

Prompt Engineering in Gen AI: A Game Changer for the professional development of Statistical Programmers and Data Scientists

Kevin Lee, Clinvia, Philadelphia, PA

ABSTRACT

In the rapidly evolving field of clinical trials, prompt engineering emerges as a game-changer for statistical programmers and data scientists. This presentation explores how professionals can harness Generative AI like ChatGPT to enhance productivity and efficiency.

We'll delve into the concept of "Prompt it" replacing "Google it," and examine what prompt engineering entails. The discussion covers techniques such as zero-shot learning, few-shot learning, and chain of thought, demonstrating how to optimize outcomes in content development, information retrieval, code generation, and data analysis. We'll also explore specific use cases in biometrics and address the importance of result validation.

By mastering prompt engineering, data scientists can unlock the full potential of AI models, enabling them to navigate the complexities of clinical trials more effectively. Data scientists will also gain practical strategies for leveraging this transformative technology in their professional development.

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 \$900B 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 Programming.
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 reference and analyze its style. I want you to follow the style. # reference :
Probe Further	Provide the list of restaurants to visit in Philadelphia. 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.	$2@4 = 8$ $3@3 = 9$ $4@2 = 8$ $5@3 = 15$ $4@2 =$	<p>It seems like you have a pattern where you're using "@" as an operator. Based on the examples you provided:</p> $2@4 = 8$ (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.	<p>Explain why SAS programmers need to learn prompt engineering. Follow below steps:</p> <ol style="list-style-type: none"> 1. Provide the introduction 2. Provide two examples 3. Provide the conclusion with introduction and two examples 	<p>**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.</p> <p>**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.</p> <p>*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.</p> <p>**Conclusion:** In conclusion, prompt engineering is a fundamental skill for SAS programmers as it enables the creation of dynamic</p>

				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 cases 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 whose 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.

Gen AI Use Cases in Biometrics

Gen AI has emerged as a transformative force, which can revolutionize the Biometrics team with its versatile applications. From code generation to automated document creation and advanced data analysis, this technology is empowering Biometrics team to streamline processes and enhance efficiency.

Below are some examples of Gen AI use case in Biometrics, and there are a lot more use cases that could utilize Gen AI capabilities.

- Information Inquiry
 - Prompt:
 - What is CDISC? Tell me in 50 words.
 - Output:
 - CDISC (Clinical Data Interchange Standards Consortium) is a global organization that develops data standards for clinical research. Its models, like SDTM and ADaM, ensure standardized data collection, submission, and analysis, improving regulatory compliance, interoperability, and data quality in clinical trials, facilitating efficient drug development and approvals by agencies like the FDA.
- Content Development (e.g., SAP, Mock-up tables, CSR)
 - Prompt:
 - Write clinical trial **protocol synopsis** based on below "title of study".
 - "title of study" : A phase 3, randomized, double-blind, placebo controlled study to evaluate the efficacy of drug A in schizophrenia patients.
 - Output:
 - **Clinical Trial Protocol Synopsis: Phase 3 Study on the Efficacy of Drug A in Schizophrenia Patients****
 - **1. Study Title:****
 - A phase 3, randomized, double-blind, placebo-controlled study to evaluate the efficacy of drug A in schizophrenia patients.
 - **2. Objective:****

The primary objective of this clinical trial is to assess the efficacy of drug A in treating schizophrenia in comparison to a placebo. Secondary objectives include evaluating the safety, tolerability, and overall clinical benefit of drug A.

****3. Study Design:****

This study will adopt a randomized, double-blind, placebo-controlled design. Participants will be randomly assigned to either the treatment group receiving drug A or the control group receiving a placebo. The double-blind nature of the study ensures that both investigators and participants are unaware of the assigned treatment.

****4. Participants:****

The study will enroll adult participants diagnosed with schizophrenia according to standardized diagnostic criteria. Inclusion and exclusion criteria will be defined to ensure the homogeneity of the study population.

****5. Interventions:****

- Treatment Group: Participants will receive drug A in the prescribed dosage.
- Control Group: Participants will receive a placebo identical in appearance to drug A.

****6. Primary Outcome Measure:****

The primary efficacy endpoint is the reduction in the severity of schizophrenia symptoms, as measured by a standardized assessment tool, at the end of the treatment period.

****7. Secondary Outcome Measures:****

Secondary endpoints include:

- Changes in cognitive function.
- Improvement in overall functioning and quality of life.
- Incidence and severity of adverse events.
- Assessment of treatment adherence.

****8. Study Duration:****

The trial will be conducted over a specified period, with defined time points for participant recruitment, treatment administration, and follow-up assessments. The duration of the study will be determined based on the anticipated rate of symptom improvement and the need for robust data collection.

****9. Statistical Analysis:****

Statistical methods, including intention-to-treat analysis, will be employed to evaluate the primary and secondary endpoints. Sample size calculations will ensure the study's power to detect meaningful differences between the treatment and control groups.

****10. Ethical Considerations:****

The study will be conducted in accordance with ethical principles outlined in the Declaration of Helsinki and local regulatory requirements. Informed consent will be obtained from all participants, and measures will be taken to ensure participant confidentiality and safety.

****11. Data Monitoring and Safety:****

An independent Data Monitoring Committee (DMC) will oversee the trial's progress, ensuring participant safety and data integrity. Adverse events will be monitored and reported according to regulatory guidelines.

****12. Conclusion:****

This phase 3 clinical trial aims to contribute valuable insights into the efficacy and safety of drug A in the treatment of schizophrenia. The results of this study may have implications for improving the management of schizophrenia and enhancing the overall well-being of affected individuals.

- Code Conversion

- Prompt:

- Convert below SAS® codes to R codes: Please show R codes only.

- ```
data dm;
 merge demo rand;
 by usubjid ;
run;
```

- Output:

- ```
dm <- merge(demo, rand, by = "usubjid")
```

- Code Development
 - Prompt:

Write a **R script** that groups data by "trta" and "avisitn" columns in "advs.xpt" and calculates the mean of the "aval" column for each group at parmcd = 'WEIGHT'
 - Output:


```
``R
# Load necessary libraries
library(haven)
library(dplyr)

# Read the XPT file
advs <- read_xpt("path/to/advs.xpt")

# Filter the dataset for parmcd == 'WEIGHT'
advs_weight <- advs %>% filter(parmcd == 'WEIGHT')

# Group by 'trta' and 'avisitn' and calculate the mean of 'aval' for each group
mean_aval <- advs_weight %>%
  group_by(trta, avisitn) %>%
  summarise(mean_aval = mean(aval, na.rm = TRUE))

# Print the result
print(mean_aval)
``
```
- Data Analysis
 - Prompt:

Select patients whose sex is male. Use below "Patient Data". Provide me the answers and R codes only.

"Patient Data":

```
{ 'patient': '01-001', 'sex': 'male', 'age': 25, 'race': 'White'},
{ 'patient': '01-002', 'sex': 'female', 'age': 28, 'race': 'Asian'},
{ 'patient': '02-001', 'sex': 'female', 'age': 50, 'race': 'White'},
{ 'patient': '02-002', 'sex': 'male', 'age': 40, 'race': 'Black'}
```
 - Output:

The selected patients whose sex is male are:

 1. {'patient': '01-001', 'sex': 'male', 'age': 25, 'race': 'White'}
 2. {'patient': '02-002', 'sex': 'male', 'age': 40, 'race': 'Black'}

```
# Create data frame
patient_data <- data.frame(
  patient = c("01-001", "01-002", "02-001", "02-002"),
  sex = c("male", "female", "female", "male"),
  age = c(25, 28, 50, 40),
  race = c("White", "Asian", "White", "Black")
)

# Filter male patients
male_patients <- subset(patient_data, sex == "male")

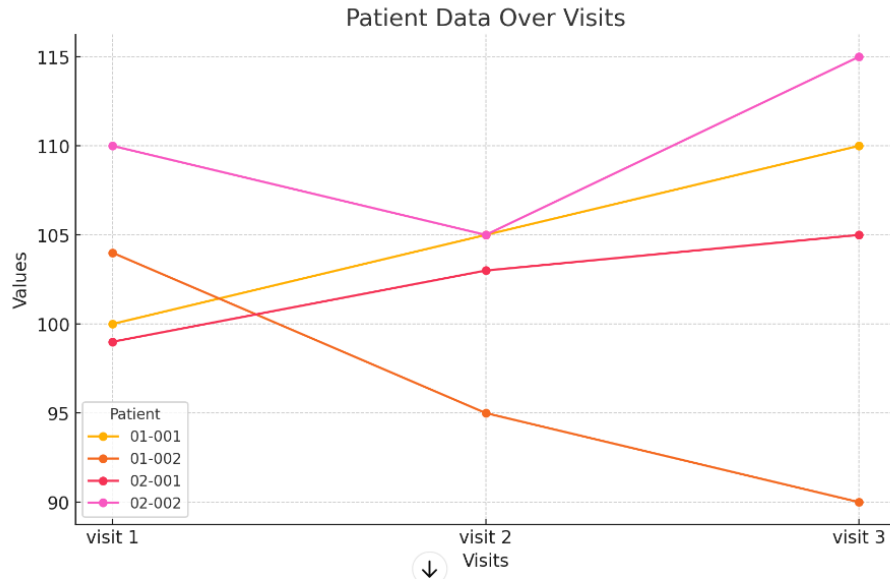
# Print results
print(male_patients)
```
- Data Visualization
 - Prompt:

Create line graph of "Patient Data"

"Patient Data":

```
{ 'patient': '01-001', 'visit 1': 100, 'visit 2': 105, 'visit 3': 110},
{ 'patient': '01-002', 'visit 1': 104, 'visit 2': 95, 'visit 3': 90},
{ 'patient': '02-001', 'visit 1': 99, 'visit 2': 103, 'visit 3': 105},
{ 'patient': '02-002', 'visit 1': 110, 'visit 2': 105, 'visit 3': 115}
```

- Output:



Above prompts and outputs have been created using free version of ChatGPT, and we could build relevant application using Gen-AI API (e.g., gpt-4o-mini).

New Process for Gen AI workflow in Biometrics

Much like an intern with exceptional potential, Gen AI models can provide valuable insights and assist in numerous tasks. However, it is crucial to approach this technology with a validation-focused mindset. The output generated by Gen AI should be regarded as a starting point, a catalyst for further exploration and refinement. The Biometrics team must implement a rigorous validation process to ensure the accuracy, reliability, and compliant use of these Gen AI-generated outputs. By treating Gen AI as a collaborative tool and subjecting its output to rigorous validation, the Biometrics team can harness its potential while maintaining the integrity, compliance and quality expected in the pharmaceutical industry.



Prompting as a Game Changer: Comparison in ChatGPT Users vs Non-Users

Recent survey evidence from a Boston Consulting Group study highlights the transformative impact of prompt-driven AI tools such as ChatGPT on knowledge work productivity and quality. In a direct comparison between ChatGPT users and non-users, individuals who leveraged ChatGPT completed

- 12.2% more tasks on average,
- finished their assigned work 25.1% faster,
- and produced outputs judged to be of 40% higher quality.

These results suggest that the performance gains are not limited to speed alone but extend meaningfully to effectiveness and outcome quality. From a prompt-engineering perspective, these findings position well-constructed prompts as a critical “game changer”: they function as a cognitive amplifier that enables users to better frame problems, explore solution spaces, and refine outputs, thereby converting AI from a passive tool into an active collaborator in complex professional workflows.

CONCLUSION

In conclusion, prompt engineering offers a transformative power for statistical programmers and data scientists. By understanding this concept and its practical applications, statistical programmers and data scientists can unlock new avenues and contribute a significant advancement for code development, content development, communications, brainstorming and much more. 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.

REFERENCES

- OpenAI Prompt Engineering in <https://platform.openai.com/docs/guides/prompt-engineering>
- Prompt Engineering Guide in <https://www.promptingguide.ai/>
- Wiki Prompt Engineering in https://en.wikipedia.org/wiki/Prompt_engineering

CONTACT INFORMATION

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

Kevin Lee
CTO at Clinvia
kevin@clinvia.com

TRADEMARKS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.