

Crafting AI Excellence: An In-Depth Guide to Model Training and Prompt Engineering

Ishmeet Singh¹ and Jahan Kulkarni[#]

¹The Hun School of Princeton, USA

[#]Advisor

ABSTRACT

This paper explores the field of AI prompt engineering, specifically understanding Large Language Model (LLM) training in order to optimize response efficacy, emphasizing the critical stages of data collection, preprocessing, and annotation. Our research outlines key principles of effective prompt engineering, including clarity, specificity, conciseness, engagement, and goal orientation. Through various experiments, we demonstrate how specificity in prompts enhances the detail and accuracy of LLM responses. We also examine the impact of techniques like "Chain of Thought" prompting paired with complementary strategies to extract even more productive responses. Finally, we provide a formula for crafting effective prompts and discuss the broader implications of prompt engineering in fields such as education and programming, showcasing its transformative potential. This comprehensive survey serves as a practical guide for navigating the complexities of AI model training and prompt engineering.

Introduction

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines, typically possessing the ability to perform tasks that practically require human cognition, such as visual perception, speech recognition, decision-making, and language translation. AI can be classified into narrow AI, which is specialized for specific tasks like facial recognition or self-driving cars, and general AI, which, though theoretical, would possess the ability to understand and perform any intellectual task a human can do. Key components of AI include machine learning (the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.)(ML), which involves training algorithms to learn from data and make decisions, deep learning (DL), which uses neural networks(a method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain.) to model complex patterns, and natural language processing (NLP), which focuses on the interaction between computers and human language. The specific type of AI we are focusing on are Large Language Models (LLMs), specifically, OpenAI's GPT-4. Large Language Models (LLMs) use advanced structures called transformers to understand and generate human language. These models are trained on a plethora of text and data, allowing them to recognize complex language patterns and handle tasks like translating languages, summarizing information, creating text, and even holding conversations. By first learning from a broad range of texts and then fine-tuning on specific tasks, LLMs show impressive abilities to understand context and adapt to various applications, significantly enhancing how humans interact with machines and driving progress in automation.

It is very important to understand how a model is made, especially if you rely on it often for help. The model doesn't think as a human, rather it looks at the annotations made in its training stage, and finds the key words in your prompt. searches through its datasets and finds those words, and then it starts guessing on what words follow one another and places a confidence level on each. Understanding this process will help you construct better prompts because you know why prompts must be very specific. In cooking food better ingredients lead to better results. Similarly, when it comes to using AI, the better question you ask the better answer you will get. In cooking it's easy to tell

the higher quality ingredient, however, in prompt engineering it becomes a little trickier. Prompt engineering is the deliberate construction of a prompt with the desire of yielding the most effective and efficient result. In simple terms, it is a science of actively structuring a question with intent to receive the most effective and efficient result from an AI for your problem.

As time goes on, AI will become more and more essential to our day to day life and we will rely on it for more and more tasks. AI has become what the iPhone was in 2007 in terms of easily accessible cutting edge-technology with the potential to entirely change the way people go about their daily lives. As we saw with the iPhone, those who learned how to use it well and earlier rather than later were able to leverage their expertise to utilize this technology to its fullest potential in essentially all aspects of life. Similarly with AI, those who grasp the power it holds will quickly surpass those who fail to learn about this up and coming technological advancement. Similar to learning the most efficient way to search Google by cutting out unnecessary transition and preposition words, educated users of AI will be able to efficiently generate more accurate, concise, and efficient results through the use of particular prompt engineering.

Methods and Results

Defining Quality Prompts

When we had a conversation with ChatGPT 4o, the model broke down the necessities of a good prompt into 5 main categories: Clarity (Ensure your prompt is clear and unambiguous, using proper grammar and punctuation to avoid misinterpretation), Specificity(Provide specific details and context to guide the AI toward generating relevant and useful responses.), Conciseness(Keep your prompt brief and to the point, avoiding unnecessary words to maintain focus and relevance.), Engagement (Make your prompt interesting by adding context or scenarios, and use questions effectively to elicit better responses.), Goal-Oriented (Have a clear goal in mind for what you want the AI to achieve, and tailor your prompt to steer the AI toward fulfilling that goal). For simple and straightforward tasks, a single, well-versed prompted is likely the most efficient way to yield the desired result. Implementing aspects of clarity, examples, and a specified end goal are crucial to assisting the model to generate a result that most closely aligns with the users intended outcome.

Below we have constructed the ideal formula for a prompt geared to answer a specific question with minimal subquestions:

Ideal Single Prompt Formula

Objective Statement

- Clearly define the goal or task
- Ex. "Write a compelling product description for a new smartphone."

Context:

- Provide background information or context
- Ex. "This smartphone has unique features such as a high-resolution camera and long battery life, targeting tech-savvy consumers."

Specific Instructions:

- Give detailed instructions on what is expected
- Ex. "Highlight the key features, benefits, and unique selling points. Mention any special offers or promotions."

Format Requirements:

- Specify the desired format

- Ex. "Provide a 150-word product description."

Constraints and Conditions:

- Mention any limitations or conditions
- Ex. "Use persuasive language. Target tech-savvy consumers."

Examples:

- Give examples of what a good response might look like
- Ex. "For instance, you might start with 'Experience next-level photography with our new high-resolution camera that captures every detail.'"

Decomposed Prompting

The nuance of prompt engineering however goes further than these 5 facets, these 5 metrics are only effective measurements for simple questions such as "How do I slice an apple?" The more subjective, complex, and multi-part questions have a completely separate set of metrics, the question is which type of prompt is better to use? To solve this question we had ChatGPT 4o write an essay using two prompts with different prompting techniques. The two prompts both tasked ChatGPT with the same challenge, however, one prompted the LLM to write the essay in one go while the other used a decomposed prompting method, asking the model to decompose an essay into its outline, then to write each individual part of the outline alone. The full essay prompt was "Write an essay about food" and the single paragraph prompt was "Write an outline for an essay about food", followed by "Write the intro paragraph". Below are the paragraphs:

Decomposed Prompt Testing Results

Introduction Paragraph from full essay: Food is not just a source of sustenance but a cornerstone of human culture and survival. It plays a multifaceted role in our lives, extending beyond mere nutrition to include cultural significance and social bonding. Understanding the importance of food helps us appreciate its value in our daily lives. Food is essential to humans not only for its nutritional value but also for its cultural significance and role in fostering social connections.

Introduction Paragraph from decomposed prompts: Food is not just a source of sustenance but a cornerstone of human culture and survival. Throughout history, food has played a crucial role in shaping civilizations, fostering social bonds, and defining cultural identities. Beyond its cultural significance, food is essential for human health, providing the necessary nutrients and energy to sustain life. As such, understanding the multifaceted importance of food is vital. This essay will explore why food is indispensable to humans, focusing on its nutritional value, cultural significance, and role in fostering social connections.

Decomposed Prompt Testing Analysis

As we can see, there is a clear distinction between the results. Both outputs cover the same information overall, however, the second paragraph contains more detail as well as a style and elegance that is absent in the simple and basic first output. We credit this to the fact that AI responds better to specificity in prompts, and while the original prompt "write an essay about food" was specific about the topic, it is not as specific as the prompt "write the introduction for an essay about food." The more specific prompt allows GPT to focus on creating a good introduction, as opposed to a good overall essay which may contain a lesser introduction and still be considered 'good'.

As demonstrated by the prompt grader, specificity is key when it comes to engineering prompts and it is essential to creating a productive prompt. While in some cases the Ideal Single Prompt Formula is very useful to extract a specific, accurate answer to a more simple and straightforward problem, the decomposed prompting method is more useful for more complex and multi-faceted challenges.

Complex Problem Prompting

This formula is designed for minor problems that only have one definitive or objective solution with very few minor sub-questions and nuances. The formula for major issues is similar, yet has some notable differences. For example, if you are tasked with having to research a complex topic such as chemistry, your first prompt is along the lines of: “Please provide an overview of the fundamental principles of chemistry, including atomic structure, chemical bonding, and the periodic table, to help me understand the basics before diving into more complex topics.” The LLM will respond with a very brief overview on each of the topics in the prompt. After this you should use the ideal prompt formula above to take a deeper dive on each of the subjects. This is because what separates AI from sources like Google is AI can hold a conversation, and when it remembers that you asked about atomic structure in the question prior, it will build on its response and expand your understanding.

Chain-of-Thought Prompting

“Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review” written by Shengxin Zhu explores the impact of employing a Chain of Thought (Asking the model to explain its process) prompting technique while tasking ChatGPT with a unique and abstract problem to solve. Specifically, using the phrase “think step by step” at the end of the question prompt yielded more detailed and thorough answers and explanations, a method labeled ‘Chain-of-Thought Prompting (CoT)’.

CoT paired with Decomposed Prompting

However, while word choice is definitely important in shaping an LLM’s response the way the user intends to receive it, the order and structure in which these words are used is equally as important. In our experiment we tested the different responses that ChatGPT generated as a result of two different lines of questioning. The first line consisted of a complex, abstract question immediately followed in the same entry box with the additional statement “think it through step by step and provide a definitive answer”, the key component of CoT Prompting. The second line of questioning in a new chat asked the exact same complex, abstract question in the exact same wording, but without the additional statement. Then, after ChatGPT’s initial response to the question, we prompted a second time as a reply with the statement “think it through step by step and provide a definitive answer”, effectively using the decomposed prompting method from earlier, but adapted to the CoT prompting method as well.

After testing with several different abstract questions, we found a pattern amongst ChatGPT’s responses. In the conversation where we prompted ChatGPT twice, we received more detailed and thorough answers and explanations. This is an interesting observation because even though we are prompting ChatGPT the exact same questions in the exact same wording, different responses are generated due to the structure in which we prompt the LLM. This is likely due to ChatGPT’s feature to adjust its response as a result of user feedback. In the situation where we only prompted ChatGPT once, the prompt served as an initial command with specific detail regarding the way in which ChatGPT was expected to answer the question. However, in the second situation, not only did we provide specific details for how we expected ChatGPT to answer the question, but by decomposing the prompt and asking the CoT as a second individual prompt, ChatGPT likely received the CoT instruction as feedback to adjust or add on to its original response. Thus, ChatGPT’s second response to the second prompt was more thorough and detailed because it didn’t just process the Chain of Thought line command as an outline, but as direct feedback to the way it had originally responded, leading the LLM to focus more heavily on that specific aspect of the prompt as opposed to accepting it as a general guideline for its response.

CoT/Decomposed Prompt Testing Results

Sample input 1:

“Which came first, the chicken or the egg? Think it through step by step”

Sample output 1:

““““

- Evolutionary Background:
 - Birds evolved from reptiles. The first birds appeared during the Jurassic period.
 - The process of evolution involves gradual changes over many generations.
- Egg-Laying Predecessors:
 - Reptiles and other species that existed before chickens laid eggs.
 - Therefore, egg-laying as a reproductive method predates the existence of chickens.
- Genetic Mutation:
 - A creature very similar to a chicken (proto-chicken) would have laid an egg.
 - A genetic mutation within this egg resulted in the first true chicken.
 - Therefore, an egg containing a chicken must have come before the actual chicken itself.

””””

Sample input 2:

“Which came first the chicken or the egg”

ChatGPT preliminary response

“Think it through step by step”

Sample output 2:

““““

- Understanding Evolution:
 - Evolution is a gradual process where species change over time through mutations and natural selection.
 - Birds, including chickens, evolved from earlier, reptilian ancestors.
- Proto-Chickens:
 - Before the existence of the modern chicken, there were birds very similar to chickens but not quite the same (let's call them proto-chickens).
 - These proto-chickens reproduce by laying eggs.
- Genetic Mutation:
 - Genetic mutations occur randomly and can be passed on to offspring.
 - At some point, a genetic mutation occurred in the reproductive cells of a proto-chicken.
 - This mutation resulted in a slight change in the DNA, creating a new genetic code that would eventually be recognized as the modern chicken.

””””

CoT/Decomposed Prompt Testing Analysis

While the two responses are similar and cover a lot of the same facts such as a general overview of the process of evolution, the second response clearly has a more in depth explanation of both evolution and its direct application to this specific problem. While the first output did state “birds evolved from reptiles”, the second output offered a clearer explanation specifying “Evolution is a gradual process [...]” and “[chickens] evolved from earlier, reptilian ancestors”. The second output also clearly defines the idea of a “proto-chicken” before analyzing its importance, as opposed to

the first output which vaguely glossed over the concept before continuing the line of reasoning. Overall, the second response uses more in depth analysis and explanation, as well as more thorough definitions and clarifications for more complex ideas and concepts. This is likely because the format of using two separate prompts is received as feedback rather than instruction, causing the LLM to focus more heavily on the command because feedback is associated with dissatisfaction from the user and thus the LLM concentrates on employing the command in its response at a higher degree. When receiving several pieces of instruction as an initial prompt, the LLM interprets all parts as having equal weight and importance, causing responses to be less concentrated on each specific instruction provided.

However, by splitting a two part prompt into two parts, the LLM can generate an initial response solely focused on the first part of the prompt, followed by a revised response that focuses more heavily on the second part of the prompt while also incorporating large parts of the initial response which was heavily concentrated on the first part of the prompt. The result is a final response that heavily concentrates on both parts of the prompt individually and cooperatively instead of all parts of the prompt as a whole. This observation is similar to the observation made about optimizing the LLM's ability to perform multiple functions by splitting the parts of the prompt into smaller sub prompts rather than a larger, singular prompt. However, the key difference is that in this particular instance instead of asking ChatGPT to perform a function such as "write me an essay", we are asking the LLM to problem solve an abstract challenge using critical thinking. Therefore, the notable conclusion beyond the observation itself, is that this method of optimizing LLM results by splitting prompts into smaller sub prompts holds true for tangible commands as well as abstract problem-solving.

Discussion

While this paper primarily focuses on prompt engineering in relation to LLMs, the principles and techniques discussed can be applied to a wide range of AI systems. The discipline of prompt engineering involves constructing inputs that guide AI models to produce the most effective and accurate outputs. This methodology is not confined to language models but extends to various AI applications, including image recognition, recommendation systems, and automated decision-making tools.

Image Recognition

Similar to how specific prompts improve LLM responses, more detailed instructions enhance the accuracy of image recognition systems. For instance, specifying "identify all red cars in this image" instead of "identify cars" helps the model focus on a narrower task, improving precision. This specificity can significantly enhance the performance of computer vision applications used in security, healthcare, and autonomous driving (Maryville University Online) (Knowledge at Wharton).

Recommendation Systems

In recommendation engines used by platforms like Netflix or Amazon, clear and specific user inputs (such as "recommend action movies with a strong female lead") yield better results compared to vague requests (like "recommend a movie"). This principle of providing detailed and context-rich prompts helps the AI system to better understand user preferences and deliver more relevant suggestions, thereby improving user satisfaction and engagement (Daffodil Software).

Automated Decision-Making

In financial AI systems, detailed and structured prompts (e.g., "analyze the impact of interest rate changes on small-cap stocks over the past five years") lead to more insightful analyses than general inquiries. This approach ensures that the AI model processes specific variables and contexts, leading to more accurate and actionable insights, which are crucial in fields such as finance, healthcare, and logistics (India STEM Foundation).

Complex Problem Prompting

This formula is designed for minor problems that only have one definitive or objective solution with very few minor sub-questions and nuances. The formula for major issues is similar, yet has some notable differences. For example, if you are tasked with having to research a complex topic such as chemistry, your first prompt is along the lines of: "Please provide an overview of the fundamental principles of chemistry, including atomic structure, chemical bonding, and the periodic table, to help me understand the basics before diving into more complex topics." The LLM will respond with a very brief overview on each of the topics in the prompt. After this you should use the ideal prompt formula above to take a deeper dive on each of the subjects. This is because what separates AI from sources like Google is AI can hold a conversation, and when it remembers that you asked about atomic structure in the question prior, it will build on its response and expand your understanding.

Conclusion

In this paper, we present an in-depth examination of the AI model training process, emphasizing the pivotal stages of data collection, preprocessing, and annotation. We delve into the essential techniques of prompt engineering, showcasing its role in crafting productive and streamlined AI responses. Through various experiments, we demonstrate how specificity and iterative prompting significantly enhance AI performance. We also explore broader applications of prompt engineering, highlighting its transformative potential in fields like education and programming. Finally, we offer a forward-looking perspective on the future of AI model training and prompt engineering, stressing the need for continuous advancements in data processing and prompt construction. In summary, prompt engineering stands as a critical method for optimizing LLM capabilities using methods such as Chain of Thought and Sub Prompting to yield more thorough responses, and we hope this paper serves as a foundation for ongoing research and innovation regarding new techniques in the field.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Daffodil Software. (n.d.). Improving AI Recommendation Systems: Techniques and Best Practices. Retrieved from <https://www.daffodilsoftware.com/blog>

India STEM Foundation. (n.d.). Advanced AI Systems in Financial Decision-Making. Retrieved from <https://www.indiastemfoundation.org>

Knowledge at Wharton. (n.d.). Enhancing Image Recognition with AI: A Comprehensive Guide. Wharton University of Pennsylvania. Retrieved from <https://knowledge.wharton.upenn.edu>

Maryville University Online. (n.d.). Applications of Image Recognition in Healthcare. Retrieved from <https://online.maryville.edu>

Zhu, S. (n.d.). Unleashing the Potential of Prompt Engineering in Large Language Models: A Comprehensive Review. Retrieved from <https://www.example.com>