MSc Computer Science & Big Data Analytics

ChatGPT: The advancement of knowledge and incorporation for its users.

15/05/2023

Word Count – 3,000

# Abstract

# Introduction

## Background

ChatGPT, widely utilized by the public, corporations, and academia, prompts the question: how valuable is it for users? OpenAI introduced ChatGPT on November 30, 2022, as a conversational AI called ChatGPT (Generative pre-trained Transformer) [1]. Using Natural Language Processing (NLP), it provides comprehensive responses to user queries [2]. ChatGPT finds practical application in various scenarios, such as resolving user uncertainties, enhancing comprehension of topics, and generating stories or poems based on given keywords. While media and academic sources have covered ChatGPT extensively, limited research focuses on user perspective due to its recent release.

## Scope for Research

This proposal aims to examine if ChatGPT enhances user knowledge and its integration into users' routines. Despite research on ChatGPT and its applications, such as solving programming bugs [4], providing education [6,8], and democratising knowledge [9], there is limited research on user perception. Tlili et al. [7] conducted a notable study addressing this gap, but due to ChatGPT's novelty, this proposed study aims to expand on it; aimed at users from the public to gain insights into their interactions and found usefulness.

This quantitative study aims to create a survey to test whether participants have used ChatGPT, if it has improved their knowledge, and if they now incorporate ChatGPT into their knowledge-seeking process.

The study can benefit academics, corporations, and the public by providing insights on the usefulness of ChatGPT compared to or alongside other knowledge-enhancing techniques like Google, YouTube, and social media platforms. It also has the potential to inspire further research in this area.

## Limitations

Due to limited time and resources, only one survey will be conducted with mostly closed questions to gather prompt and enriching feedback. Additionally, the analysis will be limited to the collected data without comparison to other datasets, due to the lack of research in this area. It is important to note that bias may be introduced in this study as participants interested in the topic may be the only ones who choose to participate.

## Ethical Issues

When the survey is conducted, the researcher will ensure informed consent from each participant, clarifying requirements, expectations, and the option to opt out. Participant data will be anonymised to protect identity, with some characteristics (such as age, gender, occupational background, and geographic location) used for research purposes only. The study allows participants of any gender, sexual orientation, race, or belief. No conflicts of interest are anticipated, and there is no expected exposure to risks or dangers for participants.

# Literature Review

## Introduction

ChatGPT – also known as Chat Generative Pre-Training Transformer – is a large language Model (LLM) Artificial Intelligence (AI) tool created by Open AI to answer a user’s request via the use of Natural Language Processing (NLP) tasks [1]. Since its inception, ChatGPT is applied in various practices such as essay writing, fact checking, and idea generation [3]. However, as well as answering these types of user request, does ChatGPT provide further education or knowledge enhancement for its users? Therefore, this literature review aims to examine currently available research – or lack thereof - on whether users get to enhance knowledge further in that area from the perspective of the user, and if so, is ChatGPT being incorporated more into their day-to-day activities an effect.

## Providing knowledge enhancement and day to day incorporation for its users

A study seen from Surameery and Shakor [4], examines the use of ChatGPT in solving programming bugs, both in terms of debugging assistance and explanations. Furthermore, the capabilities and representation of knowledge from this AI tool were explored, concluding that ChatGPT could be a benefit for developers in bug fixing. Additionally, this paper mentions how ChatGPT can be a part of a comprehensive toolkit to solve problems and provide explanation alongside other tools. However, this paper focused on the application use of the software rather than gaining perspective from the user, which could overlook many insights on the user’s enhancement from bug fixing. Furthermore, although this helps to answer whether a user could incorporate ChatGPT into day-to day activities, the limitation of this paper is that no further research was conducted to explore this concept for users.

Aljanabi [5] explains how ChatGPT can provide possibilities and future directions for users and the software itself. Additionally, describing how ChatGPT has potential to provide user personalisation, and alluding to integration for different working fields; yet this paper provides more of a general view on possibilities of the tool, than to see how users interact with it.

An excerpt from Liberman [6] suggests ChatGPT as a tool to explore sources of information for users, including for students as a primary user, to assist with understanding context of a subject. Although, Liberman does go further to say that ChatGPT is not perfect, and there are some areas where it may not provide accurate information for its user. Again, this source provides only a highlight of how it can provide knowledge enhancement and detail is missing to explain this further.

Tlili et al [7] shows a more comprehensive case study on the use of ChatGPT, and its perception from the public, as well as initial user interactions and experiences. A series of analyses were conducted including Twitter network and sentiment analysis on tweets relating to ChatGPT. Additionally, interviews were conducted on participants, quoting that ChatGPT could be used as a learning aid, or idea generation. The paper concluded, stating that ChatGPT is a powerful educational tool, with limitations such as a focus on early adopters of the tool, with a small number of participants. Finally, this paper has performed some great groundwork for this proposal in question and hoping to extend the results from this study further by providing additional insight since its publication.

A study produced by Kung et al [8] explored the performance of ChatGPT on the United States Medical Licensing Exam (USMLE). Although this research’s primary focus was on testing performance on the exam above, a discussion is made on the results on whether the AI tool can assist in knowledge enhancement of users in medical education who are taking such an exam by gaining insight into areas that needs development. However, limitations of this research keep it focused on ChatGPT and its output from USMLE, although alluding to explore both use of knowledge enhancement for medical students taking the exam and further incorporation into day-to-day use. Additionally, no medical student participation is shown against this study to explore the improvement of scoring in the USMLE.

Yue et al [9] explores the use of ChatGPT in providing financial knowledge to non-financial users. The study goes further by stating that ChatGPT has the potential to empower individuals in gaining financial knowledge to help make informed decisions. Additionally, tests were performed on how the AI tool, explains definitions of financial terms and metrics, as well as explaining SHAP scores based on different publication approaches. Whilst the study goes in depth testing the capabilities of ChatGPT and the responses it provides and how this can help others gain new knowledge in the financial space, there is no evidence of testing these prompts against participants and to determine from them whether this information provided is something that can be understood and learnt, or if this can be incorporated into a laypersons day-to-day routine to expand their knowledge further.

## Conclusion

The purpose of the literature in question was to evaluate the current research available on participants who have used ChatGPT, and whether a) it has enhanced their knowledge either generally or within a specific realm, or b) if so, if ChatGPT has been incorporated into their day-to-day activities. It appears though this review, that the use of ChatGPT is still novel, since its release from Open AI late 2022; and whilst at least one paper has made some progress in this area, the belief is that there is a gap within the research space on the accommodation of participants and understanding their interactions rather than separate stand-alone studies of the researcher or topic alone.

# Research Hypothesis and Questions

The following information below represents a set of two alternative hypotheses to be explored as part of the study, where the first hypotheses has an impact on the second hypotheses.

## 3.1. Research Hypotheses

1. The utilization of ChatGPT increases the likelihood of enhancing a user's knowledge on a specific topic.”
2. “Users who perceive ChatGPT as enhancing their knowledge are more likely to incorporate ChatGPT into their routine for knowledge acquisition compared to those who do not perceive ChatGPT as enhancing their knowledge."

Based on these hypotheses proposed, there are potential leading questions that could be answered to re-confirm or extend the response further.

## 3.2. Research Questions

As well as the hypotheses to explore, there is potential to ask additional research question to provide further context in this area.

1. Is ChatGPT being used by the participant?
2. If ChatGPT is being used by the participant, did the use of it enhance a user’s knowledge based on what was submitted?
3. If ChatGPT did enhance a user's knowledge, has ChatGPT been adopted for routine knowledge acquisition by the user?

# Methodology

The goal of this research proposal is to establish whether ChatGPT – if used – has enhanced a user’s knowledge; and if so, has ChatGPT then since been incorporated into a user’s routine for knowledge acquisition.

## Research Design

This dissertation proposal proposes a quantitative study for its speed and ability to address specific questions, allowing for future repeatability and measurement of attitude changes [10-11]. The quantitative method offers objectivity, generalisability to a broader context – enabling a foundation to build further research upon, and the ability to measure the potential cause and effect between users' knowledge enhancement of ChatGPT and its incorporation into daily use. Statistical inferences can also be used to test the hypotheses [11-12]. Qualitative research, although valuable for building themes, providing context, and gaining insights from participants, will be a good candidate for future studies once this preceding research is complete. Due to time and resource constraints, a mixed method approach is not feasible, making the quantitative study the preferred approach.

### Quantitative Research Method

Quantitative research will utilise a survey study methodology to collect information via a questionnaire, asking specific, mainly closed-ended questions and analysing the results [14]. The purpose of the survey is to gather participant feedback on ChatGPT's usage, knowledge enhancement, and adoption. Conducted online, it offers access to a large participant group and ensures an optimal response rate compared to physical submissions [14]. The survey shall be cross-sectional as the research is novel, to establish responses from one point in time [15], that can be compared against in future research.

## Preparation

Before the research can be conducted, the survey needs to be created ready for distribution, aimed to be done within the research project timeline.

## Population

The population of this study consists of all individual responses who have used or interacted with ChatGPT, regardless of demographic. However, user characteristics will be collected to allow for analysis and aid future research. Additionally, research will be carried out as a population-based study – to gather extensive information on the general population and expedites research and analysis. Therefore, at this time a sample will not be selected, and to provide results based on the whole population instead.

## 4.3. Data Collection

As this suggested research involves creation of an online survey, an online survey instrument will be used. This instrument shall be created via a software tool known as Google Forms. With this method, a link will be shared on multiple platforms such as LinkedIn, Discord, Facebook, or email where applicable. The steps projected in this data collection process will be similar to the following:

1. Share a link to individuals across multiple social platforms to access the Google Form survey created. Also, a link to the created Google form will be distributed via email should it be required.
2. The individual answers questions and makes their submission.
3. At the end of the time allotted to receive responses, data shall be exported from the Google Form tool to a downloaded csv file.
4. From there, the csv file will be imported into a python script to produce relevant statistics and analysis.

## Data Analysis and Results

### Management and storage

As described in the previous section, data shall be collected via a Google Forms survey, then exported into a CSV (Comma-Separated Values) file, which then is imported for analysis and statistical generation.

### Software / tools used for Data Analysis

The CSV file will be imported into Python, a user-friendly and interpretable high-level programming language [16]. Python supports data analysis through libraries like pandas [21], SciPy [22], matplotlib [23], and seaborn [24], facilitating statistical analysis and data visualization for the collected data.

### Data Analysis Steps

Whilst the series of analysis may depend on the amount of data collected, the following list of tasks are expected to be executed to present findings and derive insight form the data collection:

#### 1 – Conversion of values

Most participant data collected will consist primarily of categorical responses, including yes or no choices and demographic information selected from drop-down lists (e.g., age bracket, gender, occupational background). Consequently, data conversion is anticipated to transform categorical values into numerical representations, utilising transformative logic.

#### 2 – Descriptive Analysis

Initial analysis will summarise data points using tables and graphs. Examples include total participant count, percentages of population with specific characteristics. Additionally, this analysis will address the main research questions:

1. How many participants found ChatGPT enhanced their knowledge? What percentage of the population reported knowledge enhancement, and the percentage where it did not.
2. Among participants who reported knowledge enhancement, how many now use ChatGPT for knowledge acquisition? This analysis will provide percentages for both the subset and the entire population.

This analysis will also explore participant characteristics for potential insights.

#### 3 – Pearson’s Correlation Coefficient

#### Pearson correlation assesses the strength and direction of a linear relationship between two variables, indicating their closeness [19-20]. This study examines the association between ChatGPT users who have experienced knowledge enhancement and the integration of ChatGPT into their daily knowledge acquisition routine. The `dataframe.corr('pearson')` function in pandas [21] will be employed to generate a correlation table. A positive correlation is anticipated, signifying that users who enhance their knowledge with ChatGPT are inclined to incorporate it into their regular knowledge-seeking practices.

#### 4 - Hypotheses Testing

To assess the statistical significance of the survey results and investigate the two hypotheses, we will employ the binomial test in python using the scipy.stats library package. The null hypothesis, in both cases, will assume that only 50% of the participants reported an improvement in their knowledge and the inclusion of ChatGPT as a tool for regular knowledge acquisition. Should more than half of the responses demonstrate this trend, the null hypothesis will be rejected, and the alternative hypothesis, which aligns with the research theories, will be accepted.

### Results from Analysis

The results shall be displayed in both tabular, graphical – in terms of data visualisations and textual formats for the reader dependant on the context of what is being shown.

## Project Planning and Timescales

The below table gives a expected outline on how long the proposed research project will take, with tasks in their relevant order.

|  |  |
| --- | --- |
| **Task** | **Estimated time to complete** |
| Survey Creation and preparation | ½ Week |
| Survey Distribution | 1 Day |
| Await responses from participants | 1 Month |
| Collection and analysis of responses | 1 Week |
| Finish Writing Dissertation | 1 Week |
| Proof-reading final Dissertation | 1/2 Week |

## Risk Analysis

The table below shows the risks that should be considered with this research, the likelihood of occurrence, the severity of how it could impact the project, and what mitigation measure could be deployed to combat it.

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood** | **Severity** | **Mitigation** |
| Unable to get survey responses via social media platforms | Moderate | Moderate | Curate a email distribution list to send survey |
| Not enough responses before the planned deadline | Moderate | Moderate | Allocate additional research time and repost request to participate in survey |
| Not enough detail found in some responses | Low | Moderate | Remove entries as not enough data to analyse |
| Responses cannot be collected via google forms | Low | High | Allocate additional research time and find alternative survey tool to send out to participants |
| Computer programming software failed / unable to perform analysis | Low | High | Find alternative programming software such as R to perform analysis. |

# Ethical Approval

# References

[1] E. Gregersen, “ChatGPT | Definition & Facts | Britannica,” [*www.britannica.com*](http://www.britannica.com), May 10, 2023. https://www.britannica.com/technology/ChatGPT (accessed May 13, 2023).

[2] M. Rahman, H. J. R. Terano, N. Rahman, A. Salamzadeh, and S. Rahaman, “ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples,” *Journal of Education, Management and Development Studies*, vol. 3, no. 1, pp. 1–12, Mar. 2023, doi: https://doi.org/10.52631/jemds.v3i1.175.

[3] N. M. S. Surameery and M. Y. Shakor, “Use Chat GPT to Solve Programming Bugs,” *International Journal of Information technology and Computer Engineering*, vol. 3, no. 31, pp. 17–22, Jan. 2023, doi: https://doi.org/10.55529/ijitc.31.17.22.

[4] M. Aljanabi and ChatGPT, “ChatGPT: Future Directions and Open possibilities,” *Mesopotamian Journal of Cyber Security*, vol. 2023, pp. 16–17, Jan. 2023, doi: https://doi.org/10.58496/mjcs/2023/003.

[5] M. Lieberman, “What Is ChatGPT and How Is It Used in Education?,” *Education Week*, Jan. 04, 2023. Accessed: May 13, 2023. [Online]. Available: https://www.edweek.org/technology/what-is-chatgpt-and-how-is-it-used-in-education/2023/01

[6] A. Tlili *et al.*, “What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education,” *Smart Learning Environments*, vol. 10, no. 1, Feb. 2023, doi: https://doi.org/10.1186/s40561-023-00237-x.

[7] T. H. Kung *et al.*, “Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models,” *medRxiv*, vol. 2, Dec. 2022, doi: https://doi.org/10.1101/2022.12.19.22283643.

[8] T. Yue, D. Au, C. C. Au, and K. Y. Iu, “Democratizing Financial Knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology,” *papers.ssrn.com*, Feb. 02, 2023. https://ssrn.com/abstract=4346152 (accessed May 13, 2023).

[9] T. Yue, D. Au, C. C. Au, and K. Y. Iu, “Democratizing Financial Knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology,” papers.ssrn.com, Feb. 02, 2023. https://ssrn.com/abstract=4346152 (accessed May 13, 2023).

[10] Formplus, “15 Reasons to Choose Quantitative over Qualitative Research,” Formpl.us, Jul. 10, 2019. https://www.formpl.us/blog/quantitative-qualitative-research (accessed May 19, 2023).

[11] T. Williams, “Why Is Quantitative Research Important?,” GCU, Jun. 14, 2021. https://www.gcu.edu/blog/doctoral-journey/why-quantitative-research-important#:~:text=The%20purpose%20of%20quantitative%20research (accessed May 19, 2023).

[12] S. McLeod, “What’s the Difference between Qualitative and Quantitative research?,” *Simply Psychology*, Apr. 06, 2023. https://www.simplypsychology.org/qualitative-quantitative.html (accessed May 19, 2023).

[13] K. Hammarberg, M. Kirkman, and S. De Lacey, “Qualitative Research Methods: When to Use Them and How to Judge Them,” Human Reproduction, vol. 31, no. 3, pp. 498–501, Jan. 2016, doi: https://doi.org/10.1093/humrep/dev334.

[14] S. McCombes, “Doing Survey Research | A Step-by-Step Guide & Examples,” Scribbr, May 06, 2022. https://www.scribbr.co.uk/research-methods/surveys/ (accessed May 19, 2023).

[15] J. W. Creswell and J. D. Creswell, Research Design: Qualitative, Quantitative & Mixed Methods Approaches, 5th ed. Los Angeles: Sage, 2018, p. 149.

[16] Python Software Foundation, “Welcome to Python.org,” Python.org, Nov. 18, 2019. https://www.python.org (accessed May 19, 2023).

[17] A. Gefferth, “One Hot Encoding scikit vs pandas,” Medium, Mar. 13, 2023. https://towardsdatascience.com/one-hot-encoding-scikit-vs-pandas-2133775567b8 (accessed May 20, 2023).

[18] N. Abraham et al., Coding all-in-one. Hoboken, Nj: John Wiley & Sons, Inc, 2017, p. 354.

[19] D. J. Rumsey, Statistics for dummies. Hoboken, Nj: John Wiley & Sons, 2016, pp. 297–300.

[20] D. Spiegelhalter, STATISTICS : the art of learning from data. Pelican, 2019, p. 58.

[21] W. Mckinney, “Data Structures for Statistical Computing in Python,” 2010. Accessed: May 20, 2023. [Online]. Available: https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf

[22] scipy, “SciPy documentation — SciPy v1.8.1 Manual,” docs.scipy.org, Feb. 19, 2023. https://docs.scipy.org/doc/scipy/ (accessed May 20, 2023).

[23] J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007, doi: https://doi.org/10.1109/mcse.2007.55.

[24] M. Waskom, “seaborn: statistical data visualization,” Journal of Open Source Software, vol. 6, no. 60, p. 3021, Apr. 2021, doi: https://doi.org/10.21105/joss.03021.

# Appendices