How accurate is ChatGPT's instruction to create software applications?

A research proposal to determine the helpfulness of ChatGPT in creating software applications

Anh Dung Luu
COMP SCI 7405
The University of Adelaide
Adelaide SA, Australia
a1915303@adelaide.edu.au

1 Importance of the research question

Software applications are computer programs with multiple features to serve a particular purpose for the users. The creating and developing process of an application is usually complicated and requires the amount of programming knowledge corresponding to the complexity of the application. Nevertheless, each user has their own preferences for applications and typically selects those that precisely fulfill the specific tasks without unnecessary features. In addition, the presence of malware and security risks makes users avoid using applications with low reputation no matter how suitable they are.

In 2022, OpenAI introduced ChatGPT, an Artificial Intelligence chatbot capable of assisting humans in multiple fields thanks to its massive training data from the Internet. In Software Development, ChatGPT is well known for helping developers create and debug software applications by giving advices and instructions that are easy to understand. There have been reports about individuals who did not have any coding knowledge and experience created fully functional applications with the help of ChatGPT.

However, ChatGPT's response is not always correct, which means its generated source code and guidance may not work. A clearer view of the accuracy of ChatGPT's instructions in creating an application is needed as the answer to this paper's research question.

2 Previous research

There have been researches about how ChatGPT could support developers and users in developing software. These researches pointed out the benefits as well as drawbacks and limitations of using ChatGPT and AI tools in general to generate source code. A research also focused on using ChatGPT for a particular programming language.

In this research, I will analyze ChatGPT's accuracy in a broader view and with a mindset of a normal user trying to create an application rather than an experienced developer.

3 Data Preparation

The main activity of the research will be an experiment where I will ask ChatGPT to generate source code and instructions to build a software application for a particular purpose. For each question, there may also be follow-up questions for ChatGPT to debug and fix errors. The number of follow-up questions depends on the complexity of the application.

The latest ChatGPT model (GPT-4.5) will be used to generate responses. To get access to this model, a ChatGPT Plus subscription will be purchased.

100 questions, excluding follow-up questions, corresponding to 100 different applications, will be given to ChatGPT. The questions will be prepared with the following pattern:

> "Generate code to build a <*type of application*>. <*application description*>. Instruct me how to run this application."

The application description includes all features and requirements which will be used to assess the product built from ChatGPT's response.

The questions will be divided into 5 levels of complexity. The number of questions at each level is as follows:

Level of complexity	Number of questions	Allowed number of follow-up questions
1	30	2 per question
2	20	5 per question
3	20	8 per question
4	20	10 per question
5	10	15 per question

Table 1: Number of questions and follow-up questions per level of complexity

After each question, the state of whether the application works or not and the number of follow-up questions asked will be recorded as data for analysis.

4 Analysis Plan

Recorded data will be used to calculate the accuracy of ChatGPT's instruction. The two most important measurements for calculation are the number of applications being created successfully and the number of follow-up questions asked. For each level of complexity, the accuracy of ChatGPT's instruction will be calculated using the following formula:

$$accuracy = \frac{score}{n * (allowedFU + 1)}$$

$$score = \sum_{i=1}^{n} (allowedFU - askedFU_i) * successful_i$$

where n is the number of questions of that level, $allowedFU$ is the allowed number of follow-up questions for each question, $askedFU$ is the number of follow-up questions asked during the creating process of an application, and $successful$ is 1 if the application works or 0 if not. $score$ describes how many follow-up questions would be asked; the more follow-up questions used the less $score$ would be.

Conclusion will be drawn based on each level of application complexity and will suggest applications of which level could be created with the help of ChatGPT.

5 Relation to Course Material

This research question is an exploratory question as it is trying to determine the accuracy of ChatGPT's responses in the process of creating applications. The philosophical worldview used for this question is Positivist.

The theory in this research question is a quantitative theory as it uses deductive logic in which the question is formed from existing facts on the Internet and the goal of the research is to verify a measure (which is the accuracy of ChatGPT's instructions). The strategy used for this research is a quantitative strategy which is experiment.

The research plan is designed to minimize as many threats to validity as possible. However, there are still some threats to validity which have not been solved, such as researcher's bias when making list of questions and applications' requirements.

REFERENCES

[1] Bucaioni, A, Ekedahl, H, Helander, V & Nguyen, PT 2024, 'Programming with ChatGPT: How far can we go?', *Machine Learning with Applications*, vol. 15, p. 100526.

[2] Biswas, S, 2023, 'Role of ChatGPT in Computer Programming', *Mesopotamian Journal of Computer Science*, 2023, pp. 9–15. DOI:10.58496/MJCSC/2023/002.

[3] Gewirtz, D 2025, *How to use ChatGPT to write code - and my favorite trick to debug what it generates*, ZDNET, viewed 6 April 2025, <https://www.zdnet.com/article/how-to-use-chatgpt-to-write-code-and-my-favorite-trick-to-debug-what-it-generates/>

[4] OpenAI 2025, *ChatGPT Pricing*, OpenAI, viewed 6 April 2025, <https://openai.com/chatgpt/pricing/>

[5] OpenAI 2024, *Terms of Use*, OpenAI, viewed 6 April 2025, <https://openai.com/policies/terms-of-use/>

[6] Pantelimon, FV, Posedaru, B Ștefan, Pocatilu, P, Filip, FG & Ciurea, C 2024, 'Improving Programming Activities Using ChatGPT: A Practical Approach', in *Proceedings of 22nd International Conference on Informatics in Economy (IE 2023)*, vol. 367, Springer Singapore Pte. Limited, Singapore, pp. 307–316.

[7] Timothy, M 2023, *How to Build a Web App With ChatGPT*, MUO, viewed 6 April 2025, <https://www.makeuseof.com/chatgpt-how-build-web-app/>