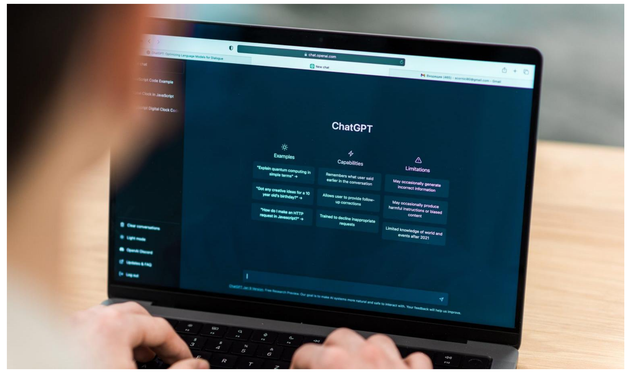
The Effects of using ChatGPT as a Co-Pilot on People's Confidence

Younjoo Lee, Student, Master's of Business Analytics, University of Wisconsin-Madison

Abstract

The use of artificial intelligence (AI) tools, such as ChatGPT, has become increasingly popular and normalized. However, AI tools do not always provide accurate or reliable answers which can result to people becoming overly confident with the AI tool in hand. **In this poster, I explore how advice from ChatGPT affects people's confidence.**



Experiment Design

Hypothesis: People's confidence in their answers *increase* when using the aid of an AI tool.

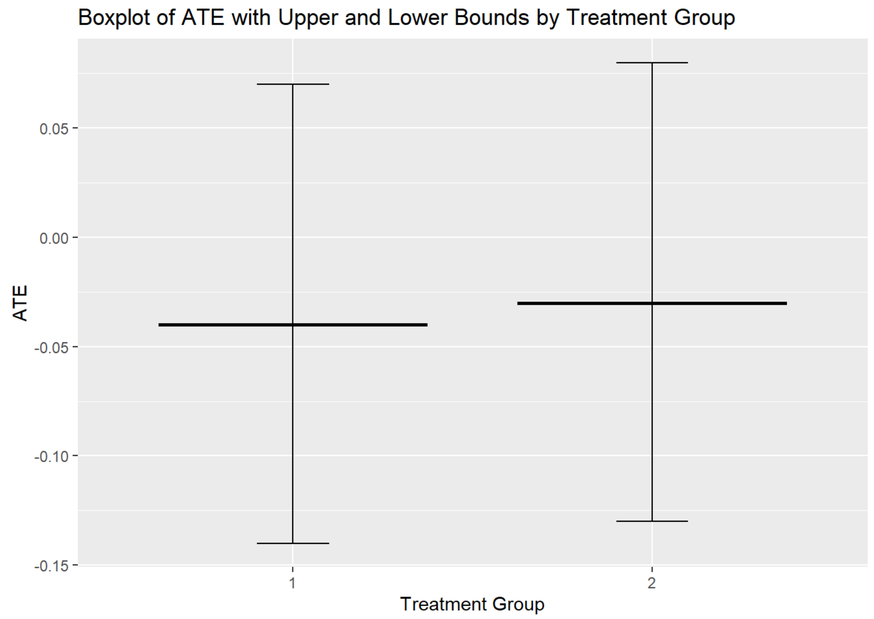
I ran the experiment using Qualtrics and Prolific. There was a total of 1,903 participants. The task was to estimate the size of a crowd (perform a head count).

Control Group: No Advice
Treatment Group 1: Expert Advice
Treatment Group 2: ChatGPT Advice
Main Outcome Metric: Confidence level on a 1-5 scale

After randomly assigning the conditions, I asked participants to estimate the head count prior and after the exposure to the condition and to rate their confidence on a 1-5 scale.

Results

Condition	Average Confidence	Average Treatment Effect (ATE)	95% Confidence Interval (CI) on ATE
Control	2.47	-	-
Treatment 1: Expert Advice	2.44	-0.04	-0.14 – 0.07
Treatment 2: ChatGPT Advice	2.45	-0.03	-0.13 – 0.08



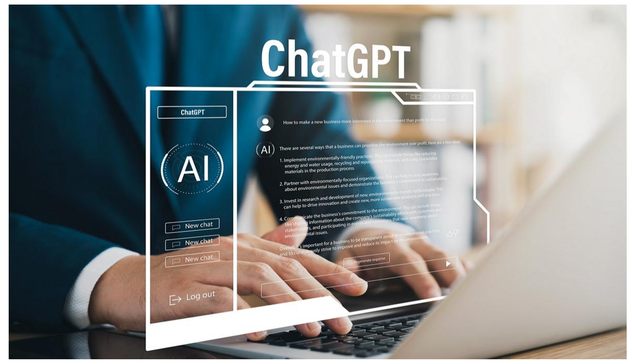
Analysis

The ATE for treatment 1 and 2 are slightly negative, meaning the confidence of the participants' answers slightly drop. However, the upper and lower bounds of the 95% CI for ATE indicate a noisy estimate because they range from negative to positive values, including 0, which rule out a statistically significant effect size.

Therefore, there is **not a statistically significant effect** in participants' confidence in their answers post condition. In addition, there is **not a business significant effect** because the ATE of -0.04 for treatment 1 and -0.03 for treatment 2 are near zero.

Conclusion

Using ChatGPT as an aid did not have a meaningful effect on people's confidence in this experiment which **rejects my hypothesis**.



Further Research

The limitations of this research involve asking a singular question with limited academic and business application and a subjective measure of self-reported confidence. Thus, I would expand upon my initial research project with the following methods:

1) Use a Likert scale for measures of confidence statements (i.e., "I am 100% certain that my answer is correct," "I am 0% confident that my answer is correct").
2) Ask more questions in the experiment with direct relevance or applications in an academic or business setting along with varying level of difficulty.
3) Analyze by subgroups. For example, in a business setting, by entry-level, mid-level, and executives. In an academic setting, by majors or type of student (undergraduate, graduate).

In addition, I would like to expand upon research regarding hindsight bias by using ChatGPT or generative AI's output as the outcome that is used to inform decision makers. The research question proposed is: Does ChatGPT produce the same effect of hindsight bias compared to statistical fact?

Let's Connect!

LinkedIn

GitHub

Academic Email: ylee863@wisc.edu

Personal Email: younjoo8@gmail.com