MSc Computer Science & Big Data Analytics

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

24/05/2023

Word Count – 3,251

# Abstract

ChatGPT, an AI conversational tool, has gained widespread adoption in various applications, raising questions about its value for users. This research proposal aims to investigate the extent to which ChatGPT enhances users' knowledge and whether it becomes an integral part of their knowledge acquisition routine. While research on ChatGPT and its applications is progressing, this study focuses on understanding user perspectives. A comprehensive literature review has been conducted to identify existing research and knowledge gaps that this study aims to address, contributing to the current understanding of the topic. The proposed investigation will employ a quantitative survey study to gather participants' views on their use of ChatGPT. The collected data will be analysed to examine users' perception of ChatGPT and explore potential relationships with participant characteristics. The results will be presented in tabular and graphical formats, providing insights that may either support or challenge the hypotheses stated.

**Keywords—ChatGPT, knowledge enhancement, knowledge acquisition**

# Introduction

## Background

ChatGPT, widely utilised 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 [3]. 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 integration into users' routines. Despite research on ChatGPT and its applications, such as solving programming bugs [4-6], providing education [7-8, 10-11], and democratising knowledge [9], there is limited research on user perception. Tlili et al. [10] conducted a notable study addressing this, but due to ChatGPT's novelty and small scope, 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 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 and beyond. No conflicts of interest are anticipated, and there is no expected exposure to risks or dangers for participants.

# Literature Review

## Introduction

ChatGPT – a large language Model (LLM) Artificial Intelligence (AI) tool created by Open AI which answer requests via Natural Language Processing (NLP) tasks [1], is applied in various practices. However, 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 into their day-to-day activities.

## Research found upon ChatGPT Evaluation and its uses.

A study seen from Surameery and Shakor [4], examines the use of ChatGPT in solving programming bugs, understanding if the tool aids in debugging assistance, explanation, and corrections for software development practices. Additionally, a comparison between ChatGPT and Integrated Development Environments (IDE’s) testing effectiveness between practices. This study is extended further from works by Sobania et al [5] and Bang et al [6]. The study – in silo – provides some context about how ChatGPT can be used and against other tools but provide no specific examples as to what bugs were explored and how they were solved. Additionally, there was no research conducted on software developers using ChatGPT to solve any programming problems nor a perception of how this has helped them.

With Sobania et al [5], an evaluation performed back in January 2023 against ChatGPT on responses regarding if a piece of code has a bug and how it should be fixed; concluding a mostly positive but mixed results on performance – particularly where verification is needed, but better responses compared with other tools once context is applied on either side of the query. This study extends research from Surameery and Shakor [4], showing detailed examples of prompts used and responses from ChatGPT stating “would yield ChatGPT to be a viable tool that would help software developers in their daily tasks”. This excerpt from the research was not the focus and more a foundation that can be built upon, where the research proposed helps at least partially address the gap the above research presents.

Finally, Bang et al [6] evaluates ChatGPT against a series of public datasets covering different tasks, and how it compares with other LLM’s – stating a good performance, but not without the risk of external hallucination – where a confident response is received without any justification from its training.

Based on all studies presented above, they all have a similar theme to test ChatGPT and its application with evaluation against different metrics, other LLM’s or other tools. More importantly, a gap is seen throughout the literature so far in that, whilst ChatGPT has been tested on performance and use in silo for a particular use or topic, no views from an array of users have been collected to understand whether users have used or adopted its practice on the scenarios presented above, giving space for this proposed research to fill that gap.

## Research on ChatGPT and the enhancement of user knowledge and incorporation

Aljanabi [7] that extend the notion of using ChatGPT, enabling possibilities and future directions for users and the software itself. This paper provides a great introduction into the use of ChatGPT and the opportunities it presents such as the potential for user personalisation, and alluding to integration for different working fields, and enhancing lives in meaningful ways. However, as Alunjabi [7] frames this work as an introductory piece into the subject, no specific examples tested either in a standalone study or across a sample of participants to assess whether the claims made here are proven.

In a similar theme, an excerpt from Liberman [8] suggests ChatGPT as a tool to explore sources of information for users, focused on students as the primary user, assisting with understanding context of a subject. Although, Liberman does go further to say that ChatGPT may not provide accurate information in all areas for its user. Again, this source does provide a highlight on what ChatGPT is and how it can be used to assist with context, but little information of how it provides knowledge enhancement and detail is lacking to explain this further, which is understood considering the context of the source.

Whilst the sources above provide more of an introduction into ChatGPT as a field to explore, the next set of sources delve further, albeit with limited range. A study produced by Kung et al [8] explored performance of ChatGPT on United States Medical Licensing Exam (USMLE). Although the research’s primary focus was testing ChatGPT performance against the exam, a discussion was made on results on whether ChatGPT can assist knowledge enhancement for users taking the exam by providing insight into areas that needs development. However, limitations of this research keep it focused on ChatGPT and its output from USMLE, other than alluding exploration into knowledge enhancement for medical students taking the exam and further incorporation into day-to-day use. Therefore, a gap in the study is represented by no medical student participation shown against this study to explore the improvement of scoring in the USMLE.

A separate study found from Yue et al [9] explores the use of ChatGPT in providing financial knowledge to non-financial users, testing queries on how it can explain financial terms and metrics such as alpha, beta, earnings to price (EP) or Illiquidity to a user based on different context and publication approaches. This study is beneficial in that ChatGPT has the potential to help individuals gain financial knowledge to assist in making informed decisions. However, there is no evidence of testing these exercises against participants to test the claims made, one whether information provided is understood and helpful and adds to their day-to-day routine to expand their knowledge further.

To summarise above, some evidence is found to exploring the notion of ChatGPT against users for knowledge enhancement, and a good introduction into possibilities that could be held here, although there seems to be a gap in measuring user engagement of the tool, instead focusing into more of what the tool can do.

## Research on ChatGPT with user perception

Only one piece of research came through thus far, which looks to understand the use of ChatGPT from a user perspective, which is conducted by Tlili et al [10] via case study on the use of ChatGPT, and its perception from educational users, including 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 for idea generation.

The study found above, whilst not much other research is available showing user perceptions on ChatGPT, it provides a great foundation to build upon There are limitations with this research however, such as a small group of participants involved in the study, a focus within education rather than a broader perspective, and qualitative , not quantitative research done, therefore whilst themes are generated, no way to measure user’s perspective.

## Conclusion

The purpose of this literature review was evaluating 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 research 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 have 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. However, a draft of the survey is created with a link and related screenshots found in the **Appendix.**

## 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 like 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 an 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 | 2 Weeks |
| 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 an 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

As this research will involve the perception of participants view in form of a quantitative survey study, there are some potential items that need to be considered. Participants' rights will be prioritized throughout the study, and all procedures will adhere to ethical guidelines outlined by The Institute of Electrical and Electronics Engineers (IEEE) [25] within their code of ethics [26].

Participants will be provided with a clear explanation of the study's purpose, benefits this research aims to bring and procedures involved. Informed consent will be obtained from each participant before participation, ensuring they can make an informed decision. Consent will be documented through the online survey platform, with participants required to indicate their voluntary participation by selecting the option “I agree to these terms and wish to participate”, before the survey can begin.

Participants' confidentiality will be maintained. Data collected will be stored and anonymised. Only the researcher will have access to the data, which will be used solely for research purposes. No personal identifying information will be expected in the survey results but will be separated from the survey responses should this arise during analysis.

The study poses minimal risks to participants. However, measures will be taken to mitigate any potential risks, such as providing clear instructions on the survey process to avoid any confusion. The researcher involved in this study have no conflicts of interest that could compromise the objectivity or integrity of the research.

## Action to Obtain Ethical Approval

To get ethical approval for this research project, this proposal shall be sent to the dissertation supervisor in the first instance, due to the ethical consideration outlined. Whilst this study involves the use of humans to get their views on the use of ChatGPT, there is minimal to no harm or misconduct expected towards participants who engage with the survey.

# References

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

[2] OpenAI, “Introducing ChatGPT,” OpenAI, Nov. 30, 2022. https://openai.com/blog/chatgpt (accessed May 16, 2023).

[3] 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.

[4] 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.

[5] D. Sobania, C. Hanna, M. Briesch, and J. Petke, “An Analysis of the Automatic Bug Fixing Performance of ChatGPT,” Jan. 2023. Accessed: May 24, 2023. [Online]. Available: https://arxiv.org/pdf/2301.08653.pdf

[6] Y. Bang et al., “A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity,” arXiv:2302.04023 [cs], vol. 1, Feb. 2023, Accessed: May 24, 2023. [Online]. Available: https://arxiv.org/abs/2302.04023

[7] 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.

[8] 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

[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] 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.

[11] 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.

[12] 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).

[13] 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).

[14] 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).

[15] 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.

[16] 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).

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

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

[19] 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).

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

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

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

[23] 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>

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

[25] 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.

[26] 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.

[27] IEEE, “History of IEEE,” @IEEEorg, 2019. https://www.ieee.org/about/ieee-history.html (accessed May 21, 2023).

[28] IEEE, “IEEE Code of Ethics,” ieee.org, 2020. https://www.ieee.org/about/corporate/governance/p7-8.html (accessed May 21, 2023).

# Appendix – Draft Survey

The below link shall take you the draft survey currently created, in the hope – with some potential amendments -to be used as the official vehicle to collect the information for this research.

<https://forms.gle/8WXM3qcAMBZizoL88>

Please also see below screenshots of the current survey layout:

A screenshot of a phone

Description automatically generated with medium confidence

Figure First section of survey, with description, log of email address, and ask first question.

A screenshot of a phone

Description automatically generated with medium confidence

Figure Second section of survey on ChatGPT use.

A screenshot of a computer screen

Description automatically generated with low confidence

Figure Third section relating to demographics.

A picture containing text, electronics, screenshot, web page

Description automatically generated

Figure Section shown if participant is declared as a student.

A screenshot of a chat

Description automatically generated with low confidence

Figure Section Shown if participant declared as a working professional.

A screenshot of a web page

Description automatically generated with low confidence

Figure Final window confirming submission once submitted.