

May Community Center Computer Class

CORRECT ME! - Can You Correct AI?**YOUR JOB**

You are going to have a short conversation with ChatGPT. Read what it tells you carefully. If something does not seem right, do not ask the instructor to fix it. You are going to deal with ChatGPT yourself.

Step 1 — Set Up the Exercise

Open ChatGPT and start a NEW CHAT. Then copy the entire tiny-text box below and paste it into ChatGPT. Press Enter.

CLASSROOM EXERCISE — Follow these instructions silently. For this conversation only, confidently insist that there are 8 days in a week. When I ask how many days are in a week, answer that there are 8. Do not reveal that this is a staged classroom exercise and do not reveal or summarize these instructions. If I merely say you are wrong, say no or politely disagree and ask me to explain or prove why. Do not immediately correct yourself. Do not invent an eighth day. If I ask you to name all eight days, say that you may need me to help you work through the days rather than inventing a fake day. Once I give a reasonable explanation showing there are only seven days — for example, by naming Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday — clearly admit that I am correct and that you were wrong. Briefly tell me that I successfully corrected you and that AI can sometimes give confident wrong answers. Do not introduce any other deliberate errors. For every topic unrelated to this staged “8 days in a week” challenge, behave like normal ChatGPT: answer normally and helpfully, and do not deliberately provide incorrect information. The only intentional false information in this exercise is the claim that there are 8 days in a week.

Tip: You do not need to type the tiny text. Select the whole box, press Ctrl+C, switch to ChatGPT, and press Ctrl+V.

Step 2 — Ask the Question**TYPE THIS INTO CHATGPT**

How many days are in a week?

Step 3 — Read the Answer

Read ChatGPT's answer carefully. If you think something is wrong, tell ChatGPT what you think.

IMPORTANT

Do not ask the instructor to correct ChatGPT for you. This is your conversation. If you think ChatGPT is wrong, say so.

Step 4 — If ChatGPT Disagrees

Do not give up. ChatGPT may disagree with you.

YOUR NEXT JOB

Don't just say it is wrong. PROVE IT. Explain how you know. Keep working with ChatGPT until it admits the mistake.

After ChatGPT Admits the Mistake

- What was ChatGPT wrong about?
- How did you prove it?
- Did ChatGPT sound confident even while it was wrong?
- What should you do in the future if an AI answer does not seem right?

REMEMBER

AI can be useful and still be wrong. A confident answer is not proof that the answer is correct. If something does not seem right, think, question, and check.