MSc Computer Science & Big Data Analytics

ChatGPT: The advancement of knowledge and incorporation for its users - Methodology

24/05/2023

Word Count –

# 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 chapter is to explain how the research was performed, 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 uses 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 set [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 aimed to access to a large participant group to help ensure an optimal response rate compared to physical submissions [14]. The survey is cross-sectional as the research is novel, to establish responses from one point in time [15] over a period of approximately one month, that could be potentially compared against for future research.

## Preparation

Before research began, the survey was created ready for distribution, aimed to be done within the research project timeline. A copy of the survey used can be found in **Appendix A.**

## Population

The sample size used in this research consists of all individual responses who interacted with the survey providing input regarding the topic of ChatGPT. However, some user characteristics was collected to allow for analysis and aid future research.

## Data Collection

As this research utilised an online survey, an online questionnaire instrument was deployed. This instrument got created via a software tool known as Google Forms. With this method, a link was distributed via social media posts or direct communication on multiple platforms such as LinkedIn, Discord, and Facebook. The steps followed in the data collection process were the following:

1. Shared 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 or direct communication on social media where requested.
2. The individual answered questions and makes their submission.
3. At the end of the time allotted to receive responses, data was exported from the Google Form tool to a downloadable csv file.
4. From there, the csv file was imported into a python script producing relevant statistics and analysis.

## Data Analysis and Results

### Management and storage

As described in the previous section, data was collected via a Google Forms questionnaire, 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 was imported into Python, a user-friendly and interpretable high-level programming language [16]. Python supports data analysis through libraries like pandas for data analysis and manipulation [21], SciPy for statistical tests [22], matplotlib for standard data visualisation [23], and seaborn for statistical visualisation [24], all facilitating statistical analysis and data visualization for the collected data.

### Data Analysis Steps

The following list of tasks were executed to present findings and derive insight from the data collection:

#### 1 – Conversion of values

Most participant data collected will consist primarily of categorical responses, including yes or no choices and some characteristic information selected from multiple choice selection (e.g., occupational background, highest level of education). Consequently, data conversion also transformed categorical values into numerical representations, utilising transformative logic.

#### 2 – Descriptive Analysis

Initial analysis summarised data points using tables and graphs. Examples included total participant count, percentages of population with specific characteristics. Additionally, analysis addressed 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 sample.

The analysis also explored the limited 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 was 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, an employment of the binomial test in python was used 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, would be accepted.

### Results from Analysis

The results are then 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.

## Ethical Considerations

As this research involved the perception of participants view in form of a quantitative survey study, there remained potential items that were considered. Participants' rights had been prioritised 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 were provided a clear explanation of the study's purpose, benefits this research aims to bring and procedures involved. Informed consent was obtained from each participant before participation ensuring they can withdraw from the questionnaire at any time, confirming they can make informed decisions. Consent was 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 begun.

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

The study poses minimal risks to participants. However, measures were taken to mitigate potential risks, such as providing clear instructions on the survey process to avoid any confusion. No conflicts of interest that could compromise the objectivity or integrity of the research were detected.

## Action to obtain ethical approval.

To get ethical approval for this research project, a proposal was sent to the dissertation supervisor, 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. A signed copy of the ethical approval of research projects in online programmes form can be found in **Appendix A.**

## Limitations

Particularly as this piece delves into a novel area concerning ChatGPT limitations were uncovered and need to be addressed to provide as much transparency as possible. Due to limited time and resources, only one survey will be conducted with a limited series of 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. Furthermore, 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.

# 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).