

# Effectiveness of Generative AI on the Development of Graphic Software Artifacts

Ellen Mezera

School of Enterprise Computing and Digital Transformation, TU Dublin, Ireland

X12345678@myTUDublin.ie

## Introduction

The surge in Artificial Intelligence research, particularly in Generative AI following the release of models like GPT-3 in 2022, has prompted extensive exploration of its potential for enhanced productivity in various industries, especially in coding tasks. As AI solutions mature, software engineering practices are evolving, with a notable adoption of Generative AI for coding, while the application of these technologies to tasks involving image generation is still in its early stages. This research sets out a methodology leveraging Large Language Models to generate graphic software artifacts. More specifically it applied GPT-4's text-to-text capabilities in generating UML Diagrams supported by Mermaid as a rendering tool. The resulting diagrams were validated by Subject Matter Experts to assess results produced by GPT-4.

## Research Question 1

**RQ-1** How can OpenAI's GPT-4 model be leveraged to generate software architecture diagrams?

This research sets out a methodology on how ChatGPT-4 can be used for software diagrams generation using its text-to-text capabilities, while text-to-image are not yet mature. Moreover, it applies this methodology on the generation of UML Class and Sequence diagrams.

The methodology is applied in an experiment where Chat-GPT is asked to generate diagrams for three use case scenarios around a Online Shopping System.

## Research Question 2

**RQ-2** What is the efficacy of OpenAI's GPT-4 model generating software architecture diagrams based on the evaluation from Subject Matter Experts?

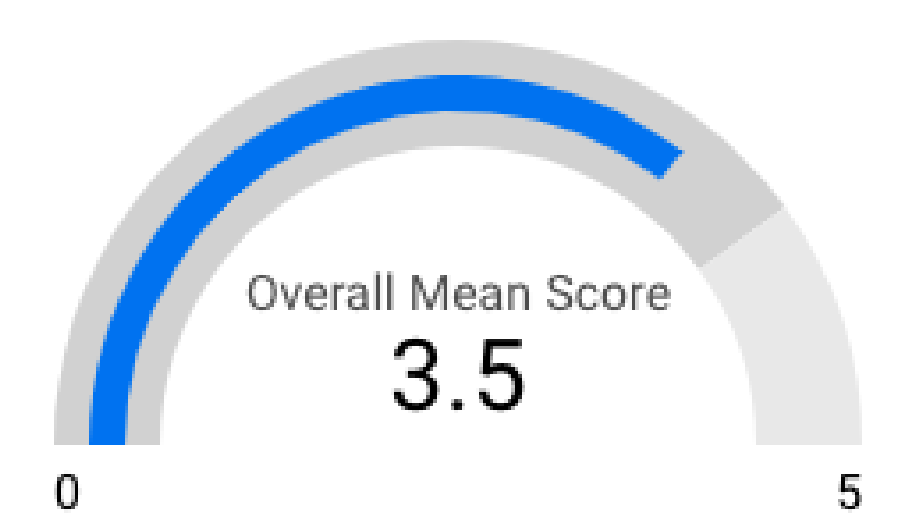
In order to validate the results obtained from the experiment created to address RQ-1, Subject Matter Experts assessed the results via a validation questionnaire, scoring 5 categories: Accuracy, Clarity, Completeness, Technical correctness and Usefulness.

## Results

Class and Sequence diagrams were generated by GPT-4 for three use case scenarios around a Online Shopping platform in ascending order of complexity. The six diagrams were scored by Subject Matter Experts via a in 5-point Likert scale questionnaire.

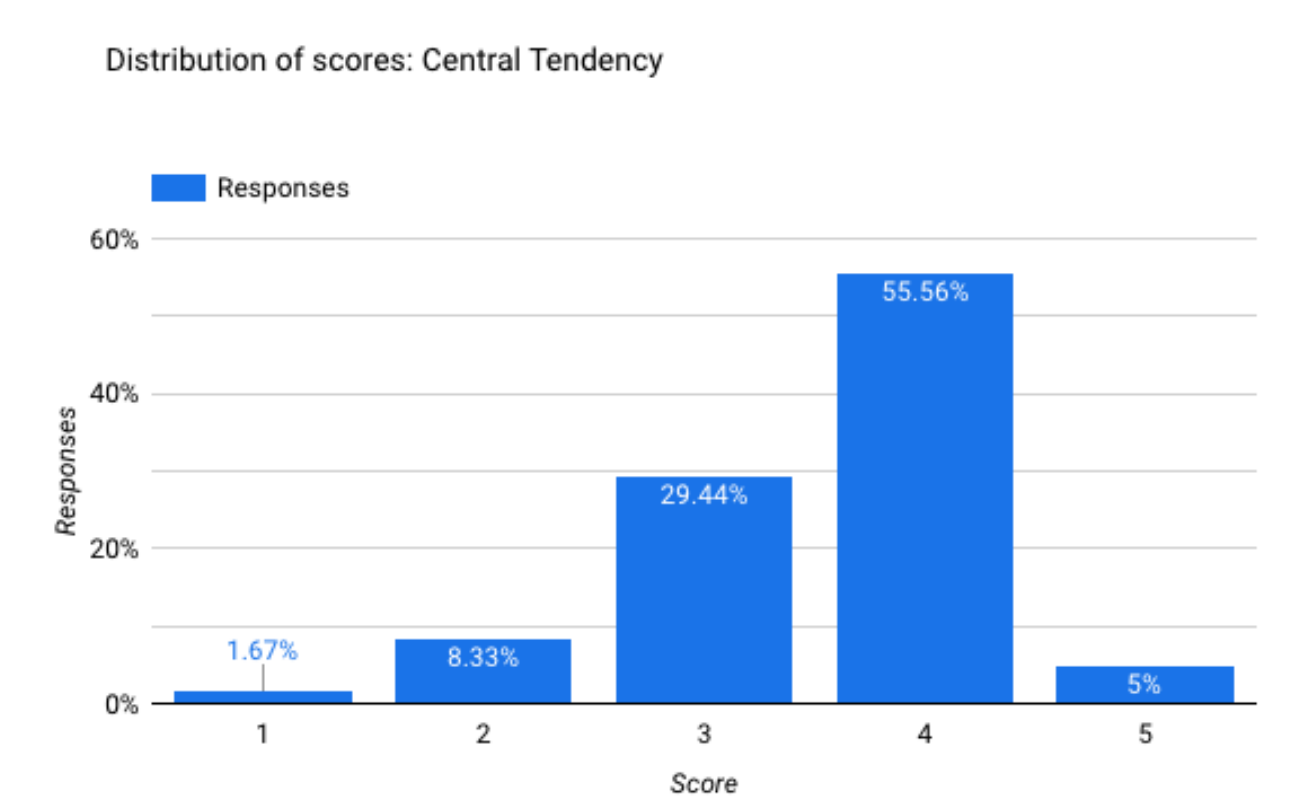
### 1. Effectiveness of ChatGPT in UML Diagram Generation:

The overall Mean Score of 3.54 out of 5 indicates a moderate level of effectiveness of ChatGPT in generating UML diagrams through the methodology applied in this experiment.



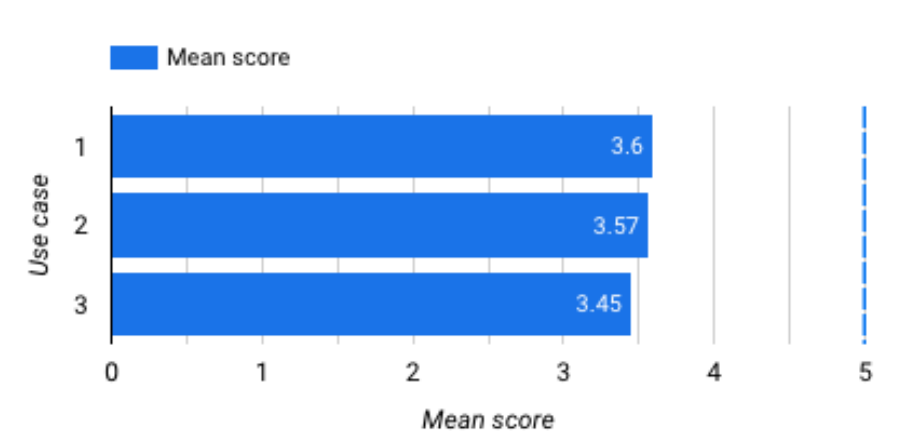
### 2. Distribution of scores:

The majority of scores given by evaluators through the questionnaire across all diagrams are concentrated between 3 and 4, with 4 having its majority share. The central tendency of scores indicates that evaluators found the diagrams to be above average in terms of the categories accessed.

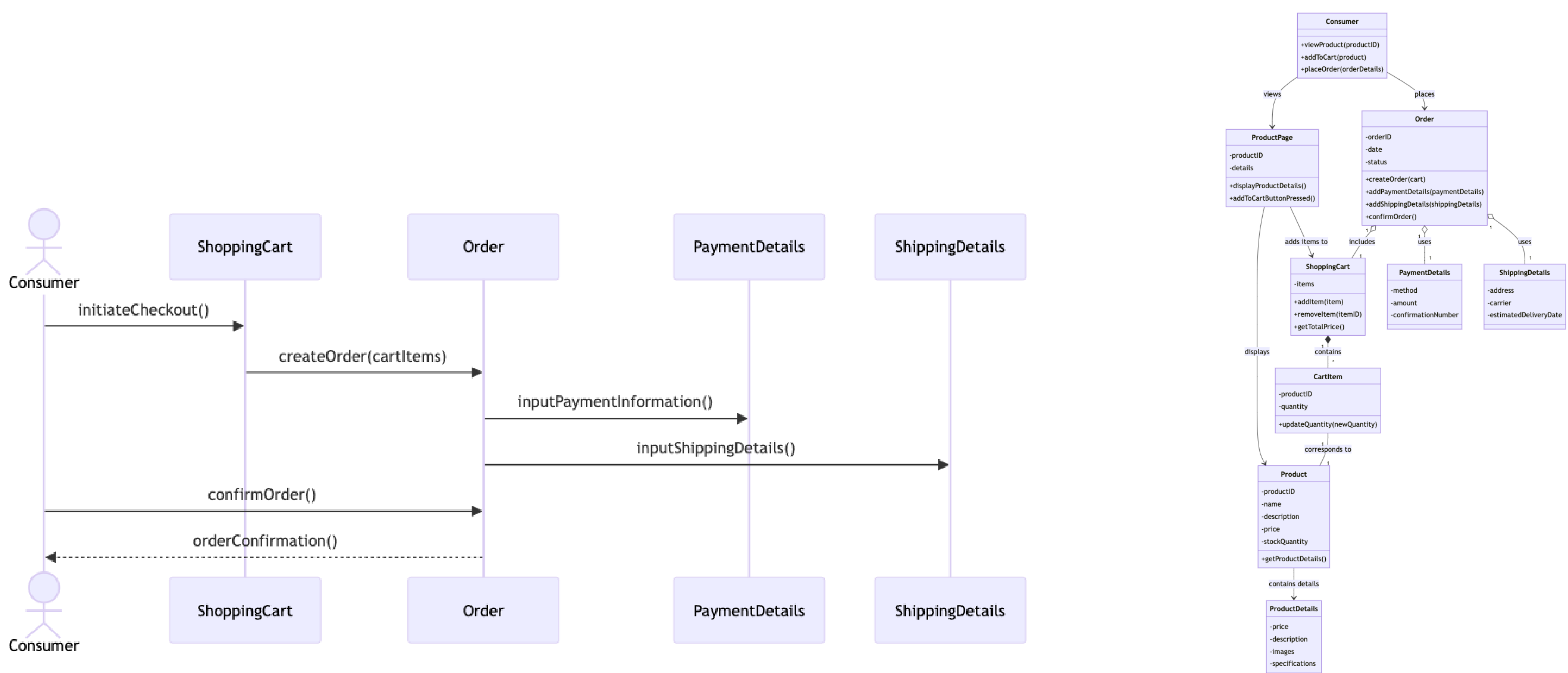


### 3. Performance across Different Complexity Levels:

The three use cases scenarios selected for the experiment were in an escalating complexity order. The mean scores outline a drop as use cases grow in complexity, however the drop is quite minimal. The data suggests a possibility of limitations of the LLM's capability to effectively handle higher UML diagrams of higher complexity.



## UML Class and Sequence Diagrams generated by ChatGPT-4



## Conclusions and Future Work

This research underscores the growing importance of exploring Generative AI's role in Software Engineering tasks, particularly in generating software architecture diagrams. While the study reveals that ChatGPT-4 demonstrates moderate efficacy in producing UML Class and Sequence Diagrams, further investigations should extend to other diagram types, involve a more diverse group of evaluators, and integrate a feedback loop for continuous improvement. Additionally, ethical considerations must be carefully addressed as Generative AI becomes increasingly integrated into real-world scenarios.

## QR Code for Recording

