

# UC Riverside

## UCR Honors Capstones 2023-2024

### Title

SEARCHING FOR REFINEMENT

### Permalink

<https://escholarship.org/uc/item/9pb3h1cx>

### Author

Chow, Josalyn S

### Publication Date

2024-07-24

# SEARCHING FOR REFINEMENT

By

Josalyn Syut-Yin Chow

A capstone project submitted for Graduation with University Honors

May 24, 2024

University Honors  
University of California, Riverside

APPROVED

Dr. Rich Yueh  
School of Business – Department of Information Systems

Dr. Thomas Kramer  
School of Business – Department of Marketing

Dr. Richard Cardullo, Howard H Hays Jr. Chair  
University Honors

## ABSTRACT

This research tests whether the responses from utilizing ChatGPT can have more refined outputs compared to regular search engines when participants receive prompt-coaching. Receiving prompt-coaching means participants are provided tips and guidance that will, in theory, enhance their ChatGPT searching abilities. There will be four groups of participants—the control group, the Google search-engine users group, the ChatGPT users without prompt-coaching, and ChatGPT users with prompt-coaching. All participants will be randomly placed into a group and be asked to take a pre-survey, conduct a search activity on a randomly generated topic, complete a scored quiz on the same topic, and fill out a post-survey. Although all groups will complete a different search activity and be quizzed on different topics, the biggest differences in the search task will be between the control group, who will only partake in a search activity without the quiz, and the ChatGPT group with prompt-coaching, who will receive an extra section of guidance on how to enhance their inputs for improved ChatGPT answers before completing their search activity. There are three main hypotheses within the study. The first and second hypothesis states that ChatGPT users will perform better on the quiz and perceive the quiz to be easier than the control group, the Google-only users, and the ChatGPT users without prompt-coaching; the last hypothesis states that ChatGPT users will perceive a higher sense of flow than the control groups during their search compared to the control group, the Google-only users, and the ChatGPT users without prompt-coaching. Results revealed a statistically significant difference between the control group and the ChatGPT with prompt-coaching group, demonstrating that utilizing ChatGPT properly after realizing effective tips can improve participant learning and knowledge. However, no other statistically significant differences were discovered between any other remaining conditions. Additionally, the lack of statistical

significance between the quiz scores of the ChatGPT users compared to the other conditions and the responses of participants on the match questions indicate they did not perceive the questions to be easier. Lastly, the absence of enjoyment and satisfaction found in the post-survey implied an inability to enter the state of flow.

## ACKNOWLEDGMENTS

I would like to acknowledge my two incredible mentors who provided me with great assistance throughout the journey of my capstone, Dr. Rich Yueh and Dr. Thomas Kramer. Thank you for the numerous hours dedicated to not only helping me refine my work, but guiding me to the completion of my project – I could not have done it without you both. I would also like to acknowledge my parents and my younger sister, who provided me with encouragement and support during times of uncertainty and struggle with my project. Their encouragement motivated me to continue working harder and bring this project to life. Finally, I am deeply grateful to all who expressed their support, assisted me during the process of the project, and crossed the finish line with me in the end.

## TABLE OF CONTENTS

Abstract.....	2
Acknowledgements.....	4
Introduction.....	6
<i>Hypotheses</i> .....	11
Methods.....	13
<i>Participants</i> .....	13
<i>Materials</i> .....	14
<i>Procedures</i> .....	16
Results.....	20
Discussion.....	22
<i>Limitations</i> .....	24
Conclusion.....	25
References.....	27

## INTRODUCTION

### *Literature Review*

In November 2022, ChatGPT, a form of generative artificial intelligence (AI), was released for public use. ChatGPT is an artificial intelligence chatbot developed by OpenAI that allows users to have conversations of desired length, language, style, format, and detail, specializing in understanding the nature of human-like conversations. Originally, the knowledge of ChatGPT was limited to world events leading up to September 2021. As of 2024, ChatGPT has knowledge of information and world events pertaining to January 2022. Although ChatGPT has many functions for varying users, it is commonly used in education by all types of students and staff.

More specifically, aside from a usability standpoint, the type of users and willingness to use ChatGPT must be taken into consideration. Currently existing research involves many studies utilizing ChatGPT within the medical field, meaning that the users who test ChatGPT are often medical students and educators. In a medical study surveying medical students, residents, and faculty, ChatGPT was asked to create questions about using the AI chatbot for education and research. The results revealed that only 40% of participants had tried using ChatGPT, and from that percentage, medical students and residents were significantly more likely to use ChatGPT for research compared to the faculty. In fact, 64.2% of medical trainees used ChatGPT while only 31.1% of medical staff tried the chatbot ([Hosseini, 2023](#)). Even though the participants only expressed a “somewhat” interest with slight uncertainty on using ChatGPT, the AI tool is still new technology that will most likely grow in popularity as improvements are made. In addition, while usage level has not yet been tested in academic institutions, the study demonstrates that students remain a large group using ChatGPT. If ChatGPT is going to be utilized in an

educational setting, especially individuals who are in the process of learning, it is important to ensure that the answers generated by the chatbot are quality information by measuring its performance and abilities.

Although ChatGPT is a relatively new AI tool, experts have already begun conducting extensive research regarding its training and algorithm. In fact, the current research in the field heavily encompasses topics that discuss the performance of ChatGPT, the results the chatbot produces, the ethics of ChatGPT, and the overall pros and cons, particularly in an academic and medical environment. One such benefit produced from the rise of ChatGPT is its performance ability, specifically regarding the accuracy in question responses. In the context of studying the most frequently asked andrology-related questions collected from health forums, hospital websites, and social media platforms, 108 out of the 136 questions fed into ChatGPT replied with outputs that met the set criteria ([Caglar et al., 2023](#)); that is equivalent to an 87.9% accuracy. Even though there was a 9.3% error, researchers noted that all responses were answered at least partially correct. In addition, the reproducibility rate was 90%. Such high statistical accuracy and consistency seems to demonstrate a strong reliability factor within ChatGPT. Furthermore, ChatGPT has several other beneficial capabilities beyond simply providing the user with accurate information, such as its ability to remain efficient, provide instantaneous exchanges, utilize case-based learning, translate passages of other languages, and simulate the roles of others ([Smith et al., 2023](#)). For example, in the specific context of medicine, ChatGPT is able to act the role of a patient and converse with the user through instantaneous exchanges to practice a real-life situation of communication between a patient and a doctor. The AI chatbot can also hold casual conversation and simplify technical terms, making difficult and complex concepts much



easier to understand. ChatGPT has also been determined to assist educators in the classroom, such as generating high-quality word banks that can be used for student assessments ([Han et al., 2023](#)). With such features utilized in an academic setting, ChatGPT can assist students in improving their subject skills, learning efficiency, provide more educational availability, and motivate students to progress their learning ([Smith et al., 2023](#)).

However, there are mixed reviews with the credibility, accuracy, and reliability of the AI chatbot tool. Despite its many capabilities, many of ChatGPT's abilities require refinement and enhancements. For instance, while ChatGPT is able to translate foreign languages, words in certain languages may have several or ambiguous definitions, resulting in difficulty capturing the true meaning of the statement ([Smith et al., 2023](#)). The difficulty ChatGPT has may cause issues with the sentence structures or produce grammatical issues; even though ChatGPT may function better than Google Translate, it still has its errors. Additionally, because ChatGPT was trained not only through large amounts of existing online data, but also partially by using supervised learning, the outputs generated can lead to potential biases. Supervised learning is a method of perceptron learning that requires human intervention and utilizes back propagation, which is where prediction errors are fed into the network as inputs to make adjustments and corrections so future responses can be improved. However, such learning methods and information extracted from the internet used to train the AI model are all produced by humans, which creates the possibility of subjectivity within the responses. Due to the subjective answers, some responses produced may be considered biased or discriminatory towards minority groups with a priming effect based on its wording ([Smith et al., 2023](#)).

Another predicament lies within ChatGPT's tendencies in generating nonsensical information or answers. Although the responses may sound accurate, there is no way of determining how ChatGPT has arrived at certain conclusions. Known as hallucination, the term explains that there are times when AI generates "statistically plausible" information that is actually inaccurate and misleading in machine learning ([Han et al., 2023](#)), despite being based on potentially real existing research. Relying on ChatGPT for knowledge but failing to know when the chatbot is being truthful due to lack of knowledge poses a huge credibility challenge for the AI. Even more than generating half-truths, there are moments when ChatGPT answers with completely fictitious responses. In one study, researchers wanted to test ChatGPT's accuracy rates by asking the chatbot to first write an article on "facial identification using masks" and then provide a list of sources that it referenced. However, when the researchers were reviewing the list, they were unable to find any of the references within the field of existing literature ([Guleria et al., 2023](#)), meaning that the sources were falsified. When entered into GPTZero, the article was identified as AI written text, but the references were not; this indicates that GPTZero is unreliable because it believes the references are authentic while they are actually unidentifiable in the existing literature when checked. In addition, the article failed to detect any plagiarized content using the "Urkund Plagiarism Detection software" ([Guleria et al., 2023](#)). The results from GPTZero and the plagiarism detection software tell the researchers that everything in ChatGPT is originally generated content, which is why it makes sense that the sources would be "original," or fake, as well. While there are some studies that have shown that ChatGPT can generate accurate and real references, it cannot be fully and completely trusted. During another study, researchers found that eight out of the ten, or 80%, of the reference books generated linked to existing books online ([Giray, 2023](#)), but the other 20% of inaccuracy must be accounted for; moreover, many of the

generated web pages did not lead to any real sites ([Giray, 2023](#)). Although a 20% inaccuracy rate may appear to be a small worry, the severe implications of mistakenly citing a fake reference in an academic environment or noting down overall wrong information must be considered. The inaccuracy of the references not only demonstrates a lack of reliability and trustworthiness, but also shows that artificial intelligence has significant room for improvement before it becomes a revolutionary piece of technology that can be utilized successfully without checks.

While there is in-depth research on the benefits and drawbacks of ChatGPT, the field lacks research on potential methods that can help enhance the accuracy, credibility, and reliability of the AI chatbot that are unrelated to the involvement of developing algorithmic improvements. Although there are few experimental research studies, there are, however, theories on steps the user can take to improve their search outputs. Known as the prompt engineering, it introduces a “user-friendly method to optimize interactions with generative AI models, particularly for beginners” ([Lo, 2023](#)). To maximize the potential of coherent AI-generated content that can better adhere to the needs of the user, prompt engineering is an important skill to master. An intentional form of prompt engineering is a theory called the CLEAR framework model, which stands for concise, logical, explicit, adaptive, and reflective. In the five aspects, “concise” represents maintaining clarity and discarding superfluous words while “logical” emphasizes structure and comprehension within the prompt. The “explicit” aspect focuses on the specifications and precise inputs in order to receive the most optimal outcome, and “adaptive” highlights the way prompts can be flexible and creatively customized in terms of word choice and phrases. Lastly within the CLEAR framework, “reflective” discusses analyzing the user’s past prompts and making adjustments where improvement is necessary. Evaluating prompt

history is vital to ensuring that users can learn how to write future inputs that will produce further enhanced responses ([Lo, 2023](#)).

As mentioned previously, the CLEAR framework only remains as suggestions to assist users in improving their prompt inputs, but has never been tested. Because ChatGPT research mainly discusses benefits and issues within the actual chatbot, my study aims to fill gaps in the field where users can be the change. In fact, I question whether inconsistencies and differences in outputs between various studies can be attributed to the way participants are writing their prompts. Although I cannot claim that providing guidance and learning tips can correct the errors and perfect reliability within ChatGPT, I believe that users will see increased refinement in their outputs if they implement the given methods. The overall goal of my study is to assess whether prompt-coaching similar to the CLEAR framework can enhance user search quality.

### *Hypotheses*

In this study, there are three sets of hypotheses (H1, H2, H3) and three sub-hypotheses within each set (a, b, c).

**H1a:** ChatGPT users will perform better on the quiz than the control group.

**H1b:** ChatGPT users will perform better on the quiz than the Google-only users.

**H1c:** ChatGPT users with prompt coaching will perform better on the quiz than ChatGPT users without prompt coaching.

**H2a:** ChatGPT users will perceive the quiz to be easier than the control group.

**H2b:** ChatGPT users will perceive the quiz to be easier than the Google-only users.

**H2c:** ChatGPT users with prompt coaching will perceive the quiz to be easier than ChatGPT users without prompt coaching.

**H3a:** ChatGPT users will perceive a higher sense of flow than control groups during their search.

**H3b:** ChatGPT users will perceive a higher sense of flow than Google-only users during their search.

**H3c:** ChatGPT users with prompt coaching will perceive a higher sense of flow than ChatGPT users without prompt coaching.

In regards to H1 and H2, both hypotheses predict that ChatGPT users will perform better and perceive the quiz to be easier than those who are utilizing a search engine. Such predictions are based on the way ChatGPT functions, as the AI chatbot is able to immediately generate information summaries of all kinds and provide explanations in simple-to-understand terms from the input. The capabilities of ChatGPT allow users direct access to answers whereas looking on the Google engine requires users to sift through the search results. In addition, searching on Google may not always provide clear answers, which can be limiting in a timed quiz. Because our prediction is that users will be able to find the answer to their search topic at a quicker rate on ChatGPT than on Google, the participants will be able to learn the information faster, maintain higher accuracy on the quiz, and have better overall performance results. Being able to find answers with less difficulty should also allow participants in the ChatGPT condition to perceive the quiz as easier than those in the Google search engine condition.

The flow theory that H3 refers to was developed by Csikszentmihalyi, which explains the state that individuals enter when they become fully focused and deeply immersed within an activity or task. H3 predicted that the ChatGPT users groups will perceive a higher sense of flow than the other groups due to an important factor of flow, known as challenge-skill balance. Posited by Csikszentmihalyi as one of the basis factors for flow theory, challenge-skill balance refers to whether “our abilities are well matched to the opportunities for action” ([Csikszentmihalyi, 1996, p. 111](#)). For example, if a challenge is too difficult for the individual’s skillset, then the individual may feel anxiety or frustration; on the other hand, if a challenge is overly underwhelming for an individual, they can easily become bored. Simply put, when the challenge fails to meet the skill level of the participant, they will be unable to enter the state of flow. The belief that the ChatGPT user groups in H3 will perceive a higher sense of flow stems from the idea that the AI chatbot can tailor responses to participant skill level of learning. With a simple input command, ChatGPT can re-generate answers to be simpler or with deeper complexity in order to fit the understanding of the user despite the topic the participant must research.

## **METHOD**

### **Participants**

A total population of 98 Business 101 Information Technology students from the University of California, Riverside were sampled as test subjects for the study. The average age of the participants was 21. All participants were within the School of Business and are either junior or senior standing, with the exception of one freshman and one sophomore. More specifically, 81.1% of participants were college juniors, and 16.7% were college seniors. The other 2.2% were either college freshmen or sophomores.

## **Materials**

The material(s) required for this research study include a survey questionnaire designed on Qualtrics, an online survey software.

## ***Pre-Survey***

The pre-survey was created to assess how familiar participants felt with using technology and the internet. There were three main parts pieced together to form the pre-survey, which were the demographic questions, the Innovativeness and Technology Avoidance scale, and the Internet Self-Efficacy scale. The demographic questions were asked to better understand the age range and type of student the participant belonged within, which could provide insight into the generation of technology they grew up with. The other two sets of pre-survey questions were used to determine the relationship the participant has with technology and the internet.

The questions from the Innovativeness and Technology Avoidance scale were extracted directly from the research paper, *Technisierte Umwelten als Handlungs Und Er lebensräume älterer Menschen*, which translates to technological environments as areas of action and experience for older people ([Mollenkopf & Kaspar, 2004](#)). The German item scale was selected from another study researching interindividual differences in the role of self-efficacy and technology avoidance between older technology experts versus non-experts that referenced Mollenkopf and Kaspar ([Jokisch et al., 2020](#)).

The last set of questions regarding the Internet Self-Efficacy scale measures the feelings that the participant holds in regards to their understanding of the internet and technology. More specifically, the questions ask about “confidence” in their abilities. The survey questions were also directly extracted from an article called Internet Self-Efficacy and the Psychology of the Digital Divide, which had a goal of writing a survey on internet self-efficacy with high construct validity and was successful in doing so ([Eastin and LaRose, 2000](#)).

### ***ChatGPT Tutorial***

The ChatGPT tips tutorial was created by researching multiple sources that have experimented with the abilities of the AI chatbot and documented their findings and experiences based on its outputs. Tips that were most commonly found to be repeated in several sources were compiled together and included as part of the tutorial guide.

### ***Topic Selection***

The three selected topics for participants to research and take a scored quiz on were the Big Five psychological personality traits, climate change, and the insula, or insular cortex, in the fronto-parietal opercular region of the brain. All the topics were selected with intention; while they were random, they were chosen with levels of difficulty in mind. As a commonly discussed topic and simple concepts, the Big Five was selected for the easiest level. Since climate change is a topic that is often discussed but still brings a level of unawareness, it was chosen for medium difficulty. Finally, as a highly specific neuroscience topic that is mainly understood after being taught, the insular cortex was selected for the most difficult level. The topic shown to each



participant is randomly generated each time to ensure that the different difficulty levels were tested on all types of students.

### ***Post-Quiz Survey***

The post-quiz survey was designed to measure flow through engagement, focus, and satisfaction. More specifically, the questions were created with Csikszentmihalyi's theory in mind on how to enter the state of flow. One aspect of flow theory states that all the senses of an individual must be fully immersed and concentrated on achieving a goal ([Csikszentmihalyi, 2008](#)), which is why the post-quiz survey includes questions regarding focus and engagement. To gain a better understanding of how focused the participant was, I also included a question asking about external distractions and surrounding environmental factors. Like previously mentioned, entering flow also means having the optimal challenge-skill balance ([Csikszentmihalyi, 1996, p. 111](#)). The challenge-skill balance questions were intended to understand whether the topic and quiz questions matched their researching abilities. Lastly, the last section of post-survey quiz questions asks about satisfaction and enjoyment levels when completing the search task. Pleasure and enjoyment are both feelings that the individual attains when they reach a state of flow as the task involves a sense of calmness, feedback, and sense of control ([Csikszentmihalyi, 1990](#)).

### **Procedures**

Before participants were assigned to conditions, they were presented with the consent form and a set of pre-survey questions. The pre-survey consists of demographic, background, and use of technology questions.

### *Condition 1*

If the participant is placed into the first condition, also acting as the control group, the participant will be provided a link with instructions on how to play the game, Wordle. Their first step is to play Wordle for five minutes and report the amount of Wordles they were able to solve correctly within the given time. When the participant completes the first step, they are randomly assigned one of the three topics to be quizzed on. Once they complete the quiz, the participant should continue to the next page to answer the post-quiz survey that asks questions to measure their feelings of engagement, focus, and enjoyment with the study.

### *Condition 2*

If the participant is placed into the second condition, they will be directed to a page that instructs them to use Google as their search engine when they are presented with two minutes to search about a quiz topic. The page will also provide the participant instructions to answer quiz questions after their search to the best of their ability and follow any other instructions provided afterwards. Below the instructions will be a single question with five options asking the participant how often they use Google for research—never, rarely, sometimes, often, and always. Once the participant proceeds to the next page, they are randomly assigned one of the three quiz topics to research on Google. Guidelines to assist the participants with their search are provided along with the two-minute timer that begins counting down. As soon as the timer finishes, the participant should continue to the next page and complete the five-question quiz to the best of their ability based on the search results they produced. After the participants complete the scored quiz and move to the next page, they will be asked to paste the first three searches they used to understand the quiz topic better. If the participant used less than three searches, they may leave

the other boxes blank. Under the search questions, the participant will be asked if more than three searches were used, and the total number of searches used if there were more than three. Finally, in the last section, participants will be presented with a post-quiz survey that asks questions to measure their feelings of engagement, focus, and enjoyment with the study.

### *Condition 3*

If the participant is placed into the third condition, in addition to those questions, the participant is asked whether they have used ChatGPT before, how many times they have used ChatGPT in the past, and how familiar they feel using the AI tool. Upon completing the pre-survey, they will be directed to a page that provides a tutorial on creating and utilizing ChatGPT for searching information. After the tutorial, the participant will be instructed to prompt-search about a randomly assigned topic they are provided for two minutes using ChatGPT. On the following page, guidelines to assist the participants with their prompt-search are provided along with the two-minute timer that begins counting down. As soon as the timer finishes, the participant should continue to the next page and complete the five-question quiz to the best of their ability based on the search results they produced. After the participants complete the scored quiz and move to the next page, they will be asked to paste the first three prompts they used to understand the quiz topic better. If the participant used less than three prompts, they may leave the other boxes blank. Under the prompt-search questions, the participant will be asked if more than three prompts were used, and the total number of prompt-searches used if there were more than three. Finally, in the last section, participants will be presented with a post-quiz survey that asks questions to measure their feelings of engagement, focus, and enjoyment with the study.

#### *Condition 4*

If the participant is placed into the fourth condition, in addition to those questions, the participant is asked whether they have used ChatGPT before, how many times they have used ChatGPT in the past, and how familiar they feel using the AI tool. After completing the pre-survey, they will be directed to a page that provides a tutorial on creating and utilizing ChatGPT for searching information. Before the official search begins, the participant is pre-assessed by answering a question on how they would choose to prompt-search information on ChatGPT. Afterwards, they are provided with a guide for prompt-coaching tips that teaches the user how to utilize ChatGPT to its full potential as well as write prompts that will refine and produce optimal outputs. When they read through all the tips, the same pre-assessment question will be asked again along with the correct answer and an explanation. Upon completing the ChatGPT prompt-coaching guide, the participant will be instructed to prompt-search about a randomly assigned topic they are provided for two minutes using ChatGPT. On the following page, guidelines to assist the participants with their prompt-search are provided along with the two-minute timer that begins counting down. As soon as the timer finishes, the participant should continue to the next page and complete the five-question quiz to the best of their ability based on the search results they produced. After the participants complete the scored quiz and move to the next page, they will be asked to paste the first three prompts they used to understand the quiz topic better. If the participant used less than three prompts, they may leave the other boxes blank. Under the prompt-search questions, the participant will be asked if more than three prompts were used, and the total number of prompt-searches used if there were more than three. Finally, in the last section, participants will be presented with a post-quiz survey that asks questions to measure their feelings of engagement, focus, and enjoyment with the study.

## RESULTS

The independent variable is the condition the participant is placed in (control, Google search, ChatGPT with no prompt-coaching, ChatGPT with prompt-coaching), and the dependent variable is the score of the quiz. Although there were a total of 98 participants, there were only scores for 85 participants. In the scored quiz, there were a total of five questions, and the total average of scores was 3.25. Within the four conditions, the control group ( $N = 19$ ) scored an average of 2.68, the Google search group ( $N = 22$ ) scored an average of 2.91, the ChatGPT with no prompt-coaching group ( $N = 22$ ) scored an average of 3.55, and the ChatGPT with prompt-coaching group ( $N = 22$ ) scored an average of 3.77. Between all the four conditions, there was a statistically significant difference between the average quiz scores ( $p = 0.02$ ).

After running several multiple comparisons analyses, I was able to uncover which means from the quiz scores differed significantly from one another. While I predicted that the users in the ChatGPT groups (with or without prompt-coaching) would perform better than all the other conditions, that failed to be the case. Even though the ChatGPT group mean scores were higher (highest with prompt-coaching) than all the other conditions, there was no statistically significant difference between the quiz scores of the ChatGPT with prompt-coaching group compared to the Google search group ( $p = 0.106$ ) and ChatGPT with no prompt-coaching group ( $p = 0.930$ ). However, the ChatGPT with prompt-coaching group performed significantly better than the control condition ( $p = 0.032$ ), which matched the prediction of H1a.

The Innovativeness and Technology Avoidance scale from the German paper, *Technisierte Umwelten als Handlungsräume älterer Menschen*, demonstrated no differences

among all four conditions. The scale questions are presented in the form of a Likert scale, starting with “do not agree at all” (1) to “agree entirely” (5). Results show that scores from the four conditions range between 3.7 to 3.9, with control being 3.86, Google search being 3.79, ChatGPT with no prompt-coaching being 3.70, and ChatGPT with prompt-coaching being 3.91. Overall, the averaged score of 3.82 is considered to have a higher statistical significance compared to the midpoint of 3. Regarding the Internet Self-Efficacy scale, the questions were also presented through a Likert scale, but instead ranges from “Strongly Disagree” (1) to “Strongly Agree” (7). In this case, the mean is 5.28 for the control group, 5.21 for the Google search group, 5.06 for the ChatGPT with no prompt-coaching group, and 5.70 for the ChatGPT with prompt-coaching group. There is a marginal statistical significance ( $p = 0.065$ ) between the ChatGPT with no prompt-coaching versus with prompt-coaching. The average amongst the four conditions was 5.31, resulting in a high statistically significant difference in comparison to the midpoint of 4.

In addition, I also ran analyses on the feeling of enjoyment and sense of satisfaction within the participant during their completion of the task. More specifically, the participants not only had a low average of 1.26 for feeling of enjoyment when doing the task on a scale from “No” (1) to “Yes” (4), but also had no statistically significant differences among the four conditions ( $p > 0.520$ ). Next, the participants were generally equal in sense of satisfaction regarding the task, but there was also a low average of 1.57 on the same scale as enjoyment. Similar to feeling of enjoyment, there was no statistically significant difference among the four conditions in terms of satisfaction ( $p = 0.084$ ). Lastly, I looked into a variable I call “match” on a scale of “No” (1) to “Yes” (5), which are two questions related to whether the questions on the quiz matched the

participant's researching abilities, or otherwise known as the searches they made (Question 1) and whether the questions on the quiz matched their abilities to search for the answers (Question 2). Question 1 under match averaged a total of 2.34, with the control group having a mean of 2.36, the Google search group having a mean of 2.17, the ChatGPT with no prompt-coaching group having a mean of 2.58, and the ChatGPT with prompt-coaching group having a mean of 2.23. Question 2 performed similarity with an overall average of 2.45, where the mean was 2.41 for the control group, 2.29 for the Google search group, 2.79 for the ChatGPT with no prompt-coaching group, and 2.27 for the ChatGPT with prompt-coaching group. Any statistically significant difference was not found within Question 1 ( $p = .582$ ) or Question 2 ( $p = .359$ ).

## DISCUSSION

The participants in the ChatGPT groups with or without prompt-coaching scored higher on average than the other two conditions, the control and the Google search group, matching H1a, H1b, and H1c, but most groups failed to be statistically significant from each other. While the higher averages demonstrates that ChatGPT can potentially prove to be a more efficient and useful tool to search for information in a short period of time compared to traditional methods, the lack of statistical significance in most groups shows that there is not enough evidence to make that determination. However, I found that there was a statistically significant difference between the control group and the ChatGPT group with prompt-coaching. Since the search activity of the control group focused on playing as many as Wordles as possible, they were not instructed to use any form of search tool to assist their current state of learning and understanding on the topic before the scored quiz. The statistically significant difference between the control group and the ChatGPT group with prompt-coaching illustrates that using ChatGPT

as a search tool along with teaching tips on how to properly utilize the AI chatbot can actually enhance the learning and knowledge of the user compared to the control group that had a constant, stagnant state of abilities throughout the study.

In the Innovativeness and Technology Avoidance scale included within the pre-survey, no differences were found among the four conditions, meaning that all participants, no matter the group, were similar in the way they felt about technology avoidance. Furthermore, the overall average of 3.82 that is slightly higher than the midpoint of 3 signifies that the participants generally felt more favorable attitudes toward technology use rather than avoidance. While they may have mostly agreed with technological advancements rather than avoidance, their opinion on the topic was not strong. For the Internet Self-Efficacy scale in the pre-survey, the marginal significance discovered between the ChatGPT with prompt-coaching versus no prompt-coaching group implies that there was a slight noticeable difference between the confidence and comfortability level in technology use within those specific participants. Because the ChatGPT with prompt-coaching group averaged 5.70 while the ChatGPT with no prompt-coaching group averaged 5.06, the higher-average group involved with prompt-coaching seems to be more familiar and confident with technology use.

Because the average of the four conditions was 1.26 for enjoyment and no statistically significant differences were found, it means all groups felt a low sense of enjoyment when completing the study. With a low average of 1.57, the same applied to the feeling of satisfaction, meaning that the participants in all conditions did not feel a sense of satisfaction after completing the assigned task. Their lack of enjoyment and satisfaction within their instructed tasks implies that the



participants in all four conditions did not enter a state of flow as our hypothesis predicted. As for the “match” questions regarding whether the quiz questions match the participant’s searches and whether the quiz questions matched their ability to find the answers, the low score from participants on both questions indicates that the quiz level may have differed from their skills. Whether the quiz was easier or more difficult for participants, flow theory states that the challenge level of the task must match the abilities or skills of the participant in order for the individual to enter the state of flow. As a result, participants were most likely unable to experience flow during the course of the study, meaning that H2a-c and H3a-c were incorrect.

### ***Limitations***

While the data collected accurately matched H1a, H1b, and H1c, and specifically had statistical significance for H1a (ChatGPT with prompt-coaching group performs better than control group), there are certain limitations that must be considered when looking at the results. One limitation to this study was that there were only one round of participants that consisted of 98 people, which means the total population sample was lower than expected. Despite originally wanting to run multiple rounds of the study, due to complications occurring during the study, I was only able to conduct one round of surveying. In addition, because I was only able to sample one round of participants, the quiz topics were unable to be tested multiple times to ensure construct validity. More tests should be run in order to measure whether the quiz topics and questions were an accurate representation of difficulty level for participants. Another limitation regarding the participants is that the population sample were only students from the School of Business at UC Riverside. More specifically, they were sampled from the BUS 101 Information Technology Management class, which is mainly a class taken by upperclassmen (juniors and seniors).

Although there are a wide variety of business students from different concentrations who are required to take BUS 101, having a more diverse range of students from several different majors as the type of participants can more accurately and effectively understand whether the results can be generalized or applied to a larger population. Lastly, the questions in the post-survey that were used to measure flow were not a previously validated scale. The survey questions were designed based on the research of other flow scales and flow theory, but were not drawn from a scale that has been demonstrated to measure flow accurately; this means we cannot be fully certain that the enjoyment, satisfaction, and match questions can be used to conclude that the participants entered a state of flow.

## **CONCLUSION**

In conclusion, the goal of my study was to assess whether using ChatGPT can be more effective than using traditional means of searching, such as the Google search engine, especially if the users were provided prompt-coaching guidance and tips. The participants who were assigned to use ChatGPT were also predicted to perceive the task as easier and enter a state of flow compared to those who were using a traditional search engine or no search engine at all. While the results demonstrated a statistically significant difference between the control group and ChatGPT with the prompt-coaching group, no statistical significance was discovered between any other conditions. In addition, due to lack of statistical significance between the low scores of the quiz and the way participants responded that the questions did not match their researching abilities in all conditions, it does not appear that ChatGPT users perceived the quiz to be easier than the other groups. Participants also expressed a lack of enjoyment and satisfaction, most likely being unable to enter a state of flow. In order to better understand if the results can be

considered valid, more rounds of the study should be conducted to determine whether adjustments to the quiz topic or questions must be made to the survey. Furthermore, revisions to the scale of flow should be made to ensure that the scale being used is a valid measure of flow theory and the population sample should be increased as well as diversified. Overall, while parts of the study accurately matched the hypotheses, the study should be conducted again in the future to observe if the results can be replicated and understand what changes must be made to the study if there are any differences in the data.

## REFERENCES

- Calgar, U., Yildiz, O., Ozerverli, M. F., Aydin, R., Sarilar, O., Ozgor, F., & Ortac, M. (2023). Assessing the performance of chat generative pretrained Transformer (CHATGPT) in answering andrology-related questions. *Urology Research and Practice*, 49(6), 365–369. <https://doi.org/10.5152/tud.2023.23171>
- Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience*. Harper and Row.
- Csikszentmihalyi, M. (1997). *Creativity: Flow and the psychology of discovery and invention*. HarperCollins Publishers.
- Eastin, M. S., & LaRose, R. (2000). Internet self-efficacy and the psychology of the digital divide. *Journal of Computer-Mediated Communication*, 6(1), 0–0. <https://doi.org/10.1111/j.1083-6101.2000.tb00110.x>
- Gewirtz, D. (2024, January 24). *How to write better CHATGPT prompts in 5 steps*. ZDNET. <https://www.zdnet.com/article/how-to-write-better-chatgpt-prompts-in-5-steps/>
- Gindham, A. (2023, October 31). *How to write the perfect CHATGPT prompt and become a CHATGPT expert*. The Writesonic Blog - Making Content Your Superpower. <https://writesonic.com/blog/how-to-write-chatgpt-prompts>
- Giray, L. (2023). Chatgpt references unveiled: Distinguishing the reliable from the fake. *Internet Reference Services Quarterly*, 28(1), 9–18. <https://doi.org/10.1080/10875301.2023.2265369>

- Guleria, A., Krishan, K., Sharma, V., & Kanchan, T. (2023). CHATGPT: Ethical concerns and challenges in academics and research. *The Journal of Infection in Developing Countries*, *17*(09), 1292–1299. <https://doi.org/10.3855/jidc.18738>
- Han, Z., Battaglia, F., Udaiyar, A., Fooks, A., & Terlecky, S. R. (2023). An explorative assessment of CHATGPT as an aid in medical education: Use it with caution. *Medical Teacher*, 1–8. <https://doi.org/10.1080/0142159x.2023.2271159>
- Hosseini, M., Gao, C. A., Liebovitz, D. M., Carvalho, A. M., Ahmad, F. S., Luo, Y., MacDonald, N., Holmes, K. L., & Kho, A. (2023). An exploratory survey about using CHATGPT in education, healthcare, and research. *PLOS ONE*, *18*(10). <https://doi.org/10.1371/journal.pone.0292216>
- Jokisch, M. R., Schmidt, L. I., Doh, M., Marquard, M., & Wahl, H.-W. (2020). The role of internet self-efficacy, innovativeness and technology avoidance in breadth of internet use: Comparing older technology experts and non-experts. *Computers in Human Behavior*, *111*. <https://doi.org/10.1016/j.chb.2020.106408>
- Lo, L. S. (2023). The clear path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4). <https://doi.org/10.1016/j.acalib.2023.102720>
- Mirvis, P. H., & Csikszentmihalyi, M. (1991). Flow: The psychology of optimal experience. *The Academy of Management Review*, *16*(3), 636–640. <https://doi.org/10.2307/258925>

Mollenkopf, H., & Kaspar, R. (2004). Technisierte Umwelten Als handlungs- und erlebensräume älterer menschen. *Lebensformen Und Lebensführung Im Alter*, 10, 193–221.

[https://doi.org/10.1007/978-3-663-10615-9\\_10](https://doi.org/10.1007/978-3-663-10615-9_10)

Nield, D. (2024, February 22). *17 tips to take your CHATGPT prompts to the next level*. Wired.

<https://www.wired.com/story/17-tips-better-chatgpt-prompts/>

Smith, A., Hachen, S., Schleifer, R., Bhugra, D., Buadze, A., & Liebreuz, M. (2023). Old dog, new tricks? exploring the potential functionalities of CHATGPT in supporting educational methods in Social Psychiatry. *International Journal of Social Psychiatry*,

69(8), 1882–1889. <https://doi.org/10.1177/00207640231178451>

Terrasi, V. (2023, May 3). *How to write CHATGPT prompts to get the best results*. Search Engine

Journal. <https://www.searchenginejournal.com/how-to-write-chatgpt-prompts/479324/>