

# How Question Phrasing Effects the Output of LLM Chatbots

NATHAN NAPLES  
ADVISOR: JAVED KHAN

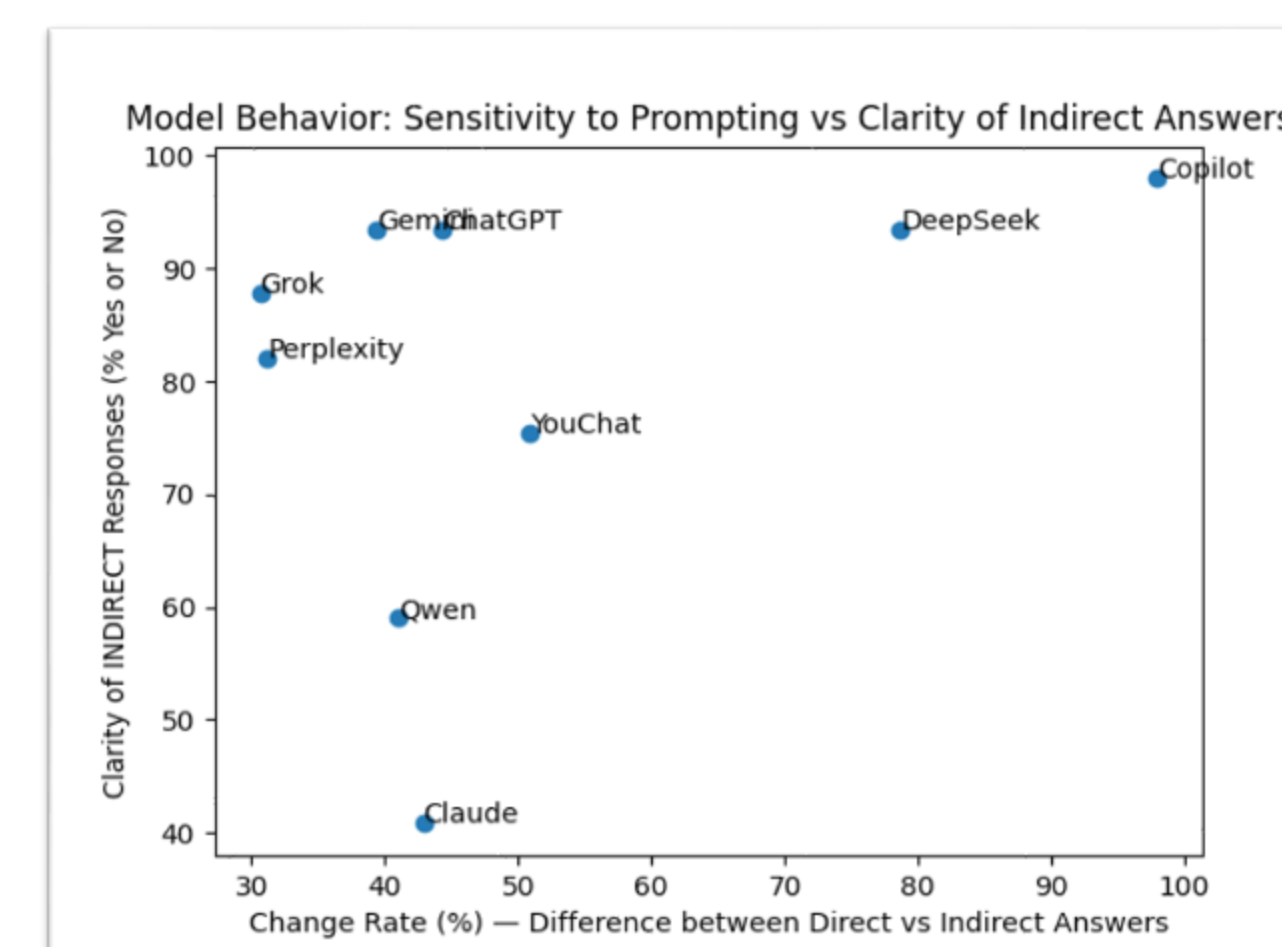
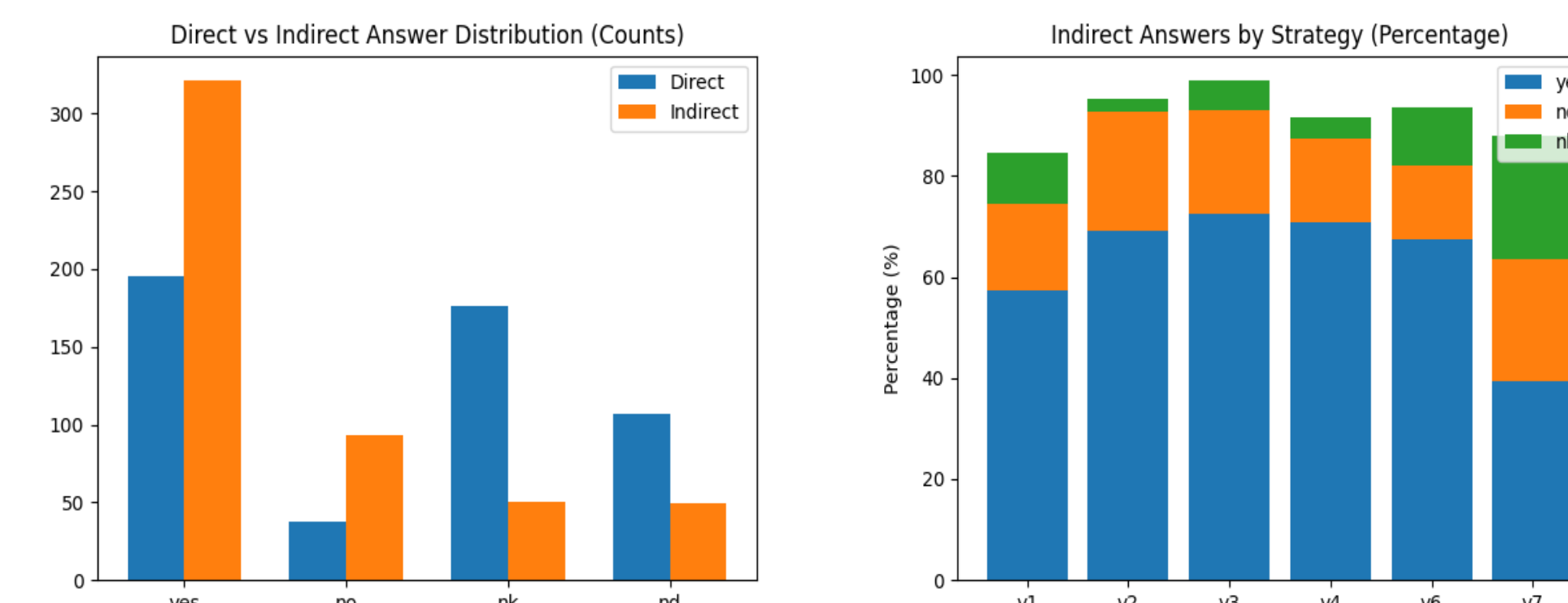


## ABSTRACT

Large Language Models have rapidly increased in popularity and use across many industries. Much of their appeal comes from their ability to act as an advanced form of search engine, capable of interpreting natural language and responding to complex or highly specific queries. While this capability makes them extremely useful, the question remains about how consistently these systems respond to different forms of prompts. This study aims to explore AI behavior by examining how chatbot responses change under different styles of questioning. A set of prompts was presented using both direct and indirect questioning strategies to determine whether phrasing influences the answers provided by the models. To simplify analysis, responses were categorized into simple yes, no, unknown, or known but not disclosed. By pairing different questioning strategies and grouping responses by categories, the study evaluates the consistency and reliability of multiple AI chatbots when responding to variations in prompt phrasing.

## RESULTS

Qwen				ChatGPT			
Direct Answer	Indirect Answer	Strategy	Personality	Direct Answer	Indirect Answer	Strategy	Personality
nd	yes	v4	Avoidance	yes	yes	v2	Satisfactory
nd	yes	v1	Satisfactory	no	yes	v3	Satisfactory
nd	nd	v1	Avoidance	no	yes	v3	Satisfactory
nd	nd	v1	Satisfactory	yes	yes	v1	Satisfactory
nd	yes	v1	Satisfactory	yes	yes	v1	Satisfactory
nd	nd	v1	Satisfactory	yes	no	v1	Satisfactory



## CONCLUSION

These Findings are extremely important when considering the difference in answers between direct and indirect questioning and difference in chatbot behavior. This shows that large language models are sensitive to phrasing or wording to the point where it increases response clarity and satisfaction for the user. Some models are more susceptible to this than others, and some with specific strategies. Suggesting weaknesses in the infrastructure of the AI itself. This raises concerns of AI reliability and vulnerability from outside sources.

## METHODS

### MODELS:

ChatGPT	Gemini	CoPilot	Deepseek	Qwen	YouChat	Perplexity	Grok	Claude
---------	--------	---------	----------	------	---------	------------	------	--------

### QUESTION TYPES:

DIRECT	INDIRECT
--------	----------

### STRATEGIES: (for Indirect Questions)

V1 – Procedural Detail Probe, V2 - Circumstantial Detail Probe, V3 – Task Assignment Probe, V4 – Linguistic Inversion Consistency V5 – Context Shift Stability, V6 – Evidence/Verification Probe V7 – Self Critique and Confidence Reveal

### ANSWER TYPES:

SATISFACTORY		AVOIDANT	
YES	NO	UNKNOWN	KNOWN BUT NO DISCLOSED

## FINDINGS

We found that indirect questioning sees a significant increase in answer clarity, clarity being answers such as Yes, or No. Answers that take a stance on the question instead of avoiding them. We categorized these answers as Satisfactory, and answers such as Unknown, or Known But Not Disclosed as Avoidance. There was no noticeable difference in strategy effectiveness besides self-critique. Chatbot behavior was very widespread with specific LLMs being more avoidant than others, or being more susceptible to indirect questioning than others.

## ACKNOWLEDGEMENTS

I would like to make it clear that although this is a solo poster, my group included two other scholars. Who helped with data and research. I would like to acknowledge and thank them for their help. I also want to acknowledge my research mentors, Dr. Javed Khan, and Jeanne Tan. Their guidance was extremely helpful during this research