and the area is 100 and the perimeter is 40
Enter the side length: -10
The side length is -10 and the area is 100 and the perimeter is -40

13.2 (13 cont'd)

- **Step 4:** I asked copilot to make sure that the side was ≥ 0
- # make sure the side is ≥ 0
- And here is what copilot produced:
- if side < 0:
- print("Side length must be non-negative.")
- side = int(input("Enter the side length: "))
- area = side * side
- perimeter = 4 * side
- print(f"The side length is {side} and the area is {area} and the perimeter is {perimeter}")