“Prompt it”, not “Google it” : Prompt Engineering for Statistical Programmers and Biostatisticians

Kevin Lee, Director of Data Innovation at Karuna Therapeutics

ABSTRACT

Since its release, ChatGPT has rapidly gained popularity, reaching 100 million users within 2 months. Even a new concept has emerged : “Prompt it” is now the new “Google it”. Research shows ChatGPT users complete projects 25% faster. The paper is written for Statistical Programmers and Biostatisticians who want to improve their productivity and efficiency by using ChatGPT prompts better.

The paper explores the pivotal role of prompts in enhancing the performance and versatility of ChatGPT or other Large Language Model. The paper shows how Statistical Programmers and Biostatistician utilize ChatGPT's capabilities and benefits such as the content development (e.g., emails, images), search for the information, Programming assistance in R, SAS and Python, Result Interpretation and many more.

The paper also elucidates the distinctive advantages of employing prompts over traditional search methods. It emphasizes the unique characteristics of prompt engineering in ChatGPT. Various techniques, such as zero-shot learning, few-shot learning, reflection, chain of thought, and tree of thought, are dissected to illustrate the nuanced ways in which prompts can be engineered to optimize outcomes. The comprehensive exploration also offers insights into how to prompt better by adding constraints, incorporating more contexts, setting roles, coaching with feedback, probing further, and introducing step-by-step instructions to ChatGPT. The paper discusses ChatGPT's functionality in modifying and resubmitting the prompt, copying the answer, regenerating the answer, and continuing the previous prompt.

The paper highlights how Stat programmers and Biostatisticians use and lead the transformative impact of prompts to be more productive and effective.

INTRODUCTION OF Gen AI

Generative AI (Gen AI) utilizes trained Machine Learning models to create entirely new content based on prompts. Gen AI is rapidly gaining importance and popularity in all the industries, and its revenue is expected to increase from $137B in 2024 to $900 in 2030.

Gen AI could generate various outputs.

- Text : Content Writing, Chatbots, Assistants, Search
- Code : Code Generation, Data Set Generation
- Image : Image Generation, Image Edit
- Audio : Voice Generation/Edit, Sound creation, Audio Translation
- Video : Video Creation/Edit, Voice Translation, Deepfake

INTRODUCTION OF ChatGPT

ChatGPT is one of the most popular Gen AI, and it is developed by OpenAI. It is trained on large corpus of text about 300B tokens, and its main strength lies in

- The ability to generate human-like response in various contents.
- The ability to understand and generate content in a wide range of domains.

What is Prompt?

The word "prompt" is the new concept and term in Gen AI, and it explains how to use the new wave of AI assistants like Gen AI such as ChatGPT. Just like "Google it" became synonymous with searching for information online, "Prompt it" is emerging as the go-to way to interact with Gen AI and ChatGPT. By providing a clear and specific prompt, we can instruct Gen AI to complete a variety of tasks, generating creative text formats, videos, images, and other contents. Essentially, the prompt is the new way of giving Gen AI instructions and guiding it towards the desired outcome.

What is Prompt Engineering?

Prompt Engineering is the skillful craft of constructing prompts that effectively guides ChatGPT to produce the desired contents. When prompts are poorly designed, they can lead to suboptimal or irrelevant outputs. Conversely, well-crafted prompts enhance the ability of Gen AI and ChatGPT to generate accurate and contextually appropriate contents.

How to Prompt better

Crafting effective prompts is the key to getting the most out of Gen AI. Here's how programmers and statisticians can elevate prompts and unlock a world of creative and informative possibilities:

- Get Specific: Instead of a vague request like "write a story," programmers and statisticians should specify the genre, characters, or plot points. The more details are provided, the more focused and relevant response will be.
- Add Constraints: Sometimes, limitations can spark creativity. Programmers and statisticians can try including limitations like word count, specific vocabulary, or a particular style (e.g., haiku poem). These constraints can push Gen AI to generate unique and surprising outputs.
- Add More Context: The more context programmers and statisticians provide, the better Gen AI can understand the situation and tailor the response accordingly.
- Set the Role: It is like giving Gen AI a specific job. Programmers and statisticians can instruct Gen AI to act as a journalist writing a news article, a marketing copywriter crafting an ad, or a creative writer composing a poem. This role-playing helps Gen AI adjust the style and voice accordingly.
- Coach with Feedback: Programmers and statisticians provide feedback on the initial responses. Programmers and statisticians let Gen AI know what we like or dislike and offer suggestions for improvement. This feedback helps Gen AI adjust and refine the initial responses to better meet the needs.
- Probe Further: If the initial response isn't quite what programmers and statisticians envisioned, we can use follow-up prompts to delve deeper. Programmers and statisticians can ask clarifying questions or provide additional details to guide Gen AI in the right direction.
- Step-by-Step Introduction: For complex tasks, programmers and statisticians can break it down into smaller steps. This helps Gen AI understand the overall goal and approach the problem in a methodical way.

Below are examples of how to prompt better.

How to prompt better	**Generic Prompt Example**	**Prompt Example with "Better Prompt"**
Get Specific	Write me a promotion request.	Write me a promotion request for Manager of Statistical Programmer.
Add Parameter or Constraints.	Write me a promotion request.	Write me a promotion request within 100 words.
Add more Context (e.g., Reasons)	Write me a promotion request.	Write me a promotion request. I have already conducted many managerial roles.
Set the role.	Write me a promotion request.	I am currently working as principal statistical programmer. Write me a promotion request.
Coach Gen AI with feedback		… Write it again in 50 words.
Provide a reference	Write me a promotion request.	Write me a promotion request. Read the following passage and analyze its style. I want you to copy the style.
Probe Further	Provide the list of restaurants to visit in Baltimore.	……. Tell me the foods to try at place 1.
Provide step by step instruction	Classify the safety issues of the texts.	Classify the safety issues of the texts. Follow these steps. 1. Identify adverse events in text. 2. Identify patient in text. 3. Identify medication in text. If text answers to "Yes" on all the steps, classify as "Safety Issue', Otherwise 'No Safety Issue'.

Prompt Engineering Technique

Prompt engineering is the art of crafting informative instructions for Gen AI. By carefully structuring text prompts, programmers and statisticians can guide Gen AI to perform a wide range of tasks, from writing SAS/R/Python codes to answering complex questions. Gen AI encompasses several key skills that influence how effectively we interact with Gen AI. These skills include zero-shot, few-shot, and chain of thought prompting, which the paper explores with specific examples in the following.

Prompt Engineering Techniques	**Purpose**	**Method**	**Prompt Example**	**Output/Response Example**
Zero-Shot	One time Query	Straight-up Query	What is 2 + 8?	2 + 8 equals 10
Few-Shot	Teaching AI models to follow specific patterns	Feeding model with examples before asking the desired questions.	2@4 = 8 3@3 = 9 4@2 = 8 5@3 = 15 4@2 =	It seems like you have a pattern where you're using "@" as an operator. Based on the examples you provided: 2@4 = 8 (2 multiplied by 4) 3@3 = 9 (3 multiplied by 3) 4@2 = 8 (4 multiplied by 2) 5@3 = 15 (5 multiplied by 3) So, following this pattern, 4@2 would be equal to 4 multiplied by 2, which is 8.
Chain of Thought	Teaching AI model the thought process it should use.	Breaking the prompt into simpler steps – step 1, step 2, step 3, step 4 and so on.	Explain why SAS programmers need to learn prompt engineering. Follow below steps: 1. Provide the introduction 2. Provide two examples 3. Provide the conclusion with introduction and two examples	**Introduction:** Prompt engineering is a crucial skill for SAS programmers, enabling them to create dynamic and interactive reports by incorporating user inputs. This practice enhances the flexibility and usability of SAS programs, allowing users to customize their analyses without delving into the underlying code. As the demand for user-centric data applications grows, mastering prompt engineering becomes essential for SAS programmers to deliver more effective and tailored solutions. **Two Examples:** *Example 1:* Consider a scenario where a SAS programmer is tasked with creating a sales analysis report. By incorporating prompt engineering, the programmer can introduce prompts for users to input specific date ranges, product categories, or geographic regions. This empowers users to customize the report according to their needs, providing a dynamic and interactive experience. Without prompt engineering, the report would be static and limited in its ability to cater to various user requirements. *Example 2:* In another instance, imagine a SAS programmer developing a healthcare analytics dashboard. By implementing prompt engineering, users can input parameters such as patient demographics, time periods, or specific medical conditions to generate personalized reports. This not only streamlines the data analysis process but also ensures that healthcare professionals can extract relevant insights tailored to their patient populations. Without prompt engineering, the dashboard would lack the adaptability needed for diverse medical scenarios.

				Conclusion: In conclusion, prompt engineering is a fundamental skill for SAS programmers as it enables the creation of dynamic and user-friendly reports. The ability to incorporate prompts allows for customization and interactivity, making the data analysis process more accessible to end-users. The examples of a sales analysis report and a healthcare analytics dashboard illustrate how prompt engineering enhances the flexibility and utility of SAS programs. As organizations increasingly prioritize user-centric data applications, SAS programmers who master prompt engineering will be better positioned to deliver solutions that meet the evolving needs of their users.

Prompt Use Cases for Statistical Programmers and Biostatisticians

Prompt Engineering unlocks a new level of efficiency and opportunities for statistical programmers and biostatisticians. Below is the list of use causes where statistical programmers and biostatisticians can leverage for their works.

- Simple Inquiry – Programmers and Biostatisticians could simply inquire for the information.
- Content Generation - Programmers and Biostatisticians could generate simple content. Examples could be "Write the email to request data transfer", "Use below data to generate patient profiles", "Summarize below contents/email chains" and many more.
- Coding - Programmers and Biostatisticians could ask Gen AI to write the codes. Examples could be "Convert below SAS codes to R codes", "Using below comments, create SAS codes", "I want to merge rand and demo by usubjid. Write SAS codes.", and many more.
- Data Analysis - Programmers and Biostatisticians could analyze simple data. Examples could be "Select patients who sex is male below Patient Data", "Select patients whose age is greater than or equal to 40 below Patient Data", "Sort Patient Data by age", "Count below Patient Data by race", and many more.

CONCLUSION

In conclusion, prompt engineering offers a transformative power for statistical programmers and biostatisticians. By understanding this concept and its practical applications, statistical programmers and biostatisticians can unlock new avenues and contribute a significant advancement for data analysis, content development, communications. As the field of Gen AI and ChatGPT continues to innovate, prompt engineering will undoubtedly play an increasingly pivotal role in the future of biometrics functions – statistical programmers and biostatisticians.

REFERENCES

- OpenAI Prompt Engineering in https://platform.openai.com/docs/guides/prompt-engineering

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
Director of Data Innovation at Karuna Therapeutics
kevin.lee@karunatx.com

TRADEMARKS