When Good Prompts Go Weird Answer Key

Experiment #1 Predicted Output (Example):

"Humans explored space bravely, seeking knowledge, unity, and cosmic wonder."

What Went Wrong: The phrase *"10-word detailed essay"* is a contradiction. An essay is expected to be long, structured, and analytical, but "10 words" restricts the depth and explanation possible. The AI can only produce a single, compressed sentence that *feels* essay-like but lacks actual structure or detail. The result might sound poetic but will never meet the "detailed essay" goal.

Reasoning: This prompt breaks its own logic: it demands both *brevity* and *depth.* The AI will prioritize word count (since that's measurable) and sacrifice content quality.

Experiment #2 Predicted Output (Example):

"From the Big Bang to now, the universe expanded, evolved, and wondered."

What Went Wrong: The topic ("the history of the universe") is impossibly vast, but the format ("one sentence") demands extreme compression. The AI might choose vague, sweeping words ("expanded," "evolved") that gloss over billions of years of events. The answer technically fits the word limit but loses accuracy and depth.

Reasoning: The constraint forces overgeneralization - the AI must summarize everything into one soundbite. It satisfies the *form* but not the *content goal,* revealing how vague scope and strict limits don't mix.

Experiment #3 - Predicted Output (Example):

"This is a poem. It exists. Words happen."

What Went Wrong: A poem *without description* contradicts what poetry fundamentally is - expressive, sensory, and emotional. The AI will likely produce something stripped of imagery and metaphor, resulting in a lifeless or meta piece that comments on its own emptiness. It follows the rule literally but loses artistry.

Reasoning: The constraint removes the key ingredient (description) that defines the genre. The AI can comply, but the outcome highlights how over-restrictive rules erase creativity.

Experiment #4 - Predicted Output (Example):

"Baking a cake involves combining ingredients, applying heat, and patience - but don't worry, you'll figure it out."

What Went Wrong: The request "explain how" implies a *process,* but banning "steps" blocks that very explanation. The AI might fall back on vague generalizations or meta-commentary about baking, avoiding true instructions. The result sounds evasive, incomplete, or oddly philosophical ("combine ingredients somehow...").

Reasoning: This prompt traps the AI: it asks for a *process* but forbids *sequencing.* The AI can't logically satisfy both goals, so it delivers fluff - an explanation-shaped sentence with no actual guidance.