1. What is Prompt Testing?

1.1 Introduction

Context

The arrival of language models such as GPT in the software industry has opened up new possibilities for human-machine interaction.

GPT (Generative Pre-trained Transformer) is an artificial intelligence-based language model that has been trained on large amounts of text data to generate coherent and relevant text in response to text

input.

Large Language Models (LLMs) made their debut with the release of OpenAI's GPT-3 in 2020. Since then, their popularity has steadily grown, until late 2022 when interest in LLMs and the broader field of generative AI skyrocketed due to significant advances in LLMs.

The latest model, GPT-4, was released in March of 2023 and it represents the next generation of this technology.

While OpenAI hasn't disclosed specific changes made to the underlying model, the company emphasizes the significant improvements over its predecessor.

According to OpenAI, GPT-4 can process up to 25K words at a time, which is eight times more than the original GPT-3 model.

Additionally, GPT-4 can understand more nuanced instructions, requests, and questions than GPT-3.5, the model used in the existing ChatGPT AI.

As of the writing of this book, major companies are announcing their AI and LLM projects, with plans to integrate them into their current software portfolios.

This integration will lead to a complete transformation in the way we interact with software, as well as how that software is developed and tested.

Vision

GPT and other language models have opened the door to a profound change in how software is developed and used.

This change will turn the entire industry upside down faster than we can imagine, and all the necessary pieces are already on the board, starting to line up.

My vision is that in the near future, we will interact with our software through models such as GPTx, which will become the preferred interface for most applications. Therefore, they will need to be tested properly in order to ensure a high-quality level.

The goal of this book is to analyze this change in the software industry from a Quality Assurance perspective.

A new way of developing and using software requires a new way of testing. A new test paradigm called **Prompt Testing**.

Assumptions

There are already a large amount of large language models available, and more will come in the near future. It is practically impossible to determine which one will prevail.

While it is not the purpose of this book to enter into that debate, we need to choose one for practical purposes to exemplify all the concepts that will be

presented in the later chapters.

Everything points to OpenAI's GPT-4 or its future versions as a good candidate. Therefore, for the remainder of the book, we will use the term "GPTx" to refer to large language models, without implying any preference.

The same applies to the LangChain framework. It will be explained in the next section. It is not the only framework available, nor is it claimed to be the best.

However, in its current state, it provides precisely what is needed to exemplify the testing concepts and philosophy of Prompt Testing.

Author's Opinion

In March 2023, Microsoft announced the integration of GPT-4 across its Office portfolio. In May 2023, Google did the same with BART and Palm-2.

Week after week, there are news about companies from different domains offering services with integrated GPT-4 and other LLMs. Duolingo, Evernote, and Khan Academy are some of the most recent examples.

I am convinced that the way we interact with software will change completely in the short term. Instead of using software tools, we will have software assistants powered by AI. We will collaborate with them instead of merely using them.

The Quality Assurance industry must react and adapt quickly. As we will see in the next chapters, GPTx will represent both a challenge and an opportunity.

An opportunity to leverage Quality Assurance with the help of AI.

Prompt Testing

In a nutshell, Prompt Testing is a new testing paradigm that establishes the foundation for testing all those systems that integrate GPTx as a user interface.

This kind of integration has the potential to revolutionize the way we interact with software. Instead of having to learn specific commands or navigate complex menus, users can simply type what they want to do in natural language, and the software can understand and respond appropriately.

This can make the software more accessible to a wider audience and reduce the learning curve for new users.