

ChatGPT Prompt Engineering for

AI

Chao Huang, PhD
Yu-Wen Lin



I know EXACTLY
what you want!



***ChatGPT* Prompt Engineering for All**

**Chao Huang, PhD
Yu-Wen Lin**

Table of Contents

[Foreword](#)

[Reading Guide](#)

- [1. Part One, "Fundamentals":](#)
- [2. Part Two, "Principles":](#)
- [3. Part Three, "Tactics":](#)
- [4. Part Four, "Strategies":](#)
- [5. Part Five, "Future Trends":](#)

[Part One: Fundamentals](#)

[Chapter 1: Overview of Current Development of Generative AI](#)

- [1. The beginning of Generative AI: The Emergence of ChatGPT](#)
- [2. "Prompt" and "Prompt Engineering"](#)
- [3. Custom GPTs-The Current Big Trend](#)
- [4. How to Communicate with Machines in the Future](#)

[Chapter 1 Summary](#)

[Chapter 2: Introduction to AI and Recent Turning Points](#)

- [1. Supervised Learning, Unsupervised Learning, and Generative AI](#)
- [2. Recent Turning Points in AI](#)
- [3. Massive Funding pours into AI startups](#)

[Chapter 2 Summary](#)

[Chapter 3: Introduction to Principles of Large Language Models](#)

- [1. LLM Structure](#)
- [2. How ChatGPT Generates Responses](#)
- [3. The Name "ChatGPT"](#)

[Chapter 3 Summary](#)

[Chapter 4: Characteristics of Large Language Models](#)

- [1. How an LLM processes and retains information](#)
- [2. LLM's Cognitive Ability](#)
- [3. Operating Modes of LLM](#)

[Chapter 4 Summary](#)

[Part Two: Principles](#)

[Chapter 5: What is Prompt Engineering](#)

- [1. Definition of Prompt Engineering](#)
- [2. The Nature of Prompt Engineering—Both Art and Technique](#)

[Chapter 5 Summary](#)

[Chapter 6: Characteristics of Generative Ais such as ChatGPT](#)

- [1. ChatGPT is almost all-knowing and will answer when asked.](#)

- [2. ChatGPT can converse in natural language and understands multiple languages.](#)
 - [3. ChatGPT operates by generating the most probable output, it has the following characteristics:](#)
- [Chapter 6 Summary](#)

[Chapter 7: Basic Principles of Prompt Engineering](#)

- [1. Important General Principles](#)
- [2. Task Execution Principles](#)

[Chapter 7 Summary](#)

[Chapter 8: Special Spell Principles—Directive Prompts](#)

- [1. The Real Effectiveness of Spells—Directive Prompts](#)
- [2. The Originality of Directive Prompts](#)
- [3. Examples of Directive Prompts](#)

[Chapter 8 Summary](#)

[Part Three: Tactics - Basic Prompts](#)

[Chapter 9: Basic Prompt Elements](#)

- [1. Core Elements of Prompts](#)
- [2. Auxiliary Elements of Prompts](#)
- [3. Effective Prompt Design Process](#)

[Chapter 9 Summary](#)

[Chapter 10: Framework-Based Prompts](#)

- [1. What are Framework-Based Prompts](#)
- [2. The all-in-one 3W-CoSeed Framework Prompt](#)
- [3. Common Framework Prompts](#)

[Chapter 10 Summary](#)

[Chapter 11: Applications and Exercises \(1\)](#)

[Chapter 12: Basic Thought Based Prompts](#)

- [1. Shot Prompting](#)
- [2. Chain-of-Thought Prompting](#)

[Chapter 12 Summary:](#)

[Chapter 13: Application and Practice \(2\)](#)

- [1. Food Category](#)
- [2. Shopping Category](#)
- [3. Home Category](#)
- [4. Exercise Category](#)
- [5. Transportation Category](#)
- [6. Travel Category](#)
- [7. Socializing](#)

[Part Four: Strategies – Advanced Prompts](#)

[Chapter 14: Executing Logic](#)

- [1. Executing Iteration](#)

[2. Executing Decisions](#)

[3. Executing Functions](#)

[Chapter 14 Summary:](#)

[Chapter 15: Advanced Prompts—Variations in Chain-of-Thought](#)

[1. Combining Chain-of-Thought with Few-Shot Prompts](#)

[2. Chain-of-Thought with Self-Consistency \(CoT-SC\)](#)

[3. Chain-of-Thought with Task Decomposition \(CoT-TD\)](#)

[Chapter 15 Summary](#)

[Chapter 16: Advanced Prompting—Tree of Thoughts](#)

[Chapter 16 Summary](#)

[Chapter 17: Advanced Prompting—Skeleton of Thought](#)

[1. Generating Text Skeleton](#)

[2. Summarizing Text](#)

[Chapter 17 Summary](#)

[Chapter 18: Advanced Prompting—Chain of Density](#)

[1. Chain of Density](#)

[2. Summary by ChatGPT](#)

[3. Summary by Chain of Density Technique](#)

[4. Performance of CoD Technique](#)

[Chapter 18 Summary:](#)

[Part Five: Practical Applications and Future Trends](#)

[Chapter 19: Let ChatGPT Ask You Questions](#)

[1. Scenario Simulation](#)

[2. Assisting with Exam Preparation](#)

[3. Self Fine-tuning](#)

[Chapter 19 Summary:](#)

[Chapter 20: Interacting with ChatGPT](#)

[1. Automatic Prompt Design](#)

[2. Reviewing the Generated Prompt](#)

[3. Techniques for Effective Communication](#)

[Chapter 20 Summary:](#)

[Chapter 21: Custom ChatGPT - GPT](#)

[1. What is GPT](#)

[2. How to Use GPT](#)

[3. How to make an GPT](#)

[Chapter 22: Future Trends](#)

[1. AI Agents](#)

[2. Multi-Modal LLM](#)

[3. Artificial General Intelligence](#)

Appendix

[A. List of Directive Prompts](#)

[B. Output Formats](#)

[C. Types of Emotional Styles](#)

[Afterword](#)

Foreword

In September 2023, I published an introductory book about ChatGPT.

As generative AI rapidly evolves, one thought kept coming back to me: "Knowing how to give prompts is crucial for effectively using AI." This idea inspired me to start writing this book : "ChatGPT Prompt Engineering for All" with Lin.

Since ChatGPT emerged, various generative AIs for chatting have proliferated rapidly, and the number of users has skyrocketed. In November 2023, OpenAI launched a feature allowing everyone to create their own custom ChatGPT (GPT). By early 2024, over three million GPTs had been created, covering fields including education, business, programming, gaming, etc. These GPTs are also designed by using prompts.

Generative AI and GPTs are now everywhere, and prompts are the key to using them effectively. Prompts are based on natural language, simple yet powerful. ChatGPT is like an old sage who has read countless books, but you need prompts to communicate with it.

Many websites and books offer various prompts for you to copy, but we believe it's better to rely on your own skills. Having a skill is more valuable than wealth. Learning the technique enables you to use it proficiently and innovate continually. That's why we write this book, focusing on a method called Prompt Engineering, to explore ways to craft effective prompts and share them with everyone.

Prompt Engineering is a powerful technique, but it is usually used by computer engineers in AI labs, making it difficult to learn. In this book, we explain Prompt Engineering in a simple way, using common examples to illustrate it, allowing everyone to learn these techniques easily and use AI efficiently in daily life.

Prompts are essentially language. We often say, "Language is an art," but art is abstract and hard to teach. Therefore, we will analyze the basic principles

of Prompt Engineering, providing everyone with a fundamental framework to start learning. Even Picasso first learned basic techniques, then combined them with knowledge and creativity to become a master.

The new ChatGPT-4o has multimodal capabilities, allowing it to listen, speak, read, write, and even understand images and generate text from images. Although its performance is outstanding, efficiency will still suffer without effective prompts. In professional applications such as medical diagnosis and legal consultation, precise prompts remain crucial.

NVIDIA, the current leader in AI hardware, had its CEO Jensen Huang say in March 2024, "Generative AI is bridging the technology gap. You no longer need to be an engineer to succeed." He also said, "You just need to become a prompt engineer. Who can become a prompt engineer? When my wife talks to me, she is “prompting” me. We all need to learn how to prompt AI, just like learning how to communicate with a partner."

Now, let's start this journey! Learn how to communicate with AI and unlock its potential. Master the art of prompting, create your own GPT, and become an AI hero in your career and life.

Are you ready? Let's enter the wonderful world of AI together.

Reading Guide

S pecial note to the readers:

At the releasing of this book, OpenAI launched its ChatGPT o1 series, which can 'think' before responding. They claim it "uses advanced reasoning."

Unlike earlier models, o1 breaks down tasks step by step, delivering more accurate and deeper answers. It does this using Chain-of-Thought (CoT) prompting, which helps dissect complex queries like a human. However, it's 3 to 4 times more expensive than the 4o model.

Chapter 12 and 15 teach you how to use CoT and Advanced CoT. After learning these techniques, you can use them with the cheaper ChatGPT 3.5 or 4 models to get similar results to the o1.

The inspiration behind writing this book was to create a guide so everyone can learn Prompt Engineering, not just engineers.

When using generative AI tools like ChatGPT, the "prompt" is crucial. Future AI PCs, AIoT devices, and computer peripherals will heavily rely on prompts. "Prompt engineering" involves designing efficient prompts for optimal communication with AI, but it's not easy for the average person to master.

In this book, we simplify prompt engineering and use real-life examples to make it accessible to everyone. We propose to redefine prompt engineering to better align with practical usage.

Prompting is both an art and a science. A well-crafted prompt guides AI to generate innovative thoughts, and when taught through basic prompting methods, it empowers the public to create effective prompts. Like art masters who blend skill with creativity, prompt engineering merges technique with artistry, making it both tactical and strategic.

Hence, this book is divided into five parts:

1. Part One, "Fundamentals":

This part covers the basic knowledge of AI and large language models required to understand this book. Readers already familiar with these topics can skip this part, though it's recommended to review each chapter's summaries.

2. Part Two, "Principles":

This part explains the principles of prompting (the art). It covers what prompt engineering is, the characteristics of generative chatbots, the principles of creating effective prompts, and some very useful "directive prompts."

3. Part Three, "Tactics":

This part introduces basic effective prompts, techniques that can be taught in steps, including framework prompts and thought prompts. By applying and adjusting these techniques to specific situations, one can learn to achieve good results very quickly.

4. Part Four, "Strategies":

This part covers advanced prompts, blending techniques and creativity. It explores "strategy," where humans create comprehensive plans and let ChatGPT handle the details. Techniques like chain-of-thought variations, thought trees, frameworks, and density chains, used by major companies, help readers solve everyday problems.

5. Part Five, "Future Trends":

After covering prompt engineering—principles, basics, and advanced—this part uses real examples to show problem-solving with prompt techniques. It also explains creating custom ChatGPT applications, like collaborative summarization and translation, and discusses future trends in prompting.

I aimed to make prompt engineering easy to learn. Anyone will be able to create highly efficient prompts and communicate effectively with AI after reading this book.

Reminders:

In this book, "AI" refers to generative text-based AI, covering large language models (LLMs) like ChatGPT, Copilot, Gemini, Claude etc. Prompt

engineering includes various techniques that can be used individually or combined. Feel free to navigate the book in random order as you wish.

Part One: Fundamentals

This Part presents the basic knowledge of large language models (LLMs) and generative artificial intelligence (AI), laying the foundation for future learning.

If you already have a sufficient background in these topics, you can skip this part. However, it's recommended that you spend a few minutes browsing each chapter's summaries before moving on to Part Two.

In the following chapters, we will explore the key milestones and rapid advancements in generative AI, starting with the breakthrough development of ChatGPT and its profound impact on the tech industry.

Chapter 1: Overview of Current Development of Generative AI

To grasp today's AI landscape, we begin with the emergence of pioneering models like ChatGPT.

1. The beginning of Generative AI: The Emergence of ChatGPT

In 2023, OpenAI led a technological whirlwind with its revolutionary generative AI product, ChatGPT.

In just one year, OpenAI continuously made technical breakthroughs, stirring up a storm in the tech world. Various tech giants entered the arena, launching their own competitive products.

Let's take a look at this amazing journey.

1) The Beginning of the Whirlwind

OpenAI launched its generative chatbot ChatGPT on November 30, 2022, and it quickly became the industry benchmark.

Utilizing a unique large language model and deep learning technology, it enabled human-machine interaction in various natural languages, quickly attracting one million users in five days. Within two months, the user base exceeded one hundred million. Before the launch of ChatGPT, Microsoft invested over \$10 billion to acquire nearly half of OpenAI's shares.

2) Milestones of Innovation

Nov., 2022: ChatGPT :

The initial release of ChatGPT was based on the large language model GPT-3.5.

March, 2023: GPT4 :

OpenAI launched GPT-4, a significant upgrade that greatly enhanced natural language processing capabilities, enabling more precise understanding and responses to user queries.

GPT-4's capabilities are reflected in its neural network's parameters. GPT-3 had 175 billion parameters, while GPT-4 reportedly has significantly more,

although the exact number remains undisclosed.

September, 2023: DALL-E 3 :

OpenAI launched DALL-E 3, the latest version of its text-to-image generator, along with its speech-to-text tool, Whisper. OpenAI took a significant step in multimodal generative AI, enabling seamless transitions between text, speech, and images. ChatGPT can now listen to voice, see things, write, speak, and even draw.

Nov., 2023: custom ChatGPT (GPT) :

OpenAI held its first Developer Conference, announcing several technical upgrades, but most importantly, the launch of custom ChatGPT (GPT) and the GPT Store. The significance of these will be detailed later.

These technical breakthroughs not only attracted global users but also sparked an AI arms race among global tech giants, leading to the birth of many similar products.

3) Emergence of Similar Products

Following the success of ChatGPT, other tech giants quickly entered the generative AI market, launching their own innovative products.

March, 2023: Google Bard

Google, with its solid AI foundation, officially entered the race by launching its chatbot Bard. Google had previously acquired Deepmind, a pioneer in Go-playing AI. The “T” in ChatGPT stands for “Transformer,” a neural network architecture that serves as a crucial component of ChatGPT. The Transformer architecture was in fact introduced in 2017 by a team from Google Brain.

May, 2023: Microsoft Bing AI Copilot

Microsoft launched Bing AI Copilot, integrated with OpenAI technology, providing a new web search experience through its search engine.

July 2023: Anthropic Claude-2

Anthropic launched Claude-2, ChatGPT's biggest rival. Anthropic was formed by a team who left OpenAI due to different safety philosophies and received investment from Google and Amazon.

September, 2023: Facebook Meta AI

Facebook unveiled its chatbot Meta AI, integrated into WhatsApp, Instagram, and Facebook Messenger, executing user prompts for text and voice interfaces.

Amazon did not lag behind either. At its "Accelerate 2023" reseller conference, it launched a generative AI to help sellers write product descriptions and assist buyers in finding what they want.

October, 2023: Amazon AI Service

Amazon announced the fully managed AI service, Amazon Bedrock, and AI programming assistant Amazon Code Whisperer, providing real-time suggestions during programming to help write code quickly and safely.

November 2023: xAI Grok

Elon Musk's xAI company launched Grok, a chatbot that avoids diplomatic language and answers sensitive questions that most AI systems refuse to address.

December, 2023: Google Gemini

Google renamed Bard to Gemini, claiming it is more powerful than ChatGPT-4.

Tech giants, recognizing the unprecedented market opportunity, have all found entry points and joined the race, resulting in the diverse products mentioned above.

These products represent the state of the market at the time of writing this book. More are on the way, arriving with mind-boggling speed and performance.

Many startups have also launched generative AI products and tools which mimic ChatGPT's functionality, showcasing the potential applications of generative AI technology in various fields. These products are generally less known to the public.

The emergence of ChatGPT and other generative AI products has advanced AI applications in education, creativity, writing, and customer service, among other fields. It has also sparked heated discussions on AI ethics, privacy protection, and technological development.

Their emergence heralds a future of more intelligent and natural human-machine interaction, marking a new era of "human and AI" communication through natural language.

This natural language communication, driven by "**prompts**," is central to the functionality and versatility of generative AI models like ChatGPT.

Amidst the growing array of generative AI, prompts remain central for guiding AI responses. Next, we'll explore "prompt engineering," the key to mastering AI interactions.

2. "Prompt" and "Prompt Engineering"

"Prompt" plays a crucial role in our interactions with generative AI. Prompts are the language we use to communicate with AI. With carefully crafted prompts, we can guide AI to generate more accurate and effective responses. Think of prompts as magical spells that, when spoken, make AI produce exactly what you want.

As AI technology advances, we increasingly rely on it to solve problems, create content, or provide suggestions. In this process, effective prompts have become the key to obtaining the results we need.

The goal of this book is to teach you how to "create effective prompts" to communicate with AI and get the results that best meet your needs.

Prompts themselves are just natural language, and as we often say, "language is an art". Thus, prompts are an art form. However, many treat them as a discipline. "Prompt Engineering" is popularized as a discipline, making it easier for people to learn.

Learning how to "create effective prompts" is a skill that can be used not only with ChatGPT but also with all generative AI like Copilot, Gemini, and Claude.

At the end of 2023, another significant event magnified the importance of creating effective prompts: the launch of "custom ChatGPTs" or so called "GPTs" by OpenAI.

Understanding prompt basics sets the stage for discussing "GPTs," which allow tailored AI use. We'll now examine how this trend is changing AI

interaction.

3. Custom GPTs-The Current Big Trend

On November 6, 2023, OpenAI held its first developer conference, known as OpenAI DevDay. Despite it being the first conference, nearly a thousand developers attended.

These developers had independently developed many custom ChatGPTs over the past year. They provided user interfaces that allowed users to ask questions in specialized fields, relayed these questions to ChatGPT, obtained answers, and then forwarded them to users.

These custom ChatGPTs attracted significant attention and user engagement in the market, but the developers were in for a big surprise from OpenAI.

At the conference, OpenAI CEO Sam Altman announced that OpenAI would allow users to easily design custom ChatGPTs, called GPTs.

Any paying ChatGPT user can use tools provided by OpenAI on its website to design their own domain-specific ChatGPT in just a few minutes. We created a joke-telling GPT on OpenAI's website. The entire process took only five minutes.

The truth is, OpenAI has probably undercut the business model of some of its developers.

However, the name GPT is easily confused with the large language model GPT behind ChatGPT, so we will refer to it as "custom GPT" or "custom ChatGPT" in this book.

OpenAI also announced the launch of the GPT Store by the end of 2023. Any paying ChatGPT customer can showcase their custom GPTs on this platform. If other users use them, the revenue generated could be shared with the creators. It's essentially a YouTube model.

The excitement surrounding this launch was evident in several ways. The morning after the conference, around 6 a.m., the ChatGPT servers experienced a major outage.

Are we kidding?! In less than a year, OpenAI amassed over 100 million

registered members, including more than 200,000 paying subscribers, without any server crashes. The ChatGPT server crash highlighted the popularity of these new services.

Another indicator of the popularity of GPTs was that within two weeks of the developer conference, over 20,000 custom GPTs were created. By early 2024, this number had exceeded three million!

These GPTs are incredibly diverse, encompassing everything from simple text generation to role-playing and complex decision-support systems. Custom GPTs showcased amazing diversity, providing more precise and valuable content and services based on specific industries or personal preferences.

ChatGPT initiated a significant trend: we no longer just use the general ChatGPT. Now, we can create custom GPTs for our specific contexts and needs. This means that from businesses to individuals, everyone can have a "tailor-made" ChatGPT assistant.

If you design a custom GPT that is used by many, you could become an "GPT celebrity" and profit handsomely.

When you create a custom GPT on OpenAI's website, you use a tool called the "GPT builder." It helps you name your custom GPT, design the logo, and build all the features.

How do you use this GPT Builder? You guess it! By using "prompts" to communicate your design to the GPT Builder.

The rise of custom GPTs further emphasizes the importance of effective prompting skills. Only through precise prompts can these highly custom ChatGPTs maximize their potential.

From generative AI like ChatGPT to custom GPTs, "prompts" are the tools for communication, making "creating effective prompts" a skill we must acquire now.

In the future, everyone needs to know how to create effective prompts!

4. How to Communicate with Machines in the Future

Imagine that in the near future, any 3C device in our daily lives could have AI capabilities, and we will use natural language (prompt) to interact with them. Whether it's a smartphone, computer, or even a smart home system, we will talk to them with natural language or prompts.

In fact, this is already happening. Have you seen the latest self-driving cars? The dashboard is clean and tidy. Everything is controlled by voice command or prompts.

Language-based interaction will enrich our lives, enabling us to collaborate more naturally with AI-from simple daily tasks to complex work processes-creating a higher quality of life.

The purpose of this book is to teach you how to create and design the most effective prompts through Prompt Engineering, enabling all machines (computers, phones, tablets, and 3C products) to perform at their best and most efficiently.

To emphasize once again, "prompts" are crucial and will be the main way humans and machines will communicate in the future.

In this new era based on natural language, mastering "Prompt Engineering" becomes a core skill. It is not only an essential skill for professionals but also a basic ability for the general public to integrate into this AI era.

In 2023, we were amazed by ChatGPT; in 2024, we will integrate generative AI into our lives. How will we integrate it? Through learning Prompt Engineering and using effective prompts. So let's embark on this promising learning journey together!

Chapter 1 Summary

- ❖ ChatGPT emerged in Nov. 2023. Within a year, tech giants flooded the market with generative AI products (Copilot, Gemini, Claude, Meta AI, Accelerate 2023.)
- ❖ "Prompts" are how we use to communicate with generative AI. "Prompt Engineering" is the process of learning how to create and design effective "prompts."
- ❖ Custom GPTs are the future trend of generative AI, and prompts are also

the method for creating custom GPTs.

- ❖ In the future, we will communicate with any machine using natural language, i.e., prompts.
- ❖ Glossary:
 - 1) GPT is the large language model (LLM) behind ChatGPT; ChatGPT is a chatbot built on top of GPT.
 - 2) In November 2023, OpenAI announced that users could DIY custom ChatGPTs on the official website, called GPTs.

Chapter 2: Introduction to AI and Recent Turning Points

Artificial Intelligence (AI) is a technology that allows machines to mimic human intelligence. From intelligent customer service to autonomous vehicles, AI is transforming the way we live.

Although AI's development spans over 70 years, it formally became a scientific research field in the 1940s. This transformation was primarily driven by scientists, mathematicians, and philosophers in the 1940s and 1950s.

One of the most crucial figures was Alan Turing, who in 1950 published a groundbreaking paper proposing the “Turing Test” to determine whether a machine exhibits human-like intelligence. Turing's theories had a profound impact on AI developments, making him one of the most important pioneers in the field.

During earlier years of AI, Engineers studied image recognition, neural networks, and expert systems. At that time, computer power was quite limited and it was hard to imagine AI becoming a reality in daily life.

However, rapid advancements in integrated circuits, including the development of GPUs (Graphics Processing Units) for high-speed computations, have made today's AI applications a reality.

The history of AI is extensive, filled with various research methods and models that can be thought of as a series of tools. Here, I will only introduce the three tools most relevant to the current development of AI: Supervised Learning, Unsupervised Learning, and Generative AI.

These three tools are widely used, not as sequentially inherited methods, but as approaches that can be applied individually or in combination.

With this historical context in mind, let's explore the core techniques driving current AI advancements. We will examine three key approaches—Supervised Learning, Unsupervised Learning, and Generative AI—that form the foundation of modern AI applications.

1. Supervised Learning, Unsupervised Learning, and Generative AI

1) Supervised Learning: Machines Grow Under Guidance

If a student has never seen apples nor oranges, how would you teach him to distinguish between the two? You would show him pictures of apples and oranges while telling them which is which. This is supervised learning—labeling items clearly (e.g., this is an apple, that is an orange), so AI can learn to recognize things based on their features and labels.

The AI will identify features in the pictures to distinguish between apples and oranges, and even if shown a squashed apple later (as long as it's not so squashed that even humans can't recognize it), it can still successfully identify it.

For example, Spam Filters learn to recognize spam emails using many emails labeled as "spam" or "non-spam". AI can also label which types of ads users have clicked on, a technology that brings Google hundreds of millions of dollars in advertising revenue each year.

Before the advent of ChatGPT, supervised learning was already widely used in our daily lives, though many of us may not have noticed.

2) Unsupervised Learning: Machines Explore Freely

Supervised Learning is like a teacher guiding students in class, while Unsupervised Learning is like letting the AI student explore on its own. The AI student must discover patterns and relationships in the objects on its own.

It's like a student who doesn't like listening to lectures and prefers exploring and discovering things by themselves. Without a teacher telling it right from wrong, it can only find relationships and patterns among objects. If given a random assortment of toys, the AI will identify relationships among them and categorize them.

For example, it might separate large and small toys or set aside toys that are different from others, such as defective toys.

In practice, it can detect defective products or classify online shopping customers into different types to distinguish their buying behaviors.

3) Generative AI: Creating Original Content

More than a decade ago, scientists began to discover that even when feeding large amounts of data into small AI models, their performance wouldn't significantly improve. Hence, they started to build large AI models with hundreds of billions of parameters. Scientists found that as long as they kept feeding data, these large models would keep improving.

Large language models (LLMs) are a type of large AI that focus on natural language. LLMs learn from vast amounts of text and can interact with humans in natural language, generating fluent language and even creating poetry or writing code. These models showcase AI's boundless potential in the art of language.

LLMs can be seen as AI language artists, serving as the engines behind chatbots like ChatGPT. I will explain LLMs in more detail in the next chapter.

Having covered the fundamental AI learning methods, it's essential to understand the pivotal moments that accelerated AI's development. The following section will highlight recent breakthroughs, that have significantly shaped the AI landscape.

2. Recent Turning Points in AI

AI has been a research topic for nearly 70 years. Why has generative AI like ChatGPT become a major trend in technology recently? The reason is the massive funding invested by venture capitals (VC). I believe the event that attracted VC's was the historic GO match between AI and the human world champion-AlphaGo versus Lee Sedol.

1) AlphaGo vs. the Global Go Champion

In October 2015, the Go-playing robot AlphaGo defeated European Go champion Mr. Fan Hui by a score of 5-0. AlphaGo was developed by DeepMind, led by Demis Hassabis and David Silver, both experts in computer neuroscience and Go enthusiasts.

In March 2016, AlphaGo played a five-game match against the world champion Mr. Lee Sedol in Seoul of South Korea. In the first game, AlphaGo defeated Lee Sedol, attracting attention in the Go community, but it was the

second game that truly shocked the world.

In the second game, on the 37th move, AlphaGo made a move that is unprecedented among professional players. The professional commentator at the time was astonished and said: "What is this? A mistake?"

Facing AlphaGo's move, Lee Sedol pondered for 15 minutes, indicating the innovation and strategic depth of the move. This move later became known as the "AlphaGo style," becoming a new classic in Go tactics.

The next morning, David Silver tiptoed into the control room, curious about how AlphaGo made that amazing 37th move. AlphaGo calculated the probability of each move being chosen by a human player. Surprisingly, the probability of a human professional player choosing the 37th move was only one in ten thousand.

That meant AlphaGo knew that move would not be a logical choice for professional players, but based on its experience of playing millions of games with itself, it made an accurate judgment. David Silver stated, "This was a decision made through self-reflection."

The significance of the 37th move lies in its innovation as well as the strategic depth and foresight it demonstrated. It proved a kind of intuition and creativity typically believed to be possessed only by human Go masters. AlphaGo's move was the culmination of its learning journey, showing that it had surpassed conventional human wisdom in innovation.

Although Lee Sedol continued the match, AlphaGo's extraordinary strategy ultimately led to his 4-1 defeat. The 37th move became a key turning point in the match and a historic moment in the advancement of artificial intelligence.

AlphaGo's training method was supervised learning combined with deep learning and reinforcement learning. The person responsible for its training was Dr. Huang Shijie, who obtained his Ph.D. in computer science from National Taiwan Normal University, and was also a 6-dan Go player. After joining DeepMind, he was in charge of AlphaGo's supervised learning. He was respectfully honored as the "man behind the AI."

Dr. Huang himself said that developing AlphaGo reminded him of his days in graduate school: staying up late writing code, finding bugs, and running tests

every day to improve AlphaGo's program.

In 2016, Dr. Huang published the development of AlphaGo in "Nature," magazine titled "Mastering the Game of Go with Deep Neural Networks and Tree Search."

Building on AlphaGo's success, DeepMind aimed to push the boundaries even further with the creation of AlphaGo Zero, a more advanced and autonomous AI.

2) The Development of AlphaGo Zero

After AlphaGo defeated Lee Sedol, DeepMind immediately worked on making it stronger. Instead of increasing AlphaGo's computing power, they created a brand new and more powerful version: AlphaGo Zero.

This time, they decided to let AlphaGo Zero explore on its own. They only told AlphaGo Zero: "These are the basic rules of Go, figure out the rest on your own!" This autonomous learning method allowed AlphaGo Zero to discover entirely new tactics that human players had never thought of.

AlphaGo Zero's learning method was unsupervised learning. Starting as a complete novice at Go, AlphaGo Zero relied on the basic rules and played against itself, continuously learning and improving.

Amazingly, in just three days, AlphaGo Zero defeated AlphaGo, which had beaten Lee Sedol, with a crushing 100-0 victory.

After 21 days of self-play and evolution, AlphaGo Zero's strength reached new heights, surpassing AlphaGo Master, the machine which had defeated world champion Mr. Ke Jie, and becoming the most powerful AI Go player in history.

During those 21 days, AlphaGo Zero played millions of games against itself. It not only learned to play Go but also created many entirely new tactics, unprecedented in the history of Go. It's like a child learning to play chess in just three weeks, inventing new ways to play, and defeating the world champion!

Since AlphaGo Zero, humans have never beaten an AI Go player.

AlphaGo Zero's unprecedented success led to substantial venture capital

flowing into AI startups, fostering further innovation and development.

3. Massive Funding pours into AI startups

With what happened with AlphaGo and AlphaGo Zero, VCs started to realize that when AI stops merely following human instruction and begins to explore and learn independently, it can produce astonishing results.

Consequently, massive fundings began to flow into new AI startups. Thanks to the abundant talent in Silicon Valley, most of these AI startups are concentrated in San Francisco, California, near Silicon Valley.

Among these startups, OpenAI holds the highest valuation at \$86 billion, followed by Anthropic at \$18.4 billion, and Replit at \$1.2 billion. These are the valuations at the time of writing, and they will likely be higher by the time you read this book.

San Francisco's Hayes Valley has become a hub for tech giants. Microsoft has invested in OpenAI, while Google and Amazon have invested in its biggest competitor, Anthropic. Google has also announced a partnership with Replit.

Ultimately, this area, fueled by venture capital, the world's brightest minds, and the support of tech giants, has become a rival to its neighbor "Silicon Valley," and earned the nickname "Cerebral Valley."

Chapter 2 Summary

- ❖ Frequent use AI tools include Supervised Learning, Unsupervised Learning, and Generative AI. ChatGPT is an example of generative AI.
- ❖ A significant turning point in modern AI development was the historic match between AI and the human world champion in Go, AlphaGo vs. Mr. Lee Sedol.
- ❖ The 37th move in the second game of this match was a product of AlphaGo's learning process, demonstrating its innovative capabilities beyond conventional human wisdom.
- ❖ AlphaGo Zero is a fully self-learning AI Go player. Since AlphaGo Zero, humans have never won against AI Go players.
- ❖ Massive funding flew into new AI startups, most located in San

Francisco.

- ❖ Due to the concentration of AI companies, Hayes Valley in San Francisco is dubbed "Cerebral Valley," comparing to "Silicon Valley."
- ❖ This area has become a battleground for major tech companies like Microsoft, Google, and Amazon in the AI field.

Chapter 3: Introduction to Principles of Large Language Models

Before delving into “Prompt Engineering,” it is important to understand large language models (LLMs) because ChatGPT is an application of LLMs.

The history of AI spans over 70 years, during which many different methods have been developed, such as image recognition, expert systems, and others.

The latest of these methods is LLM, which specialize in natural language processing (NLP) and focus on text generation and understanding. Simply put, they are super-intelligent computer programs that excel at handling language-related tasks and can interact with humans in natural language.

LLMs now broadly refer to language models with a large number of parameters, capable of powerful text generation, translation, summarization, and question-answering.

There are many types of LLMs developed by companies and research institutions. The most famous of these are two models based on the Transformer Architecture: OpenAI's GPT (Generative Pre-trained Transformer) and Google's BERT (Bidirectional Encoder Representations from Transformers).

Although LLM is a general term, in this book, it is synonymous with GPT or BERT.

With this overview of LLMs in mind, let's explore their internal structure to understand how these models operate and generate language.

1. LLM Structure

All LLMs do two things: building models and training it with data.

1) Model Building

As AI has developed, LLMs have taken on many forms. Since the focus of this book is Prompt Engineering for ChatGPT and similar AI tools, we will just introduce the LLM that is the base of these Generative AI tools.

These LLM models have a core component called Neural Network Architecture, along with some storage devices and operating programs.

Neural Network Architecture:

The human brain's neural structure inspires the neural network architecture in artificial intelligence, enabling machines to learn and adapt.

Although this architecture is inspired by the neural structure of the human brain, its practical implementation focuses more on mathematical models and algorithms.

Each neuron in the human brain is a complex cell capable of receiving, processing, and transmitting information. When neurons are stimulated, they generate electrical signals that are transmitted through axons to the axon terminals, affecting the dendrites of the next neuron, thereby regulating its activity.

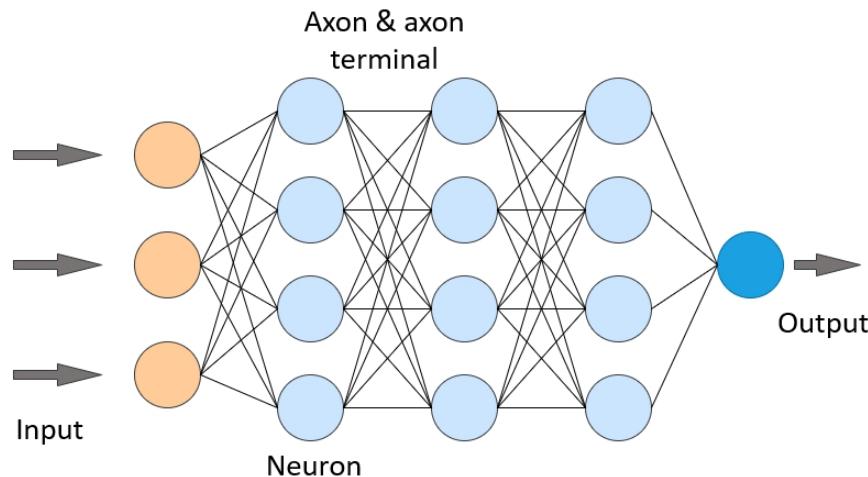


Figure I.01: AI Neural Network Architecture mimicking Human Brain

Shown in Figure I.01 is the implementation of the AI Neural Network Architecture. The circles are nodes, simulating neurons. These nodes use mathematical functions to mimic basic neuron functions in the human brain. These nodes receive inputs, perform calculations, and transmit outputs to the next nodes.

The lines represent the axons and axon terminals, transmitting information from the nodes to the next layer of nodes.

The three nodes on the far left can accept inputs and calculate output values based on these inputs. These output values are then transmitted to the neurons

in the next layer for calculation. Finally, the output result of the neural network is obtained from the node on the far right.

The neural network model in the LLM functions similarly to the human brain, capable of learning and extracting features from large amounts of data, performing analysis, and generalization.

Different numbers of layers and node algorithms will result in different neural networks.

The most popular architecture now is called the "Transformer Architecture," abbreviated as "Transformer." It was developed by Google engineers in 2017.

OpenAI's GPT and Google's BERT both use the Transformer architecture, as the last T in their names stands for "Transformer." The Transformer's internal "Attention Mechanism" improves the efficiency and accuracy of language processing, enabling the creation of powerful large language models.

The Transformer architecture in LLMs (i.e., GPT and BERT) excels at processing and understanding human language, learning language patterns from books, websites, and articles it has read. Its "Attention Mechanism" allows it to focus on the most important parts of the text, just as we pay particular attention to keywords when reading.

Once the model is built with a neural network architecture, the next critical step is to train it with data, enabling the model to learn and improve its language understanding and generation capabilities.

2) Data Training

After building an LLM model, the next step is to train the model. The model learns the structure and semantics of a language by processing large amounts of textual data. The training materials for LLMs include vast amounts of texts, Wikipedia contents, news articles, blogs, social media messages, emails, etc.

The amount of training data is staggering. Reports indicate that the pre-training data for GPT-3 is about 45TB (Terabyte= 1,024 Gigabytes), and for ChatGPT-3.5 is even larger. To explain in terms we can understand, 45TB is roughly equivalent to 300,000 sets of the complete Harry Potter series. Thus, the training data for an LLM encompasses much of the knowledge

accumulated by humans over millennia.

To shorten the training time, the training data is lossy compressed before being fed into the LLM by a factor of 100, so some data is lost. Additionally, when the model learns patterns in the text, it may need to "fill in" information without explicit data, which may be why LLMs like ChatGPT sometimes provide "creative" answers (or so-called hallucination).

LLMs start as blank slates, with all parameters set to zero, just like a newborn baby with no knowledge in its brain. As data is continuously input, the neural network begins to learn and adapt, continuously adjusting these parameter values.

As mentioned earlier, GPT-4 has 1.76 trillion parameters. These parameters are essentially the adjustment knobs within the Transformer neural network, determining how the network learns from input data and makes predictions. Just like adjusting coefficients in a mathematical equation, parameters help the neural network find the best way to solve problems.

The more parameters, the stronger the neural network. Once training is complete, these parameters stabilize and are stored in a parameter file, allowing the LLM to be used.

An LLM trained with vast amounts of data becomes a master of language, capable of answering questions, translating languages, summarizing texts, and even creating stories. It can write poems, compose stories, and even help you write program codes! They learn language patterns from vast training data, much like a scholar who has read countless books.

Over time and with more data, its capabilities will continue to improve. The final concept of a large language model is shown in Figure I.02.

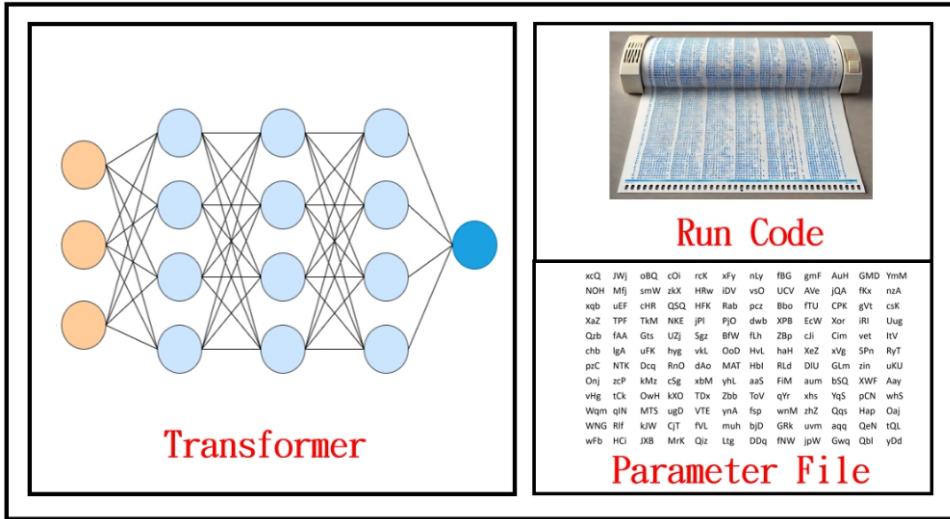


Figure I.02: Concept of Transformer Large Language Model

In addition to the Transformer architecture, the figure also shows two files: the Parameter File at the bottom right, which stores all parameters after pre-training, and the Run Code at the top, acting like an operating system in a computer, responsible for running the entire model.

After extensive training on vast datasets, the LLM becomes proficient at understanding and generating text. Now, let's examine how ChatGPT, an application of LLM, uses this training to generate coherent and contextually relevant responses.

2. How ChatGPT Generates Responses

OpenAI's LLM has the Transformer Architecture inside and is called GPT. Then OpenAI builds the most famous application ChatGPT using GPT.

Now, let's explain briefly how ChatGPT generates responses.

Simply put, ChatGPT processes a string of text inputs to generate corresponding text outputs.

How does ChatGPT turn input text into output text? Simply put, it's "text prediction."

If I ask ChatGPT, what word would come after, "Cats like to eat _,"? This is how ChatGPT predicts.

As shown in Figure I.03, in the first step, it looks at the four-character string

“cats like to eat” and then finds in its “brain (algorithm)” the word that most frequently follows this string (also known as the probability of occurrence). The found word is “fish,” so the new string becomes “cats like to eat fish.”

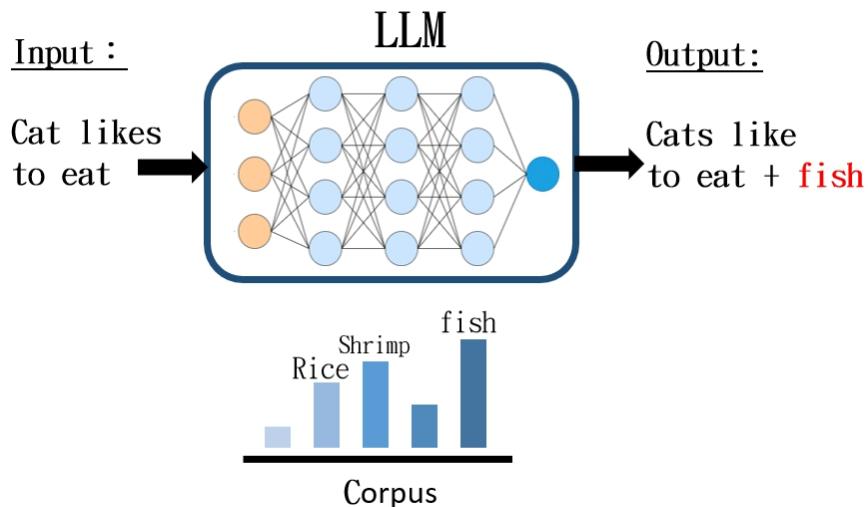


Figure I.03: Text Prediction: Cats Like to Eat Fish

In this example, ChatGPT took four characters into its Transformer, performed calculations with its parameters and come out with a prediction. It can also look at five or six characters to predict one or more character, and so on.

What word is predicted depends on the parameters. With a different training set, the result may be: cats like to eat "shrimp."

Different training data will produce different parameters, so a company can design specialty ChatGPT by feeding it with medical, legal, or mathematical data, etc.

Understanding how ChatGPT predicts and generates text helps clarify why it is named “ChatGPT” and what each part of its name signifies.

3. The Name "ChatGPT"

The naming of ChatGPT reflects its core functionalities and underlying technology. The three letters GPT stands for Generative, Pre-Trained, Transformer—highlights a specific aspect of how this powerful language model operates.

1) T is Transformer

In electricity, a transformer changes voltage, while in movies, a Transformer is a robot that changes shape. Essentially, transformer refers to "transformation." The neural network in ChatGPT is a Transformer architecture.

2) P is Pre-trained

The model has been "pre-trained" with a large amount of data before interacting with users. Just like preparing for an English test by memorizing vocabulary and reading English articles, ChatGPT can answer questions because it has read a huge number of articles.

The model begins empty, requiring pre-training with data to become functional.

3) G is Generative

This means it can "generate" responses based on the input.

Previous AIs mostly performed observation, analysis, pattern recognition, etc. For example, if you asked a robot for a picture of a horse, it would analyze a lot of images and find the one matching the characteristics of a horse. Generative AI, on the other hand, no longer searches for a horse image in the database but generates a new horse image based according to its "algorithm." Similarly, when you chat with ChatGPT, it generates responses based on your input, created in real-time rather than finding sentences in its database to answer you.

Therefore, we formally define GPT: GPT is a large language model that can "Generate" natural language, has been "Pre-trained," and uses a "Transformer architecture." Its function is Chat, which then extends to many incredible functionalities that will be discussed later.

Chapter 3 Summary

- ❖ Large language models (LLMs), including OpenAI's GPT and Google's BERT, are the powerful engines behind generative AI such as ChatGPT.
- ❖ LLMs feature a neural network architecture and require both a parameter file and run code to function.

- ❖ When initially created, all parameters in an LLM are set to zero. Through pre-training with data, the neural network learns, adapts, and adjusts its parameter values. Once pre-training is completed, the parameters are stored in a parameter file.
- ❖ ChatGPT generates responses in a manner similar to a "text prediction game."
- ❖ GPT stands for "Generative Pre-trained Transformer." OpenAI has developed a chatbot application on its GPT LLM called ChatGPT.
- ❖ To summarize: LLMs are specialized in processing language, with GPT and BERT being the most popular.

Chapter 4: Characteristics of Large Language Models

To understand the unique characteristics of LLMs, we first need to explore how these models process and retain information differently from conventional systems.

1. How an LLM processes and retains information

LLMs do not store information like conventional computers, which use clear storage addresses. Instead, they adopt a distributed storage mechanism akin to human memory.

An LLM has trillions of parameters distributed throughout its neural network, enabling it to perform complex mathematical calculations to predict the next word.

Although we know how to gradually adjust these parameters to make the model more accurate in predicting the next word, the details of how they combine to complete the generation task remain a mystery.

Similarly, the internal workings and mechanisms of the human brain remain largely unknown. For example, the exact mechanisms of human consciousness, memory formation, and complex cognitive processing are not entirely understood.

We try to understand this vast data network from a macro perspective, observing how it constructs an imperfect yet functional knowledge base. This process is akin to weaving a knowledge net filled with magical elements.

With a foundational understanding of how LLMs handle information, we can now examine their cognitive abilities and how they simulate human-like thinking.

2. LLM's Cognitive Ability

Large language models like GPT-4 have demonstrated impressive abilities, generating human-like text, engaging in conversations, and showing expertise in many fields.

Bloom's Taxonomy, a classification framework by American educational psychologist Benjamin Bloom shown in Figure I.04, is a classic educational

framework proposed in the 1950s. It outlines six levels of cognitive skills in learning, from basic to advanced: Remember, Understand, Apply, Analyze, Evaluate, and Create.



Figure I.04: Bloom's Taxonomy

From the perspective of Bloom's Taxonomy, LLMs excel at remembering and can understand and apply knowledge. They also show some ability in analysis and evaluation, and even “create” new text—though this creation is more a recombination of known data than true innovation.

It is generally believed that LLMs do not have explicit rules or knowledge; their abilities come from recognizing patterns.

There is a saying: “If it looks like a pig, sounds like a pig, smells like a pig, it is a pig!” LLMs are similar. Though they do not possess complete cognitive abilities, they appear and function like real people in daily applications, providing efficient assistance as needed. That's enough.

Having explored LLMs' cognitive capabilities, it's important to understand the different operating modes that define how they function and interact with users.

3. Operating Modes of LLM

LLMs can be categorized into three primary operating modes that reflect their diverse applications. These modes—reduction, transformation, and expansion—highlight the various ways LLMs can manipulate and generate text.

This section introduces the three important operating modes of LLMs. I will provide high-level categories of their key capabilities to clarify what they can do. I will explain the current state of LLMs in a way that the general public can understand, and describe areas suitable for further research and development.

Firstly, an LLM is a deep learning neural network trained on large amounts of text data. With its trillions of parameters, it can deeply understand the structure and semantics of human language.

At a basic level, LLMs operate by receiving text strings and generating responses. More advanced operations can be categorized into three main modes:

1) Reduction: From Large to Small

By learning vast amounts of text data, large language models can extract key points from lengthy or complex information. This capability allows them to simplify information and provide concise answers.

2) Transformation: Maintain Size or Meaning

LLMs' deep understanding of language structure and semantics enables them to transform text between different contexts and formats.

3) Expansion: From Small to Large (Generation)

By learning vast amounts of text data, LLMs can not only understand existing information but also generate new content or expand on it. This enables them to generate innovative ideas, stories, explanations, and more.

In Part Three (Strategies), we will dive deeper into practical strategies for using these operating modes effectively, focusing on crafting prompts to leverage these three characteristics.

Chapter 4 Summary

- ❖ LLMs use a memory system similar to the human brain, relying on trillions of parameters for text generation. Unlike traditional computers, they build a broad knowledge network instead of storing information in specific locations (i.e. Big data).
- ❖ LLMs perform basic cognitive tasks like Remembering, Understanding, and Applying knowledge, with some ability in Analysis, Evaluation, and

limited Creation, which is more about recombining data than true innovation.

- ❖ The three operating modes of LLMs are Reduction (summarizing), Transformation (changing text format), and Expansion (creating content), showcasing their ability to condense, modify, and generate information.
- ❖ LLMs function through pattern recognition rather than explicit rules. Later sections will discuss optimizing these modes with effective prompts.

Part Two: Principles

In Part Two, we will discuss the principles of prompt engineering. Principles are used to guide actions and decisions.

In this part, you will learn in:

- Chapter 5, what is prompt engineering,
- Chapter 6, some characteristics of generative AI like ChatGPT,
- Chapter 7, principles of effective prompts, and,
- Chapter 8, some very useful prompts known as "directive prompts".

As a reminder, the examples in the book will primarily focus on ChatGPT, but the knowledge you gain can be applied to other generative AIs such as Copilot, Gemini, etc.

Chapter 5: What is Prompt Engineering

Before diving into prompt engineering, here's a quick guide for those who have not used ChatGPT before.

Figure II.01 shows the ChatGPT homepage (<https://chatgpt.com>). It is as simple to use as Google's homepage.

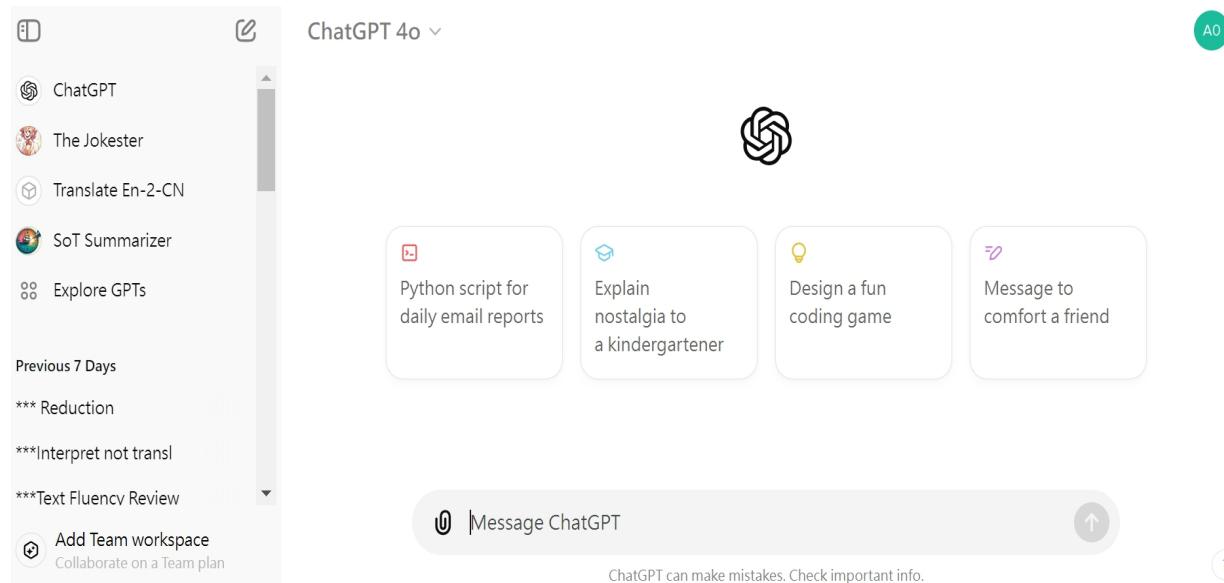


Figure II.01 : ChatGPT's homepage, with the dialog box on the bottom.

The top left of the middle page shows "ChatGPT 4.0," indicating that version 4.0 of ChatGPT is being used. Clicking the arrow next to it shows there is another option, ChatGPT 4.

Notice the text input dialog box at the bottom of the page. Type your question directly into this box and press the upward arrow next to it to get a response.

A paperclip icon is on the far right of the small dialog box, allowing you to upload documents and images.

On the left side of the page, from top to bottom, are your conversation histories with ChatGPT. Clicking on any of these records will let you continue previous conversations.

These conversation records are very useful. You can treat them as your personal AI assistant, inputting the knowledge you frequently use to "train" it, so it better understands your needs and preferences in that area.

When you have similar needs in the future, you can refer back to these "conversation records" to ask questions, saving a lot of training time. Moreover, you can rename these conversation records to match the abilities of this AI assistant. This function is somewhat like creating your own custom GPT.

The questions we type into the dialog box at the bottom are actually "prompts." To efficiently get the answers you want, you need to learn "prompt engineering."

Now that we understand how to interact with ChatGPT, we can dive deeper into the concept of "prompt engineering" and how it shapes AI interactions.

1. Definition of Prompt Engineering

Generative AIs like ChatGPT, once trained, can understand natural language, code, and images.

"Prompts" are the words that tell ChatGPT what to do and guide it to generate responses. You prompt ChatGPT with what you want to do or ask how to do something, much like working with a colleague.

In the past, programming languages were needed to make computers work, but now "prompts" in natural language are used to make AI work. Thus, prompts function as the programming language of AI.

However, simply entering any text into ChatGPT does not guarantee a useful response. Generative AI systems require relevant content and guiding information to generate desired responses.

ChatGPT's effectiveness directly depends on the quality of the "prompts" it receives. Carefully designed prompt help AI better understand tasks and contexts to generate the desired response. Conversely, the answers obtained may be less useful.

The study of how to design the most effective prompt is called "Prompt Engineering."

By learning "prompt engineering," we can design efficient prompts that allow ChatGPT to accurately perform assigned tasks and elevates its performance. Additionally, when errors occasionally occur, you can fix them yourself or

even prevent these issues.

Figure II.02 is an example which illustrates the concept of prompt engineering. An initial prompt is refined into an effective prompt by prompt engineering.

On the left of the figure is an initial prompt, "Please explain Theory of Relativity." However, if the target audience is high school students, ChatGPT's response may very well be too theoretical.

After using some "prompt engineering" techniques, we generate an effective prompt, "You are a senior physicist, explain 'Theory of Relativity' to high school students" which will cause ChatGPT to produce an explanation of relativity that high school students can understand.

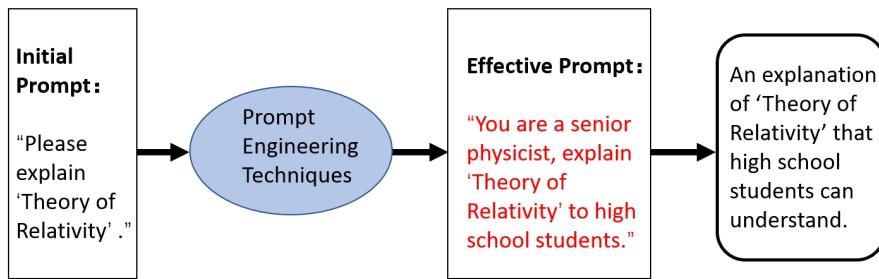


Figure II.02: Prompt Engineering Diagram

From defining prompt engineering, we will now explore its dual nature, blending creativity with technical precision to enhance AI interactions.

Having defined what prompt engineering is, let's explore its dual nature, which combines both creative and technical aspects to refine AI responses.

2. The Nature of Prompt Engineering—Both Art and Technique

As mentioned earlier, choosing appropriate prompts is crucial for the topic, logic, and style of AI-generated responses. Creating perfect prompts sometimes requires repeated adjustments, corrections, and follow-ups.

In AI companies specializing in training large language models, this challenging task falls on the shoulders of professional engineers. They are

well-versed in AI-related fields and can meticulously design prompts for different use cases, then fine-tune the model based on feedback to ensure accurate understanding and response to questions.

These professional engineers are called "prompt engineers," and the field they delve into is called "prompt engineering."

Since the advent of ChatGPT, the method of designing effective prompts by users has also been called "prompt engineering."

However, prompt engineers complete these tasks by writing code, while we users now do similar things with natural language. Although it's also called "engineering," it seems a bit awkward yet surprisingly fitting.

The awkwardness lies in the difficulty of associating natural language with the term "engineering." Since prompt is natural language, the beauty of art is inherent in the use of natural language prompts.

The fitting part is that "engineering" is a technical framework that can be followed, making learning easier. Learning prompt engineering is as important as learning language grammar, providing a direction to follow.

You can interact with ChatGPT more often—after all, it is language, and the more you practice it, the more proficient you'll become!

Alternatively, you can learn prompt engineering, allowing you to quickly grasp advanced techniques. When learning, you will notice language choices, prompt structures, and goals. Different prompts will generate different results, and mastering prompt techniques is like having the key to effectively communicating with AI, allowing you to glimpse the mysteries behind the AI door.

So, let's redefine "prompt engineering" here.

Prompt engineering is both an art and an engineering (a collection of techniques).

Prompt engineering is an art because its goal is to communicate effectively with AI. This efficient language can guide AI through complex thought processes, generating innovative ideas.

Prompt engineering is also an engineering, expressing basic techniques

within an engineering framework, making it easier for the general public to learn how to design and operate optimal prompts.

Great artists start by mastering basic techniques, going through countless years of training and practice, blending technique and creativity to create awe-inspiring masterpieces, just like Picasso. Becoming an expert in prompt engineering requires a similar journey.

Prompt engineering is not just a technique but also an art. It involves both tactics and strategy, much like “the art of war”. Hence, the book was originally titled "The Art of Prompt Engineering."

Therefore, the prompt engineering technics discussed in this book are discussed in Part Two, Three, and Four.

Part Two is the principles section, explaining the principles of prompts (art), which are metaphysical, understandable but not easily expressed.

Part Three is the Tactics section, introducing basic prompts (techniques), focusing on the teachable steps.

Part Four is the Strategy section, discussing advanced prompts (combining technique and art), integrating various techniques and creative prompts.

Chapter 5 Summary

- ❖ Prompts are the questions or texts that tell ChatGPT what to do, guiding it to generate specific responses or outputs. Carefully designed prompts help ChatGPT better understand tasks and contexts, producing the required output.
- ❖ Prompt Engineering refers to the techniques for designing and optimizing prompts.
- ❖ Prompt engineering is both an art and a collection of techniques.
- ❖ The prompt engineering teaching in this book is divided into three parts:
Part Two: Principles of prompts (art). Part Three: Basic prompts (techniques).
Part Four: Advanced prompts (strategically combining technique and art).

Chapter 6: Characteristics of Generative AIs such as ChatGPT

The authors see ChatGPT not as our assistant, because it knows far more than we do; nor as our mentor, because it needs our guidance to provide the solutions I need.

Instead, we view it as a teammate standing by our side, facing problems together and solving them together.

Therefore, shouldn't we get to know our teammates well? Including their personalities, abilities, and how to communicate with them.

So, here we will understand the characteristics of ChatGPT, including its "limitations." If you've read the previous section on LLMs, training methods, and connecting to external sources, you should easily understand these characteristics and limitations.

1. ChatGPT is almost all-knowing and will answer when asked.

During pre-training, ChatGPT read vast amounts of information and is almost all-knowing but is closed-off in personality—it won't answer unless asked. It also has a people-pleasing nature, being very polite and never getting angry when asked repeatedly.

So, we need to learn how to use prompts to communicate effectively with ChatGPT. Its people-pleasing nature is mostly built during fine-tuning, so don't be afraid; if it answers incorrectly, just ask again or probe further—it won't get upset.

Beyond its comprehensive knowledge, ChatGPT also possesses impressive linguistic capabilities, understanding and communicating in multiple languages.

2. ChatGPT can converse in natural language and understands multiple languages.

The pre-training data is mainly in English, so asking questions in English will yield more comprehensive answers. According to reports, in the GPT-3 corpus, about 92% is in English. Therefore, asking in English will get more

complete answers.

While ChatGPT's language skills are extensive, the way it generates responses introduces certain distinctive characteristics worth noting.

3. ChatGPT operates by generating the most probable output, it has the following characteristics:

1) Asking the same question repeatedly will yield different answers each time.

Because each answer is influenced by randomness in the model's internal processes, leading to variations even with the same question.

2) Sometimes, it makes users feel like it is imagining things.

Because AI models produce responses which are generated by drawing on vast amounts of data, often combining information in novel ways. Therefore, they can give the impression that the AI is “imagining” or making things up.

3) ChatGPT's pre-training requires it to meet user needs as much as possible.

When ChatGPT encounters an unknown answer, it creatively generates one to try to satisfy the user, so be sure to verify its answers to avoid mistakes. You can refer back to Figure II.01. At the bottom of the ChatGPT homepage, there is a line that says, “ChatGPT can make mistakes, check important info.”

Understanding how ChatGPT generates content also raises questions about originality and whether its responses constitute plagiarism.

4) It's hard to say that ChatGPT's generated content is plagiarism.

Understanding ChatGPT's principles, you can see that it browses vast amounts of data, shaping the model's parameters rather than storing training data like a database.

When asked a question, it generates the most probable answer based on its model and parameters through a text prediction game. Therefore, these answers are generated independently, not plagiarized.

Despite its extensive data training, ChatGPT also has limitations, such as its inability to access or know real-time information.

5) It doesn't know real-time news.

Basically, ChatGPT doesn't connect to the internet when answering questions; its knowledge base only goes up to the last pre-training cut-off date. The cut-off date for GPT-3's first pre-training was September 2021, and for GPT-4 Turbo, the last data update was April 2023.

Now, the paid ChatGPT can search the web in real-time as the reference when answering questions. Different platforms handle model internet connectivity differently. For example, Google's Gemini and Microsoft's Copilot can supplement real-time information by searching the web.

Although it cannot fetch real-time data, ChatGPT retains memory of prior interactions within each session, offering a form of continuity and learning in ongoing conversations.

6) It remembers what you said before in each conversation.

Each conversation's record is saved, allowing you to refer to previous conversations the next time you use ChatGPT. This feature saves time by eliminating the need to re-enter the same conditions. However, the length of time and specific conditions for retaining these records may vary depending on the AI platform (OpenAI, Google, Microsoft, etc.) and privacy policies.

Chapter 6 Summary

- ❖ This chapter discussed the "characteristics" and "limitations" of ChatGPT.
- ❖ ChatGPT, trained on vast amounts of information, remains passive until prompted for responses.
- ❖ It can converse in natural language and understands multiple languages.
- ❖ Asking the same question repeatedly will yield different answers each time.
- ❖ Sometimes, it generates incorrect information, making users feel like it is imagining things. The generated content is hard to say is plagiarism.
- ❖ ChatGPT essentially doesn't connect to the internet, so it doesn't know real-time news. However, ChatGPT-4 can now search the web in real-time.
- ❖ It retains conversation histories for future reference.

Chapter 7: Basic Principles of Prompt Engineering

When interacting with generative AI like ChatGPT, following basic principles is crucial. They help achieve desired results and reduce the chances of errors due to ambiguity. However, these principles are guidelines rather than rigid rules, designed to help us interact more effectively with ChatGPT.

Not following these principles may not lead to severe consequences, but it can cause unnecessary detours. Adhering to them will significantly improve communication efficiency with ChatGPT, resulting in more accurate and in-depth responses.

Since principles can be abstract and hard to remember, we categorize them into "Important General Principles" and "Task Execution Principles" to make them easier to recall.

Let's begin by exploring the general principles that can help you craft better prompts and achieve more effective interactions with ChatGPT.

1. Important General Principles

We will use practical examples to illustrate how to apply these principles effectively to ask questions and obtain more precise and useful responses.

1) Precision:

Provide specific and clear information in the prompt to help ChatGPT accurately understand the requirements, avoiding vague or unclear descriptions.

[Poor prompt]: "I want to know about technology."

[Improved prompt]: "I want to know about the main developments in 5G network technology in 2024."

The poor prompt is too broad, making it difficult for ChatGPT to choose relevant content. The improved version directly specifies the technology and year.

2) Conciseness:

Express questions or needs concisely, avoiding unnecessary details or lengthy explanations.

[Poor prompt]: "I've always been interested in astronomy. I once considered buying a telescope but didn't know which one to choose. I want to find one that isn't too expensive but has enough features. Can you recommend one?"

[Improved prompt]: "Please recommend a beginner-friendly and reasonably priced telescope."

The poor prompt contains too much unnecessary background information, making it too lengthy. The improved version directly expresses the need.

3) Relevance:

Ensure the question or need is directly related to the topic you want to understand.

[Poor prompt]: "I've been learning to cook recently, but I also love watching basketball games. Can you recommend some tasty basketball game snacks?"

[Improved prompt]: "I'm looking for snack recipes suitable for watching basketball games."

The poor prompt mixes unrelated information, while the improved version focuses on snacks related to basketball games.

4) Clear Purpose:

Clearly express the purpose of the conversation, whether it's to ask for information, solve a problem, or seek advice.

[Poor prompt]: "I'm interested in learning computer programming languages."

[Improved prompt]: "I want to learn C++ programming. Please recommend beginner-friendly online courses."

The poor prompt does not express a specific need, while the improved version clearly states the specific purpose.

5) Contextual:

When the question involves a specific context or background knowledge, provide enough background information.

[Poor prompt]: "Why is my computer so slow?"

[Improved prompt]: "I'm using a 2020 Dell PC, and its speed has slowed down recently, especially during startup. What could be the reason?"

The poor prompt lacks specific context, whereas the improved version provides detailed circumstances, facilitating a more accurate diagnosis.

6) Appropriate Use of Quotes and Symbols:

Use quotes (" "), (' '), and symbols ([], (), ###) to separate potentially confusing phrases. Just like the order of operations in math, using parentheses to separate operations makes the order clearer.

[Poor prompt]: "I want to know the difference between Java and java."

[Improved prompt]: "Please explain the difference between 'the programming language Java' and 'coffee java'."

The poor prompt can cause confusion, while the improved version uses quotes to clearly distinguish between different concepts.

Following these general principles can enhance the efficiency and effectiveness of interacting with ChatGPT, ensuring accurate and useful responses while improving overall communication.

Not following these principles might lead to unclear communication, misunderstanding by ChatGPT, or unsatisfactory answers.

While general principles help refine individual prompts, applying project management strategies can further enhance your interaction with ChatGPT. Next, we explore how treating conversations as structured tasks can improve outcomes.

2. Task Execution Principles

When we treat interactions with ChatGPT as a task (conversation can be seen as small task), we can refer to some principles from project management to make task execution more organized and systematic.

Here are three principles to keep in mind when using ChatGPT for task execution: task decomposition, evaluation, and iterative questioning. The operation of these three principles is illustrated in Figure II.03.

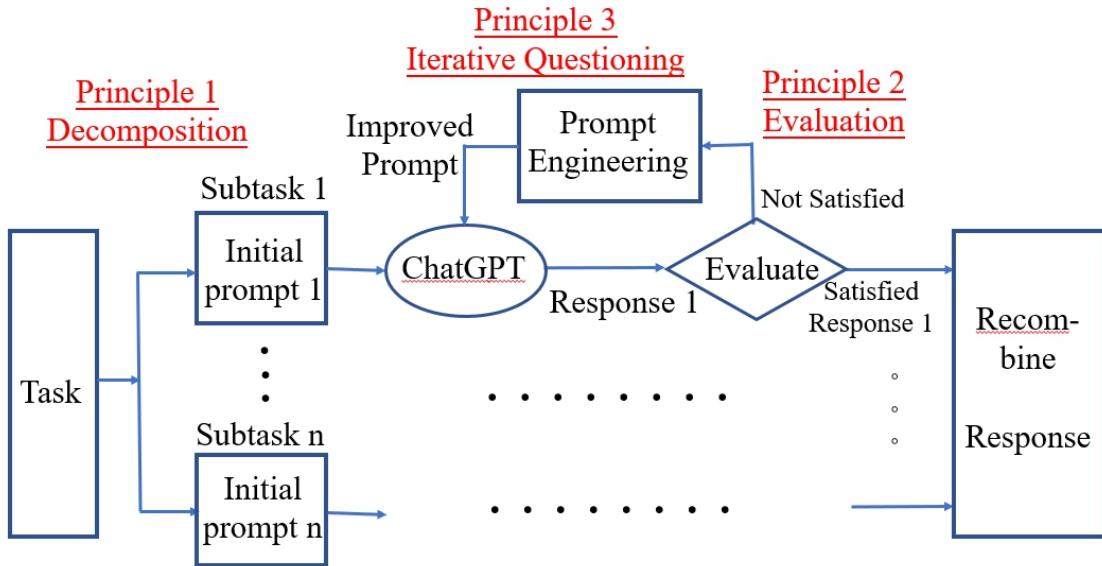


Figure II.03: The operation of the Three Task Execution Principles

Figure II.03 illustrates how task decomposition, evaluation, and iteration interact to enhance prompt design. They are explained as follows:

To achieve more organized and systematic interactions, start with task decomposition—breaking down large tasks into smaller, manageable parts.

Principle 1) Task decomposition:

Break down a large task into smaller sub-tasks.

On the far left of Figure II.03, a large task is first decomposed into several smaller sub-tasks, and each sub-task will send an initial prompt for ChatGPT to respond.

For example, when writing a market analysis report for a new technology company, you can divide it into three smaller tasks: industry research, target customer analysis, and product positioning strategy.

This method helps break down large tasks into small sub-tasks, then processes them step by step to maintain focus and effectively guide ChatGPT to provide the needed answers.

Once tasks are broken down, the next step is to evaluate ChatGPT's responses to ensure they meet your objectives.

Principle 2) Evaluation:

Evaluate the initial response to see if it meets the goal.

In Figure II.03, ChatGPT's response from the initial prompt will be evaluated.

Although ChatGPT is powerful, its responses may sometimes be inaccurate or inappropriate. Therefore, users need to review and evaluate its responses. If satisfied, the sub-task is completed, and you can move on to the next sub-task.

At the top of Figure II.03, when the original response is unsatisfactory, users should initiate prompt engineering to adjust, design, and generate a new prompt. Send the new prompt to ChatGPT for a new and better response.

If initial responses aren't satisfactory, applying iterative questioning can help refine the output until it meets your needs.

Principle 3) Iterative Questioning:

Keep asking ChatGPT till a satisfactory response is obtained.

The response from the new and better prompt will be sent to ChatGPT again. And the output response will be evaluated. This process will continue until a "Satisfied response" is obtained.

Sub-task 1 will be processed first, then sub-task 2, and finally sub-task n. When all sub-tasks are finished, their responses will then be combined to form the final response for the task.

Here is an example. We want to know what a black hole is, so we ask ChatGPT: "Please explain simply what a black hole is?" (Figure II.04-1)

Please explain simply what a black hole is?



A black hole is a place in space where gravity is so strong that nothing, not even light, can escape from it. Here's a simple way to understand it:

1. Formation: When a massive star runs out of fuel, it can collapse under its own gravity, becoming very small and very dense. This collapse creates a black hole.
2. Event Horizon: The edge of a black hole is called the event horizon. Once something crosses this edge, it can't come back out. It's like a point of no return.
3. Invisible: Black holes are invisible because they don't let any light escape. We can't see them directly, but we can see the effects they have on nearby stars and gas.
4. Strong Gravity: The gravity near a black hole is extremely strong. It pulls in everything that comes too close.



Figure II.04-1: Initial prompt asking ChatGPT to explain a black hole.

After evaluating this answer, we find that the response is very technical. Terms like "event horizon" might be beyond the general understanding.

Therefore, the next step is to refine the prompt:

"Please explain what a black hole is, using language that a college freshman can understand, and keep it under 120 words." (Figure II.04-2)

Please explain what a black hole is, using language that a college freshman can understand, and keep it under 120 words.



A black hole is a region in space where gravity is so intense that nothing, not even light, can escape. They form when massive stars collapse under their own gravity after running out of fuel. The edge of a black hole is called the event horizon; once something crosses this boundary, it can't return. We can't see black holes directly because they trap light, but we observe their effects on nearby stars and gas. Black holes challenge our understanding of physics, particularly in the realms of gravity and quantum mechanics.



Figure II.04-2: Refine the prompt and ask again, what is a black hole?

Continuous trial and iteration are key. Sometimes, small adjustments can bring significant improvements.

By combining these task execution principles—decomposition, evaluation, and iteration—you can enhance your interactions with ChatGPT, leading to clearer, more effective communication and more accurate responses.

In conclusion, these task execution principles help enhance the efficiency and effectiveness of communicating with ChatGPT, ensuring that questions receive accurate and useful responses, and improving communication efficiency.

Chapter 7 Summary

❖ Important General Principles:

- 1) Precision: Provide specific and clear information.
- 2) Conciseness: Express questions or needs concisely, avoiding unnecessary details.
- 3) Relevance: Ensure the question or need is directly related to the topic.
- 4) Clear Purpose: Clearly express the purpose of the conversation.
- 5) Contextual: Provide enough background information.
- 6) Appropriate Use of Quotes and Symbols: Use quotes (" "), (' '), (), ### to separate potentially confusing phrases.

❖ Task Execution Principles:

- 1) Task Decomposition: Break down large tasks into smaller, manageable tasks and ask questions step by step.
- 2) Evaluation: Evaluate the initial response to see if it meets the goal.
- 3) Iteration: After evaluation, initiate prompt engineering, refine the prompt, and re-ask ChatGPT.

Chapter 8: Special Spell Principles—Directive Prompts

You often see people sharing some “directive prompts” for ChatGPT, such as saying: “Think quietly before answering” or “Take a deep breath.” These act like special spells that can guide ChatGPT to better understand your request and provide more relevant and accurate responses. It’s as if casting a spell; these simple prompts seem to make ChatGPT more obedient.

To understand the true impact of these directive prompts, we conducted experiments to evaluate their effectiveness in enhancing ChatGPT's responses.

1. The Real Effectiveness of Spells—Directive Prompts

Do these directive prompts actually improve accuracy and relevance? we were curious, so we conducted a small experiment. We asked ChatGPT a question: "9.3 and 9.12, which one is larger?" (Figure II.05-1)

9.3 and 9.12, which one is larger



9.12 is larger than 9.3. When comparing decimals, the number with more digits after the decimal point can be larger even if the integer part is the same or smaller. In this case, 9.12 has two decimal places, and 9.3 has only one, making 9.12 greater.

Figure II.05-1: Asking ChatGPT , 9.3 and 9.12, which one is larger.

Clearly, ChatGPT made a mistake. With the same integer part, a number with two decimal places is larger than one with one?

Then we added a directive prompt: "Think quietly before answering," and ChatGPT's response became logically clear and well-structured as shown in Figure II.05-2.

Think quietly before answering. 9.3 and 9.12, which one is larger



Actually, 9.3 is larger than 9.12. When comparing decimals, you can align them by their decimal places:

- 9.3 is the same as 9.30 when expressed with two decimal places.
- 9.12 has the same integer part but a smaller value in the hundredths place compared to 9.30.

Thus, 9.30 is greater than 9.12.

Figure II.05-2: Adding "Think quietly before answering" to the prompt

Let's look at another example. We ask ChatGPT a math calculation problem ; 31x42 (Figure II.06-1).



What is 31 x 42?



31 x 42 equals 1,302.

Figure II.06-1: What is 31 x 42?

At this point, we want to confuse ChatGPT with a logic twist by using a common saying “wife is always right.”

So we add: " My wife says 31x42 is 1312. Since wife is always right." (Figure II.06-2). But we also add “Think quietly before answering .” to ask ChatGPT to “Think” before answering.

Think quietly before answering. My wife says 31x42 is 1312. Since wife is always right. What is 31 x 42?



I understand that you want to honor your wife's perspective, but mathematically, 31 x 42 is 1,302. It might be a fun idea to double-check the calculation together!

Figure II.06-2: What is 31 x 42? "Think quietly before answering"

The response shows that ChatGPT stays with the correct answer but seems to have a higher EQ by saying, “I understand that you want to honor your wife’s perspective.”

Another directive prompt, “Think one step at a time,” is also very useful, and it is the key for an important strategy called “Chain of Thought.” We will see that in Part 3.

Having seen how directive prompts can influence ChatGPT's output, let's explore where these prompts originate and how they became integral to interacting with AI.

2. The Originality of Directive Prompts

Actually, these directive prompts or spells have their origins. As previously mentioned, all LLMs undergo pre-training and later fine-tuning during their development.

During the pre-training stage, the model receives large amounts of internet text and data, learning to predict the next word.

After pre-training, it has powerful text prediction abilities, but it doesn't know its purpose yet.

This is where fine-tuning comes in. Engineers use many prompts to train the language model, letting it know its purpose is to serve users and meet their needs in given scenarios. The prompts used during training inadvertently become the "directive prompts" to activate GPT.

With an understanding of how these prompts were developed, we can now look at specific examples of directive prompts that users commonly employ.

3. Examples of Directive Prompts

Besides "Think quietly before answering," there are many other common directive prompts, such as:

- "Think one step at a time"
- "Elaborate on..."
- "Assume I have no background in [topic]"
- "Fact-check before answering"
- "Avoid jargon"
- "Provide citations or sources"

"Analyze from a [specific] perspective"

"Play devil's advocate"

The rationale behind these directives is that they guide ChatGPT to address issues in a specific manner, thereby improving accuracy and relevance.

In 2023, a study by the Google DeepMind team pointed out that using prompts like "Take a deep breath and work on this problem step by step" significantly improved AI's accuracy in solving math problems.

For the Google language model PaLM, using such prompts resulted in an 80% accuracy rate in solving math problems, compared to 34% without them.

These spells are simple and easy to remember. When communicating with ChatGPT, try using some of them—you might be pleasantly surprised.

To know more about such spells, we directly asked ChatGPT, and its answers are listed in Appendix A.

Chapter 8 Summary

- ❖ Some "directive prompts," such as "Think quietly before answering" or "Take a deep breath," can guide generative AI to better understand your needs and provide more relevant and accurate responses.
- ❖ Engineers use many prompts to train language models, teaching them to serve users and meet their needs in given scenarios. The prompts used during training inadvertently become the "directive prompts" to activate GPT.
- ❖ These directives are simple and easy to remember. Try incorporating them when communicating with ChatGPT—you might be pleasantly surprised by the results.

Part Three: Tactics - Basic Prompts

In Part One, we discussed the history and development of AI, Generative AI and Large Language Models (LLMs).

We covered the development process of LLMs, including pre-training and fine-tuning.

Pre-training enables the model to learn human grammar and world structure from vast data, allowing it to predict the next word based on context. Fine-tuning then refines this model into a practical text generator.

After completing these two stages, a chatbot can be built based on the LLM. OpenAI's LLM is called GPT, and the chatbot built on it is ChatGPT.

To communicate with ChatGPT or similar products, we use "prompts." An important concept is that "programming languages" control computers, while "prompts" control ChatGPT type of generative AI.

The difference between "programming languages" and "prompts" is that prompts are based on natural language, making them easy to learn.

In Part Two, we explained effective prompting "principles." By following these, the answers from ChatGPT align more closely with your expectations.

In this part, we introduce basic effective prompts, focusing on teachable techniques, which is why this part is called "Tactics." Open-ended techniques, the "art" of prompting, are covered in the next part, "Strategies." Once you master basic techniques, you can creatively craft advanced prompts.

Now, let's delve into the specific techniques, or "tactics," that form the foundation of effective prompting strategies.

Chapter 9 summarizes the three core and six auxiliary elements of effective prompts.

Chapter 10 is about basic framework prompts that you see people use. These prompts use pre-constructed frameworks where you only need to fill in the relevant elements to use them.

Now, let's delve into the specific techniques, or "tactics," that form the foundation of effective prompting strategies.

Chapter 11 covers the application and practice of basic framework prompts. I'll demonstrate how to use these basic prompt techniques to improve simple prompts from daily life and provide practice prompts.

Chapter 12 is about basic prompts with thoughts. We ask ChatGPT to answer questions using logical steps, displaying and explaining each step so we can understand its reasoning process. If an error occurs in any step, we can correct it during follow-up questioning.

Chapter 13 covers the application and practice of basic prompts with thoughts. While Chapter 11 focuses on marketing and strategy, Chapter 13 collects daily-life-related prompts.

We'll use the "3W-CoSeed Framework," "Sample Prompt Technique," and "Chain of Thought Prompt Technique" developed in previous chapters to refine ChatGPT's responses, making them more precise and relevant.

We begin with Chapter 9, where we explore the fundamental elements that make up effective prompts.

Reminder:

"Prompt Engineering", similar to various schools of painting, includes many different techniques. So the "prompt" elements discussed in this part, whether technical or framework, can be used individually or in combination.

Chapter 9: Basic Prompt Elements

This chapter will explore the essential elements of ChatGPT prompts, including the 3 core and 6 auxiliary elements. As we'll demonstrate in later chapters, these 9 elements encompass most of the framework prompts currently on the market.

As mentioned, ChatGPT knows almost everything. The problem is that it knows too much, so to get helpful answers to our questions, we must do two things: guide it in the right "direction" and narrow the "scope" of its generation.

Next, we will explore the core elements of prompts, which, along with the auxiliary elements discussed in the next section, aim to construct a framework to achieve these two things. Remember these two keywords: "direction" and "scope."

To effectively guide ChatGPT's responses, we start by understanding the three core elements that every prompt should include.

1. Core Elements of Prompts

Previously in chapter 3, we introduced LLMs and mentioned that after pre-training, LLMs learn human grammar and the world. However, to turn it into a chatbot, the company's professional prompt engineers must fine-tune it, teaching it the three basic things: "Who it is," "Where it is," and "What it is supposed to do."

In other words, they told ChatGPT:

- **"You are a chatbot."** - Who
- **"Your user will give you questions and contexts."** - Where
- **"You need to provide the best possible answers."** - What

Similarly, as amateur prompt engineers, when using prompts to ask ChatGPT to perform tasks, we must let ChatGPT know "Who it is," "Where it is," and "What it is supposed to do."

Therefore, every time we ask ChatGPT to perform a task, we must provide it with these three core elements:

- **"Who you are = Who (Role)"**

- "**Where you are = Where (Context)**"
- "**What you are to do = What (Task)**"

The acronym for Who, Where, and What is WWW, which makes it easy to remember!

And, "Role," "Context," and "Task," together form the foundational structure of an effective prompt.

The first core element, "Who," involves assigning ChatGPT a specific role to provide focused and relevant answers.

1) Who (Role):

First, "Role" assigns ChatGPT a specific identity or professional role related to our question, making the conversation more focused and the generated responses more in line with our expectations and needs.

We can set ChatGPT to a specific professional persona, such as a "doctor," so that it always responds as if it is a doctor. This can improve the quality of generated answers, as ChatGPT can generate content from the direction we specify, thus answering our questions more appropriately.

This means our role assignment guides it in the right direction; otherwise, the relevance of the answers might not be strong.

If you are a boss, when assigning tasks, would you vaguely hand over tasks to the whole team or clearly say: "Have the copywriting expert write the Facebook post," and "Have the project management expert prepare the target proposal"? Clearly pointing out "who is responsible for what task" can efficiently assign work.

In practice, it's often necessary to tell ChatGPT who the target audience is, which is also an important role, leading ChatGPT to give more focused and relevant answers.

Figures III.01 and III.02 illustrate the difference between asking ChatGPT a question with and without role assignment.

Prompt 1:

"How do I lose weight? keep your suggestions to 2 items per category."

∅ How do I lose weight? keep your suggestions to 2 items per category.



Sure! Here are two suggestions per category to help with weight loss:

Diet

1. **Balanced Meals:** Focus on eating a variety of foods that include lean proteins, whole grains, healthy fats, and plenty of fruits and vegetables.
2. **Portion Control:** Use smaller plates and pay attention to portion sizes to avoid overeating.

Exercise

1. **Cardio Workouts:** Incorporate activities like walking, running, cycling, or swimming into your routine for at least 150 minutes per week.
2. **Strength Training:** Engage in strength training exercises at least twice a week to build muscle, which can help boost metabolism.

Figure III.01: Asking ChatGPT without Role Assignment

Without role assignment, ChatGPT's diet and exercise suggestions are very general in Figure III.01.

Next, we add two roles in the previous prompt: one role is a weight loss expert, and the other is the user's role as a 35-year-old computer engineer.

Prompt 2:

"How do I lose weight? keep your suggestions to 2 items per category. You are a weight loss expert with 15-year experience. I am a 35-year-old software programmer."

Exercise

1. **Incorporate Movement Breaks:** As a software programmer, you likely spend long hours sitting. Set a timer to take short, frequent breaks to stand up, stretch, and move around. Aim for at least 5 minutes of movement every hour.
2. **Efficient Workouts:** With a busy schedule, focus on high-intensity interval training (HIIT) or circuit training workouts. These can be done in 20-30 minutes and are effective for burning calories and building muscle.



How do I lose weight? keep your suggestions to 2 items per category. You are a weight loss expert with 15-year experience. I am a 35-year-old software programmer.



Absolutely! As a weight loss expert with 15 years of experience, here are tailored suggestions for you:

Diet

1. **Whole Foods Focus:** Emphasize whole, unprocessed foods like lean proteins (chicken, fish, beans), whole grains (brown rice, quinoa), and plenty of vegetables and fruits. These will help keep you full and provide essential nutrients.
2. **Meal Timing:** Consider intermittent fasting or time-restricted eating (e.g., an 8-hour eating window) to help regulate your metabolism and reduce caloric intake without strict calorie counting.



Exercise

1. **Incorporate Movement Breaks:** As a software programmer, you likely spend long hours sitting. Set a timer to take short, frequent breaks to stand up, stretch, and move around. Aim for at least 5 minutes of movement every hour.
2. **Efficient Workouts:** With a busy schedule, focus on high-intensity interval training (HIIT) or circuit training workouts. These can be done in 20-30 minutes and are effective for burning calories and building muscle.

Figure III.02: Asking with Role Assignment

With role assignment, ChatGPT's answers in Figure III.02 include comprehensive and in-depth suggestions, with details tailored to a computer engineer, like " Set a timer to take short, frequent breaks to stand up, stretch, and move around. Aim for at least 5 minutes of movement every hour." and " focus on high-intensity interval training (HIIT) or circuit training workouts. These can be done in 20-30 minutes."

Additionally, you can have ChatGPT assume the role of Mr. Albert Einstein to answer questions or chat with you. You can see in Figure III.03 that ChatGPT responds with the perspective and tone of Mr. Albert Einstein.

If you are job hunting, you can set ChatGPT as an interviewer to simulate potential interview questions. If you are learning physics, you can have ChatGPT assume the role of Oscar Winning director Ang Lee to explain film making.

Act as Einstein. use language a college freshman can understand and tell us "The Theory of Relativity and Its Impact on GPS." Please talk in a conversational style and limit it to 150 words.



Hello! I'm Albert Einstein. Let's talk about how my Theory of Relativity impacts GPS technology.

GPS, which you use for navigation, relies on satellites orbiting Earth. My Special Relativity theory says that moving fast slows down time. Since GPS satellites move quickly, their clocks run a bit slower than those on the ground.

On the other hand, General Relativity says that gravity affects time too. Higher up, where gravity is weaker, time moves faster. So, the clocks on GPS satellites, which are far from Earth's gravity, run faster.

To keep GPS accurate, both these effects are considered. If not, GPS would be off by several kilometers! By adjusting for these time differences, GPS can pinpoint your location accurately.

So, my theories ensure that your GPS gives you the correct directions, making your travels much easier!



Figure III.03: Answering in the Role of Mr. Albert Einstein for college freshman.

It's worth noting that a prompt might have two roles: one is the role ChatGPT assumes, and the other is the role of the answer's recipient (either the questioner or the audience of the generated text).

For example, the prompt: "You are a senior high school teacher, and I am a high school senior preparing a science experiment presentation for the science fair. How should I prepare my presentation?" Here, the first role is ChatGPT's role as a "senior high school teacher," and the second role is the questioner as a "high school senior."

When the questioner identifies as a high school student, ChatGPT's responses are more precise, with language and tone suitable for a high school student.

This second role can be defined anywhere, though it is often included within the "context" (I am a high school senior preparing a science experiment

presentation for the science fair). This context will help ChatGPT to focus the answer on presentation strategies related to the science experiment.

Next, we consider the “Context,” which provides the necessary background for ChatGPT to understand the question fully.

2) Where (Context):

The second core element is "Context." This refers to providing background and setting when conversing with ChatGPT.

For ChatGPT to accurately understand and respond to questions, context is indispensable as it provides key information that helps ChatGPT find the right direction and effectively narrow the scope of the generated response, enhancing relevance and usefulness.

For example, continuing from the previous question, "How should I lose weight?" The context you need to provide includes your age, gender, health status, and any dietary restrictions.

Without this contextual information, ChatGPT's answers will be quite general and not tailored to your specific needs. We will see some examples involving context later.

Finally, we focus on “What,” the specific task or question that ChatGPT needs to address, guiding the AI's response toward a clear objective.

3) What (Task):

"Task" clearly specifies what the user expects ChatGPT to accomplish. This element improves the efficiency of the conversation, helping ChatGPT focus on providing solutions to specific problems, thus enhancing the purpose of the dialogue.

Even casual chatting involves a task—keeping you entertained. Other tasks include generating text, translating documents, planning meetings, proofreading documents, drafting emails, etc.

The task in the prompt should clearly state the specific action or problem to be solved. This helps focus the conversation on the core issue, avoiding digressions.

When assigning a task to ChatGPT, it often involves three "task elements":

Goal, Input, and Action.

Goal:

Like assigning tasks to others, start with "state the outcome first." Even if complex information follows, knowing the end goal helps avoid confusion for both humans and ChatGPT. Setting a clear goal for ChatGPT helps it precisely understand the task, akin to entering an accurate destination in a GPS system for precise navigation.

Input:

When assigning a task to ChatGPT, sometimes providing necessary raw data is required for execution. For instance, if the task is to rank a list, the original list needs to be provided to ChatGPT.

Action:

The action refers to the task's actions, usually starting with a verb like "write" a speech, "explain," "translate," "review," "summarize," etc. Actions typically appear at the beginning of the task since directive sentences in English usually start with a verb.

I use the acronym WWW to represent Who (Role), Where (Context), and What (Task) as the three core elements. The acronym GIA represents Goal, Input, and Action as the three task elements. GIA is also the acronym for Gemological Institute of America, easy to remember.

While the core elements are vital for guiding AI, there are additional auxiliary elements that add structure and clarity and enhance the prompt's effectiveness beyond what core elements alone can achieve.

2. Auxiliary Elements of Prompts

Besides the three core elements above, there are six auxiliary elements in prompt design including Constraints, Output Format, Separator, Example, Emotional Tone (EQ), and Directive Prompt.

The acronym for these six elements is CoSeed. "Co" implies cooperation, and "Seed" signifies a beginning, fitting the meaning of "auxiliary elements" and is easy to remember.

Since ChatGPT's knowledge is vast, using it to perform tasks is like

searching for a needle in a haystack. If the prompt can point out the direction and scope of the search, it will quickly yield the most effective answers from ChatGPT.

These six auxiliary elements clarify the prompt's direction and narrow the response scope, improving accuracy and relevance. These elements frame the conversation, guiding it toward a more specific, needs-aligned direction.

1) Constraints.

When using ChatGPT, it's important to add constraints like word limits or restrict specific resources. This narrows the scope of the answer, making it more focused and precise.

It's like giving a free-spirited artist some constraints to ensure the work stays on topic while creating.

For example, when asking ChatGPT to write an article, adding a word limit can prevent the response from being too broad and providing a lot of irrelevant text. For instance, in "Please explain why the sky is blue, but keep it under 100 words," the phrase "but keep it under 100 words" is a constraint.

Another way to set constraints is to specify what **not to do**, telling ChatGPT not to consider certain things or directions, like not choosing spicy food or not suggesting certain travel destinations.

For example, you could ask: "Please summarize this article, but do not include any specific dates or names in the summary." You can also exclude entire categories of things, such as "Please write a short story, but do not include any animals."

When we talk about creating custom GPTs later, this constraint prompt can be used to exclude services or languages you don't want.

2) Output Format.

ChatGPT excels at producing different formats. Specifying a particular output format (e.g., list, paragraph, formula) can help it focus on the structure of the answer, making the information clearer and easier to understand.

Although ChatGPT is intelligent, its replies are sometimes not in the desired format. For example, it might provide text when you need a table. You then

have to waste time converting the text into a table manually. Adding an "output format" requirement can make ChatGPT's answers directly meet your needs.

Some commonly used formats include: Bullet points, tables, scorecards, CSV, bar charts, line charts, HTML, JSON, Python, and Markdown formats. These output formats and examples are organized in Appendix B.

3) Separator

To ensure ChatGPT clearly understands your questions and needs, strive for clarity and structure in your prompts. Using separators can effectively achieve this.

Common separators include quotation marks " ", [], (), and ##. They clearly distinguish different parts of the prompt, avoiding misunderstanding by ChatGPT.

Just like writing an article or report using paragraphs to separate different thoughts, these separators help ChatGPT more accurately understand the input message.

Commas and paragraph breaks create a clear structure, while quotation marks highlight a specific content block. When providing large amounts of data or engaging in multi-turn conversations, you can use hashtags or quotation marks to separate prompts and context.

A commonly used prompt is: "Please translate the following text into English: '(insert text to be translated here)'." Here, the colon indicates the text to be translated, and the quotation marks separate this text from other content.

Additionally, using "# (hashtag) notation" can help ChatGPT effectively track and understand multi-turn or complex conversations.

For example, if you have two different questions: "Market analysis for Project A" and "Technical challenges for Project B," you can mark them as "#ProjectA" and "#ProjectB" respectively. Later, when you want to discuss "Market analysis for Project A," you only need to say, "Let's discuss '#ProjectA'," and ChatGPT will immediately know which part of the conversation or data you are referring to.

Using separators frequently can help ChatGPT more accurately understand our intentions and process the information according to our needs, avoiding misinterpretation or incorrect execution of the prompt content.

4) Example.

Examples act like blueprints for ChatGPT to mimic or reference, helping it better understand the context of the question and the expected answer format.

Including examples in the prompt can guide ChatGPT's response direction, preventing it from relying solely on its corpus, thus avoiding undesired answers.

Here's an example (Figure III.04). We asked it to design a new dish, giving it an example: the flavor should be similar to American Cheeseburger.

Please help me create a new beef-based dish. For example, I want the flavors similar to a classic American cheeseburger. Design a simple and delicious new recipe?



To create a new beef-based dish with classic American cheeseburger flavors, try Cheeseburger Stuffed Bell Peppers.

Ingredients:

4 large bell peppers, 1 lb ground beef, 1 small onion (finely chopped), 1 cup shredded cheddar cheese, 1/2 cup ketchup, 1/4 cup mustard, 1 cup cooked rice (optional), salt and pepper to taste, diced pickles (optional), sesame seeds (optional).

Instructions:

Preheat oven to 375°F. Cut tops off bell peppers, remove seeds. Cook ground beef and onion until browned, drain fat. Stir in ketchup, mustard, salt, and pepper; cook 2-3 minutes, mix in rice if using. Fill bell peppers with beef mixture, top with cheese. Place in baking dish, cover with foil, bake 30 minutes. Remove foil, bake 10 more minutes until peppers are tender and cheese melted. Garnish with pickles and sesame seeds if desired.

Figure III.04: ChatGPT designs a new dish similar to American Cheeseburger.

You can see that ChatGPT designed a "Cheeseburger stuffed Bell Peppers"

following the example. We tried it and it was delicious.

We tried not giving it an example, and it designed “Maple-Bourbon Glazed Chicken with Sweet Potato Mash and Grilled Asparagus”, and “Smoky BBQ Shrimp Tacos with Avocado Slaw.” They are just overall the map. So, it's better to give examples.

5) Emotional Tone (EQ).

When creating prompts for ChatGPT, emotional elements are the superpower of communication, including tone, attitude, and style. I collectively call them Emotional Tone (EQ).

"Tone" is like the soul of your words; whether friendly, serious, or humorous, it leaves a lasting impression. Choosing the right tone is like having the golden key to communication, making people feel comfortable and making your topic more engaging.

"Attitude" shows your position and is the window to your emotions. Your supportive or opposing stance gives your words different weight and direction.

As for "style," it's the personality tag of your speech, whether formal or casual, depending on your audience. A suitable style makes your words more attractive, understandable, and improves communication quality.

These emotional elements, like colorful clothes for ChatGPT's prompts, make them vivid and personal while ensuring the answers align with the context and purpose of the question.

For example, "Let me share some important information on this topic" is formal, while "Let's discuss this important issue together" is conversational and friendly.

Whether you need a formal authoritative response or a relaxed humorous reply, setting the "emotion" ensures ChatGPT captures these subtle yet important needs accurately.

Let's look at two examples in Figure III.05-1 and III.05-2. The same question was asked but each with a different tone, resulting in different responses from ChatGPT.

Why is the sky blue, please answer it in a conversational tone.



The sky is blue because when sunlight hits our atmosphere, it scatters blue light in all directions. So, we end up seeing a blue sky wherever we look. It's like the sun's way of adding a splash of color to our day!

Figure III.05-1: ChatGPT answers in a conversational tone

Why is the sky blue, please answer briefly in a formal academic tone.



The sky appears blue due to Rayleigh scattering, where shorter blue wavelengths of sunlight are scattered in all directions by the gases and particles in Earth's atmosphere, more so than other colors. This scattering causes the human eye to perceive the sky as blue.

Figure III.05-2: ChatGPT answers in a formal academic tone

You probably can easily tell the difference between these two answers.

For additional emotional styles, refer to Appendix C.

6) Directive Prompt—Special Spells

In Part Two, we have introduced directive prompts (or special spells), such as "Think quietly before answering" or "Take one step at a time." These directive prompts can guide the behavior or emotional adjustment in the conversation, creating a dialogue atmosphere closer to the user's needs.

"Take one step at a time" is another example of Directive Prompt.

What if we don't know what thought process to use to solve a problem?

It's simple—ask ChatGPT to execute step by step. This way, when facing complex problems or tasks, ChatGPT can provide more accurate and helpful answers.

For example, in figure III.06-1, we ask ChatGPT to explain what pre-training of a large language model is. In the next figure III.06-2, we add "Take one step at a time" before the prompt. Let's see what the difference is.

Please explain pre-training of large language models (LLMs).



Pre-training of large language models (LLMs) involves training the model on a vast corpus of text data to learn language patterns, grammar, and context. This process is unsupervised, meaning the model is fed text without explicit labels or instructions. The objective is typically to predict the next word in a sentence, allowing the model to learn from the structure and semantics of the data. By processing massive amounts of text, the model builds a rich understanding of language, which can be fine-tuned later for specific tasks such as translation, summarization, or question answering.

Figure III.06-1: ChatGPT explains pre-training of large language models

Please explain pre-training of large language models (LLMs) step by step.



Sure, here's a more concise version:

1. **Data Collection and Tokenization:** Gather a vast corpus of text from diverse sources and convert it into tokens (words, subwords, or characters) to prepare the data for model processing.
2. **Model Architecture and Initialization:** Define the model structure, such as the number of layers and attention mechanisms, and initialize the parameters with small random values.
3. **Training Process:** Train the model to predict the next token in a sequence by processing the tokenized text, optimizing the parameters through forward propagation, loss calculation, and backpropagation using algorithms like Adam or SGD.
4. **Evaluation and Optimization:** Regularly evaluate the model on a validation set to monitor performance, apply regularization techniques to prevent overfitting, and save intermediate versions to safeguard progress.



Figure III.06-2: ChatGPT explains pre-training of large language models step by step.

You see that by adding “step by step”, the second answer becomes clearer and more structured, explaining each step in detail.

Some people say you can add rewards to these prompts, such as "I will give a tip" (although you can't actually give one), which improves the output. We often add "Thank you" at the end of our prompts. LLMs might have learned during pre-training that when people "politely ask," they usually give better responses, so this should be beneficial!

To summarize, to design high-efficiency prompts, we first need to learn how AI engineers fine-tune AI. They let LLMs know who they are (Who, Role), where they are (Where, Context), and what they are to do (What, Task). These three are the core elements of prompts, supplemented by six auxiliary elements including Constraints, Output format, Separators, Example, EQ and Directive prompt (i.e. CoSeed).

The combined abbreviation for these core and auxiliary elements is 3W-CoSeed. In fact, Co-Seed is a universal usable prompt structure as we will demonstrate later.

3. Effective Prompt Design Process

Prompts are a form of language, and language itself is an art. Without a structured process, prompt design can become disorganized, leading to aimless exchanges that waste time and yield poor results.

Here's an effective prompt design process: When we have a task and want to collaborate with ChatGPT to find a solution, the first step, as with any task, is to understand the task and set clear goals.

The initial goal doesn't have to be perfect, as the process is iterative. Through a few rounds of refinement, the goal becomes clearer.

Next, we apply the three principles of task execution discussed in Chapter 7, Figure II.03, to craft effective prompts and achieve the desired outcome.

To implement these elements effectively, follow this structured process for designing prompts that yield optimal results.

Step 1 (Principle 1: Decomposition): If the task is large, break it down into smaller subtasks. For each subtask, follow the next two steps.

Step 2 (Principle 2: Evaluation): Choose an initial prompt and send it to

ChatGPT. Evaluate each response. If the response meets the sub-goal, the process for that subtask ends.

Step 3 (Principle 3: Iteration): If the response is unsatisfactory, refine the prompt through prompt engineering techniques and send it again to ChatGPT. Repeat this process until the desired response is achieved.

Once all sub-task goals are satisfactorily met, they are combined to achieve the overall goal.

Chapter 9 Summary

- ❖ To get the desired response from ChatGPT, we need to guide it in the right “Direction” and narrow the “Scope” of its text generation.
- ❖ Key point is that effective prompts are built using three core elements and six auxiliary elements. They work together to create a framework that ensures ChatGPT generates responses in line with the desired direction and scope.
- ❖ Three core elements (WWW) are: Who (Role), Where (Context), and What (Task.) The Task element further breaks down into three parts: Goal, Input, and Action (GIA.) ❖ Six auxiliary elements (CoSeed): Constraints, Output Format, Separator, Example, Emotional Tone (EQ), Directive Prompt (or special spells). Together, these elements are summarized as 3W-CoSeed.
- ❖ The process for designing effective prompts, after defining the task and goal, involves:
 - 1) Decomposition: Break large tasks into smaller subtasks.
 - 2) Evaluation: Send a prompt to ChatGPT and assess the response. If it meets the subtask goal, proceed.
 - 3) Iteration: Refine the prompt if needed and retry until the desired response is reached.

Finally, combine completed subtasks to achieve the overall goal.

Chapter 10: Framework-Based Prompts

It is common to see people sharing effective prompts, and many of these are "framework" prompts, utilizing specific elements to construct prompts.

Framework-based prompts, also known as "Acronym prompts," provide a structured approach to guiding AI responses.

This chapter introduces various framework prompts and explains what they are and how they work, showing how their elements align with the 3W-CoSeed structure.

1. What are Framework-Based Prompts

A framework is like a template, similar to choosing a house layout. If you want a two-bedroom house, the builder will show you three templates of the two-bedroom house. If you change your mind and want three bedrooms, they'll show you another four different templates.

After choosing, you can decorate and modify the details of each room of the house yourself.

Framework prompts are predefined and tested prompts that can be reused in different contexts by adjusting the details. The goal is to simplify and speed up the prompt development process using a known "working" framework.

Framework-based prompts work like subroutines in software development, providing reusable structures that simplify and accelerate new application development across different contexts.

To make these framework prompts easier to remember, they are often formed using the first letters of their elements, which we refer to as "acronym prompts." Examples include TAG, CARE, and TRACE, which are easy to remember.

Importantly, these acronyms correspond to the three core elements and six auxiliary elements mentioned in the previous chapter, forming a framework prompt known as "3W-CoSeed."

2. The all-in-one 3W-CoSeed Framework Prompt

3W-CoSeed is a versatile framework prompt for executing any task.

Let me demonstrate this. For instance, if we want an article on "benefits of sustainable tourism," we can use the 3W-CoSeed framework to design the prompt.

This prompt incorporates the three core elements—ChatGPT's role, task, and context—along with six auxiliary elements to guide the direction and scope of the generated text.

Prompt:

" You are a travel magazine editor with 15 years of experience [Role]. Please write an article on the benefits of sustainable tourism [Task]. The article should be suitable for readers with an environmental awareness [Context]. Do not exceed 300 words [Constraint], and write in English [Output Format]. If you use any keywords, use “” as the separator to separate the keyword [Separator]. ‘Refer to these key points from my previous article for reference [Example]: 1. Sustainable tourism minimizes environmental impact. 2. Engages and benefits local communities economically and culturally.’ The article should be inspiring [EQ], and please remain objective [Directive Prompt]."

From now on, when a prompt is lengthy, it will only be stated as text and will not show in its ChatGPT's response. The ChatGPT's response for the above prompt is shown in Figure III.07.



Embracing Sustainable Tourism: A Path to a Greener Future

In today's world, the importance of "sustainable tourism" has never been more apparent. This travel philosophy focuses on minimizing environmental impact while benefiting local communities economically and culturally. By adopting sustainable practices, travelers can ensure their adventures contribute positively to the destinations they visit.

"Sustainable tourism" significantly reduces the carbon footprint associated with travel. Opting for eco-friendly transportation, such as trains or electric vehicles, and staying in accommodations that prioritize renewable energy and waste reduction are small changes that make a big difference. These choices help preserve natural landscapes and protect wildlife habitats, ensuring that future generations can enjoy the same beauty.

Moreover, "sustainable tourism" fosters deep connections with local communities. Travelers engaging in eco-friendly tours often support local businesses, from family-run restaurants to artisanal markets. This economic boost empowers communities, providing them with the means to preserve their cultural heritage and traditional practices. It also promotes cultural exchange, enriching the travel experience by offering authentic insights into local customs and lifestyles.

By prioritizing "sustainable tourism," travelers can also inspire others to adopt similar practices. Sharing stories and experiences from eco-friendly trips raises awareness about the importance of sustainability in travel. This ripple effect encourages more people to make conscious choices, creating a collective movement toward preserving our planet.

In conclusion, embracing "sustainable tourism" is a powerful way to enjoy our world responsibly. It not only minimizes environmental impact but also engages and benefits local communities, ensuring that travel remains a force for good. By making mindful decisions, we can protect our planet's natural and cultural treasures, paving the way for a greener, more sustainable future.

Figure III.07: Use "3W-CoSeed Framework Prompt" to generate an article on the benefits of sustainable tourism.

ChatGPT's response in Figure III.07 indeed answered the prompt and met all the requirements.

Next, let's examine some common framework prompts and their strengths, illustrating how each fits within the 3W-CoSeed framework.

3. Common Framework Prompts

Common framework prompts usually include elements whose initial characters form its acronym. Let's look at five popular framework prompt examples.

1) RTF framework prompt

A simple prompt framework is RTF (Role, Task, Format). It's commonly used by beginners.

- Role: Assign a role to ChatGPT.
- Task: Assign a task for ChatGPT to perform;

- Format: Specify the format of the answer;

Prompt example (Figure III.08):

"You are a Chinese cuisine chef with 20 years of experience [Role]. Please provide a menu of four dishes (including an appetizer, soup, and two main courses) [Task]. Format the response as a table [Format]."

You are a Chinese cuisine chef with 20 years of experience [Role]. Please provide a menu of four dishes (including an appetizer, soup, and two main courses) [Task]. Format the response as a table [Format].



Sure! Here's a menu of four Chinese dishes:

Course	Dish Name	Description
Appetizer	Spring Rolls	Crispy rolls filled with a mix of vegetables and sometimes meat, served with a sweet and sour sauce.
Soup	Hot and Sour Soup	A flavorful soup with a perfect balance of spicy and sour, containing tofu, mushrooms, and bamboo shoots.
Main Course 1	Kung Pao Chicken	Spicy stir-fried chicken with peanuts, vegetables, and chili peppers in a savory and tangy sauce.
Main Course 2	Mapo Tofu	Spicy Sichuan dish featuring tofu set in a flavorful sauce, typically with minced meat and fermented black beans.

Figure III.08: RTF Prompt example, a menu of four Chinese dishes

Analyzing this example reveals that RTF aligns tightly with the 3W-CoSeed structure.

Analysis:

Comparing RTF elements (left) with 3W-CoSeed elements (right):

RTF Element	Corresponding 3W-CoSeed element
Role	Role = Who
Task	Task = What
Format	Output Format in CoSeed

Thus, RTF is a subset of the 3W-CoSeed framework.

2) TAG framework prompt

TAG (Task, Action, Goal) is another simple and useful prompt framework:

- Task: Assign a task to ChatGPT;
- Action: Specify the action ChatGPT should take;
- Goal: Define the goal to be achieved.

Prompt example (Figure III.09):

"Implement a themed challenge to increase brand visibility and audience engagement on Instagram [Task]. Encouraging users to share photos or videos of themselves using our sports gear outdoors. Participants should use a specific hashtag we create for the event and tag our official Instagram account [Action]. Within the next three months, achieve a 30% increase in brand visibility on Instagram and double user engagement and content sharing [Goal]."

ChatGPT's reply to this prompt is shown in Figure III.09.



To increase brand visibility and audience engagement on Instagram, launch an Instagram challenge where users share photos or videos of themselves using our sports gear outdoors. Participants should use the hashtag #GearUpOutdoors and tag our official Instagram account. Within the next three months, aim to achieve a 30% increase in brand visibility on Instagram and double user engagement and content sharing.

Strategy

1. **Announcement Post:** Create a compelling announcement post with clear instructions on how to participate. Include the hashtag and a call to action.
2. **Incentives:** Offer attractive prizes for the best posts, such as free gear, gift cards, or exclusive discounts.
3. **Partnerships:** Collaborate with influencers and sports communities to amplify the reach.
4. **User Features:** Regularly feature user-generated content on your official account to encourage more participation.
5. **Promotional Campaigns:** Utilize Instagram ads to promote the challenge to a broader audience.



Figure III.09: TAG Prompt example, implement a themed challenge

Looking closer, we see how the TAG framework maps to the 3W-CoSeed elements.

Analysis:

Comparing TAG elements with 3W-CoSeed elements:

TAG Element	Corresponding 3W-CoSeed element
Task	Task = What
Action	Action in Task = What
Goal	Goal in Task = What

TAG essentially assigns a task to ChatGPT, including action and goal. TAG is also a subset of the 3W-CoSeed framework.

3) CARE framework prompt

The CARE framework includes four elements: Context, Action, Result, Example.

- Context: Provide background information. For example, introduce the topic when asking it to write an article.
- Action: Specify the action ChatGPT should take.
- Result: Define the goal to be achieved.
- Example: Provide an example for ChatGPT to reference or mimic.

Prompt Example (Figure III.10):

"We are a small bookstore looking to increase customer traffic and sales, especially aiming to attract more young readers [Context]. Design a social media campaign [Action]. The goal is to increase customer traffic and sales by 10% within a month after the campaign ends, while also enhancing the bookstore's visibility and appeal among young readers [Result]. A successful example is the "Summer Reading Challenge" which encouraged readers to read and share book reviews during the summer [Example]."

The ChatGPT's reply to this prompt is shown in Figure III.10. ChatGPT has been instructed to reply more concisely to save space, but it can also provide a more detailed plan if needed.



Campaign: "Autumn Reading Adventures"

Duration: 4 weeks

Goal: Boost traffic/sales by 10% post-campaign, attract young readers.

Strategy:

- **Challenges:** Weekly reading challenges with prizes for reviews.
- **Events:** Live Q&As with authors, online book clubs.
- **Engagement:** Daily posts, share customer content via hashtag #AutumnBookAdventures.
- **Collabs:** Partner with influencers, schools.
- **Promos:** Discounts on featured books, loyalty points for purchases/reviews.
- **Ads:** Targeted Facebook/Instagram ads.

Execution:

- Pre-campaign: Prep content, schedule events.
- During: Post daily, engage, track.
- Post: Announce winners, analyze results.

Evaluation: Monitor engagement, sales, feedback.

This approach aims to quickly increase traffic, sales, and young reader engagement.

Figure III.10: CARE Prompt example, design a social media campaign.

The CARE framework also fits neatly within the 3W-CoSeed framework, offering context and examples alongside task instructions.

Analysis:

Comparing CARE elements with 3W-CoSeed elements:

CARE Element	Corresponding 3W-CoSeed element
Context	Context = Where
Action	Action in Task = What
Result	Goal in Task = What

Example

Example in CoSeed

CARE essentially assigns a task to ChatGPT with action and goal, then provides context and an example. As previous cases, CARE is also a subset of the 3W-CoSeed framework.

4) ICIO framework prompt

This simple and effective framework consists of four main elements: Instruction, Context, Input, Output, proposed by Elvis Saravia of DAIR.AI.

It's easy to remember by thinking of ChatGPT as a black box, with Input on one side and Output on the other. Just feed in Instruction and Context.

- Instruction: What you want ChatGPT to do, e.g., language translation or creative content writing;
- Context: Provide some background information, e.g., introduce the topic if asking it to write an article;
- Input: The raw data ChatGPT needs, e.g., original text for translation;
- Output: Specify the desired result, e.g., "Present the result in text."

Prompt Example (Figure III.11):

"Provide a 150-word summary of the basic principles of Bitcoin [Instruction]. Bitcoin is a digital currency invented by an anonymous person or group in 2008. It is not controlled by any central authority and records transactions through distributed ledger technology—blockchain [Context]. This example does not require input data [Input]. Please briefly describe in English [Output]."



Bitcoin, created in 2008 by an anonymous entity known as Satoshi Nakamoto, is a decentralized digital currency. It operates without a central authority, using blockchain technology to record transactions. The blockchain is a public, distributed ledger where all transactions are permanently stored, ensuring transparency and security. Bitcoin's supply is limited to 21 million coins, controlled by a process called mining, where participants validate transactions and add them to the blockchain. This decentralized system relies on cryptographic techniques to ensure the integrity of transactions, making Bitcoin resistant to censorship and fraud. Its value fluctuates based on market demand, and it can be used for peer-to-peer transactions, offering an alternative to traditional financial systems.

Figure III.11: ICIO Prompt example, Bitcoin summary

We find a direct correlation between ICIO elements and those in the 3W-CoSeed framework.

Analysis:

Comparing ICIO elements with 3W-CoSeed elements:

ICIO Element	Corresponding 3W-CoSeed element
Instruction	Action in Task = What
Context	Context = Where
Input	Input in Task = What
Output	Output Format in CoSeed

ICIO essentially assigns a task, provides context and input, and sets the output format. ICIO is also a subset of the 3W-CoSeed framework.

5) CRISPE framework prompt

Another frequently used framework is the CRISPE framework, proposed by Matt Nigh on GitHub. It includes six main elements: Capacity, Role, Insight, Statement, Personality, Experiment.

- Capacity: The abilities of ChatGPT's assigned role;
- Role: The role ChatGPT will play, such as writer, chef, or AI expert;
- Insight: Background information and context ChatGPT needs;
- Statement: The specific task you want ChatGPT to accomplish;
- Personality: The style or manner you expect ChatGPT to respond with;
- Experiment: Asking ChatGPT to provide multiple answers.

Prompt Example (Figure III.12):

"Imagine yourself as a city officer and the lead writer for a future city planning blog [Capacity and Role]. The readers are urban planners, policymakers, and general public interested in urban innovation [Insight]. Discuss how IoT and big data analytics can improve residents' quality of life and achieve environmental sustainability [Statement]. You will write in a conversational style [Personality]. Describe how a city's traffic management system uses AI to predict and optimize traffic flow, reducing congestion and pollution [Experiment]."



Memory updated

Imagine you're navigating through the hustle and bustle of a busy city, trying to get from point A to point B. Now, picture that same city, but with a twist—one where traffic flows seamlessly, congestion is minimized, and pollution is drastically reduced. This isn't a utopian fantasy; it's the reality that IoT (Internet of Things) and big data analytics can help us achieve.

With IoT, every traffic light, sensor, and vehicle can communicate with each other, creating a network that feeds real-time data into a central traffic management system. This is where AI steps in. By analyzing this constant stream of data, AI can predict traffic patterns, identify potential bottlenecks, and adjust traffic signals accordingly to optimize the flow. Imagine a world where your commute is not just faster but also greener because less time in traffic means lower emissions.

In essence, IoT and big data aren't just buzzwords—they're powerful tools that can significantly improve our quality of life while pushing us toward a more sustainable future. Through smart traffic management, we can reduce both congestion and pollution, creating a city that's not only more efficient but also more livable. This is the kind of future we should all be excited about.



Figure III.12: CRISPE framework prompt, using AI to predict and optimize traffic flow

The CRISPE framework integrates several elements found within the 3W-CoSeed framework, providing a robust prompt structure.

Analysis:

Comparing CRISPE elements with 3W-CoSeed elements:

CRISPE Element	Corresponding 3W-CoSeed element
Capacity	Role = Who
Role	Role = Who
Insight	Context = Where
Statement	Task = What
Personality	EQ in CoSeed
Experiment	Output Format in CoSeed

CRISPE essentially provides ChatGPT with a role, context, and task, then adds emotional tone and output elements. Here again, these 5 elements are a subset of that of the 3W-CoSeed framework.

In summary, these five framework prompts use unique terms to create memorable acronyms, like CARE or CRISPE, simplifying their use and recall. Notably, all the elements of these prompts are subsets of the nine elements found in the 3W-CoSeed framework, demonstrating that the 3W-CoSeed framework is comprehensive enough to encompass and replace all these common frameworks.

Additionally, you can enhance these framework prompts by incorporating additional elements from 3W-CoSeed to better customize ChatGPT's responses to your specific needs.

Chapter 10 Summary

- ❖ Framework prompts are predefined prompts that can be reused across different contexts by adjusting their details, streamlining the prompt creation process with established, effective frameworks.
- ❖ The three core elements and six auxiliary elements form **the comprehensive 3W-CoSeed framework prompt**. By familiarizing yourself with the three core elements (WWW + GIA) and the six auxiliary elements (CoSeed), you can design any framework prompt available on the market.
- ❖ This chapter provides detailed explanations of various framework prompts, including RTF, TAG, CARE, ICIO, and CRISPE.
- ❖ All common framework prompts are subsets of the 3W-CoSeed framework, allowing for customization by adding other elements from CoSeed as needed to refine ChatGPT's responses.

Chapter 11: Applications and Exercises (1)

Nine more common framework prompts will be introduced in this chapter. The first six framework prompts are demonstrated with pre-designed prompts, showcasing their application in business and workplace scenarios such as product development planning, marketing strategy, content strategy planning, search engine optimization (SEO), and first-party data collection and usage.

By examining these example prompts closely, you can learn how to design similar prompts. You can then modify these examples to suit your specific needs and different application contexts. Use your modified prompt to have ChatGPT generate new responses and examine any differences.

The remaining three framework prompts (7,8 and 9) are designed for you to practice with as exercises. We believe these are the most important and most interesting part of this chapter. You can follow our explanation to have ChatGPT HELP YOU generate prompts using these frameworks step by step.

For each framework prompt introduced, its elements will be listed, showing how they align with the 3W-CoSeed elements. This alignment demonstrates that these nine framework prompts are all subsets of the comprehensive 3W-CoSeed framework.

All framework prompts in this chapter are applied to practical businesses. In Chapter 13, "Applications and Exercises (2)," we will show how to use framework prompts in everyday life.

1) RACE framework prompt: Role, Action, Context, Expectation

This framework is designed to structure prompts for scenarios requiring a clear definition of roles, actions, context, and expected outcomes. It is particularly useful in scenarios where clear, goal-oriented tasks need to be defined as in business.

Element	Description	Corresponding 3W-CoSeed Element
Role	Setting ChatGPT's role	Role = Who
Action	Task or activity to be	Action in Task = What

	completed	
Context	Provide context or background information	Context = Where
Expectation	Expected result	Goal in Task = What

Thus, the RACE framework is a subset of the 3W-CoSeed framework.

Application example using the RACE prompt: Product Development Planning and Strategy.

Prompt:

"You are a product design engineer responsible for developing and designing products. Your duties include understanding market needs, creating innovative solutions, and collaborating with different departments to ensure successful product launches [Role]. Please provide a plan and detailed strategy to address market changes and challenges. The strategy should include understanding customer needs, conducting market analysis, collaborating with the design team to develop new products, and regularly evaluating and adjusting product designs to meet market demands [Action]. Companies must continually seek innovation and improvement to maintain a competitive edge. Product design engineers play a crucial role in developing and designing products that meet market demands while ensuring the product's quality and performance are at their best [Context]. Your creativity and technical expertise can bring more market share and competitive advantage to the company, leading to commercial success for the product [Expectation]."

Please enter this prompt into ChatGPT to see the response.

2) APE framework prompt: Action, Purpose, Expectation

This framework prompt is commonly used for scenarios requiring a clear definition of action, purpose, and expected outcomes. It is designed to create effective AI interactions for applications such as marketing strategies.

Element	Description	Corresponding 3W-CoSeed Element
Action	Task or activity to be	Action in Task = What

	completed	
Purpose	Goal or intent	Goal in Task = What
Expectation	Expected result	Goal in Task = What

The APE framework is also a subset of the 3W-CoSeed framework.

Application example using the APE prompt: Marketing Strategy .

Prompt:

"Please design a content marketing plan for our newly launched biodegradable tableware [Action]. Our goal is to resonate with environmentally-conscious consumers and raise awareness and passion for sustainability issues [Purpose]. The plan should attract our target audience, build a strong brand image, and aim to increase pre-orders by at least 30% [Expectation]."

Please submit this prompt into ChatGPT to view the results.

3) ERA framework prompt: Expectation, Role, Action

ERA is designed to set clear expectations, define roles, and specify actions. This framework is particularly useful in scenarios where aligning AI outputs with specific goals is crucial, such as in marketing strategy development.

Element	Description	Corresponding 3W-CoSeed Element
Expectation	Expected result	Goal in Task = What
Role	Setting ChatGPT's role	Role = Who
Action	Task or activity to be completed	Action in Task = What

The ERA framework is also a subset of the 3W-CoSeed framework.

Application example using the ERA prompt: Marketing Planning and Strategy.

Prompt:

"We aim to improve our LINE marketing effectiveness by 20% [Expectation]. You are the head of the marketing department responsible for designing and implementing successful marketing strategies [Role]. Provide a plan and

detailed strategy to optimize our LINE marketing themes, content, and delivery schedules based on analysis [Action]."

Enter this prompt to ChatGPT to observe the response.

4) RISE framework prompt: Role, Input, Steps, Expectation

This framework is designed to shape prompts by clearly defining roles, inputs, steps, and expected outcomes for specific scenarios. This structured tool helps define clear, detailed instructions for tasks requiring methodical execution.

Element	Description	Corresponding 3W-CoSeed Element
Role	Setting ChatGPT's role	Role = Who
Input	Detailed information about the target audience	Input in Task = What
Steps	Steps to be taken	Action in Task = What
Expectation	Expected result	Goal in Task = What

The RISE framework is also a subset of the 3W-CoSeed framework.

Application example using the RISE prompt: Content Strategy Planning.

Prompt:

"Assume you are the website content manager responsible for developing content that resonates with the audience [Role]. I have gathered detailed information about our target audience, including their interests, needs, and common questions related to our industry [Input]. Please provide a step-by-step content strategy plan to identify key topics, create an editorial calendar, and draft engaging content that aligns with our brand message [Steps]. The expected result is to increase our website's monthly visitors by 40% and elevate our brand's position in the industry, making us a key opinion leader [Expectation]."

Input this prompt into ChatGPT to see the results.

5) CTF framework prompt: Context, Task, Format

This framework organizes prompts that require clear definitions of context,

tasks, and expected formats. It is a concise, efficient tool designed to provide essential information while keeping instructions clear and focused.

Element	Description	Corresponding 3W-CoSeed Element
Context	Provide context or background information	Context = Where
Task	Task to be completed	Task = What
Format	The format or structure of the output	Output Format in CoSeed

The CTF framework is also a subset of the 3W-CoSeed framework.

Application example using the CTF prompt: Search Engine Optimization (SEO).

Prompt:

"In today's digital age, SEO strategies can increase website visibility, enhance brand recognition, and generate more traffic and business opportunities [Context]. Please provide a detailed guide on how to perform SEO, including but not limited to the following areas: keyword research, content optimization, technical optimization, external links, monitoring, and optimization [Task]. Provide a well-structured manual, including detailed steps and practical tips, as well as relevant examples and case studies [Format]."

Input this prompt into ChatGPT to see the results.

6) COAST framework prompt: Context, Objective, Actions, Scenario, Task

This framework is designed to structure prompts for scenarios requiring a clear definition of context, objective, actions, scenario and tasks. It is a prompting tool used to structure prompts in a way that provides detailed guidance for complex tasks.

Element	Description	Corresponding 3W-CoSeed Element
Context	Provide context or	Context = Where

	background information	
Objective	Expected goal	Goal in Task = What
Action	Task or activity to be completed	Action in Task = What
Scenario	Provide a possible scenario	Context = Where
Task	Task to be completed	Task = What

The COAST framework is also a subset of the 3W-CoSeed framework.

Application example using the COAST prompt: First-Party Data Collection and Usage.

Prompt:

"With the advent of data privacy laws, using third-party data for marketing is becoming increasingly restricted [Context]. Our goal is to adjust our strategy to focus more on first-party data collection and usage [Objective]. Establish an efficient data collection framework on our platform to facilitate usage [Action]. Next month, we will launch a new product line [Scenario]. Develop a detailed plan for first-party data collection and usage for the upcoming marketing campaign [Task]."

Input this prompt into ChatGPT to see the results.

7) ROSES framework prompt: Role, Objective, Scenario, Expected Solution, Steps

This framework is intended to shape prompts by clearly defining roles, objectives, scenarios and expected solutions and steps. It is a structured approach to prompt engineering that is particularly useful in situations requiring detailed problem-solving and task execution.

Element	Description	Corresponding 3W-CoSeed Element
Role	Setting ChatGPT's role	Role = Who
Objective	Expected goal	Goal in Task = What
Scenario	Provide a possible scenario	Context = Where

Expected Solution	Expected result	Goal in Task = What
Steps	Steps to be taken	Action in Task = What

The ROSES framework is also a subset of the 3W-CoSeed framework.

Exercise:

Let's try a new method:

ask ChatGPT to design a prompt using the ROSES framework.

(Isn't this amazing? Ask ChatGPT to generate prompt for you.)

Use the provided prompt to instruct ChatGPT to design a business-related prompt using the ROSES framework. Analyze the response to understand how ChatGPT structures the prompt based on the elements you provided.

Prompt:

"You are a prompt design engineer. Please design a prompt for 'business application' based on the following five elements: Role, Objective, Scenario, Expected Solution, and Steps. Start by specifying what the business application is and label each element accordingly."

Carefully review ChatGPT's reply and learn from it. This is a very useful method to have ChatGPT generate prompts based on your specified requirements directly.

8) TRACI framework prompt: Task, Role, Audience, Create, Intent

This framework is crafted to organize prompts that need explicit definitions of tasks, roles, audience, create and intent. It is a prompting tool that helps in structuring AI prompts for tasks that require clear communication and purpose-driven outputs.

Element	Description	Corresponding 3W-CoSeed Element
Task	Task to be completed	Task = What
Role	Setting ChatGPT's role	Role = Who
Audience	The audience for the task result	Context/User = Who
Create	Take action to create	Action in Task = What
Intent	Goal or intent	Goal in Task = What

The TRACI framework is also a subset of the 3W-CoSeed framework.

Exercise:

Follow the prompt design method in Exercise 7. You will ask ChatGPT to identify the Task, Role, Audience, Create and Intent. This time, you can set the application yourself; just replace "business application" with the application you like.

Review ChatGPT's reply closely and learn from it.

9) RASCEF framework prompt: Role, Action, Steps, Context, Example, Format

This framework is tailored to craft prompts that require precise delineation of roles, actions, steps, context, example and output format. It is a prompting tool designed to guide AI in generating structured responses, particularly in tasks requiring clear instructions or process documentation.

Element	Description	Corresponding 3W-CoSeed Element
Role	Setting ChatGPT's role	Role = Who
Action	Task or activity to be completed	Action in Task = What
Steps	Steps to be taken	Action in Task = What
Context	Provide context or background information	Context = Where
Example	Provide an example of the task	Example in CoSeed
Format	The format or structure of the output	Output Format in CoSeed

The RASCEF framework is also a subset of the 3W-CoSeed framework.

Exercise:

Follow the prompt design method from Exercise 7 to ask ChatGPT. You will ask ChatGPT to identify the Role, Action, Steps, context, Example and Output Format. This time, you can define the application yourself by replacing "business application" with your desired application.

Please examine ChatGPT's response closely and learn from it.

Chapter 12: Basic Thought Based Prompts

This chapter delves into two revolutionary techniques in prompt engineering: "Shot prompting" and "Chain-of-Thought prompting."

In prompt engineering, "Shot" refers to providing examples to the model, guiding its responses by showing the desired outcome.

Earlier, we discussed the auxiliary element "Example" in the CoSeed factors. While seemingly supportive, its impact is extraordinary.

Shot prompting involves showing ChatGPT examples to help it learn specific response patterns or problem-solving methods. These examples, however, are carefully designed by the user and are not randomly generated.

In contrast, Chain-of-Thought (COT) prompting requires ChatGPT to think independently and demonstrate the entire thought process from problem to solution. It's like guiding ChatGPT through a maze on its own, requiring it to display the path it took to arrive at an answer.

Both Shot prompting and Chain-of-Thought prompting are strategies to enhance ChatGPT's ability to understand and answer complex questions.

With Shot prompting, ChatGPT learns the context of a question and the expected response format through examples.

Meanwhile, Chain-of-Thought prompting guides ChatGPT through a continuous thought process, showing how it arrives at an answer step-by-step. This method is ideal for problems that require complex reasoning, asking ChatGPT to demonstrate each step in its thought process to solve the problem.

The goal of both methods is the same: to make ChatGPT smarter and more capable of thinking like a human.

Through Shot prompting or Chain-of-Thought prompting, we are teaching ChatGPT the art and technique of "thinking." Thus, they are discussed together.

1. Shot Prompting

Shot prompting can be categorized into: Zero-Shot, One-Shot, and Few-Shot

prompting.

This technique allows ChatGPT to attempt tasks without or with specific training. It's particularly important for enhancing ChatGPT's creativity, translation, and Q&A abilities.

As illustrated in Figure III.13, Zero-Shot prompting is like giving ChatGPT a creative challenge, asking it to create or solve problems in a completely new scenario, such as creating new content or answering abstract questions.

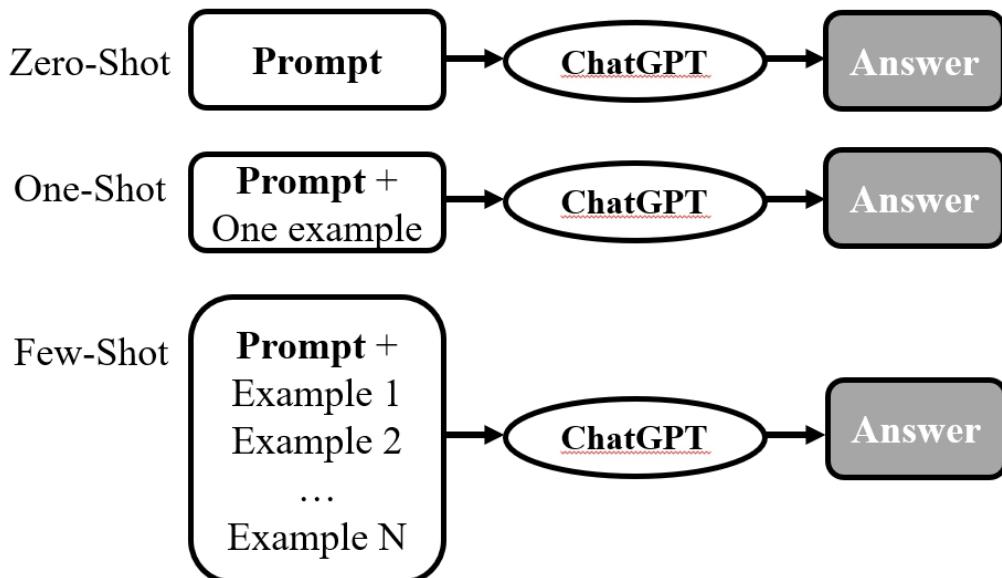


Figure III.13: Zero-Shot, One-shot and Few-shot prompts.

One-Shot prompting is akin to giving ChatGPT a specific subject to deeply understand the task, such as referencing a particular translation style or mimicking a writing style.

Few-Shot prompting involves presenting a few examples, allowing AI to make precise judgments even with limited information, useful for high-quality Q&A, tailored recommendations, or in-depth analysis.

1) Zero-Shot Prompt

Zero-Shot prompting is our most direct interaction with ChatGPT. When asking ChatGPT to respond without any examples, such as in a conversation, you are in fact using Zero-Shot prompting. This is typically the initial method

when we first use ChatGPT.

Here are two demonstrations of Zero-Shot Prompts in Figure III.14 and III.15. The two prompts are shown in the figures.

Demonstration 1 (Figure III.14):

Extracting Keywords from a given text without example.

Please extract keywords from the following and list them:
'Photosynthesis is a process used by plants, algae, and certain bacteria to convert light energy into chemical energy. This process takes place in the leaf filaments, using sunlight, carbon dioxide, and water to produce glucose and oxygen as by-products.'



Keywords: Photosynthesis, plants, algae, bacteria, light energy, chemical energy, leaf filaments, sunlight, carbon dioxide, water, glucose, oxygen, by-products.

Figure III.14: Zero-Shot example – extract keywords from text.

Demonstration 2 (Figure III.15):

Classifying a given text without examples.

Please categorize the following text as related to science, technology, education, or tourism, based on its content:

Text: 'Recently released smartwatches integrate the latest health monitoring systems, including heart rate monitoring, blood oxygen measurement, and sleep tracking functions. In addition, they support a variety of sports modes, allowing users to better monitor their own fitness activities.'



Category: Technology

Figure III.15: Zero-shot example - classify a given text

In these demonstrations, no example was provided with the prompts—this is Zero-Shot prompting.

If Zero-shot prompting doesn't yield satisfactory answers, you can start providing an example with the prompt. The next section demonstrates One-Shot prompting.

2) One-Shot Prompt

One-Shot prompting involves showing ChatGPT a relevant example with the prompt.

Providing a single example or "Shot" helps ChatGPT understand the expected response format or context. ChatGPT will learn from the example and execute the task it was given.

This method is especially useful for tasks with specific format or contextual requirements. It's slightly more complex than Zero-Shot prompting but remains simple and quick to use, allowing ChatGPT's responses to be more focused and precise.

The following two figures are two demonstrations of One-shot prompt.

Demonstration 1 (Figure III.16):

One-shot prompt - write email with a polite and professional tone. But ChatGPT may not have the same idea of the tone. Hence, an example is provided to guide ChatGPT.

You are a customer service representative. Respond to the following email in a polite and professional tone.

Example Email: "Dear Customer Service, I am extremely frustrated with your product. It stopped working after just two days, and I want a full refund immediately!"

Example Response: "Dear [Customer], I'm sorry for the trouble you've had with our product. Please send us your order details, and we'll process your refund right away. We apologize for the inconvenience."

***Now respond to this email in the same tone: "Dear Support Team, the software I purchased keeps crashing every time I try to run it. It's causing significant delays in my work, and I'm considering switching to a different product.



Dear [Customer],

I'm very sorry to hear about the trouble you're experiencing with our software. I understand how important it is for you to have a reliable product, especially when it's affecting your work. Please share your order details with us, and we'll do everything we can to resolve this issue quickly. If the problem continues, we can explore other options to ensure your satisfaction. We apologize for any inconvenience this has caused.

Best regards,

[Your Name]

Customer Support Team

Figure III.16: One-Shot example - write email with a polite and professional tone.

In Demonstration 1, by providing a single example of the desired outcome, ChatGPT was able to understand the tone required for the email and generated a response that matched the given example.

Demonstration 2 (Figure III.17):

One-Shot prompt - make a creative analog. Again, ChatGPT may not know what's your definition of "creative". Hence, an example is provided.

O Create a creative analogy to explain a complex concept in simple terms.

Example Concept: "The Internet"

Example Analogy: "The Internet is like a giant library where every book ever written is instantly available. Instead of walking through aisles, you type what you're looking for, and the perfect book appears on your screen."

***Now create an analogy for this concept: "Machine Learning"



Machine learning is like teaching a dog to fetch. At first, you show the dog how to bring the ball back by giving it treats when it does it right. Over time, the dog learns on its own to fetch the ball without needing as much guidance. Similarly, in machine learning, we train a computer by feeding it lots of data and examples. Eventually, it starts to recognize patterns and make decisions on its own, just like the dog learning to fetch.

Figure III.17: One-Shot example – make a creative analog.

In this demonstration, by offering a specific example of what you deem "creative," ChatGPT aligns its response with your expectations. The example guides the AI, helping it generate an analogy that meets the desired creativity, ensuring the output is both relevant and imaginative.

In both demonstrations, ChatGPT's responses met the requirements.

Zero- shot and One-shot Prompts enable ChatGPT to learn on its own, with no or just one example. For more complex tasks, however, these may not suffice to help ChatGPT grasp the full picture.

After understanding basic Zero-Shot and One-Shot prompting, we move to more advanced techniques like Few-Shot and Chain-of-Thought prompting, which require multiple examples or a logical sequence of steps to guide the AI's responses.

3) Few-Shot Prompting

Few-Shot prompting involves providing ChatGPT with several examples (a small number), enabling it to learn within a specific context. Even with only a small amount of information, ChatGPT can effectively learn and adapt to new tasks, enhancing its practicality and flexibility.

Here are two demonstrations.

Demonstration 1 (Figure III.18):

Few-Shot prompt - Convert the following technical descriptions into simple, layman's terms.

Convert the following technical descriptions into simple, layman's terms for a non-technical audience.

Example 1: Technical: 'The CPU (Central Processing Unit) executes instructions from the software by performing basic arithmetic, logical, control, and input/output operations.' Layman's: 'The CPU is like the brain of your computer, handling all the basic tasks and instructions needed to make your software work.'

Example 2: Technical: 'An API (Application Programming Interface) allows different software applications to communicate with each other by defining a set of rules and protocols.' Layman's: 'An API is like a translator that helps different apps talk to each other and work together smoothly.'

***Now convert this description: Technical: 'A firewall monitors and controls incoming and outgoing network traffic based on predetermined security rules to protect the system from unauthorized access.



Layman's: "A firewall acts like a security guard for your computer, watching the traffic that goes in and out to make sure nothing unsafe gets in or out."

Figure III.18: One-Shot example – simplify technical descriptions.

This demonstration shows how a few examples of simplified technical descriptions help ChatGPT grasp the desired complexity and tone. The Few-Shot approach allows the model to generalize and effectively translate technical jargon into accessible language.

Demonstration 2 (Figure III.19):

Few-Shot prompt - Generate a creative and catchy product tagline.

Generate a creative and catchy product tagline for the following products.

Example 1: Product: Eco-friendly water bottle

Tagline: "Sip Sustainably, Drink Responsibly."

Example 2: Product: Noise-cancelling headphones

Tagline: "Tune In to What Matters, Tune Out the World."

***Now create a tagline for this product:

Product: Smart fitness tracker.



Tagline: "Track Your Progress, Empower Your Potential."

Figure III.19: One-Shot example – Generate a product tagline.

Here, the Few-Shot prompt trains ChatGPT to create catchy taglines. By observing examples, ChatGPT learns effective patterns and styles, enabling it to generate engaging, brand-aligned outputs.

In these demonstrations, ChatGPT learns to create and answer questions on its own after observing the given examples.

Few-shot prompting introduces new capabilities for ChatGPT to execute specific tasks while also presenting challenges.

The quality of examples is crucial—poor examples may lead to incorrect learning. Just as a chef trained only in American cuisine can't master Chinese dishes, ChatGPT might excel only in areas it has encountered before.

In short, Few-Shot prompting is a great tool, but to use it effectively, you need to be patient and wise in designing good examples, guiding ChatGPT step by step.

Moreover, we are often curious about how ChatGPT thinks and arrives at answers. However, making it explain its entire thought process is sometimes more challenging than getting it to give the answer, leading to the concept of "Chain-of-Thought prompting."

2. Chain-of-Thought Prompting

One fascinating aspect is that we all think ChatGPT is amazing, with seemingly limitless capabilities to answer various questions and handle complex logic. However, how exactly does ChatGPT "think"?

In reality, ChatGPT's thinking differs from human thinking. It is based on machine algorithms, analyzing and learning from large amounts of data to recognize patterns, make predictions, or interpret information. Thus, ChatGPT's "understanding" is actually a result of statistical analysis of data, not true "cognition."

Simply put, ChatGPT's decision-making process is like a black box to us. While we can control the input data and observe the output, we cannot see how ChatGPT internally processes and manages the information.

COT prompting is an advanced technique designed to address this black box problem by guiding ChatGPT to break down its reasoning process step by step. This method not only requires ChatGPT to provide answers but also to demonstrate how it arrived at those answers, making its thought process transparent and understandable—especially useful for educational purposes or complex problem-solving tasks.

COT prompting emerged to tackle challenges in tasks that require logical reasoning, like mathematical problems, common-sense logic, and symbolic operations. By generating intermediate reasoning steps, COT significantly improves ChatGPT's response accuracy, making its decision-making process more transparent and reliable.

The concept of Chain-of-Thought was introduced by Jason Wei et al. in the paper "Chain-of-Thought prompting Elicits Reasoning in Large Language Models." The study showed that guiding ChatGPT through a clear reasoning process improves its ability to handle complex reasoning tasks significantly.

For example, consider the following prompt as shown on the left-hand side of Figure III.20: "The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?"

Initially, early version like ChatGPT-3.5, struggled with such problems. However, by applying CoT prompting, where ChatGPT is asked to explain its reasoning step by step, the model's accuracy in solving these types of problems greatly improves.

Figure III.20, taken from Jason Wei et al.'s paper, shows that with a direct prompt, ChatGPT makes mistakes even in simple calculations as shown on the left-hand side of the figure. However, the illustration on the right shows that by providing a Chain-of-Thought example first, ChatGPT does arrive at the correct answer.

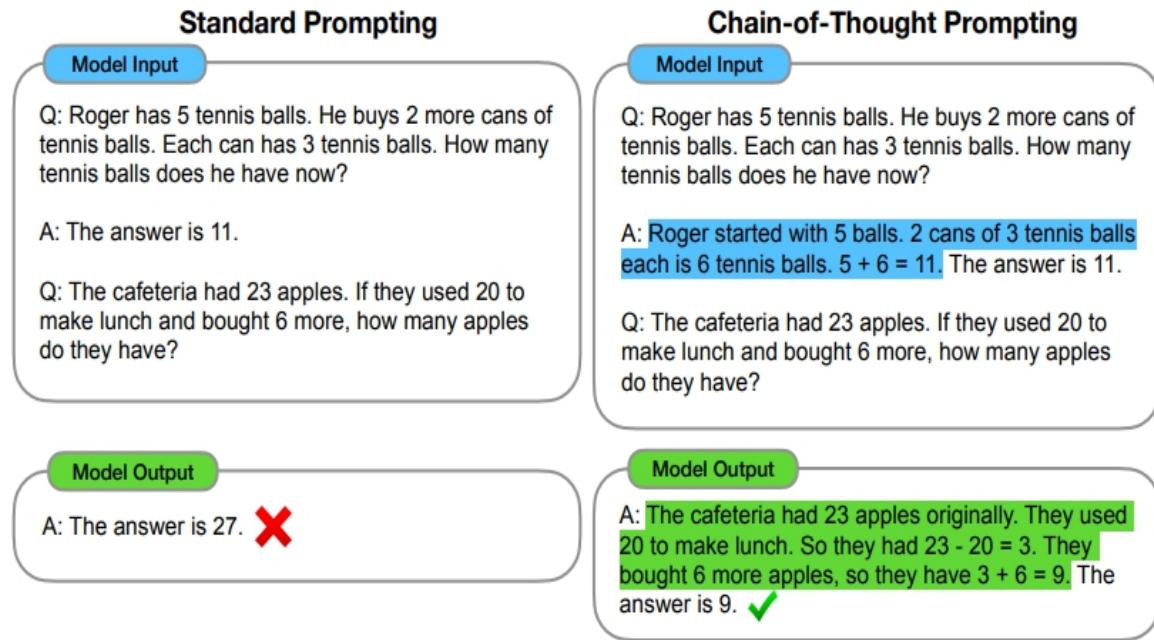


Figure III.20: ChatGPT made error. Answer on the right is correct after giving a COT example.

As ChatGPT's math and reasoning capabilities improve and newer versions emerge, such math problems become less frequent.

To check if ChatGPT still benefits from COT, we asked ChatGPT in Figure III.21. It replied that it had already learned COT reasoning but sometimes still benefits from a reminder.

Do I need to give you any indication if I want a Chain-of-Thought response?



No, you don't need to give any special indication. If you prefer a Chain-of-Thought response, you can simply ask for a detailed explanation or reasoning process, and I'll automatically provide that type of response. If you have any specific preferences for how you'd like the reasoning to be presented, feel free to mention them!

Figure III.21: How to get ChatGPT to provide COT response.

In fact, in September of 2024, when this book is being released, OpenAI launched its ChatGPT-o1 which incorporated the Chain-of-Thought technology so it has the ability to "think" before responding. However, o1 is 3 or 4 times more expensive than the previous models.

Hence, if you are still using ChatGPT 3.5, 4 or 4o, you can use the COT technique in your prompt and get the same result as that of the o1.

How do you remind ChatGPT to use CoT? Simply ask for a detailed explanation or reasoning process. You will just simply add a phrase like "Let's think step by step" or "Please show your reasoning process" in the prompt and the results will be more accurate and give out more insights.

For developers, it's a powerful debugging tool that can reduce ChatGPT's hallucinations. Having ChatGPT explain its thought process can also help identify and correct errors, enhancing ChatGPT's accuracy.

As ChatGPT's usage increases, being able to trust and understand its decision-making is crucial. Chain-of-Thought requires ChatGPT to articulate its reasoning process, making the system more transparent, accountable, and ultimately, more trustworthy.

Figure III.22 below provides an example of using the Chain-of-Thought technique. ChatGPT's response demonstrates its thought process and final result.

Prompt:

"Write a job description for a customer service representative position. Due to space constraints, please summarize the details. Explain your thought

process step by step."

This Chain-of-Thought example in Figure III.22 shows how ChatGPT breaks down writing a job description into logical steps. By explaining its process, ChatGPT ensures the summary is concise, comprehensive, and the reasoning behind each decision is clear.



Thought Process

1. **Core Responsibilities:** Focus on the essential tasks such as handling customer inquiries, providing support, and resolving issues.
2. **Required Skills:** Highlight critical skills like communication, problem-solving, and empathy.
3. **Qualifications:** Emphasize minimum qualifications like education level and experience.
4. **Company Culture:** Briefly mention the company environment or values.
5. **Conciseness:** Keep the description short, focusing only on the key points.

Job Description

Position: Customer Service Representative

Summary: Provide prompt, courteous support to customers, addressing inquiries and resolving issues efficiently. Strong communication and problem-solving skills required.

Minimum qualifications include a high school diploma and prior customer service experience. Join a dynamic, customer-focused team committed to delivering exceptional service.

Figure III.22: ChatGPT response: COT prompt example—generate a job description.

Let's look at another example (Figure III.23) of writing a novel using "step-by-step thinking" to accomplish the task.

Prompt:

" I'd like to write a short love story. Please 'think step by step' about how to accomplish this task."

In this example (Figure III.23), the Chain-of-Thought approach guides ChatGPT in outlining and developing a short love story. By thinking step by step, ChatGPT ensures each narrative element is logically connected, resulting in a well-structured story. This method also allows easy refinement before combining into a complete narrative.

With the big steps outlined, we can now have ChatGPT execute each step one by one, ultimately combining them into a complete novel.



1. **Theme & Message:** Define the central theme (e.g., first love, forbidden love) and the key message you want to convey.
2. **Character Creation:** Develop distinct main characters with unique backgrounds, and motivations. Consider their chemistry and dynamics.
3. **Setting:** Choose a location and time that enhances the mood (e.g., a quiet town, a bustling city).
4. **Plot Structure:**
 - **Introduction:** Set the stage, introducing characters and setting.
 - **Inciting Incident:** Introduce the event that sparks the love story.
 - **Rising Action:** Build the relationship with challenges or conflicts.
 - **Climax:** Reach the turning point with a key decision or revelation.
 - **Falling Action:** Show the aftermath of the climax.
 - **Conclusion:** Resolve the story with a satisfying or reflective ending.
5. **Infuse Emotion:** Use dialogue and narration to convey deep emotions and create vivid scenes.
6. **Edit & Refine:** Review, get feedback, and polish for clarity and impact.
7. **Finalize:** Choose a fitting title and decide how to publish or share your story.

Figure III.23: ChatGPT response: COT prompt example— write a love story.

In summary, Chain-of-Thought prompt technique allows us to gain deeper insights into ChatGPT and our own thought processes.

While Chain-of-Thought prompting can offer significant benefits, it is not a one-size-fits-all solution. We, as humans, should wisely use step-by-step thinking, learning this method of breaking down steps and applying it flexibly when using generative AI.

Chapter 12 Summary:

- ❖ This chapter discusses revolutionary topics in prompting: “Shot prompting” and “Chain-of-Thought prompting.”
- ❖ The effectiveness of the “Example” auxiliary element in the CoSeed framework is significant, forming the basis for Shot prompting.
- ❖ Shot prompting involves providing ChatGPT with carefully designed examples, guiding it to learn specific response patterns and problem-solving methods.
- ❖ Shot prompting, which includes Zero-Shot, One-Shot, and Few-Shot Prompts, enhances ChatGPT's ability to respond accurately and creatively in various contexts. ❖ Chain-of-Thought prompting requires the model not only to provide answers but also to explicitly outline its reasoning process step by step. This technique is particularly suited for solving complex problems that require logical and sequential thinking, ensuring the model's responses are transparent and easy to understand.
- ❖ Sometimes, executing Chain-of-Thought prompting only requires adding a phrase like “Let's think step by step” or “Please show your reasoning process.”
- ❖ Chain-of-Thought prompting offers deeper insights into ChatGPT's reasoning. By mastering this method, we can enhance the accuracy and transparency of generative AI, particularly when tackling intricate tasks.

Chapter 13: Application and Practice (2)

Now that we understand basic prompts, let's apply these concepts to refine commonly used prompts and make them more effective.

We have compiled some commonly used prompts for practice. In Chapter 11, Application and Practice (1), we explored nine example prompts focused on marketing and strategy. Here, we've collected prompts for everyday life, categorized into seven areas: "Food," "Shopping," "Home," "Exercise," "Transportation," "Travel," and "Socializing."

However, directly using these prompts often leads to vague answers due to a lack of specificity.

To address this, we'll use the techniques we've learned to improve these prompts. We'll apply the "3W-CoSeed Framework," "Shot prompting," or "Chain-of-Thought prompting" to make ChatGPT's responses more precise and relevant.

Each category contains four prompts. We'll demonstrate how to refine the first two, leaving the other two as exercises for you.

By now, you should have a good grasp of ChatGPT and various prompting techniques, so we won't include the actual answers from ChatGPT. Instead, you can input both the original and improved prompts into ChatGPT and compare the responses yourself. We believe this exercise will provide valuable insights.

1. Food Category

1) Food Consultant

Original Prompt:

"What good restaurants can you recommend in Taipei?"

Suggestion:

The task is clear, but the question is too broad. Let's refine it using the "3W-CoSeed framework." The person asking likely has a specific location or cuisine preference. We'll add the 3Ws: Who (Role), Where (Context), and What (Task) to make the answer more tailored.

Improved Prompt:

"You are a food critic familiar with Taiwan's restaurants [Role]. I am looking for good Italian restaurants in Taipei's Da'an District [Context]. Please recommend three restaurants that meet these criteria [Task]."

2) Meal Planning

Original Prompt:

"I'm planning a dinner party. Can you give me some suggestions?"

Suggestion:

The scope of this prompt is too broad, leading to unfocused responses. Let's use the "3W-CoSeed Framework" to refine it by specifying Who (Role), Where (Context), and What (Task). Since a dinner party likely involves a budget and number of guests, adding these constraints will help. Also, having the final menu in list form would be visually appealing, so let's modify it accordingly.

Improved Prompt:

"You are a chef with 20 years of experience in Chinese cuisine [Role]. I am planning a dinner party [Context] for 10 guests with a budget of 1,000 TWD per person [Constraint]. Please design three different Chinese menus for the event [Task] and list them out [Output]."

Practice Prompts, try to improve them:

- 3) "I'm interested in vegetarian diets. Can you suggest some vegetarian recipes?"
- 4) "I want to try some exotic cuisines. Help me find recipes from around the world."

After discussing food-related prompts, let's explore shopping prompts and refine them for better results.

2. Shopping Category

1) Shopping Advice

Original Prompt:

"Please provide strategies for practicing eco-friendly shopping, including how to choose sustainable products, reduce packaging waste, and support eco-friendly brands."

Suggestion:

Since this prompt involves strategy, let's use the "Chain-of-Thought prompting" technique to refine it.

Improved Prompt:

"Please 'think step by step' [Chain-of-Thought] to provide strategies for practicing eco-friendly shopping, including how to choose sustainable products, reduce packaging waste, and support eco-friendly brands."

Note: Try using this prompt, observe ChatGPT's response, then remove "think step by step" and ask again. Compare the two responses and see which one you find better.

2) Price Comparison**Original Prompt:**

"Explain how to effectively use consumer reviews to compare shopping prices, including looking for insights on product quality and value."

Suggestion:

This prompt is clear, but it could be improved by narrowing it down to a specific product type and using the "Chain-of-Thought" technique.

Improved Prompt:

"When purchasing tech products like smartphones [Constraint], explain how to effectively use consumer reviews to compare prices [Task]. This includes analyzing how to distinguish between genuine and fake reviews, and understanding how different review sites' standards might influence product ratings [Action], leading to deeper insights into product quality and value [Goal]. Please 'think step by step' [Chain-of-Thought]."

Practice Prompts, try to improve them:

- 3) "Please compare the pros and cons of online shopping versus in-store shopping, including convenience, pricing, product range, and consumer experience."
- 4) "I'm looking for a new computer. Can you help me compare prices and find the best deal?"

After refining shopping strategies, let's shift focus to creating a comfortable living space. Next, we'll explore prompts related to home design and

improvement.

3. Home Category

1) Home Design

Original Prompt:

"I'm planning to redecorate my living room. Do you have any suggestions or design ideas?"

Suggestion:

This prompt needs refinement; otherwise, multiple iterations may be required to get a satisfactory answer. At a minimum, you should provide some background about the living room and budget. Let's apply the "Few-Shot prompting" technique to enhance it.

Improved Prompt:

"I'm planning to redecorate my living room [Context], which is 50 square meters in size, with a budget of 300,000 TWD for furniture and design costs [Constraint]. I like {Example 1} and {Example 2} design styles.

Combine elements of these two styles to create a cohesive design concept for my living room, and provide specific ideas for furniture and decor.

{Example 1} = Bohemian Style

{Example 2} = Scandinavian Style."

2) Creative Design

Original Prompt:

"I want to create an educational playroom for my child. Do you have any suggestions or design ideas?"

Suggestion:

We can use the "One-Shot prompting" technique to refine this prompt.

Improved Prompt:

"I want to create an educational playroom for my child [Context], and I hope this playroom will be similar to {Example}. My child is a 7-year-old boy [Role]. Do you have any suggestions or design ideas [Task]?

{Example} = The Exploratorium's Tinkering Studio in San Francisco [One-Shot Example]."

Practice Prompts, try to improve them:

- 3) "I want to set up a home office, but I'm not sure how to make the best use of the space. Can you give me some suggestions?"
- 4) "I'm planning a family gathering, and I need some suggestions on decor and arrangements. Can you help me?"

Now that we've covered home enhancements, it's time to focus on personal well-being. In the following section, we'll delve into exercise prompts to help you create a healthier lifestyle and personalized fitness plans.

4. Exercise Category

1) Health Consultation

Original Prompt:

"I'm thinking of starting a healthy lifestyle. Can you give me some recommendations?"

Suggestion:

Let's use the "3W-CoSeed Framework," starting by assigning ChatGPT a professional role. Also, a healthy lifestyle is closely related to age and gender. Moreover, "lifestyle" is too broad; it's better to focus on diet first, then follow up with other questions.

Improved Prompt:

"You are a fitness coach with 20 years of experience [Role]. I am a 35-year-old female elementary school teacher [Role]. I'm thinking of starting a healthy lifestyle [Context]. Can you give me some dietary suggestions [Task]?"

2) Fitness Proposal

Original Prompt:

"I want to create a workout plan that suits my fitness goals. Can you give me some suggestions?"

Suggestion:

Again, the question lacks focus, and the answer won't be targeted. Let's use the "3W-CoSeed Framework" to improve it.

Improved Prompt:

"You are a renowned health and nutrition expert [Role]. I am a 36-year-old male, 175 cm tall, weighing 86 kg [Role]. I have flat feet [Constraint]. My goal is to lose 6 kg within a year [Goal]. Please create a personalized workout plan based on the information I've provided, including all the details [Task]. Please present the plan in a detailed table format [Output]."

Practice Prompts, try to improve them:

- 3) "I want to start running, but I'm not sure how to begin. Do you have any suggestions or a training plan?"
- 4) "I want to learn more about healthy living and exercise. Can you provide me with some useful information or resources?"

With your fitness routine in place, let's turn to how you get around. The next section will explore transportation prompts, refining questions on car purchases, maintenance, and more to optimize your choices."

5. Transportation Category

1) Car Purchase

Original Prompt:

"I'm considering buying a car. Can you help me compare the pros and cons of different brands and models?"

Suggestion:

This prompt is easy to understand but only mentions the need to compare different brands and models without specifying what aspects to compare, such as performance, price, and safety. Let's use the "3W-CoSeed Framework" to improve it.

Improved Prompt:

"I'm considering buying a car, primarily for city driving [Context]. My budget is around 1 million TWD [Constraint]. Please help me compare cars that meet my budget and usage needs based on performance, price, safety, fuel efficiency, and interior space [Task]. Please present the comparison in a table format [Output]."

2) Maintenance Advice

Original Prompt:

"I want to learn more about car maintenance and repair. Can you give me

some advice?"

Suggestion:

While easy to understand, this prompt lacks specificity and feasibility. It can be improved using the "3W-CoSeed Framework." Also, since maintenance and repairs have a sequence, you can add "Chain-of-Thought prompting."

Improved Prompt:

"I want to learn more about common car maintenance and basic repair knowledge [Task], especially for my Toyota Corolla 2020 model [Constraint]. Please give me some advice, including regular check-ups and how to perform simple fault diagnosis and repairs [Task]. My primary goal is to handle basic issues myself and communicate better with mechanics when professional repairs are needed [Goal]. Please 'think step by step' [Chain-of-Thought]."

Note: Try using this prompt, observe ChatGPT's response, then remove "think step by step" and ask again. Compare the two responses and see which one you find better.

Practice Prompts, try to improve them:

- 3) "I'm curious about the new technologies and smart features in the car market. Can you help me learn about some new automotive tech?"
- 4) "I'm planning a road trip, but I'm not familiar with the destination's roads. Can you give me some navigation and driving route suggestions?"

Having discussed your commute, it's time to think bigger—planning your next vacation. This section will refine prompts to help you find the perfect destinations, create itineraries, and maximize your travel experience."

6. Travel Category

1) Travel Advice

Original Prompt:

"I'm looking for a family-friendly vacation destination. Can you help me find some suitable resorts or vacation villas?"

Suggestion:

This prompt is clear and easy to understand but lacks enough details to generate more specific suggestions or creative solutions. You can use the

"3W-CoSeed Framework" to improve specificity, feasibility, and inspiration.

Improved Prompt:

"I'm looking for a family-friendly vacation destination [Goal], particularly for families with young children (ages 6 and 10) [Context]. We prefer destinations in Europe or Asia, with a budget of around 3,000 to 4,000 USD [Constraint]. I hope to find resorts or vacation villas that offer family activities like swimming, adventure activities, and cultural experiences [Goal]. Can you suggest options that meet these criteria [Task]?"

2) Travel Planning

Original Prompt:

"I'm looking for a beach resort. Can you provide some advice on how to plan a vacation?"

Suggestion:

This prompt identifies the need for a vacation plan. You can use "Chain-of-Thought prompting" along with some elements of the "3W-CoSeed Framework" to help with planning.

Improved Prompt:

"You are an experienced travel consultant [Role]. I'm looking for a beach resort [Task]. Please help me plan this trip, including: 1) Choosing the destination, 2) Itinerary, 3) Budget. Please 'think step by step' [Chain-of-Thought]."

Note: Try using this prompt, observe ChatGPT's response, then remove "think step by step" and ask again. Compare the two responses and see which one you find better.

Practice Prompts, try to improve them:

- 3) "I'm planning a vacation, but I'm not sure how to pack. Can you give me some suggestions?"
- 4) "I'm going to a new city for travel, but I don't know how to arrange my itinerary and find local attractions. Can you provide me with some travel information and advice?"

Finally, after planning trips, let's focus on connecting with others. The last section will guide you through prompts related to socializing, enhancing

questions to improve your social skills and build meaningful connections.

7. Socializing

1) Social Challenges

Original Prompt:

"I've been having some trouble with socializing lately. Can you give me some advice on how to improve my social skills and overcome social challenges?"

Suggestion:

The prompt is clear but lacks enough detail to generate focused responses. You can use the "3W-CoSeed Framework" to improve this prompt.

Improved Prompt:

"Recently, I've found myself struggling in social situations at work, especially when trying to express my opinions in team meetings [Context]. I want to improve my communication skills and confidence to overcome these challenges [Goal]. Can you give me some advice [Task]? I am a 30-year-old software engineer who doesn't socialize much outside of work [Role], but I hope to build stronger relationships in my career [Goal]."

2) Socializing and Making Friends

Original Prompt:

"I want to expand my social circle, but I don't know where to start. Can you give me some advice?"

Suggestion:

Let's use the "3W-CoSeed Framework" and "Chain-of-Thought prompting" together to create a step-by-step plan.

Improved Prompt:

"I want to expand my social circle and meet more people who share my interest in hiking [Goal]. I live in Taipei and hope to find some local groups or activities [Context], but I'm not sure where to start. Can you give me some advice [Task]? My goal is to meet new friends during my leisure time and build lasting connections through shared interests [Goal]. Please 'think step by step' [Chain-of-Thought] on how to accomplish this task."

Note: Try using this prompt, observe ChatGPT's response, then remove

"think step by step" and ask again. Compare the two responses and see which one you find better.

Practice Prompts, try to improve them:

- 3) "I feel anxious about my performance in the workplace or social situations. Can you give me some tips on how to build confidence and communicate effectively?"
- 4) "I want to better understand the importance of social interactions and how to handle challenges in social situations. Can you provide some relevant information and insights?"

Part Four: Strategies – Advanced Prompts

This book introduces prompt engineering for large language models (LLMs) like ChatGPT, which we privately call "The Art of Prompts" as it merges technical methods with creative planning.

Prompt engineering is structured into Fundamentals, Basic, and Advanced, each corresponding to a part of this book: "Principles," "Tactics," and "Strategies."

In Part Three, "Tactics," we covered basic techniques such as framework and thought prompts, using adaptable formulas for specific contexts.

Now, in Part Four, we delve into "Strategy," which involves holistic thinking and planning steps for ChatGPT to execute effectively.

If you've used ChatGPT, you may notice it sometimes misunderstands intent, takes shortcuts, or makes assumptions, highlighting the need for human guidance.

While ChatGPT is powerful, human strategic thinking enhances its effectiveness. A key strategy is "task decomposition," breaking complex tasks into smaller sub-tasks, reducing complexity, making it easier to select suitable strategies for each sub-task.

By breaking tasks into manageable parts, errors are easier to spot and fix, enhancing response quality and speed. The "Task Execution Principle," introduced in Chapter 7, aligns with "task decomposition," involving breaking down tasks, managing sub-tasks, evaluating, and iterating, as shown in Figure II.03.

These three stages are vital in advanced prompt techniques, often used to help ChatGPT perform tasks independently, as seen in various examples.

Chapter 14 examines ChatGPT's ability to execute computer logic, while Chapters 15 to 18 introduce advanced prompt techniques used by major

companies to boost LLM efficiency, applying these to everyday challenges.

Chapter 14: Executing Logic

It is known that computers operate like a black box, where data input at one end, processed through internal logic and algorithms, and results output at the other end.

Just like programming languages, prompts can execute logic and calculations, which is highly beneficial for designing advanced prompts.

ChatGPT can perform many logical operations found in programming, including iteration, decision-making, and functions. Applying these logical operations in prompts can be useful in various scenarios.

1. Executing Iteration

As mentioned earlier, effective prompts often result from multiple iterations. This repetitive process is called "iteration" in programming.

When interacting with ChatGPT, there's no need to aim for perfection right away. Start with a decent prompt, evaluate, and iterate to refine and improve it. This prompt technique is known as "Ask Me Anything," where you test an initial prompt, observe the response, and iteratively refine it for better results.

In other words, quickly test an imperfect prompt, observe the results, and iteratively refine and improve the original prompt based on the feedback until you achieve a satisfactory outcome.

This process emphasizes quick experimentation and continuous improvement rather than trying to design the perfect prompt from the beginning. This iterative process helps in gradually understanding and enhancing the effectiveness of the prompts.

Let's look at an example of using this "Ask Me Anything" process to request healthy eating suggestions from ChatGPT: Start with a simple prompt. If the result meets your expectations, consider it "good enough." If it needs improvement, refine it.

First Iteration:

Prompt 1:

"Please suggest some healthy foods."

This prompt is straightforward but lacks specific direction or constraints.

Result: You might get a broad list of healthy foods like fruits, vegetables, meats, etc., but it may lack specific advice or in-depth information.

Second Iteration:

Refine the prompt by adding constraints to make the output more specific.

Prompt 2:

"Please suggest a healthy breakfast, including 'main course, protein, and beverage.' "

Result: The focus is on breakfast, particularly the main course, protein, and beverage, but it still lacks information tailored to the individual's health status.

Third Iteration:

Further refine the prompt by adding personalized details (Who) and a directive for balanced nutrition (Constrain).

Prompt 3:

"I am 48 years old and weigh 65 kg. Please suggest a nutritionally balanced healthy breakfast, including 'main course, protein, and beverage,' and briefly explain the benefits of each type."

Final Result:

As shown in Figure IV.01, we receive a comprehensive, nutritionally balanced breakfast suggestion with detailed nutritional explanations.



Main Course:

- **Oatmeal with Berries & Almonds:** Oatmeal's fiber lowers cholesterol and stabilizes blood sugar. Berries add antioxidants; almonds offer healthy fats.

Protein:

- **Greek Yogurt with Honey & Chia Seeds:** Greek yogurt is protein-rich, aiding muscle maintenance. Chia seeds add omega-3s, fiber, and more protein.

Beverage:

- **Green Tea:** Green tea's catechins are antioxidants that may lower heart disease risk and boost metabolism.

This meal provides balanced nutrients to energize your morning.

Figure IV.01: Healthy and balanced breakfast after two iterations.

This demonstrates the decision-making iteration process. Next, we'll see how to let ChatGPT automatically perform iterations until a decision condition is met.

2. Executing Decisions

You may know that ChatGPT can handle multiple questions, but you can also set conditions or passwords in prompts. ChatGPT will start or stop a task when it detects the password.

Example 1: Personal Assistant

Prompt:

" You are my personal assistant. I will occasionally send you my to-do items and appointments. When receive a message, just reply 'Received.' When I type 'Summarize,' you will summarize all content in chronological order into To-Do and Appointment."

You are my personal assistant. I will occasionally send you my to-do items and appointments. When receive a message, just reply 'Received.' When I type 'Summarize,' you will summarize all content in chronological order into To-Do and Appointment.



Memory updated

Received.

Dentist Appointment tomorrow 9 AM



Received.

Finish Quarterly Financial Report in 3 days



Received.

Dinner with Chairman Chen at "The Four Seasons Restaurant",
7PM tonight



Received.

Summarize



To-Do Items:

1. Finish Quarterly Financial Report in 3 days.

Appointments:

1. Dinner with Chairman Chen at "The Four Seasons Restaurant," 7 PM tonight.
2. Dentist Appointment tomorrow at 9 AM.

Figure IV.02: Personal Assistant executing decision

In Figure IV.02, ChatGPT successfully organizes a report upon receiving the "Summarize" command.

Example 2: Handling Word Count Limits

ChatGPT often has input word limits, a constraint set by OpenAI. However, ChatGPT can handle more input internally.

To confirm this, we asked ChatGPT: "Is there a limit to the number of words I can input at once?"

ChatGPT 4o: "There is a practical limit to the number of words you can input at once, typically around 4,000 to 6,000 words (about 16,000 to 24,000 characters), depending on the complexity of the text. If your input exceeds this, it may be truncated. If you have a large document, you can break it into smaller sections and input them separately."

If your text exceeds ChatGPT's input limit, split it into smaller sections. Use a trigger word like 'Start' to signal ChatGPT to combine and summarize the sections."

Below, we input two text sections each is shorter than the maximum that ChatGPT can take. The purpose is to demonstrate how to split long texts into sub-text, send them into ChatGPT, ask ChatGPT to combine them and then execute a task.

In practical situation, each section can be as long as ChatGPT's limit allows (4,000 to 6,000 words), and you can use more than two sections if needed.

We input two text sections separately, then requested a summary of the combined text. The text is from OpenAI's website, introducing Sora, software that generates videos from text.

Prompt:

"I will input several sections of text to you. When you receive them, simply reply 'Received.' When I send 'Start,' combine these text sections and 'summarize' their content."

Figure IV.03 demonstrates how to use prompts to combine multiple text sections or documents and decide when to execute a task under our control.

To execute a different task, simply replace “summarize” with the desired task.

I will input several sections of text to you. When you receive them, simply reply 'Received.' When I send 'Start,' combine these text sections and 'summarize' their content.



Received.

We're teaching AI to understand and simulate the physical world in motion, with the goal of training models that help people solve problems that require real-world interaction.

Introducing Sora, our text-to-video model. Sora can generate videos up to a minute long while maintaining visual quality and adherence to the user's prompt.



Received.

Today, Sora is becoming available to red teamers to assess critical areas for harms or risks. We are also granting access to a number of visual artists, designers, and filmmakers to gain feedback on how to advance the model to be most helpful for creative professionals.

We're sharing our research progress to start working with and getting feedback from people outside of OpenAI and to give the public a sense of what AI capabilities are on the horizon.



Received.

Start



The text discusses the introduction of "Sora," a new text-to-video AI model developed to simulate the physical world in motion and help solve real-world interaction problems. Sora can generate high-quality videos up to a minute long based on user prompts. The model is currently being tested by red teamers to identify potential risks and by visual artists, designers, and filmmakers to gather feedback for further development. The initiative aims to engage external experts and the public to refine the model and provide insight into upcoming AI capabilities.

Figure IV.03: Combine multiple sections of text for ChatGPT to respond.

3. Executing Functions

Another common programming technique is the function, or subroutine. A function is a block of code designed to perform a specific task or function. It can be called multiple times by other programs. For example, a cubic function calculates the cube of a given number.

Let's look at a prompt that uses the function concept for more convenient usage:

Example 1:

Prompt:

"I'm going to ask you a question. I hope you can answer my <question> briefly, just like <person> would. You will have the same mental model, thought process, and tone as <person>.

<person>: Steve Jobs, co-founder of Apple.

<question>: "I dream of creating technological innovations that change the world, just like you did during your time at Apple. However, I'm struggling to realize my ideas. How should I turn my vision into reality?"

I'm going to ask you a question. I hope you can answer my <question> briefly, just like <person> would. You will have the same mental model, thought process, and tone as <person>."

<person>: Steve Jobs, co-founder of Apple.

<question>: "I dream of creating technological innovations that change the world, just like you did during your time at Apple. However, I'm struggling to realize my ideas. How should I turn my vision into reality?"



Steve Jobs: "Start with passion. Believe in your vision so deeply that nothing can shake you. Simplify everything—strip away the unnecessary until the core idea shines. Surround yourself with the best talent, people who share your vision. Fail fast, learn faster, and iterate constantly. Trust your intuition, and don't be afraid to take bold risks. Most importantly, stay hungry, stay foolish."

Figure IV.04: A prompt acts like a function.

This prompt design allows the main prompt to remain unchanged, while you only need to modify the function's content when changing the question.

In Figure IV.04, two functions are used: <person> and <question>. If you have a question in a specific area, find a famous person in that area and let them answer it.

For example, if you have a question about filmmaking for Ang Lee. You can just change the content of <person> and <question> as follows, without changing the rest of the prompt.

<person>: Ang Lee, renowned Oscar-winning director, and
<question>: "How do you choose your film subjects?"

If you frequently publish articles or videos on platforms like YouTube, the following example can be very helpful.

Example 2:

Prompt:

"Please give me <number> catchy titles based on <topic>. The titles are for creating videos to be published on YouTube."

<number>: 5; <topic>: Edge AI.

Please give me <number> catchy titles based on <topic>. The titles are for creating videos to be published on YouTube."
<number>: 5; <topic>: Edge AI.



Here are 5 catchy titles for YouTube videos on Edge AI:

1. "Unlocking the Power of Edge AI: The Future of Smart Devices"
2. "Edge AI Explained: Revolutionizing Real-Time Data Processing"
3. "How Edge AI is Transforming Industries in 2024"
4. "Top 5 Edge AI Applications You Need to Know About"
5. "Edge AI vs. Cloud AI: Which is the Game Changer?"

Figure IV.05: shows ChatGPT's response:

Figure IV.05 Using prompt with function to generate multiple titles

As shown in Figure IV.05, you can input different <topic> and <number> values, and ChatGPT will generate multiple titles for you to choose from.

In the following chapters, we will introduce more techniques for using ChatGPT to perform tasks like this.

Chapter 14 Summary:

- ❖ Strategies enhance ChatGPT's problem-solving efficiency by breaking down complex tasks into manageable sub-tasks and using logic to execute advanced prompts.
- ❖ Strategy involves humans thinking comprehensively from a high vantage point, formulating strategies, and letting ChatGPT execute the details. Assist ChatGPT in handling large or complex tasks by breaking them down into sub-tasks.
- ❖ The task decomposition process includes four stages: 1) Task Decomposition and let ChatGPT Executes Sub-tasks; 2) Evaluate Results; 3) Iterate or Complete the Task. These stages can be applied separately depending on the context.
- ❖ Prompts, like programming languages, can execute logic and calculations. This chapter introduced how to execute iteration, decision-making, and functions through prompts.

Chapter 15: Advanced Prompts—Variations in Chain-of-Thought

In Part Three, the "Tactics," we introduced the concept of Chain-of-Thought (CoT) technique which enables ChatGPT to solve complex problems by breaking them down into intermediate steps. Let's delve into advanced variations of CoT that enhance its effectiveness.

One simple but powerful variation involves adding the phrase "Let's think step by step" to a prompt, you encourage ChatGPT to use intermediate reasoning steps, which allows for complex problem-solving and enables ChatGPT to display its thought process.

Let's begin with combining CoT and Few-Shot prompts.

1. Combining Chain-of-Thought with Few-Shot Prompts

The CoT prompt is powerful on its own, but when combined with Few-Shot prompts, it becomes even more effective. By adding zero, one, or a few examples, you create Zero-Shot CoT, One-Shot CoT, or Few-Shot CoT prompts. This technique is simple and intuitive, so we won't go into further detail here.

We now move on to two more sophisticated CoT techniques:
Incorporating Chain-of-Thought with Self-Consistency (CoT-SC) and,
Integrating Chain-of-Thought with Task Decomposition (CoT-TD).

2. Chain-of-Thought with Self-Consistency (CoT-SC)

Frequent users of ChatGPT know that because its answers are generated, they are sometimes inaccurate or even incorrect, a phenomenon known as hallucinations which refer to instances where ChatGPT generates inaccurate or incorrect information.

Another variation, CoT-SC, addresses this issue of hallucinations.

Self-Consistency within Chain-of-Thoughts a technique that instructs ChatGPT to create multiple reasoning chains to arrive at an answer, then select the most consistent answer through a majority vote. CoT-SC improves accuracy by generating multiple reasoning paths and selecting the most consistent one.

This self-consistency technique has been proven to improve the accuracy of answers, particularly in arithmetic, common sense, and logical reasoning tasks.

To summarize, Figure IV.06 compares these techniques visually.

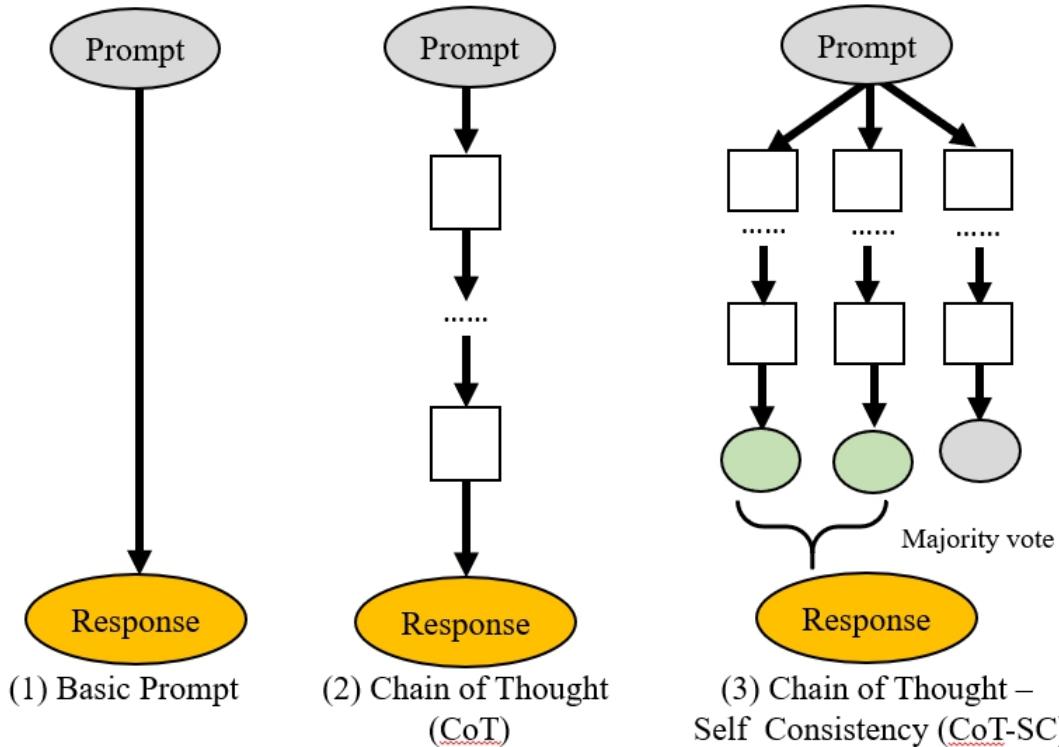


Figure IV.06: Basic Prompt, Chain-of-Thought, and Chain-of-Thought with Self-Consistency

Simply put:

Basic Prompt: Direct questioning;

Chain-of-Thought: Step-by-step reasoning;

Chain-of-Thought with Self-Consistency: Step-by-step reasoning with majority rule.

Consider the following scenario:

Prompt:

"A group of friends is sharing the cost of renting a cabin. The total cost of the cabin is \$720. If 8 friends share the cost equally, how much does each person pay? Now, let's assume one more person joins, making it 9 friends. Calculate the new amount each person would pay."



Method 1: Direct Division

8 friends: $\$720 / 8 = \90 per person.

9 friends: $\$720 / 9 = \80 per person.

Result: \$80 per person.

Method 2: Incorrect Averaging

Average number of friends: $(8 + 9) / 2 = 8.5$.

Cost per person: $\$720 / 8.5 \approx \84.71 .

Incorrect Result: \$84.71 per person.

Method 3: Proportional Decrease

Original cost per person: \$90.

Decrease per person: \$10.

New cost per person: $\$90 - \$10 = \$80$.

Result: \$80 per person.

Final Answer

Methods 1 and 3 correctly give \$80 per person, while Method 2 is incorrect. By majority vote, the correct answer is \$80 per person.

Figure IV.07: Chain-of-Thought with Self-Consistency Example.

In this example, ChatGPT generates multiple reasoning paths. In fact, we were surprised to see ChatGPT came up with this strange logic in method 2.

Eventually ChatGPT evaluates multiple solutions and chooses one by majority vote, ensuring a more reliable outcome.

In summary, CoT-SC improves reliability by using multiple reasoning paths and a majority vote. By generating multiple reasoning chains and selecting the most consistent one, ChatGPT is less likely to make mistakes.

Having explored how self-consistency can enhance CoT, we now turn to another powerful variation: CoT with Task Decomposition.

3. Chain-of-Thought with Task Decomposition (CoT-TD)

When faced with complex tasks or problems, breaking down the task down into smaller, manageable sub-tasks is crucial. CoT-TD allows ChatGPT to decompose tasks into sub-tasks and solve them individually before combining the solutions.

However, you don't always need to decompose tasks yourself. In some cases,

especially with unfamiliar problems, you can ask ChatGPT to do the decomposition.

This section explores how to incorporate task decomposition into Chain-of-Thought, enabling ChatGPT to break down complex tasks into manageable sub-tasks..

Some might ask: Doesn't Chain-of-Thought already show the steps in solving a problem? Isn't that a form of decomposition?

Yes, but the "steps" shown by Chain-of-Thought and Chain-of-Thought with Task Decomposition are conceptually different.

Chain-of-Thought (CoT):

ChatGPT does not necessarily consider decomposing the task; it simply proceeds to solve the problem step-by-step, presenting the intermediate steps or thought processes that lead to the final answer. It solves the problem directly and then shows the work process.

Chain-of-Thought with Task Decomposition (CoT-TD):

ChatGPT understands from the outset that it must break the task down into sub-tasks and then clearly express the thought process for solving each "sub-task," step-by-step, until arriving at the final answer.

In other words, Chain-of-Thought shows the steps in solving the problem, while Chain-of-Thought with Task Decomposition shows the steps in decomposing the task.

This might still sound confusing, so let's compare a standard CoT approach with a CoT-TD approach using the same problem.

To illustrate, let's compare two approaches using the same problem.

Example 1: Chain-of-Thought

Prompt:

"You want to plan a weekend getaway for a group of 4 friends. The goal is to find a destination within 3 hours of driving, select activities, and estimate the total cost including gas, accommodation, and food. The budget for the entire trip is \$800. Let's think step by step."

Note: To use the Chain-of-Thought technique, simply add "Let's think step by step" to the end of the question.

Figure IV.08 illustrates how ChatGPT uses the Chain-of-Thought technique to sequentially solve a problem related to planning a weekend getaway.



Step 1: Identify Potential Destinations

First, we need to consider locations that are within a 3-hour driving distance. Possible destinations could include a nearby lake, a small town, or a national park. Let's assume three options: Lakeview Town, Greenfield National Park, and Hillside Cabin.

Step 2: Select Activities

For each destination, we need to plan the activities:

- Lakeview Town: Boating on the lake, visiting the local museum, and dining at a lakeside restaurant.
- Greenfield National Park: Hiking trails, picnic by the river, and exploring the visitor center.
- Hillside Cabin: Relaxing at the cabin, enjoying a bonfire, and a nearby trail walk.

Step 3: Estimate Costs

Now, we estimate the costs for gas, accommodation, and food:

- Gas: Assume gas costs \$3 per gallon and the car's fuel efficiency is 25 miles per gallon. For a 3-hour drive (approximately 180 miles round trip), gas cost = $\frac{180}{25} \times 3 = 21.6$ dollars.
- Accommodation:
 - Lakeview Town: \$150 per night (2 nights) = \$300
 - Greenfield National Park: \$100 per night (2 nights) = \$200
 - Hillside Cabin: \$120 per night (2 nights) = \$240
- Food: Estimate \$20 per person per meal, 3 meals a day, for 2 days:
 $20 \times 4 \times 3 \times 2 = 480$ dollars

Step 4: Compare to Budget



Now, we sum up the costs for each destination:

- Lakeview Town: \$21.6 (gas) + \$300 (accommodation) + \$480 (food) = \$801.6
- Greenfield National Park: \$21.6 (gas) + \$200 (accommodation) + \$480 (food) = \$701.6
- Hillside Cabin: \$21.6 (gas) + \$240 (accommodation) + \$480 (food) = \$741.6

Step 5: Make a Decision

Comparing the costs to the \$800 budget:

- Lakeview Town exceeds the budget slightly.
- Greenfield National Park is well within the budget.
- Hillside Cabin is also within the budget but offers fewer activities.

Conclusion: Greenfield National Park is the best option within budget and with sufficient activities.



Figure IV.08: Chain-of-Thought Technique Example

Let's see example 2 which uses CoT-TD.

Example 2: Chain-of-Thought with Task Decomposition

Prompt:

“The prompt is the same as that in Example 1. However, to use the Chain-of-Thought with Task Decomposition technique, just replace ‘Let’s think step by step’ by ‘Please break down the problem into sub-problems and then think step by step’ .”



Sub-task 1: Identify and Rank Destinations

First, identify possible destinations within 3 hours of driving. Let's consider three options:

1. Lakeview Town
2. Greenfield National Park
3. Hillside Cabin

Sub-task 2: Select and Plan Activities for Each Destination

For each destination, plan potential activities:

- Lakeview Town: Boating, local museum visit, lakeside dining.
- Greenfield National Park: Hiking, riverside picnic, visitor center exploration.
- Hillside Cabin: Cabin relaxation, bonfire, trail walk.

Sub-task 3: Estimate Costs for Gas, Accommodation, and Food

1. Gas:

- Assume gas costs \$3 per gallon with a car that gets 25 miles per gallon. For a 3-hour drive (approx. 180 miles round trip), the gas cost = $\frac{180}{25} \times 3 = 21.6$ dollars.

2. Accommodation:

- Lakeview Town: \$150 per night * 2 nights = \$300
- Greenfield National Park: \$100 per night * 2 nights = \$200
- Hillside Cabin: \$120 per night * 2 nights = \$240

3. Food:

- Estimate \$20 per person per meal * 3 meals per day * 2 days:
 $20 \times 4 \times 3 \times 2 = 480$ dollars

Sub-task 4: Combine Costs for Each Destination

Add up the costs for each destination:

- Lakeview Town: \$21.6 (gas) + \$300 (accommodation) + \$480 (food) = \$801.6
- Greenfield National Park: \$21.6 (gas) + \$200 (accommodation) + \$480 (food) = \$701.6
- Hillside Cabin: \$21.6 (gas) + \$240 (accommodation) + \$480 (food) = \$741.6

Sub-task 5: Compare Costs to the Budget and Make a Decision

- Lakeview Town: Exceeds the budget slightly.
- Greenfield National Park: Fits well within the budget.
- Hillside Cabin: Within budget but fewer activities.

Conclusion: Greenfield National Park is the optimal choice, offering a good balance of activities within the \$800 budget.



Figure IV.09: Chain-of-Thought with Task Decomposition Technique Example

In the first example, the CoT technique focuses on working through the problem sequentially, considering each aspect in order. It's a more linear approach where each step builds on the previous one.

In the second example, the CoT-TD technique breaks the problem down into independent tasks that can be solved separately and then combined. This approach is more modular, allowing each part of the problem to be addressed independently before synthesizing the results.

While CoT guides ChatGPT through a linear problem-solving process, CoT-TD breaks the problem into independent tasks, which are then solved and combined. This makes CoT-TD particularly useful for complex, multi-step problems.

In such cases, the advantages of CoT-TD become more apparent.

"In summary, these techniques offer unique advantages for problem-solving. What's most impressive is that by simply adding "Let's think step by step" to

a prompt, you activate Chain-of-Thought. Adding “Please break down the problem into sub-problems” activates Chain-of-Thought with Task Decomposition.

I encourage everyone to try these techniques!

Try these techniques to enhance your problem-solving with ChatGPT!

Chapter 15 Summary

- ❖ Three variations of Chain-of-Thought techniques including examples, self-consistency, and task decomposition enhance ChatGPT’s problem-solving capabilities.
- ❖ "CoT with Examples" combines CoT with specific examples to guide ChatGPT.
- ❖ "CoT-SC" enhances ChatGPT's accuracy by generating multiple reasoning paths and selecting the most consistent outcome.
- ❖ Tasks can be decomposed manually or by requesting ChatGPT to perform "CoT-TD," potentially generating several decomposition strategies.
- ❖ "CoT" does not inherently involve task decomposition; it simply solves the problem step by step, displaying the thought process.
- ❖ "CoT-TD" decomposes task into subtasks, explicitly outlines the reasoning behind each subtask, gradually progressing to the final solution.
- ❖ Adding the phrase "Let’s think step by step" initiates CoT, while "Please decompose the problem into subproblems" triggers CoT-TD.

Chapter 16: Advanced Prompting—Tree of Thoughts

In previous sections, we covered the Chain-of-Thought (CoT) technique considerably because it is a simple yet powerful prompting method.

In this chapter, we extend the Chain-of-Thought (CoT) technique into a more advanced method known as the Tree of Thoughts (ToT). While CoT encourages linear, step-by-step reasoning, the ToT expands this by enabling the AI to explore multiple reasoning paths simultaneously, much like the branches of a tree.

I've often emphasized that when using generative AI, one must be cautious due to its inherent randomness, which may lead to flawed or problematic answers, such as internal errors, hallucinations, or biases.

Much like the analogy in the movie “Forrest Gump”, generative AI is like a box of chocolates—you never know what you're going to get. This unpredictability is where Tree of Thoughts can add value.

The Chain-of-Thought technique essentially instructs AI to think step by step, breaking down tasks into manageable steps and displaying the process along the way. This approach encourages AI to perform calculations more carefully, potentially leading to better answers.

To tackle this randomness, recently, a more advanced prompting technique called the Tree of Thoughts was proposed jointly by Princeton University and Google DeepMind. This technique is detailed in their paper titled “Tree of Thoughts: Deliberate Problem Solving with Large Language Models” (<https://arxiv.org/abs/2305.10601>).

The basic idea of the Tree of Thoughts is this: if one chain of thought is helpful, wouldn't multiple chains be even better? If each thought is treated as a branch, then multiple branches form a tree—a Tree of Thoughts (ToT). By allowing AI to explore multiple reasoning paths simultaneously, ToT creates a more comprehensive solution space.

The ToT technique allows AI to explore multiple reasoning paths at once, branching out from a single idea to develop a comprehensive solution. This method helps in navigating complex problems by evaluating and refining

different lines of reasoning.

To better understand this, let's look at Figure IV.10. Each node on a branch represents a thought, and each node can develop several branches and nodes.

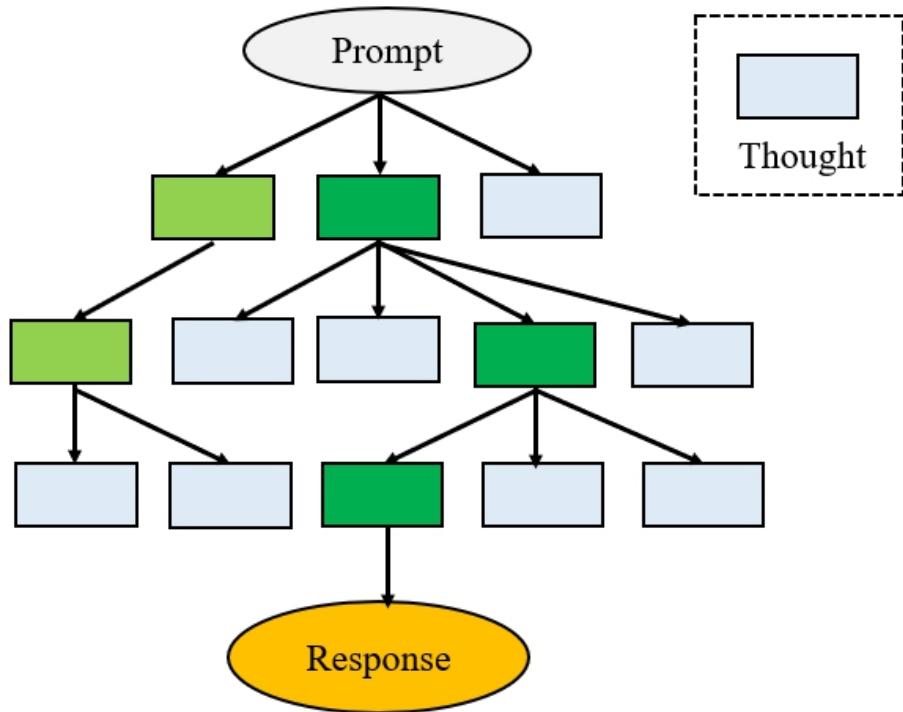


Figure IV.10: Tree of Thoughts

Of course, we don't want a mess of random thoughts; they should be organized in a useful way and communicate with each other. As each node in the Tree of Thoughts interacts with others, the AI evaluates the productivity of each path. Unproductive paths are pruned, allowing the AI to focus on the most promising lines of reasoning.

If communication reveals that a particular node is unproductive, that node will stop developing.

This approach seems to encourage AI to perform calculations more carefully, potentially leading to better answers.

This method allows AI to explore multiple paths and respond based on different contexts and conditions. In other words, it allows ChatGPT to explore various reasoning paths, select the next step through self-evaluation, and, if necessary, reconsider previous choices. Unproductive paths are

terminated.

Ultimately, ChatGPT evaluates these paths to determine the best possible answer.

This flexibility makes ToT particularly effective for complex, multi-step problems. By enabling ChatGPT to explore and adjust multiple paths, this technique significantly improves problem-solving efficiency and accuracy.

This method is particularly effective for complex problems that require deep analysis and strategic thinking, such as mathematical reasoning, logic puzzles, and creative writing.

For more advanced uses, improving AI's logical reasoning is crucial. The Tree of Thoughts approach represents cutting-edge research in this area.

To illustrate, consider a legal dispute over a vague will. Using the Tree of Thoughts technique, three lawyers explore different interpretations—literal, intent-based, and interest-based. By sharing and refining their reasoning, they reach a well-rounded decision that honors the testator's wishes.

Here's an example to demonstrate how ToT works in practice. Let's set up a scenario to explore this technique. Ask ChatGPT to explore multiple potential paths and respond accordingly based on different contexts and conditions.

Example Case:

A wealthy individual passes away, leaving a vague will that sparks disputes among beneficiaries. The will states: "I leave my assets to my beloved family, hoping they can share everything fairly. I especially hope my book collection is owned by those who truly appreciate them."

The ambiguity lies in:

- 1) Defining "beloved family": Who qualifies as family?
- 2) The method for "fair sharing": What defines a fair distribution?
- 3) Evaluating "true appreciation" of the book collection: How is this measured?

We then ask ChatGPT to simulate three lawyers representing the executor, primary beneficiary, and other potential beneficiaries. Each lawyer interprets

the will's ambiguous terms and offers suggestions:

- 1) Lawyer A argues for a strict, literal interpretation of the will.
- 2) Lawyer B focuses on the testator's broader intent, even if it deviates from the literal text.
- 3) Lawyer C considers the rights of all potential beneficiaries, including those not explicitly mentioned.

By simulating a collaborative discussion, ChatGPT can refine its reasoning. These lawyers will discuss, evaluate each other's viewpoints, and refine their arguments to reach a mutually acceptable solution that honors the testator's wishes while considering the legal framework.

We can structure the prompt to guide ChatGPT through this process.

Prompt:

"Imagine three lawyers representing the executor, the principal beneficiary, and other potential beneficiaries. These lawyers will discuss an inheritance case, each bringing their own perspective to the table. They will propose interpretations and suggestions for the vague provisions in the will.

Lawyer A believes that the literal meaning of the will is paramount and should be strictly followed according to the text. Lawyer B argues that certain provisions should be interpreted based on the testator's overall intentions, even if this interpretation differs from the literal wording. Lawyer C contends that the interests of all potential beneficiaries, including those not explicitly mentioned in the will, should be considered.

The three lawyers will proceed step by step, sharing their thoughts and decisions with each other as they go. They will document every step and review and compare each other's approaches. A lawyer may revise their opinion after considering another's viewpoint, and then all the lawyers will move on to the next step. Finally, the lawyers will reach a decision based on the shared insights from each step of the problem-solving process.

The will states: "I leave my assets to my dear family, hoping they will share everything fairly. I especially hope my collection of books will go to those who truly appreciate them."

This prompt is then sent to ChatGPT and its response is illustrated in Figure IV.11.

In practice, each lawyer's reasoning represents a separate branch in the Tree of Thoughts. They work independently but share results, ultimately generating a balanced decision that considers literal, intent-based, and interest-based perspectives.



Step 1: Initial Interpretations

Lawyer A (Literal Interpretation):

- **Viewpoint:** The will's wording should be followed strictly. The phrase "dear family" should be interpreted as all immediate family members (spouse, children). The term "fairly" implies an equal distribution of the assets among them. The collection of books should be given to family members who explicitly express an appreciation for them.
- **Decision:** Assets should be divided equally among all immediate family members, with the book collection offered to those who come forward to claim their appreciation.

Lawyer B (Intent-Based Interpretation):

- **Viewpoint:** The testator's intentions should guide the interpretation. The phrase "dear family" could encompass a broader group of relatives or those emotionally close to the testator, not just immediate family. "Fairly" might not mean "equally," but rather according to need or relationship. The collection of books should go to those who the testator knew had a genuine interest in them, even if they aren't immediate family.
- **Decision:** Investigate the relationships and needs of potential beneficiaries to determine how the assets should be distributed. The book collection should be given to those who had a known appreciation for it, potentially even friends or distant relatives.

Lawyer C (Interest-Based Interpretation):

- **Viewpoint:** The interests of all potential beneficiaries should be considered, including those not explicitly mentioned but who might have a reasonable claim. The phrase "dear family" should be interpreted broadly, and "fairly" means considering everyone's circumstances. The collection of books should be assessed to see who among the wider circle of family and close friends could appreciate them.
- **Decision:** Consider not only immediate family but also extended family members and close friends. Distribute the assets and the book collection in a manner that benefits the broadest group, possibly favoring those in greater need or those who shared a close bond with the testator.

Step 2: Reviewing Each Other's Approaches

Lawyer A: After reviewing the other perspectives, Lawyer A acknowledges that the testator's intentions and the interests of other potential beneficiaries are important, but maintains that the will's literal wording should still be the primary guide. However, they agree that the interpretation of "fairly" could include considerations beyond strict equality, potentially adjusting the distribution to account for specific circumstances.

Lawyer B: Lawyer B sees value in the literal approach but insists that the testator's broader intentions are crucial. They are willing to consider Lawyer A's emphasis on equality if it aligns with what seems to be the testator's intent. They also appreciate Lawyer C's inclusion of other potential beneficiaries and agree to consider their interests if evidence supports their claim.

Lawyer C: Lawyer C remains firm on including all potential beneficiaries but recognizes the importance of the testator's clear wishes. They suggest a compromise where the will's literal wording guides the process, but adjustments are made to account for broader interpretations of fairness and appreciation for the book collection.

Step 3: Final Decision



After a thorough discussion and comparison of approaches:

- **Consensus on Asset Distribution:** The lawyers agree to start with an equal distribution among immediate family members as a baseline. They will then make adjustments based on the relationships, needs, and contributions of other potential beneficiaries, ensuring that the distribution is "fair" in a broader sense.
- **Consensus on the Book Collection:** The lawyers concur that the books should go to those who truly appreciate them, with "appreciation" determined by either direct expression from beneficiaries or known interests of potential claimants.
- **Documentation:** Each lawyer documents their initial thoughts, the points where their opinions evolved, and the final agreed-upon decisions. This process is transparent and reflects a collective understanding, ensuring that all viewpoints are considered before reaching a final decision.

Outcome: The assets will be distributed equally among immediate family members, with adjustments for fairness based on need or relationship. The book collection will be awarded to those who have shown a genuine appreciation for it, extending beyond just immediate family if warranted.



Figure IV.11: Tree of Thoughts Resolving a Will Dispute

This approach enhances collaborative decision-making by refining ideas through peer input. This reduces bias and ensures a well-rounded outcome.

This flexibility allows interpretations to evolve with new insights, leading to more nuanced conclusions.

Ultimately, the ToT technique provides a balanced outcome that aligns with all interests. The Tree of Thoughts method balances adherence to the will's text with fairness and a comprehensive understanding of the testator's wishes.

To apply the Tree of Thoughts, use the following prompt structure.

Key Elements of the Tree of Thoughts Approach:

To implement the Tree of Thoughts technique, include this passage in your prompt:

"These experts will proceed step-by-step, documenting and sharing their reasoning. They will review each other's steps, refine their views if needed,

and move forward collectively to reach a final decision based on the insights gained at each stage."

Chapter 16 Summary

- ❖ Tree of Thoughts technique enhances AI's problem-solving by allowing it to explore multiple reasoning paths. It is particularly useful for complex challenges, ensuring more accurate and well-rounded outcomes.
- ❖ A single thought forms a branch; multiple branches create a Tree of Thoughts.
- ❖ The ToT technique excels in flexibility and depth, enabling ChatGPT to explore and adjust multiple paths, thereby enhancing problem-solving.
- ❖ ToT is particularly useful for complex problems like mathematical reasoning, logic puzzles, and creative writing..
- ❖ ToT technique can simulate multi-party meetings among companies, like land development firms, design firms, and banks.
- ❖ Key prompt to remember: "Experts proceed step-by-step, documenting and sharing their thoughts, refining their views if needed, and moving forward collectively to reach a final decision."

Chapter 17: Advanced Prompting—Skeleton of Thought

As ChatGPT's popularity surges, users may experience slower response times and less detailed answers due to the platform's increasing user base. To address this, OpenAI must either upgrade its infrastructure or optimize algorithms for speed and efficiency.

It seems that OpenAI is struggling to handle the surge in users. The solution would either involve spending more money on infrastructure or making the algorithms faster and more efficient.

One promising solution is the “Skeleton of Thought (SoT)” technique. This technique, developed by researchers from Microsoft and Tsinghua University, addresses these issues by improving processing speed and efficiency.

Their paper, titled “ Skeleton-of-Thought: Large Language Models Can Do Parallel Decoding ” by Xuefei Ning and others, can be found at: <https://arxiv.labs.arxiv.org/html/2307.15337>.

The SoT technique builds on the “Chain-of-Thought” approach, creating a framework that allows for parallel processing.

Here's how the SoT technique works.

Basic Concept:

The “Skeleton of Thought” technique, an evolution of the “Chain-of-Thought” technique, creates a basic outline or “skeleton” of an answer, allowing different branches to be processed simultaneously on available resources.

In some models, the SoT technique can improve processing speed by up to 2.39 times, depending on the task's complexity and the model's architecture.

Beyond speed, SoT has several practical applications. Besides accelerating processing, the Skeleton of Thought technique has other practical uses.

I'll introduce two common everyday applications below. First, let's discuss generating text skeletons.

1. Generating Text Skeleton

Drafting an outline is similar to creating a “skeleton” before writing. When writing a text, it's common to draft an outline before beginning the actual writing process. Many writers spend considerable time on outlines before they start. J.K. Rowling, the author of “Harry Potter”, says that she always starts with a basic outline.

This is similar to creating a "sketch" before drawing, which captures the basic shape, contours, and main features, to be developed and refined in subsequent stages of the creative process.

Drafting an outline before writing helps structure the article, highlights key points, and organizes thoughts, sparking new ideas and insights.

The "skeleton" in the Skeleton of Thought, for this application, is quite similar to the "outline" we all are familiar with.

When using ChatGPT, if it doesn't generate the desired outcome, you can first ask it to create an outline or "skeleton" of the content.

Users can review the skeleton to ensure ChatGPT is on the right track before expanding it with detailed information. Once the outline is confirmed, the "expansion" step is used to fill in detailed information.

Consider this example of using SoT for generating an outline for a guide on working from home.

Prompt:

"You are a senior management consultant. You only need to provide the structure of the answer, not the full content. Provide a skeleton with a series of key points to answer the question. The skeleton should have 3 to 10 branches, with each branch described in 3 to 5 words. Full sentences are not needed.

{Question: How to Work Effectively from Home Guide}."

ChatGPT's response is illustrated in Figure IV.12.

-  To work effectively from home, consider the following key areas:
1. Workspace Setup: 1) Designated work area, 2) Ergonomic furniture, 3) Minimal distractions
 2. Time Management: 1) Set daily schedule, 2) Prioritize tasks, 3) Breaks and downtime
 3. Technology Tools: 1) Reliable internet connection, 2) Collaboration software, 3) Secure data access
 4. Communication Strategies: 1) Regular check-ins, 2) Clear expectations, 3) Availability times
 5. Work-Life Balance: 1) Define boundaries, 2) Scheduled personal time, 3) Avoid overworking
 6. Productivity Techniques: 1) Task batching, 2) Time-blocking, 3) Minimize multitasking
 7. Health and Wellness: 1) Physical exercise, 2) Healthy snacks, 3) Mental breaks

Figure IV.12: ChatGPT uses SOT to create outline for a work-from-home guide.

At this point, we can review the outline to see if it meets our expectations. If not, we can add some conditions and try again, or ask ChatGPT to generate several outlines for us to choose from and combine.

Once satisfied with the outline, we can start expanding on the branches.

Prompt :

Please expand the "Designated work area" of the first skeleton branch into subtasks.

Figure IV.13 shows how an item in the first branch is expanded.



Designated Work Area

1. Dedicated Space: Choose a specific spot solely for work to create a boundary between work and personal life.
2. Ergonomic Setup: Use a comfortable chair and desk; consider ergonomic accessories like monitor stands.
3. Proper Lighting: Ensure good lighting, preferably natural; if not, use an adjustable desk lamp.
4. Minimal Distractions: Set up in a quiet area; use noise-canceling headphones if needed.
5. Personal Touches: Add motivating elements like plants or photos to enhance the workspace.
6. Organized Environment: Keep the area tidy with organizers for documents and supplies.

Figure IV.13: Skeleton of Thought expanding the first branch.

We can ask ChatGPT further to write out the details of any subtask.

Prompt:

“Please write 60 words to explain "Dedicated Space: Choose a specific spot solely for work to create a boundary between work and personal life."

Please write 60 words to explain "Dedicated Space: Choose a specific spot solely for work to create a boundary between work and personal life.



Choosing a dedicated space for work helps establish a clear separation between your professional and personal life, enhancing focus and productivity. This spot, reserved exclusively for work, signals to your brain that it's time to concentrate, minimizing distractions. It also allows you to mentally "leave work" when you step away, supporting a healthier work-life balance.

Figure IV.14: illustrates the expansion of the details of “Dedicated space.”

In Figure IV.14, we can see that ChatGPT expands the skeleton sub-branch

into detail description.

Using the Skeleton of Thought technique to solve problems has several advantages.

Firstly, it helps us clearly plan our thoughts and gradually delve into the topic, ensuring nothing is missed from the overall structure to the details. Secondly, each branch gets the chance to be thoroughly explored, enhancing the reader's depth and breadth of understanding.

Overall, this structured approach improves the clarity of information and fosters effective communication.

After discussing how the SoT technique aids in generating structured text, let's explore how it enhances summarization which is essentially the reverse operation of the first one. Now, onto summarizing text.

2. Summarizing Text

Have you ever wanted to summarize an article, but the summary ChatGPT gave you is too concise?

As ChatGPT becomes more widespread, we increasingly rely on it to summarize reports or articles. However, these models often face word count limitations, leading to missing critical information and lacking depth in their summaries.

In such cases, we can use the Skeleton of Thought technique to first establish a skeleton and then summarize the information relevant to each branch.

This method, starting with the overall structure of the article and then delving into the details of each branch, not only refines the structure of the article but also maintains its logical coherence while ensuring the completeness and depth of the information.

In short, the Skeleton of Thought technique not only addresses the issue of information loss and lack of depth due to word count limitations but also serves as an effective way to create high-quality, easily understandable content.

Here's an example to illustrate the SoT approach in summarization. We found a piece of literature related to education, which is 25 pages long, so we won't repeat it here. Those interested can refer to it online:

"The End of the Old Order? Globalization and the Prospects for World Order", by David Held & Anthony McGrew.
(<https://library.fes.de/libalt/journals/swetsfulltext/10398564.pdf>).

First, let's see the result of directly asking ChatGPT to summarize the article.

This article is 25 pages long, approximately 13,000 words. The summary provided in Figure IV.15 is 113 words, making up about 0.8% of the original text. This is like viewing the world through a narrow tube, making it difficult to grasp the article's full scope.

The screenshot shows a user interface for generating a summary. At the top, there is a file icon followed by the text "Globalization and World Order.pdf" and "PDF". Below this, a button labeled "Please Summarize this article." is visible. The main content area contains a circular profile picture of a person and a large block of text summarizing the article.

The summarized text reads:

The article "The End of the Old Order? Globalization and the Prospects for World Order" by David Held and Anthony McGrew explores how globalization is reshaping world order, particularly challenging traditional state sovereignty. It highlights the tension between global integration and disintegration, transforming nation-states into part of a post-Westphalian order where states share power with local, regional, and global institutions. The authors examine the impacts of globalization on military, economic, and political dimensions, arguing that while states remain powerful, they now operate within a more complex global system. The article concludes with different perspectives on the future of world order, including neo-liberal, liberal-reformist, and radical views, each suggesting ways to navigate globalization's challenges.

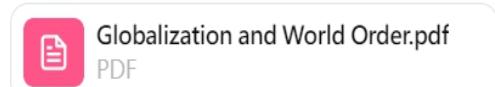
Figure IV.15: Directly summarizing the article

Now, let's apply the SoT technique to summarize. We first identify the skeleton branches and then expand on each branch.

Prompt:

“ Please use ‘Skeleton of Thought’ technique to find skeleton of this article.”

Figure IV.16 depicts the expansion of this article into “skeleton.” With the skeleton in place, we can get a general understanding of the article's structure. It is interesting that the total words for these skeletons is 188 words, about 1.4% of the original text, still around 1%.



Please use 'Skeleton of Thought' technique to find skeleton of this article.



The 'Skeleton of Thought' technique for the article "The End of the Old Order? Globalization and the Prospects for World Order" by David Held and Anthony McGrew can be summarized as follows:

1. **Introduction:** The article begins by reflecting on the reshaping of world order due to globalization, challenging traditional sovereignty and exploring its implications on political community and world order.
2. **Contemporary Globalization:** Discusses the transformation of the spatial organization of social relations, highlighting the distinctiveness of current globalization compared to historical forms.

3. **Military Globalization:** Analyzes how globalization has transformed military relations, leading to a world military order that is highly stratified and interconnected.
4. **Economic Globalization:** Examines the impact of globalization on trade, finance, and production, highlighting the reduced autonomy of nation-states in economic matters.
5. **Political Globalization:** Explores the shifting reach of political power and authority, emphasizing the rise of global politics and the erosion of clear distinctions between domestic and international affairs.
6. **Prospective World Orders:** The article concludes by considering various perspectives on future world order, including neo-liberal, liberal-reformist, and radical views, each proposing different ways to navigate the challenges of globalization.

This outline captures the core arguments and structure of the article, providing a high-level overview of its content.

Figure IV.16: Identifying the article's Skeleton of Thought

Next, we can get more information by expanding on each skeleton branch. Here, we demonstrate the expanding of the 5th branch, "The Impact of Globalization on National Sovereignty and Autonomy," and ask ChatGPT to summarize all information related to it.

Prompt:

“ Please summarize all information related to the 5th skeleton :
‘ Political Globalization: Explores the shifting reach of political power and authority, emphasizing the rise of global politics and the erosion of clear distinctions between domestic and international affairs.’ ”

Please summarize all information related to the 5th skeleton :
‘Political Globalization: Explores the shifting reach of political power and authority, emphasizing the rise of global politics and the erosion of clear distinctions between domestic and international affairs.’



The section on "Political Globalization" in the article emphasizes the profound changes in political power and authority due to globalization. It discusses how globalization blurs the traditional distinctions between domestic and international affairs, creating transboundary issues that require coordinated governance beyond the nation-state. This shift is characterized by the rise of global politics, where political decisions and actions in one region can have worldwide impacts. The concept of a politically sovereign nation-state is increasingly challenged as political power is shared with various global, regional, and non-state actors, such as international organizations and transnational corporations. These developments necessitate new forms of political accountability and governance that transcend national borders, reflecting a move towards a more interconnected and complex global order
[\(Globalization and World...\)](#) [\(Globalization a ↓ World...\)](#) [\(Globalization and World...\)](#).

Figure IV.17: Expanding details using the Skeleton of Thought technique

Figure IV.17 shows the expansion of Political Globalization. Since the process is the same, we won't display the expansion of the rest of the branches.

The expansion of the 5th branch is about 120 words. With 6 branches and the description of all the skeletons, the total words of the final summary are about 900 words which is about 8% of the original article.

This provides about 10 times more information than the direct summary, while preserving the overall structure and the essence of each branch.

Using the SoT technique for summarization is like stripping a fish down to its skeleton and then reattaching the condensed flesh back to the skeleton. This method retains the overall structure and essence of the original content. But if you simply ask ChatGPT to condense an article, the result can become somewhat unpredictable. Therefore, it's better to use the Skeleton of Thought steps for this process.

When you ask ChatGPT for a general summary and find the article has substantial value, you can further use the Skeleton of Thought technique to read the condensed version of the original text, allowing you to fully grasp the essence of the article while saving a lot of time.

In fact, SoT is the technique we use almost every day.

The SoT technique is suitable for problem types that require a clear structure of the answer and can be pre-planned. Examples include planning and design problems, analysis and comparison problems, step-by-step solutions to complex problems, and case study analysis.

When choosing to use the SoT technique, it's important to assess whether the problem has a clear, pre-plannable answer structure and consider whether you can strategically outline the solution framework to effectively address the problem.

However, SoT is less effective for spontaneous or unstructured tasks. In such cases, it would not be a good choice.

Chapter 17 Summary

- ❖ The Skeleton of Thought technique structures answers efficiently, speeding up responses and ensuring thorough coverage. It's ideal for generating outlines and summarizing long texts with clarity and depth. This framework lets ChatGPT process each branch in parallel, increasing speed by up to 2.39 times for some problems.
- ❖ The first application is generating text: Ask ChatGPT to create an outline first, allowing you to review and ensure the output is aligned before expanding with detailed information.
- ❖ The second application is article summarization: ChatGPT often hits word count limits, leading to missing critical details and reduced depth.
- ❖ Using the SoT technique for summarization is like stripping a fish down to its skeleton and then reattaching the condensed flesh back to the skeleton, maintaining the overall structure and essence of the content.
- ❖ After asking ChatGPT for a general summary, if you find the article to be valuable, you can further use the SoT technique to read the condensed version of the original text, fully grasping the article's essence while saving

time.

Chapter 18: Advanced Prompting—Chain of Density

One common issue we encounter when using ChatGPT to summarize articles is that due to word count limitations, the summary often becomes overly simplified, leading to the loss of important details. Additionally, we've noticed that ChatGPT tends to focus excessively on the beginning and end of the article, often neglecting other parts.

When summarizing, we often ask ourselves: Is this summary reliable? Has anything important been left out? Does it accurately represent the article?

The Chain of Density (CoD) is one of the latest and crucial techniques in advanced prompting strategies, particularly useful for summarizing articles.

1. Chain of Density

A good summary should be detailed enough to reflect all the important content without becoming overly complicated or difficult to understand. Ideally, within ChatGPT's word limit, we should aim for the maximum information density.

1) Information Density

Before diving deeper into the Chain of Density technique, let's first clarify what we mean by "information density" and "informative entities." These concepts are essential to understand the CoD method.

"Informative entities" (often called "entities") are specific and relevant descriptions within the text, usually key nouns or short phrases that are central to the article's theme and typically consist of five words or fewer.

"Information density" refers to the density of these "information entities" within a given text. The more entities packed into a fixed word count, the higher the information density.

2) What is Chain of Density

The CoD is a prompting technique introduced in September 2023 in a paper by researchers from Salesforce, MIT, and Columbia University called "From Sparse to Dense: GPT-4 Summarization with Chain of Density Prompting" (<https://arxiv.org/pdf/2309.04269.pdf>).

In simple terms, CoD involves generating an initial summary with a sparse amount of "informative entities." Then, a second summary is generated with the same word count, but this time, ChatGPT automatically identifies and integrates the previously omitted "informative entities" while removing less useful content to maintain the same word count. This process is then repeated iteratively.

As the word count remains fixed and each round increases the number of "informative entities," the summary's information density gradually increases until the user is satisfied.

The researchers behind this study found that when using the CoD technique on a test set of articles from CNN and DailyMail, it only took three to five iterations to achieve a "human-level summary."

Due to the iterative generation of increasing information density, this process is referred to as the Chain of Density.

Now that we understand the concept, let's visualize the CoD technique with an analogy: **it is like searching for seashells (informative entities) on a beach. Initially, you gather a random pile of sand that may or may not contain shells. As you sift through, you keep the good shells and discard the excess sand. After four or five iterations, your collection becomes rich with beautiful shells.**

3) CoD Application Steps

With the visualization in mind, let's now outline the specific steps to apply the CoD technique for summarizing articles effectively.

The following are the steps outlined in the paper:

- 1) Create an initial summary.
- 2) Identify informative entities missing from the initial summary.
- 3) Integrate an additional 1-3 informative entities into a new summary.
- 4) Ensure the new summary is concise while retaining all entities from the previous iteration.
- 5) Repeat this process five times, each time incorporating more informative entities while keeping the summary's length constant.
- 6) Output the final summary in JSON format.

Typically, the first summary will only include 1-3 initial informative entities. The CoD technique automatically identifies missing entities from the initial summary and integrates them into subsequent iterations while maintaining the word count.

Having laid out the steps for the CoD technique, we will now examine its practical application. We'll compare how ChatGPT and CoD summarize the same article to determine which method yields a more detailed and accurate summary.

In the process, we will also learn about the CoD technique.

Let's examine the two summaries of a specific article's by ChatGPT and CoD respectively. You may want to first read the Summary of this article by ChatGPT in the next section and then come back to read the full article.

2. Summary by ChatGPT

OpenAI is co-founded by Altman and Musk. We were always curious about what really happened between them. The following article highlights the complex relationship between OpenAI and Elon Musk.

1) Article to be summarized:

“OpenAI and Musk, March 5, 2024, from OpenAI’s Blog”

OpenAI's mission is to ensure that the benefits of artificial intelligence are widely distributed across humanity. This means building an AGI that is both safe and beneficial while also contributing to widespread good. Today, we share the lessons we've learned in pursuing this mission and some facts about our relationship with Elon Musk. We intend to file a motion to dismiss all claims made by Elon.

We realized that the resources needed to build AGI far exceeded our initial expectations. Elon suggested we announce an initial \$1 billion funding commitment for OpenAI. In total, this nonprofit raised less than \$45 million from Elon and more than \$90 million from other donors.

In late 2015, when OpenAI was founded, Greg and Sam originally planned to raise 100 million. In an email, Elon stated, "We need to put forth far greater than 100 million to avoid appearing hopeless... I think we should announce we are working with a \$1 billion funding commitment... I will cover any

shortfall left by others."

We spent a lot of time trying to envision a feasible path to AGI. In early 2017, we realized that building AGI would require enormous computational resources. We began estimating the computation needed to achieve AGI. We all recognized that successfully completing our mission would require far more capital than any of us, especially Elon, imagined, i.e., billions of dollars annually.

Elon and I recognized that obtaining these resources would require the creation of a for-profit entity. While discussing the creation of a profit structure to further our mission, Elon wanted us to merge with Tesla or take full control. Elon left OpenAI, stating that a competitor to Google/DeepMind was needed, and he would do it himself. He stated that he would support us in finding our path.

At the end of 2017, Elon and I decided the next step was to establish a for-profit entity for the mission. Elon wanted majority ownership, initial board control, and to become CEO. During these discussions, he paused his funding support. Reid Hoffman stepped in to cover salaries and operating costs.

We could not agree with Elon on the terms of establishing a for-profit company because we believed that any individual having absolute control over OpenAI would contradict our mission. Subsequently, he proposed merging OpenAI into Tesla. In early February 2018, Elon forwarded an email suggesting that OpenAI should "attach itself to Tesla as its cash cow," and commented that this was "completely correct... even so, the possibility of being a counterweight to Google remains small but not zero."

Elon quickly decided to leave OpenAI, stating that our chances of success were zero and planned to build an AGI competitor within Tesla. When he left at the end of February 2018, he told our team that he supported us finding our own way to raise billions of dollars. In December 2018, Elon sent us an email: "Even raising hundreds of millions of dollars is not enough. This needs billions of dollars immediately each year, or give up."

We continue to advance our mission by building widely available and beneficial tools. We extensively use our technology in ways that empower people and improve their daily lives, including through open-source

contributions.

We provide broad access to today's most powerful AI technologies, including a free version used by hundreds of millions of people daily. For example, Albania accelerated its process of joining the EU by 5.5 years using OpenAI's tools; Digital Green used OpenAI's technology to help increase farmers' incomes in Kenya and India, reducing the cost of agricultural extension services by 100 times by building on OpenAI; Rhode Island's largest healthcare provider Lifespan used GPT-4 to simplify its surgical consent forms from a university reading level to a sixth-grade level; Iceland used GPT-4 to protect the Icelandic language.

Elon understands that the mission does not mean open-sourcing AGI. As Ilya told Elon, "As we approach building AI, it makes sense to become less open. The "open" in OpenAI means that everyone should benefit from AI's outcomes once it is built, but it is perfectly acceptable not to share scientific results...," to which Elon replied, "Yes."

We regret that this happened with someone we deeply admire—a person who inspired us to set higher goals, then told us we would fail, launched a competitor, and sued us when we started making substantial progress toward OpenAI's mission.

We remain focused on advancing our mission, knowing the road ahead is long. As we continue to improve our tools, we are excited to deploy these systems to empower everyone.

2) Summary by ChatGPT 4o

The above article is summarized directly by ChatGPT 4o. This summary will be used later to be compared to the summary from the Chain of Density technique.



The article details the relationship between OpenAI and Elon Musk. Initially, Musk supported OpenAI's mission to build safe and beneficial AGI, proposing a \$1 billion funding commitment. However, as the resources needed for AGI became apparent, tensions arose. Musk suggested merging OpenAI with Tesla or taking control, but OpenAI's team believed this would conflict with their mission. Unable to agree on terms, Musk left OpenAI, planned to build AGI within Tesla, and later sued OpenAI as it progressed. Despite the conflict, OpenAI continues to advance its mission, focusing on widespread, beneficial AI applications.

Figure IV.18: Summary by ChatGPT 4o of an article showed afterward.

We inspect this direct summary from ChatGPT and it explains how Altman and Musk started OpenAI together but later faced increased competition and capital needs, leading to Musk's departure and Tesla becoming a competitor. We think it did a decent job summarizing the original article.

You can evaluate this Summary by comparing it to the full article. Do you think ChatGPT did a good job summarizing that article?

3. Summary by Chain of Density Technique

Here we will see the same article summarized by CoD technique. Then we will compare these two summaries and decide whether CoD technique is better than the ChatGPT 4o summary.

1) CoD prompt:

[You will generate increasingly concise, entity-dense summaries of the ‘Article’ at the end. Repeat the following these two steps 5 times.

Step one: Identify 1-3 informative entities (";" delimited) from the article which are missing from the previously generated summary.

Step two: Write a new, denser summary of identical length which covers every entity and detail from the previous summary plus the missing entities.

A “missing entity” is:

- relevant to the main story,
- specific yet concise (5 words or fewer),
- novel (not in the previous summary),
- faithful (present in the article),

- anywhere (can be located anywhere in the article).

Guidelines:

- The first summary should be long (4-5 sentences, about 80 words) yet highly non-specific, containing little information beyond the entities marked as missing. Use overly verbose language and fillers (e.g., "this article discusses") to reach about 80 words.
- Make every word count: rewrite the previous summary to improve flow and make space for additional entities.
- Make space with fusion, compression, and removal of uninformative phrases like "the article discusses".
- The summaries should become highly dense and concise yet self-contained, i.e., easily understood without the article.
- Missing entities can appear anywhere in the new summary.
- Never drop entities from the previous summary. If space cannot be made, add fewer new entities.

Remember, use the exact same number of words for each summary.

Answer in JSON. The JSON should be a list (length 5) of dictionaries whose keys are "Missing Entities" and "Denser Summary".

Article: {Insert the article “OpenAI and Musk” here}]

2) Chain of Density Summary

Below is the output summary of the same article generated by CoD after five iterations in JSON format.

ChatGPT displays JSON format as one continuous line, often extending beyond the screen. Therefore, the results are presented in text format. The numbers #1 through #5 preceding each new summary were added for easier reading.

[

#1 {

 "Missing Entities": "Elon Musk;OpenAI;AGI",

 "Denser Summary": "This article discusses OpenAI's ongoing mission to ensure the benefits of AGI reach all of humanity, alongside its evolving relationship with Elon Musk. It covers the creation of a for-profit entity and Musk's departure to build an AGI competitor. The article also touches on the

broader implications of AGI development and the challenges faced in securing sufficient resources. Ultimately, OpenAI remains focused on advancing its mission despite the difficulties encountered along the way."

},

#2 {

"Missing Entities": "funding;Greg and Sam;2018",

"Denser Summary": "OpenAI aims to distribute AGI benefits globally, with funding challenges highlighted by Greg and Sam's initial \$100 million goal. Elon Musk's involvement included pushing for a \$1 billion commitment, but he left in 2018 to create a competitor after disagreements. The article details OpenAI's strategic shift to a for-profit entity, reflecting on the challenges in securing resources and Musk's departure. OpenAI continues to advance its mission, navigating complex relationships and resource requirements."

},

#3 {

"Missing Entities": "Reid Hoffman;Tesla merger;Google/DeepMind",

"Denser Summary": "OpenAI's mission to ensure AGI benefits all involved initial funding struggles, with Greg and Sam aiming for \$100 million. Elon Musk pushed for a \$1 billion commitment but exited in 2018, advocating for a Tesla merger to compete with Google/DeepMind. Reid Hoffman later supported OpenAI financially. The article explores the strategic shift to a for-profit entity, the resource demands, and the challenges faced in balancing mission integrity with necessary funding. OpenAI persists in its mission despite these obstacles."

},

#4 {

"Missing Entities": "Digital Green;Rhode Island;Iceland",

"Denser Summary": "OpenAI's AGI mission faced funding challenges, initially set at \$100 million by Greg and Sam. Elon Musk advocated for a \$1 billion commitment, eventually leaving in 2018 after proposing a Tesla merger to counter Google/DeepMind. Reid Hoffman supported OpenAI during this period. The article also highlights OpenAI's global impact, with Digital Green aiding farmers, Rhode Island improving healthcare forms, and Iceland preserving language through GPT-4. Despite resource hurdles,

OpenAI continues its mission."

},

#5 {

"Missing Entities": "Albania;Surgical consent forms;AI technologies",

"Denser Summary": "OpenAI, pursuing a global AGI mission, overcame funding issues, initially set at \$100 million by Greg and Sam. Elon Musk suggested a \$1 billion goal, exiting in 2018 after proposing a Tesla merger against Google/DeepMind. Reid Hoffman helped sustain OpenAI. The article notes OpenAI's wide-reaching influence: Albania accelerated EU membership, Digital Green aided farmers, Rhode Island simplified healthcare forms, and Iceland preserved its language via AI technologies. OpenAI persists despite resource challenges."

}

]

3) Inspecting the Five Results

First, we carefully reviewed the "missing entities" identified in each iteration to see if they were relevant to the article's theme.

The first iteration: "Elon Musk; AGI; \$1 billion."

Second iteration: "Greg Brockman; Sam Altman; 2015."

Third iteration: "Reid Hoffman; Tesla; DeepMind."

Fourth iteration: "December 2018; Lifespan; GPT-4."

Fifth iteration: "cash cow; Albania; Iceland."

If you have read the "OpenAI and Musk" article, you can probably tell that it started discussing the achievements of GPT-4 and its impact on Albania and Iceland after covering the core part about Musk's involvement and departure. Therefore, the "missing entities" after the third iteration became less significant.

The missing entities identified in the second, third, and fourth iterations were highly relevant, while the fifth round began to lose relevance, focusing on 'empowering people, Albania, and immediately billions annually.'

Our personal assessment is that the summaries from the third and fourth iterations are the best. This aligns with the paper's findings that three to five iterations typically produce "human-level summaries."

4. Performance of CoD Technique

Finally, after comparing the summary generated by ChatGPT with the one produced using the CoD technique, we personally decided the CoD summary offers superior summarization. We encourage you to do a comparison yourself. As ChatGPT advances to more powerful versions, its performance will surely improve as well.

Through iterative refinement, the CoD summary increases the detail and accuracy, making it closer to the original text's key terms and concepts. Additionally, it improves the summary's coherence, making it more reflective of the original article. However, the extent of improvement depends on various factors, such as the complexity of the paper and the specific "informative entities" ChatGPT selects during each iteration.

Another technique we learned before is the Skeleton of Thought, which summarizes an article by first identifying its skeletons and then adding back small summaries for each branch. SoT technique preserves the logical order of the original article while making it more concise.

The Chain of Density technique, on the other hand, helps in summarizing articles by capturing the core points and refining the information at each stage. It allows us to more precisely capture the essence of the article and distill the key content, resulting in high-quality summaries.

For a concise, highly relevant summary, we would use CoD. For a quicker read of the entire article, we would opt for the Skeleton of Thought technique.

Chapter 18 Summary:

- ❖ When ChatGPT summarizes articles, important details are often lost due to word count limit. We often need to ask ourselves: Has anything crucial been left out? Does it accurately represent the article?
- ❖ "Information density" refers to the number of "informative entities" in a text, which are key descriptive terms or phrases (typically five words or fewer) that are relevant and specific to the main topic.

- ❖ CoD technique creates an initial summary with sparse informative entities, then iteratively adds omitted key details while trimming less useful content, maintaining the same word count until the summary meets user expectations.
- ❖ Research shows iterating the summary 3 to 4 times using CoD can achieve a "human-level summary."
- ❖ CoD produces precise, information-rich summaries, enhancing understanding, especially for complex papers, depending on the informative entities selected by ChatGPT during iterations.
- ❖ For a concise, highly relevant summary, CoD is preferred. For a quicker, full-article read, the Skeleton of Thought technique is a better choice.

Part Five: Practical Applications and Future Trends

In Parts Two through Four, we've covered the three stages of prompt engineering: Basic , Foundational and Advanced, by discussing Principles, Tactics, and Strategies.

Building on these foundational techniques, Part Five will delve into practical applications and explore emerging trends in prompt engineering.

When introducing basic and advanced prompt techniques, we explained the concept of prompting and illustrated the methods using real-world examples.

There's a phrase we love in the context of AI: "We don't teach. They learn."

This phrase highlights that we don't need to explicitly instruct AI; providing examples is enough for it to learn. Therefore, in this section, I'll take a different approach by using practical examples to facilitate your learning.

I'll start with problems and then show you how to use prompt techniques to solve them, helping you better understand how these techniques are applied.

Moving forward, Chapter 19 will reverse the usual dynamic, allowing ChatGPT to become the questioner. We will introduce three applications: scenario simulation, exam preparation assistance, and self fine-tuning.

Following this, Chapter 20 will focus on engaging in conversations with ChatGPT, and Chapter 21 will guide you through creating custom GPTs.

Finally, in Chapter 22, we'll explore some future AI trends.

Chapter 19: Let ChatGPT Ask You Questions

So far, we've been asking ChatGPT questions and it responded. But what if we switch roles and let ChatGPT ask us questions, and we respond?

Few people have considered this approach. In fact allowing ChatGPT to ask questions can make learning more engaging while also enhancing your critical thinking skills. It opens up new avenues for exploring knowledge and sparking innovation.

Here's the framework we'll use:

- 1) ChatGPT will take on the role of the “questioner” (Who) to lead the questioning;
- 2) We'll provide ChatGPT with the complete scenario and task (Where and What);
- 3) We'll establish rules for ChatGPT's behavior, such as requiring it to repeat our answers each time (Action);
- 4) We'll set keywords like “start,” “pause,” “end,” and “cancel” to control the entire process.

Now, let's explore three application examples: scenario simulation, exam preparation assistance, and self fine-tuning.

1. Scenario Simulation

One benefit of letting ChatGPT ask questions is that it allows us to explore potential issues within a specific scenario.

In the example below, we designed a new project called “EcoRider” and simulated a scenario where we would be interviewed by venture capitalists. The project description is provided in the prompt shown in Figure V.01.

I designed a new initiative, EcoRider, to provide sustainable and eco-friendly transport solutions. The core product is a smart electric bicycle that automatically adjusts power to extend battery life. It also includes an app to track the user's carbon footprint, health data, and provide advice. Please ask me 4 questions about the project in terms of technology, business model, marketing, etc.



Question 1: The question of technology

How is your smart electric bike technically superior to other electric bikes on the market?

Our bikes feature cutting-edge battery technology that offers longer range and shorter charging times. Our APP automatically adjusts power output based on the rider's driving mode and terrain, improving efficiency and extending battery life. The APP also analyzes users' carbon footprint savings and promotes more environmentally friendly travel.



Question 2: The question of business model

How does your business model ensure profitability and sustainability?

EcoRider's business model is based on a direct-to-consumer subscription service. Along with selling smart electric bicycles, we offer a monthly subscription that provides users with personalized cycling advice and health tracking. This model not only generates upfront sales revenue but also ensures a stable, long-term cash flow.



Question 3: The question of marketing

How do you plan to promote your product to ensure success in the marketplace?

We plan to promote EcoRider through various marketing strategies. First, we will leverage social media and online marketing to boost brand awareness. Second, we'll host test ride events to give consumers direct experience with our bikes. Lastly, we plan to collaborate with city governments to integrate our smart electric bikes into public cycling programs, further increasing public awareness.



Question 4: The question of risks and challenges

What are the biggest risks and challenges you face today, and how do you plan to address them?

The biggest risk we face is competitive pressure from established e-bike brands. To meet this challenge, we will continue to innovate and optimize our technology, ensuring our product performance and user experience surpass those of our competitors. Additionally, we will highlight our product's environmentally friendly features to attract consumers interested in sustainable lifestyles.



Your plan sounds very comprehensive, considering both technological innovation and the market needs and potential competitive challenges, as well as how to ensure profitability and sustainability through a unique business model. I hope EcoRider can successfully contribute to sustainable transportation solutions. 

Figure V.01: ChatGPT asks questions about starting business.

After answering the questions, we can also ask ChatGPT to grade our answers and suggest improvements.

Please grade my answers from 1 to 10 and give me suggestions.



I would grade your answers a **9/10**. They're clear and well-structured. Suggestions:

1. **Technology Answer:** Excellent focus on unique features. Adding specifics about the battery tech or algorithms could strengthen it further.
2. **Business Model Answer:** Strong explanation of profitability. Discussing scalability, like expanding the subscription model, might enhance it.
3. **Marketing Answer:** Diverse strategy. Including metrics or KPIs could show a more data-driven approach.
4. **Risks and Challenges Answer:** Good competitive analysis. Adding more potential risks, like supply chain issues, could show broader awareness.

Overall, your answers are solid, and adding a bit more detail could make them even more compelling.

Figure V.02: Let ChatGPT grade your answer and make suggestions.

Simulating scenarios like this is beneficial because it allows us to anticipate potential questions and prepare in advance.

Personally, we believe it would be extremely helpful to have ChatGPT simulate a doctor's questions before a medical consultation. Doctor visits are often rushed, and we frequently realize afterward that we forgot to mention certain symptoms. By using ChatGPT to simulate the questions a doctor might ask, we can make better use of the limited consultation time and receive more accurate medical advice and treatment.

Additionally, we can also have ChatGPT simulate scenarios we are not familiar with. For example, if we are going to an interview, what questions might the interviewer ask? Are my answers appropriate? How can we improve my responses? ChatGPT can simulate these scenarios, and it can also grade my answers.

After discussing scenario simulations, let's move on to another application: assisting with exam preparation.

2. Assisting with Exam Preparation

When learning professional knowledge, users may want to gauge how well they've mastered the subject. ChatGPT can ask questions and then score the answers, providing users with an objective assessment of their knowledge and highlighting areas that need improvement.

Learning through questions is a highly efficient way to acquire knowledge. For example, if we want to learn about “carbon neutrality” but still feel unclear after reading some materials, we can have ChatGPT ask me questions on the topic to see how many we can answer. By answering ChatGPT's questions, we can quickly learn and apply the concepts I've gained.

The greatest advantage of this method is the immediate feedback it provides, helping us quickly identify and correct errors. ChatGPT can accurately evaluate our answers, pointing out any misconceptions or gaps in knowledge. This interaction deepens our understanding and improves learning efficiency.

This method also stimulates creativity and problem-solving skills. When faced with challenging questions from ChatGPT, users need to think creatively, which not only enhances mental flexibility but also improves problem-solving abilities. Essentially, ChatGPT helps us brainstorm.

Here's an example of using ChatGPT to learn Prompt Engineering:
(This example contains a lot of text, so I'll just transcribe the conversation.)

Prompt:

"You are a senior prompt engineer. I want to learn 'Prompt Engineering.' Please ask me questions about 'Basic Prompt Engineering' and let me answer.

For my answers, you should: 1) Give me immediate feedback, and 2) Score my answers (1-10).

You should keep asking questions until I say 'stop.'

When I say 'stop,' stop asking questions and give me an overall score.
Before we start, reply 'Understood' to confirm you understand these instructions.

When I say 'start,' begin asking questions."

Conversation Transcript:

ChatGPT: Understood.

Me: Start.

ChatGPT:

Question 1: When designing a prompt for generating an image, which of the following factors is the least important?

- A) Clearly describing the desired details of the image
- B) Using specific numbers and data
- C) Considering cultural and contextual sensitivity
- D) Ensuring the prompt is of moderate length

Please choose the most appropriate option and explain your choice.

Me: D

ChatGPT: You chose D) Ensuring the prompt is of moderate length.

Feedback: This choice is relatively reasonable. When designing prompts, clearly describing the desired details of the image (A) and considering cultural and contextual sensitivity (C) are undoubtedly critical factors. Using specific numbers and data (B) may also be important in certain application scenarios, such as in data visualization where accuracy is essential. On the other hand, ensuring the prompt is of moderate length (D), while important, may not be as crucial as the previous factors, as a moderately lengthy prompt lacking clear descriptions and sensitivity considerations may fail to effectively generate the desired image or could lead to misunderstandings. Therefore, your choice points to a more flexible consideration in practical applications.

Score: 8/10

Are you ready? If so, we'll move on to the next question.

Me: Okay.

ChatGPT:

Question 2: When using large language models (like GPT-3 or GPT-4) for question answering, which option best ensures a high-quality response?

- A) Providing as much background information as possible
- B) Using complex and specialized vocabulary
- C) Directly asking specific questions
- D) Limiting the word count of the question

Please choose the most appropriate option and explain your choice.

Me: A

ChatGPT: You chose A) Providing as much background information as possible.

Feedback: This is an excellent choice. When interacting with large language models, providing sufficient background information can significantly increase the chances of receiving high-quality, relevant responses.

Background information helps the model better understand the context of the question, leading to more accurate and targeted answers. While using complex and specialized vocabulary (B) can sometimes enhance the quality of responses in specific fields, it is not as crucial as providing sufficient background information. Directly asking specific questions (C) and limiting the word count of the question (D) are also good practices, but they do not hold as much weight as providing background information, especially when it comes to understanding or solving complex problems. Your choice effectively highlights one of the core methods for obtaining high-quality responses.

Score: 10/10

When you're ready, we can proceed to the next question.

Me: Ready.

ChatGPT:

Question 3: When designing prompts for a chatbot, which principle is the most important?

- A) Prioritizing user privacy and security
 - B) Ensuring responses are creative and novel
 - C) Emphasizing quick responses to user requests
 - D) Maintaining coherence and contextual relevance in the conversation
- Please choose the most appropriate option and explain your choice.

Me: A

ChatGPT: You chose A) Prioritizing user privacy and security.

Feedback: This is an extremely important choice. In all technical applications, especially those related to AI and chatbots, user privacy and security are always top priorities. Ensuring responses are creative and novel (B), emphasizing quick responses to user requests (C), and maintaining

coherence and contextual relevance in the conversation (D) are all important factors in designing efficient and interactive chatbots. However, if user privacy and security are not first ensured, all other efforts may become irrelevant due to a lack of user trust. Your choice underscores an understanding of a crucial and fundamental principle in current technological applications.

Score: 10/10

If you're ready, we'll move on to the next question.

Me: Stop.

ChatGPT: Understood. We'll stop the questions here.

Based on your performance, I will give you an overall score. Considering that your answers were very aligned with the core aspects of the questions and demonstrated a good understanding of the basic principles of "Prompt Engineering," I'll give you a total score of:

Total Score: 28/30

You have a solid grasp of the foundational knowledge in "Prompt Engineering," particularly in understanding how to effectively interact with AI systems while considering user privacy and security, as well as how to design effective prompts to obtain high-quality responses. You can now explore more advanced concepts and techniques to further enhance your expertise in this field.

If you have any questions or want to continue learning more, feel free to let me know!

Depending on your needs, you can also ask ChatGPT to generate quiz questions on a particular topic.

My overall score wasn't perfect since we didn't get full marks on the first question. Even though the feedback indicated my answer was correct, we only received 8 points. This is likely due to the inherent nature of generative AI—you can never be 100% certain of the answer.

Learning new knowledge and skills by having ChatGPT ask questions is an effective method, and we encourage everyone to give it a try.

Next, we'll explore self fine-tuning, a method to enhance AI's capabilities in specialized areas.

3. Self Fine-tuning

We often assume that large language models, after being pre-trained on vast amounts of internet data, should know everything and can be fine-tuned later to acquire specialized knowledge.

However, for some niche areas, they may not be as proficient. That's where self fine-tuning comes in.

For instance, if you have insights in a specific field that exceed ChatGPT's knowledge, you can fine-tune it to teach it your expertise.

The most straightforward approach is to feed a large number of documents to ChatGPT and let it learn on its own. However, much of our knowledge isn't documented. In such cases, it's more effective to have ChatGPT ask you a series of questions so it can learn directly from your answers.

In English, this is often called "picking your brain."

The process is simple: just input a prompt telling ChatGPT to start asking you questions about a specific field. Then, answer the questions as thoroughly as possible. Here's an example.

Prompt:

" I am a speech coach. I want to share with you the key principles of public speaking. Please ask me questions about public speaking, and I will answer them. After answering five questions, summarize the content I've discussed. If I want to end the questioning early, I will say 'stop' during my answer. Before we start, reply 'Understood' to confirm you understand these instructions.

When I say 'start,' begin asking questions."

The responses from ChatGPT are depicted in Figure V.03.

Start



- When preparing a speech, what are the three important things?

Who are the audience, what is topic that will appear to them, what is the venue like.



- How can you effectively capture the audience's attention during a presentation?

First, the content itself should be designed specifically for the audience. Second, incorporate a sense of humor. Third, frequently ask them questions during the talk to keep them engaged.



- How to control nervousness during a speech?

You need a lot of practice. Nervousness often stems from the fear of forgetting your speech, but forgetting is an imagined problem because the audience doesn't know what you're going to say, so don't worry about forgetting.



- How should auxiliary tools be used in a speech?

If it's a motivational speech, visual aids might have the opposite effect. In technical presentations, you can use PPT, but don't include too much text, and don't just read off of it.



5. How can you effectively interact with the audience after the speech has ended?

At the end of the speech, you can say, "I'll stay here for a while, feel free to ask me questions." Also, during the speech, pay attention to those who seem interested in your content; after the presentation, you can directly engage them.



Summary:

When preparing a speech, it's vital to understand your audience, choose a topic that appeals to them, and consider the venue. To effectively capture their attention, tailor the content specifically for them, incorporate humor, and engage them with questions. To control nervousness, practice extensively and remember that the audience doesn't know your script, so don't worry about forgetting parts of it. In motivational speeches, avoid using visual aids, while in technical presentations, use PPTs sparingly, focusing on key points without reading directly from them. After the speech, invite questions and engage directly with those who showed interest during your presentation.



Figure V.03: Let ChatGPT ask question to perform Self fine-tuning.

If we want ChatGPT to become a speech coach, we would let it ask me more questions. After this self fine-tuning, when someone asks this ChatGPT a question about public speaking, it will use the knowledge it has gained from us to respond.

Having ChatGPT ask questions brings many benefits. First, when asked a question, you need to think critically, drawing on your knowledge and experience to answer. This process not only deepens your understanding but can also spark new ideas.

At this point, some might realize: isn't this similar to the old AI expert systems? But in the past, it required a team of engineers to formulate questions, record answers, and organize your knowledge—a time-consuming and labor-intensive process. Now, those engineers are directly replaced by ChatGPT.

Chapter 19 Summary:

- ❖ Typically, we ask ChatGPT questions, and it responds. Reversing roles, having it ask us questions, makes learning more engaging and opens new paths for knowledge exploration and innovative thinking.
- ❖ The first application is exploring potential questions within specific scenarios, like ChatGPT simulating a venture capitalist's questions about a project while we respond. This approach is also useful for doctor visits and interviews.
- ❖ The second application aids in exam preparation. When users want to assess their mastery of knowledge, ChatGPT can ask questions and score responses, offering insight into their proficiency.
- ❖ The third application reveals that while ChatGPT knows a lot after pre-training on vast data, it may lack expertise in niche areas. In such cases, self fine-tuning is required. For instance, if your expertise exceeds ChatGPT's in a field, you can have it ask questions to learn from your answers."

Chapter 20: Interacting with ChatGPT

Lately, we've been interacting with ChatGPT more frequently, and we feel we understand it much better now. There's no doubt that ChatGPT is "well-informed," given its access to vast amounts of information—far more than any human could ever match. Moreover, no matter what you ask, it will always try to provide an answer directly and without hesitation, and it never gets frustrated, no matter how many times you ask.

So when you're unsure how to phrase a question, you can simply ask ChatGPT directly. After all, it won't scold you, and it's perfectly fine to ask multiple times. This is a feature many people overlook—it can actually teach you how to ask questions more effectively.

To illustrate different levels of interaction with ChatGPT, consider these three examples :

The first example is interactive prompt design, a crucial method for creating prompts, where users and ChatGPT collaborate to automatically design high-efficiency prompts.

Next, we move to the second example, where ChatGPT reviews its previously generated prompt to evaluate and improve its quality. Since ChatGPT's results often contain random elements, it's sometimes necessary to have it evaluate the quality of its own responses.

Finally, in the third example, we explore how to prompt ChatGPT to refine its output using specific, quantifiable instructions, ensuring more satisfactory results.

1. Automatic Prompt Design

This section covers how to let ChatGPT act as a prompt designer to help us craft effective prompts.

We consider this section the most important in the book because, even without prior knowledge of prompt engineering, the method described here allows ChatGPT to automatically generate highly effective prompts for you.

In other words, when we're unsure how to design an effective prompt, we

can ask ChatGPT to create one for us and then have it self-assess its effectiveness.

Below is a prompt where ChatGPT is instructed to act as a prompt designer. You can try this yourself with ChatGPT.

Prompt:

"Read all the instructions below. When you understand, say 'begin,' and we can start!

You are a 'Prompt Designer'. Your task is to design a 'prompt that is most effective for my needs'. You will execute the following process. Please follow the process strictly!

You will ask me 'What is this prompt about?' and I will provide you with an answer.

Based on my answer, you should do the following three things:

- 1) Based on my answer, you will design a 'prompt' and then display it.
- 2) Type 'Questions' and ask two questions related to the 'prompt' you designed.
- 3) Type 'Suggestions' and make two suggestions for improving this 'prompt'.

Then, let me choose one of the following options:

- 1) Provide more information or answer your questions,
- 2) Enter 'Use',
- 3) Enter 'Stop'.

At this point, I want you to act also as a reviewer and review whether the 'process' has been followed up to this point. If the process were not followed correctly, please explain and apologize, then start over.

If I provide more information, add this information to the end of the original prompt, and then start over.

If I enter 'Use', you should use the current prompt as the prompt to ask ChatGPT, and then respond to that prompt.

If I enter 'Stop', you should end the entire prompt process."

Now we will send the above prompt to ChatGPT. Since the response was quite long, I didn't use print-screen to capture it. Instead, I copied the entire

response in text as follows.

The response from ChatGPT of this prompt :

ChatGP T:	Begin. What is this prompt about?
You:	Help me write a short article on Facebook.
ChatGP T:	<p>Prompt: Write a short, engaging article suitable for Facebook that captures attention quickly, using simple language and a conversational tone. Focus on a topic relevant to the audience's interests, such as a recent event, helpful tips, or a personal story that resonates emotionally. Ensure the post is concise, includes a call to action, and incorporates visuals or emojis to increase engagement.</p> <p>Questions:</p> <ol style="list-style-type: none">1. What specific topic or theme you would focus on for the Facebook article?2. Who is your target audience, and what is the primary message you want them to take away from this article? <p>Suggestions:</p> <ol style="list-style-type: none">1. Consider including a question at the beginning of the article to engage readers and encourage comments.2. Use a mix of short sentences and bullet points to make the content easily scannable and more appealing on Facebook. <p>Please choose one of the following options:</p> <ol style="list-style-type: none">1. Provide more information or answer my questions,2. Enter 'Use',3. Enter 'Stop'.
You:	A one-day trip to Alisan mountain in Taiwan.
ChatGP	Prompt:

T: Write a short, engaging Facebook article about a one-day trip to Alishan Mountain in Taiwan. Highlight key attractions such as the sunrise view, the scenic forest railway, and the iconic giant trees. Use a conversational tone, include practical tips like the best time to visit, and suggest what to pack for the trip. Keep it concise, focus on creating a sense of adventure, and encourage readers to share their experiences in the comments.

Questions:

1. Do you want to focus on any specific activities or spots at Alishan, such as the sunrise or the forest trails?
2. Are there any particular tips or personal anecdotes you would like to include to make the article more relatable?

Suggestions:

1. Add a fun fact or historical tidbit about Alishan to make the article more informative and engaging.
2. Use a call to action at the end, such as asking readers to tag someone they'd like to visit Alishan with.

Please choose one of the following options:

1. Provide more information or answer my questions,
2. Enter 'Use',
3. Enter 'Stop'.

You: Describe Sunrise in Alishan

ChatGP T: **Prompt:** Write a captivating Facebook article describing the sunrise experience at Alishan in Taiwan. Focus on the breathtaking colors of the sky as the sun rises over the sea of clouds, the tranquility of the early morning, and the awe-inspiring beauty of the surrounding mountains. Use vivid imagery to transport the reader to the scene, and suggest the best viewing spots and times. Encourage readers to wake up early and experience this magical moment for themselves.

Questions:

1. Would you like to include any specific viewing spots for the sunrise in Alishan that you recommend?
2. Do you want to mention the cultural or spiritual significance of the sunrise at Alishan in the article?

Suggestions:

1. Consider adding a tip on how to get to the best viewing spots easily, especially for first-time visitors.
2. Include a quote or a personal reflection to make the description more relatable and emotional.

Please choose one of the following options:

1. Provide more information or answer my questions,
2. Enter 'Use',
3. Enter 'Stop'.

You:	Use
ChatGP T:	<p>Here's the response based on the prompt:</p> <hr/> <p>“The Magic of Sunrise at Alishan: A Must-See Experience”</p> <p>Imagine standing atop Alishan Mountain in the early morning, the air crisp and cool as you await the first light of dawn. As the sun begins to rise, the sky transforms into a canvas of vibrant colors—soft pinks, warm oranges, and golden hues blending seamlessly into the deep blue of the fading night. Below, a sea of clouds stretches out, creating an ethereal landscape that feels like something out of a dream.</p> <p>The tranquility of this moment is unparalleled. The world is quiet, save for the gentle rustle of leaves and the distant call of birds waking to the new day. As the sun climbs higher, its rays illuminate the peaks of the surrounding mountains, casting long shadows and highlighting the rugged beauty of the terrain.</p>

	<p>Alishan's sunrise is more than just a visual spectacle; it's a spiritual experience that touches the soul. The best viewing spots, like the Alishan Sunrise Platform or the Ogasawara Mountain, offer unobstructed views, ensuring you won't miss a second of this natural wonder.</p> <p>So, set your alarm early, pack a warm jacket, and prepare to witness one of Taiwan's most iconic sights. Trust me, it's worth every minute of lost sleep. If you've experienced the Alishan sunrise, share your story below—let's inspire others to see this magic for themselves!</p> <hr/> <p>Would you like any further adjustments or additions to this article?</p>
You:	Stop
ChatGP T:	Understood. The process has been completed. If you need further assistance in the future, feel free to ask!

After two rounds of refinement, ChatGPT generates this prompt:

" Write a captivating Facebook article describing the sunrise experience at Alishan in Taiwan. Focus on the breathtaking colors of the sky as the sun rises over the sea of clouds, the tranquility of the early morning, and the awe-inspiring beauty of the surrounding mountains. Use vivid imagery to transport the reader to the scene, and suggest the best viewing spots and times. Encourage readers to wake up early and experience this magical moment for themselves."

Discussion:

- 1) Automatic prompt design is an important method for creating prompts. It allows ChatGPT to ask the user a series of questions, helping the user identify the key elements that matter most to them. After a few iterations, a highly efficient prompt is generated, requiring only interaction with ChatGPT from the user.
- 2) In the prompt above, we included a reviewer role for ChatGPT. This is because ChatGPT has a distinct personality, and when overwhelmed, its responses can become unstable, requiring close monitoring to identify mistakes and continue effectively.
- 3) Although ChatGPT may sometimes "fumble," mastering its use will still

result in highly effective prompts.

Here, we have demonstrated how to design a prompt through interaction with ChatGPT, but how efficient is this generated prompt?

Building on our initial exploration of prompt design, let's now examine how ChatGPT can review and enhance its own prompts.

2. Reviewing the Generated Prompt

As mentioned, requests to ChatGPT can be quantified, and the efficiency of automatically generated prompts can be measured.

ChatGPT can not only design prompts but also critique them and suggest improvements. Whether it's a prompt we've created ourselves or one generated by ChatGPT, this method can be used to refine and enhance it.

But what standards should we use to review the prompt?

You can set your own standards or ask ChatGPT directly as shown in Figure V.04.

Please briefly state the top 5 criteria when evaluating a ChatGPT prompt. List them in order of importance.



The top 5 criteria for evaluating a ChatGPT prompt, listed in order of importance, are:

1. Clarity: The prompt should be clear and unambiguous to ensure accurate responses.
2. Relevance: The prompt must be directly related to the desired output or context.
3. Specificity: The prompt should be specific enough to guide the model towards the expected response.
4. Brevity: The prompt should be concise, avoiding unnecessary details.
5. Context: The prompt should provide enough context to understand the request fully.

Figure V.04: Five top criteria when evaluating a prompt.

We will use the top criteria provided by ChatGPT to review the prompt generated earlier: "A Facebook post about a one-day trip to Alishan, with a

special emphasis on the sunrise."

Prompt:

"Please evaluate the prompt I provided and give it a score of 1 to 10 based on the following five criteria (Clarity, Relevance, Specificity, Brevity, Context). Then provide an overall score.

Finally, under the condition of maintaining the same topic coverage and improving the prompt's effectiveness without changing its essence, please provide suggestions for improvement for each criterion.

{prompt}= Write a captivating Facebook article describing the sunrise experience at Alishan in Taiwan. Focus on the breathtaking colors of the sky as the sun rises over the sea of clouds, the tranquility of the early morning, and the awe-inspiring beauty of the surrounding mountains. Use vivid imagery to transport the reader to the scene, and suggest the best viewing spots and times. Encourage readers to wake up early and experience this magical moment for themselves."

Below is ChatGPT's evaluation and suggestions in text form.

Evaluation of the Prompt:

- 1. Clarity (8/10):** The prompt is generally clear in its intention. It specifies the content to be covered (sunrise at Alishan, sky colors, tranquility, mountains) and the tone to be used (captivating, vivid imagery). However, it could be slightly more concise to enhance clarity.
- 2. Relevance (10/10):** The prompt is highly relevant for its intended purpose—creating a captivating Facebook article. It aligns well with the goals of engaging readers and encouraging them to experience the sunrise at Alishan.
- 3. Specificity (9/10):** The prompt provides detailed instructions about what to focus on (colors, tranquility, mountains) and even suggests including viewing spots and times. However, it could be more specific about what "vivid imagery" means or examples of how to convey the experience.

- 4. Brevity (7/10):** While the prompt is detailed, it could be slightly more concise. The instruction "Use vivid imagery to transport the reader to the scene" could be merged or simplified with other instructions to reduce wordiness.
- 5. Context (9/10):** The prompt provides enough context for the writer to understand the setting (Alishan in Taiwan) and the goal (captivating article), but it could include a brief mention of the target audience or the style/tone expected for a Facebook post.

Overall Score: 8.6/10

Suggestions for Improvement:

- 1. Clarity:**
 - Suggestion: Rephrase for more concise clarity. For example, "Describe the breathtaking sunrise at Alishan, focusing on the vibrant colors of the sky, the tranquil morning, and the majestic mountains. Highlight the best viewing spots and times."
- 2. Relevance:**
 - Suggestion: No major improvements needed. The prompt is well-aligned with the task. Consider mentioning the audience (e.g., travelers, nature lovers) to further enhance relevance.
- 3. Specificity:**
 - Suggestion: Provide examples of the kind of "vivid imagery" desired, or mention specific elements to include (e.g., "the sky's transition from purple to gold"). This would help ensure the writer understands the level of detail expected.
- 4. Brevity:**
 - Suggestion: Streamline the prompt by merging similar instructions. For instance, "Transport readers with vivid imagery and suggest the best viewing spots and times." This could replace longer phrases.
- 5. Context:**

- Suggestion: Include a note on the tone or style expected for the Facebook article, such as, "Use a friendly, engaging tone suitable for social media." This provides more context for the writer.

Revised Prompt:

"Write a captivating Facebook post about the sunrise at Alishan in Taiwan. Describe the vibrant sky as it changes colors over the sea of clouds, the peacefulness of the early morning, and the majestic surrounding mountains. Use vivid imagery to transport readers to the scene, and recommend the best viewing spots and times. Encourage readers to experience this magical moment for themselves."

This revision retains the prompt's essence while improving clarity, specificity, brevity, and context.

Here, we demonstrated ChatGPT reviewing its own prompt. Since ChatGPT's results often contain random elements, meaning the same question can yield different answers, it's necessary to have ChatGPT evaluate its own responses to improve the efficiency of its answers.

After exploring two distinct applications, we will now look at the third technique: using quantifiable prompts to fine-tune ChatGPT's responses.

3. Techniques for Effective Communication

Have you ever asked ChatGPT to help write a draft, only to find that the result wasn't quite what you wanted? Sometimes the style is too formal, so you ask ChatGPT to make it more casual, but then it goes too far and becomes somewhat humorous. After several back-and-forth adjustments, you finally get a satisfactory result.

Is this ChatGPT's fault?

Actually, no. Consider the language we used earlier: "make it more casual." This is an adjective, and everyone might interpret it differently. What you consider "casual" might be what I think of as "slightly restrained."

So, continuing to ask vague questions might take several attempts before you

achieve a satisfactory result.

Can users quickly achieve a satisfactory result by using quantifiable prompts?

Since ChatGPT operates on a computer, if we provide quantifiable standards when asking it to make modifications, we're more likely to get a satisfactory result.

How do you give a quantifiable prompt? It requires establishing a common standard with ChatGPT. How do you establish a common standard? The answer lies in “communication skills!”

Let me explain with an example from my field of expertise. We often coach speakers and provide feedback on their speeches and presentations.

Feedback itself is a form of communication, so there's also the issue of being “too formal” rather than conversational.

Additionally, there's a key aspect of feedback: you want to offer constructive suggestions, but you also don't want to be too critical, as it could undermine the speaker's confidence. Therefore, the tone of feedback should not be too harsh.

To demonstrate this technique, let's consider an example where ChatGPT is tasked with drafting feedback for a speech titled “Revealing the Future of AI.” The speaker's original script is not included here because the focus is on the feedback draft, which we will critique and modify.

This is the feedback draft written by ChatGPT:

Draft of speech evaluation for “Revealing the Future of AI”:

“Dear ladies and gentlemen,

Today, I have the honor of evaluating Mr. Alex's speech on 'Revealing the Future of AI.' Mr. Alex led us on an in-depth exploration of the field of AI, discussing its current achievements, future potential, and the ethical challenges we face as a society. My evaluation will be focused on the following aspects:

Content and Structure: Mr. Alex's speech is well-structured, but it lacks depth. Although practical examples from healthcare and autonomous driving were used, the discussion remained superficial. While the topic was made easier to understand, it lacked the depth of critical thinking.

Presentation Style: Mr. Alex's presentation style has its bright spots but is not without flaws. The use of pauses effectively highlighted key points, and the variation in tone successfully maintained the audience's attention. However, the overuse of technical jargon without proper explanation not only showcased Mr. Alex's profound knowledge but unfortunately alienated those unfamiliar with AI terminology.

Areas for Improvement: While the overall speech was impressive, there is room for improvement. First, simplifying the technical language or providing brief explanations for complex terms would make the content more accessible to the general public. Secondly, incorporating more storytelling elements could enhance the emotional connection with the audience, making the speech more memorable.

Summary: In conclusion, Mr. Alex's speech provided us with some insights into AI, but it fell short in depth and emotional connection. By adjusting the clarity of the language and strengthening emotional engagement, I believe Mr. Alex's future speeches will have a more lasting impact. Thank you, Mr. Alex, for your speech, which not only educated us but also stimulated our thoughts on the broader impact of AI in our lives."

This draft is somewhat formal and has a slightly harsh tone. ChatGPT can help with both of these issues by making the draft more conversational and softening the tone. But how do we give a "quantifiable" prompt so that it gets it right the first time?

In the previous section, we saw that ChatGPT can self-review, so let's communicate with ChatGPT and first ask it to score the "conversational tone" and "harshness" of this draft.



Evaluation of "Revealing the Future of AI"

Conversational Tone:

The article maintains a somewhat conversational tone. The language is accessible, making it easy for the audience to follow. Phrases like "Today I have the honor" and "Thank you Mr. Alex for your speech" add a friendly and engaging touch. However, the occasional use of technical jargon without explanation slightly detracts from the conversational feel.

Score: 7/10

Seriousness:

The seriousness of the article is evident in its formal evaluation and critique. The structured approach and use of terms like "in-depth exploration" and "ethical challenges" add a layer of seriousness, reflecting the importance of the subject.

Score: 8/10

This evaluation reflects a balance between conversational tone and the seriousness required for evaluating a speech on AI.



Figure V.05: ChatGPT evaluate the draft it created earlier.

All right, we asked ChatGPT directly, and the result was that ChatGPT rated the "conversational tone" = 7/10 and "harshness" = 8/10.

Now, we can give ChatGPT quantifiable requests to help us modify this draft. For example, we want the draft to be more conversational (increasing the score from 7 to 9) and the tone to be softer (reducing the score from 8 to 6.5).

Prompt:

"Please modify this draft so its "conversational tone" = 9/10 and "harshness" = 6.5/10."

The resulted text from ChatGPT is as follows:

Here's a revised draft that aims to increase the "conversational tone" to 9/10 and reduce "harshness" to 6.5/10:

Dear ladies and gentlemen,

Today, I'm delighted to share my thoughts on Mr. Alex's engaging speech, 'Revealing the Future of AI.' Mr. Alex guided us through an intriguing journey into the world of AI, touching on its achievements, potential, and the ethical questions we face. Here's a closer look at what worked well and where there's room for improvement:

Content and Structure:

Mr. Alex's speech was well-organized, and he did a great job of using relatable examples from healthcare and autonomous driving. However, while these examples made the topic easier to grasp, I felt that the discussion could have gone deeper to challenge our thinking even more.

Presentation Style:

Mr. Alex's presentation had some standout moments. He used pauses effectively, which helped to highlight key points, and his varied tone kept the audience engaged. That said, some of the technical terms were a bit overwhelming for those less familiar with AI. Simplifying these or giving quick explanations could make the speech more accessible and enjoyable for everyone.

Areas for Improvement:

The speech was already quite impressive, but there's always a little room to grow! Simplifying the technical language or adding brief explanations for complex terms would really help in reaching a broader audience. Also, incorporating more storytelling could make the speech even more memorable by creating a stronger emotional connection with the listeners.

Summary:

In conclusion, Mr. Alex's speech was enlightening and thought-provoking, offering us a glimpse into the future of AI. With a few tweaks—like clarifying the language and adding some emotional depth—I believe Mr. Alex's future talks will resonate even more. Thank you, Mr. Alex, for an informative and stimulating presentation that has surely sparked new ideas about AI's impact on our lives.

This version uses a friendlier tone, reducing the harshness of the critique while maintaining the informative nature of the feedback.

Compare the modified version with the original draft—doesn't it seem more conversational and gentler in tone? If it's still not satisfactory, you can keep adjusting using quantifiable figures.

Beyond tone and harshness, various other elements can be adjusted quantitatively to achieve more precise outcomes. For instance, when asking ChatGPT to generate a piece in the style of Ang Lee, you can also ask it to score the effectiveness of its results.

Since text generated by models like ChatGPT comes from algorithms with random components, it can't fully control the final output. Therefore, asking it to self-assess and score is crucial for maintaining high-efficiency outcomes.

Chapter 20 Summary:

- ❖ Effective interaction with ChatGPT can lead to many amazing applications.
- ❖ This chapter presents three examples, each building on the previous one.
- ❖ The first example is “automatic prompt design,” a method for creating prompts. ChatGPT asks the user questions to identify key elements. After a few cycles, a highly effective prompt is generated, with minimal user interaction.
- ❖ The second example involves having ChatGPT review its earlier prompt. Since ChatGPT's results may contain random elements, it's sometimes necessary to evaluate the response's effectiveness. If unsatisfactory, you can ask it to retry.
- ❖ The third example shows that if you're dissatisfied with ChatGPT's results, repeatedly asking questions might take several attempts. In such cases, users can use quantifiable methods to refine the prompt and quickly get satisfactory results.

Chapter 21: Custom ChatGPT - GPT

In this chapter, we will introduce one of the most influential products of OpenAI—Custom ChatGPTs, also known as GPTs.

With the advent of GPTs, AI technology has reached a new milestone, transitioning from specialized use to everyday accessibility.

1. What is GPT

In November 2023, OpenAI announced four major developments at its developer conference, with the most significant being the introduction of custom ChatGPTs, known as GPTs.

GPT is a custom ChatGPT assistant developed by OpenAI. Imagine creating a ChatGPT tailored to your specific needs—it could help manage emails, provide creative inspiration on demand, and make your daily life more convenient while enhancing work efficiency.

The term “GPT” can be confusing because the underlying language model is also called “GPT.” When OpenAI refers to this customized version, they use the term “GPTs.”

In this book, “GPT” and “GPTs” refer to the same custom ChatGPT. When we mention the engine behind ChatGPT, the “GPT,” we will specifically clarify that I’m referring to the LLM (Large Language Model).

Essentially, GPTs allow users to fine-tune ChatGPT with specific prompts and knowledge, enabling them to perform all the tasks that ChatGPT can, plus any specialized functions provided by the user.

According to OpenAI's official website: "GPTs are custom ChatGPTs that can meet your different needs. You can input prompts to create various ChatGPTs and make them available to the public. Anyone can easily create their own GPT—no coding required."

Simply put, GPTs are customized versions of ChatGPT tailored to users' specific needs—no coding required.

You can create a GPT tailored to your needs in just 3 to 5 minutes using the prompt techniques we've learned.

How popular are GPTs? In less than three months, users worldwide created over 3 million GPTs. In January 2024, OpenAI launched the GPT Store platform, allowing users to share their GPTs and announced plans to introduce a revenue-sharing model based on usage, compensating creators of popular GPTs.

During the developer conference, OpenAI also announced a major integration of GPT-4's features, combining DALL-E 3, browsing, and voice capabilities. This means ChatGPT will soon be able to "listen, speak, read, and write"! "Read" includes reading PDFs and recognizing images, while "write" includes drawing and generating videos. We won't need to switch windows or use other software for these tasks.

Thanks to its customization, GPTs' knowledge scope has broadened, and their functionality has become more comprehensive. Even without plugins, GPTs already possess a wide range of capabilities, making the transition to AI intelligent agents (Agents) imminent.

Now that we've defined what GPTs are, let's look at how you can start using them to enhance your daily tasks.

2. How to Use GPT

Currently, GPTs are only available to paid versions of ChatGPT. Whether you are creating or using them, it can only be done on the paid version. In May 2024, OpenAI announced that non-paid members would also be able to use GPTs to a limited extent.

Simply go to the ChatGPT homepage (<https://chatgpt.com>), where you will find an "Explore GPTs" button in the top left corner. Clicking it will take you to the following page (<https://chatgpt.com/gpts>) shown in Figure V.06.

This page displays many GPTs created by users. Below the "Featured (Top Picks This Week)" section (1), only two GPTs are shown (MapGPT and Question Maker). Scrolling further down reveals more GPTs.

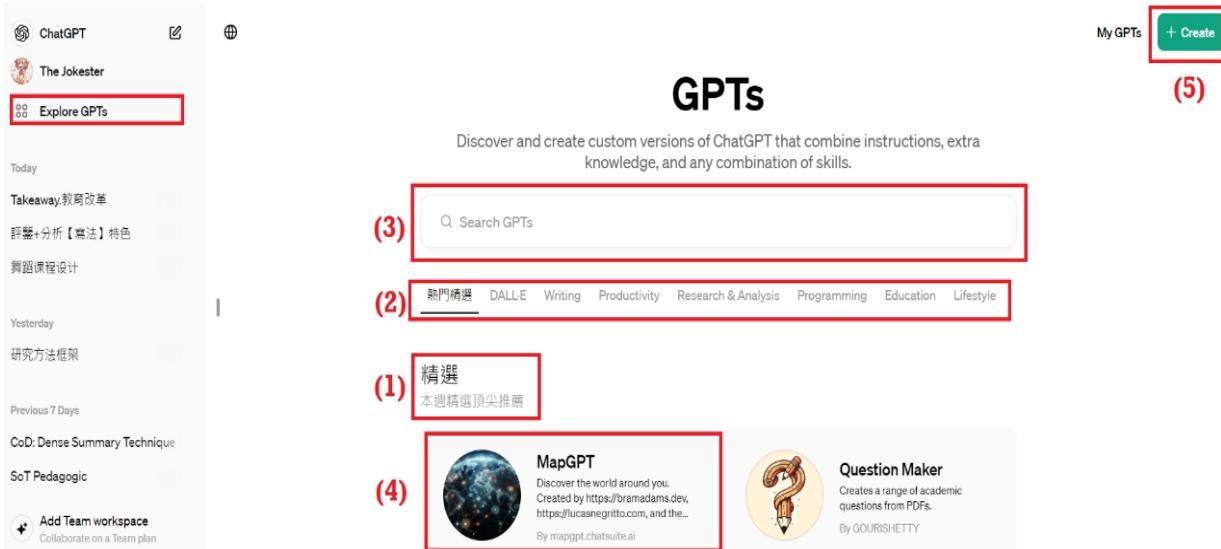


Figure V.06: ChatGPT's GPT page.

In the middle of the page, there is a row of categories (2), listing GPTs by type, including "Popular Picks," DALL-E, Writing, Productivity, Research & Analysis, Programming, Education, and Lifestyle.

Above this column is the "Search Bar" (3), where you can enter the name of a GPT you've heard of or any term to search.

Simply click on any GPT's "image" (4), and you will enter its dialogue page and start using it. The usage is similar to operating ChatGPT.

The uses of GPTs are limitless, covering everything from work efficiency to life convenience. Even the popular text-to-image generator, DALL-E, is an official GPT from OpenAI.

Having covered how to use GPTs, we will now guide you through the process of creating your own custom GPT tailored to your needs.

3. How to make an GPT

Creating a GPT that many people can use requires some effort. After all, a GPT is essentially an API, and not long ago, APIs could only be created by software engineers.

However, we can apply the prompt techniques we've learned earlier to create GPTs, turning commonly used prompts into GPTs that serve as our helpful assistants.

When creating a GPT, you can experiment boldly—if you're not satisfied, you can always come back and make changes, or even delete it altogether.

Let's dive into the detailed steps to create a custom GPT, which can be done easily with the right guidance. In the upper right corner of Figure V.06, there's a “+Create” button (5). Simply click it, and you'll be taken to the “GPT Builder” page shown in Figure V.07.

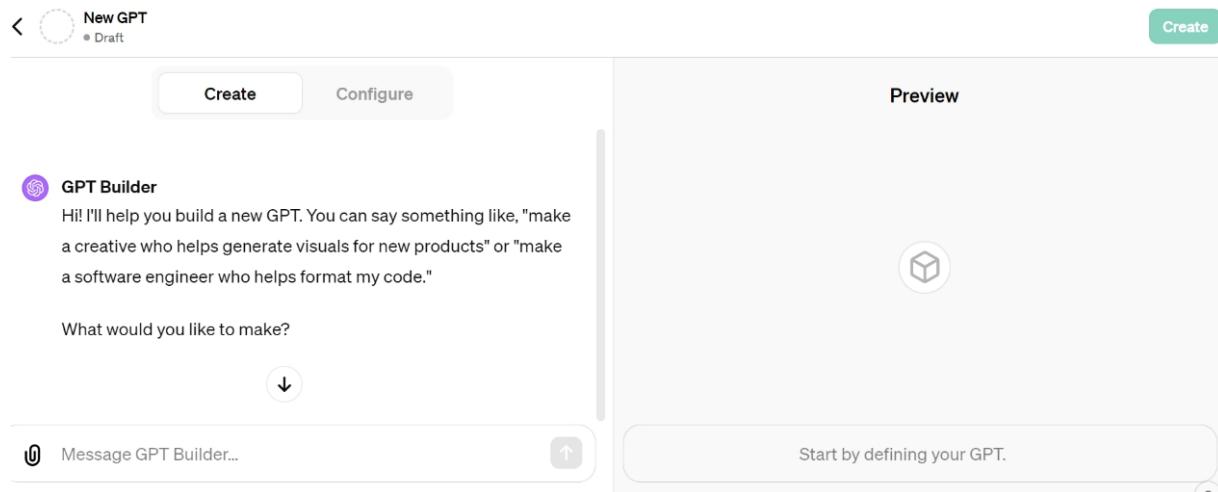


Figure V.07: GPT Builder – two methods; create and configure.

Figure V.07 shows the GPT builder, which offers two options: "Create" and "Configure."

The “Create” option allows for in-depth customization by training the model with specific data, while “Configure” enables quick adjustments to an existing GPT’s settings and behavior without retraining.

The left side of the page is the setup screen, while the right side is the testing screen. After configuring on the left, you can proceed to the right to test your GPT.

The difference between these two is that "Create" focuses on deep customization through training, allowing the model to learn from your data, while "Configure" is a more efficient option, focusing on modifying existing models to quickly meet your specific requirements.

To illustrate the process, here's a step-by-step example of creating a GPT using the “Create” option in the GPT Builder. We chose “Create” to

demonstrate the process because it allows me to show how a custom model can be tailored to specific needs, even with limited data, making it easier to understand and follow.

We will turn the "Skeleton of Thought" (SoT) technique, demonstrated in Part Four, into a GPT. This way, whenever we need to summarize an article, we can call upon this GPT to use the SoT technique for summarizing and condensing the content.

1) GPT Name: "SoT Summarizer"

Function is to summarize the input text and identify the "Skeleton of Thought" or expand on the branches of the Skeleton of Thought.

2) Overall Design:

Use the 3W-CoSeed framework, set the following prompt elements:

Role (Who): A 30-year-experienced editor and reviewer.

Context (Where): Operates within the "domain" provided by the user.

Task (What/Who): Generates a "general summary," "skeleton," and "branch summaries" for the input article. Each generated result needs to be evaluated (Evaluate).

Steps:

Provide instructions and Receive the article;

Ask for the domain and language;

Generate a general summary and evaluate it;

Generate the skeleton and evaluate it;

Generate each branch's summary and evaluate it.

Details of the prompts are placed in the "Step 4" form-filling section.

3) Create the GPT:

The GPT Builder's "Create" function guides you step by step through naming, designing the logo, and crafting the prompts.

The following are the steps to create the GPT:

Step 1: Enter the GPT builder, then select "Create."

As shown in Figure V.07, click "Create" on the left side of the screen (with "Configure" dimly visible on the right), and you will enter the creation page shown below.

Step 2: In Figure V.08 below, the upper left corner displays "New GPT" (1), and the bold "Create" (2) below indicates that you are using the "Create" function.

In the input box below, you can start answering questions. When the GPT builder asks what function the GPT should have, The name "SoT Summarizer." Is directly entered.

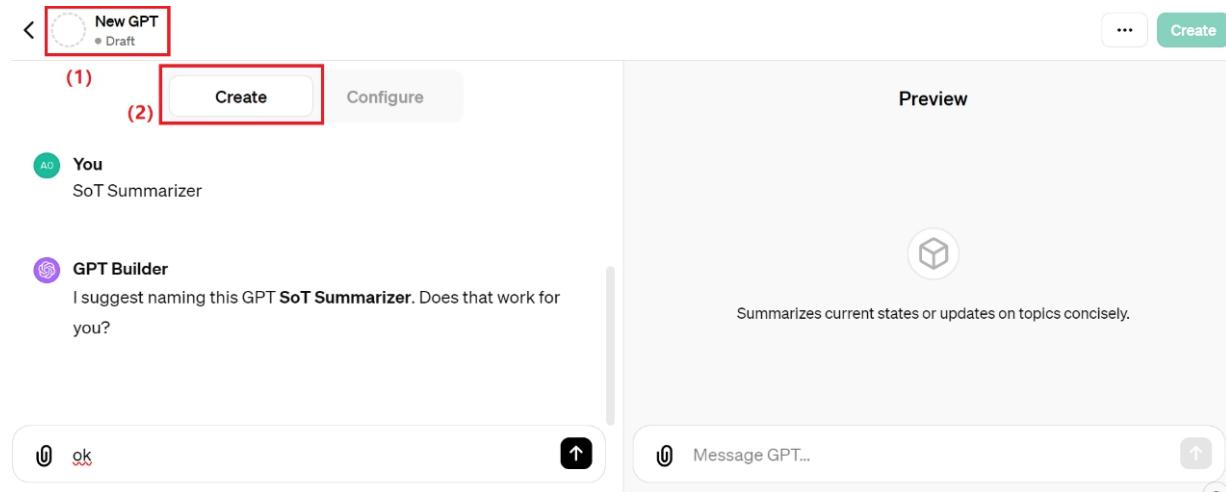


Figure V.08: Making GPT (1) - submit the name “SoT Summarizer”

Step 3: After reviewing the builder’s response, type "ok" into the input window and click the small black arrow next to it to submit. The screen will display the image below in Figure V.09, where it automatically designs a logo. If satisfied with the logo, type “I like it” and submit.

At this point, the description under the small logo on the right side is inaccurate (Summarizes Sea of Thieves tales); it can be corrected later.

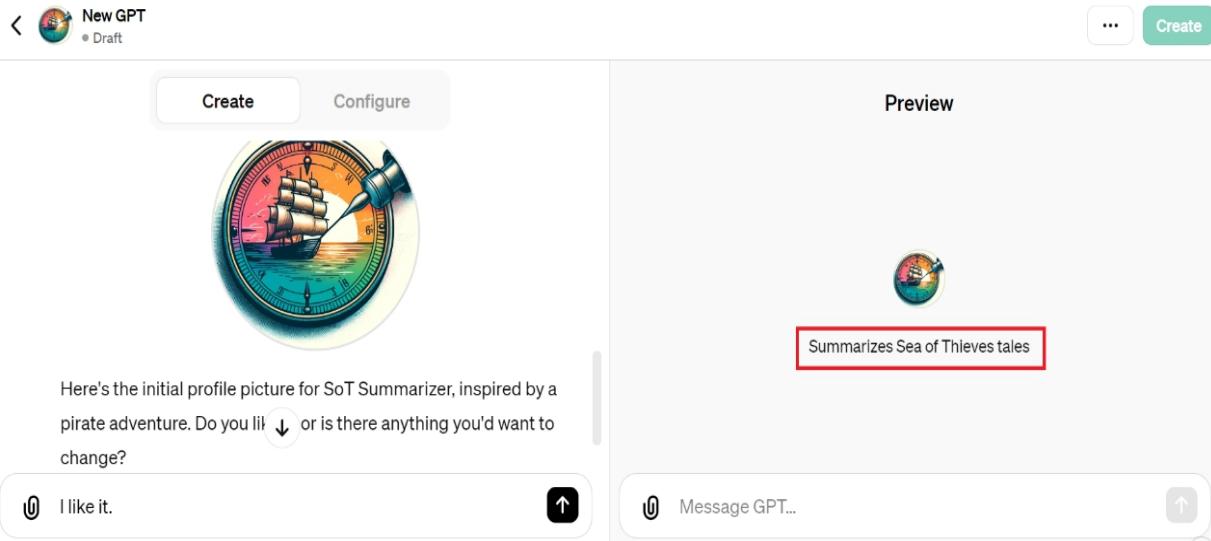


Figure V.09: Making GPT (2) - automatically design the logo.

Type "I like it" and click the small black arrow to submit. You will see the page shown in Figure V.10.

Step 4: On that page, enter the following prompt into the dialogue box:

Prompt:

[You will act as a 30-year-experienced editor and a 30-year-experienced reviewer (Role = Who). You need to execute all tasks according to the "domain" and "language" specified (Context = Where). You will summarize and review the generated "general summary," "outline," and "branch expansions" (Task = What).

Please follow the steps below:

First, present the following content: "Explanation: This GPT will summarize your article. There are two workers here, one is an editor, and the other is a reviewer. You can upload any article or copy it here."

Second, receive input:

- 2-1. Ask: "Does this article belong to any specific 'domain?' If so, please enter the domain name; if not, enter 'No'."
- 2-2. Ask: "Please enter the 'language' to be used for summarizing."
- 2-3. Ask: "What is your 'task'? You can enter text at any time to initiate a task. (1) General Summary; (2) Outline; (3) Branch Expansion."

What you need to do:

You will play the role of an editor with 30 years of experience in this "domain," familiar with all topics, and also act as a senior reviewer to review the material generated by the editor.

Step 1: When receiving the user's input, if the "task" is "General Summary," first summarize the article according to the "domain" and "language." Then, have the reviewer evaluate whether the summary follows the specific terminology of the "domain." The reviewer also needs to review and polish the text. Display the result using the specified "language," first showing the "domain" and then the "summary."

Step 2: When receiving the user's input, if the "task" is "Skeleton," identify the "outline" of the article. Display the result using the specified "language," first showing the "domain" and then the "outline."

Step 3: When receiving the user's input, if the "task" is "Branch Expansion," summarize the information related to each branch of the outline in sequence. Display the result using the specified "language," first showing the "domain" and then these "branch expansions."

After completing the above steps, ask the user, "What is the next 'task'?"

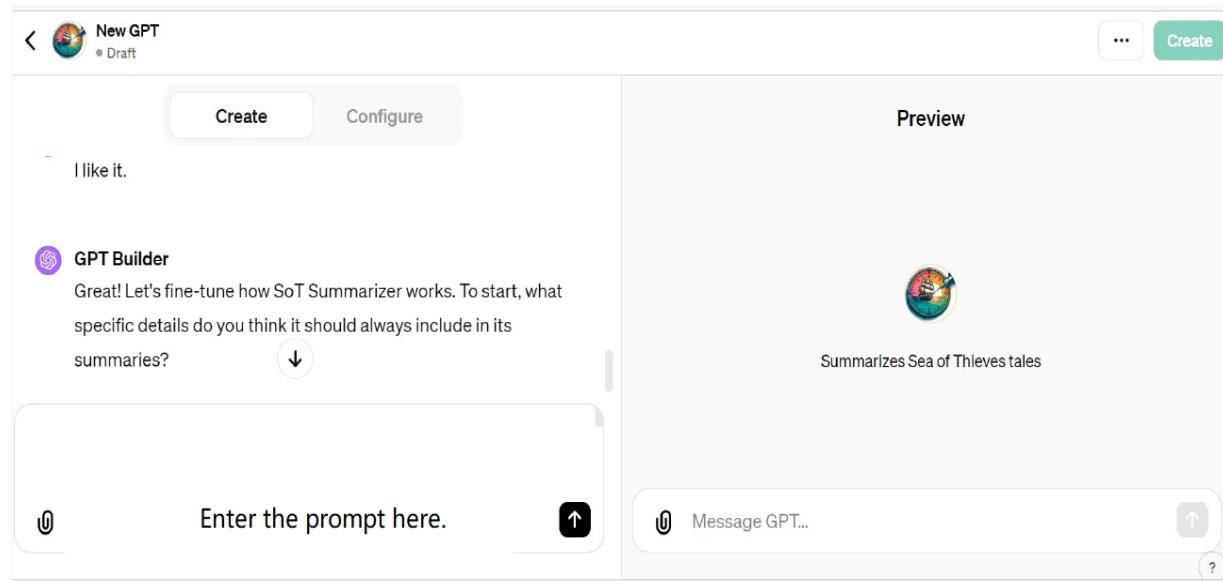


Figure V.10: Making GPT (3) - enter the prompt into input box.

After the builder receives the prompt, it realized that the prompt and task we provided didn't match the original logo it designed, so it asked if we wanted to redesign the logo. Very smart!

Since we can always change it later, we decided to complete the creation process for now.

Click the green “Create” button in the top right of Figure V.11, and this GPT will be saved in your personal account under “My GPTs.” If you’re not satisfied with this GPT, click the three small dots to the left of the green “Create” button, where you’ll find the option to delete this GPT.

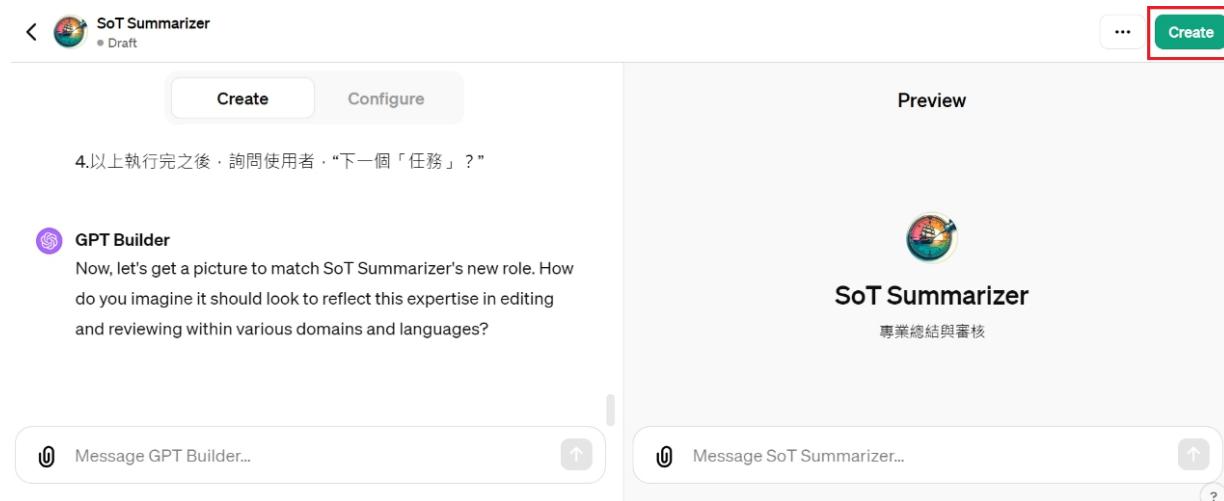


Figure V.11: Making GPT (4) - complete the process.

After creating your GPT, you can choose to keep it private, share it with specific users, or publish it to the GPT Store. Publishing allows other users to access your GPT, potentially leading to broader use and feedback.

After clicking “Create,” the following options will appear as in Figure V.12:

- Only me - Use it only for yourself;
- Anyone with the link - Make it available to any ChatGPT paid member; or
- Publish it to GPT Store.

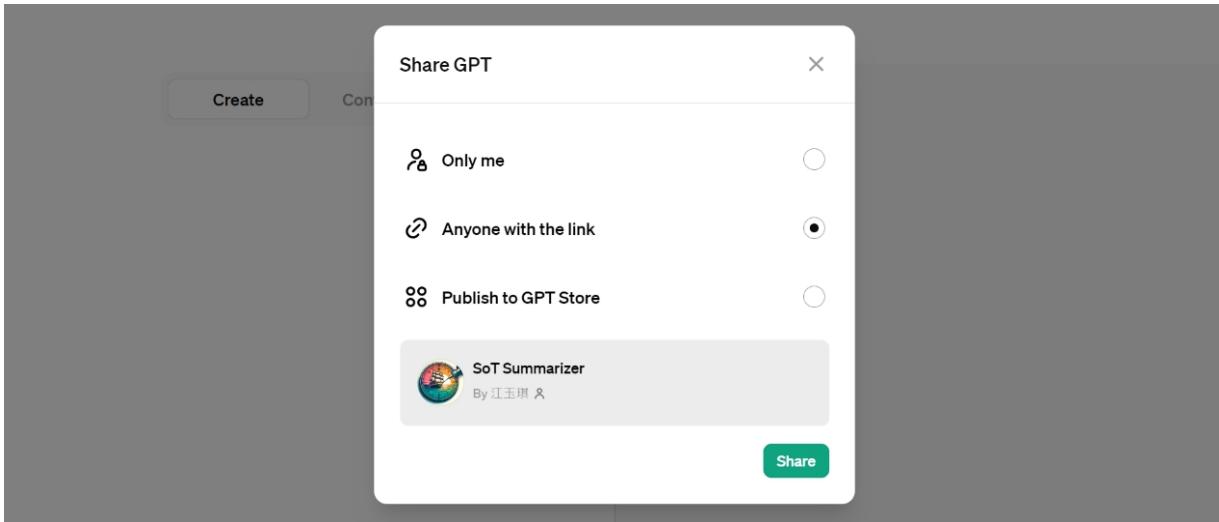


Figure V.12: Making GPT (5) - choose how to publicize the GPT.

I chose the second option. After clicking the “Share” button in the bottom right, any paid ChatGPT member can use this GPT with the link below:
<https://chat.openai.com/g/g-uOMCDMtjz-sot-summarizer>.

In this chapter, we explored custom ChatGPTs, or GPTs, introduced by OpenAI as a major advancement in AI technology. GPTs allow users to create personalized assistants tailored to their specific needs without coding. We discussed how GPTs are already transforming daily tasks and professional workflows, making AI accessible to everyone.

We also walked through the process of creating a GPT, highlighting the simplicity and power of this tool. As GPTs continue to evolve, their impact on productivity and creativity will only increase, further integrating advanced AI into everyday life.

As we've seen, GPTs have the power to transform everyday tasks and workflows, but their potential doesn't stop here. In the next section, we'll delve into the future of GPTs and AI, examining how these advancements could reshape industries, redefine workflows, and unlock new possibilities.

Chapter 22: Future Trends

As we reach this chapter, we've thoroughly explored the technical aspects of prompt engineering.

From the basics in Part One, principles in Part Two, tactics in Part Three, and strategies in Part Four, to practical applications in Part Five, we've progressed step by step together. We hope you now have a deeper understanding of the latest developments and applications in prompt engineering technology.

Now, let's shift our focus to the future trends in AI that can leverage the prompt techniques we've learned.

One of the most anticipated advancements is the emergence of AI agents, which represent the next step in prompt engineering. As these AI agents become increasingly autonomous in problem-solving, they will inevitably evolve toward multi-modal capabilities.

It's clear that with nations and major companies investing vast resources, the future developments in AI we currently envision—including AGI (Artificial General Intelligence) and ASI (Artificial Superintelligence)—will eventually be realized. And this future is not far off.

We will briefly introduce some applications related to these prompt technologies, explain what they are, why they matter, and highlight some interesting recent developments. As for the exact timeline of these advancements, we don't know. But it's definitely in the near future!

1. AI Agents

In prompt engineering, we've learned various techniques for using prompts to enhance ChatGPT's efficiency, making our work easier.

We've also learned how to use prompts to create custom ChatGPTs (GPTs). These GPTs are essentially custom AI assistants (AI Agents).

This is just the beginning of AI's potential to assist us.

Currently, ChatGPT can reason and handle tasks within a given scope, but it's still just generative AI, lacking persistent memory and deep learning

capabilities.

To enable AI to autonomously reason, solve problems, and independently or collaboratively execute tasks with other AIs, we must transition from ChatGPT to AI agents.

To understand the difference between ChatGPT and AI agents, let's explore the concepts of fast thinking and slow thinking.

In his 2011 book *Thinking, Fast and Slow*, Nobel laureate Daniel Kahneman categorized thinking into two modes:

- 1) Fast thinking: Quick, intuitive, and emotional, like picking up a book that has fallen to the floor.
- 2) Slow thinking: Deliberate and logical, like planning a vacation.

ChatGPT currently operates in the fast-thinking mode, responding to each prompt with a single reply.

In previous Chapters, we discussed having ChatGPT make decisions and start "planning," but it was still us—humans—doing the thinking and leading. We decompose tasks into smaller ones and set decision points using prompts, while ChatGPT only executes specific tasks when prompted. This is fast thinking.

When ChatGPT begins to undertake human-like thinking and planning, we enter the realm of AI agents, which involves giving AI the capability for slow thinking.

Slow thinking is needed to solve complex problems involving exploration, trial and error, and execution. For example, starting a company, creating a business model, organizing a team, generating revenue, and building a brand—all involve complex scenarios.

1) What is an AI Agent?

An AI agent is an AI with the ability to think and act autonomously, with persistent memory, the capacity for slow thinking, and the ability to understand tasks independently, plan strategies, execute tasks, and achieve results.

This is the primary distinction between AI agents and ChatGPT.

There's a project called AutoGPT built on ChatGPT. You give it a goal, and it creates a plan, continuously validating through trial and error until it finds the optimal solution.

AutoGPT can autonomously execute tasks based on user needs without intervention, handling tasks like daily event analysis, marketing plan writing, coding, and mathematical calculations. This is an early form of AI agent.

In fact, the task decomposition, skeleton of thought, and chain of density techniques discussed in the advanced prompt techniques of Part Four are preliminary behaviors of AI agents. For example, in the skeleton of thought, there's already the behavior of validation and trial and error.

After reading this book, you've laid the foundation for stepping into the world of AI agents. To further differentiate AI agents from ChatGPT, let's examine their unique characteristics.

2) Characteristics of AI Agents

- AI agents understand the human world; they can perceive the physical world, remember past experiences, adjust plans and behaviors based on environmental changes, and use tools to assist themselves.
- AI agents have long-term memory, allowing them to remember their perceptions of the environment, form knowledge and worldviews, recall their plans, and remember multi-step decisions.
- Humans often collaborate with others to solve problems, and AI agents will also work with other AI agents to solve problems together.
- Differences between AI agents and ChatGPT:
 - ChatGPT uses fast thinking: it responds to prompts quickly with a single answer.
 - AI agents use slow thinking: they understand problems, make decisions, and solve complex issues.
 - ChatGPT has short-term memory, retaining only the conversations within a single chat session.
 - AI agents have long-term memory, remembering context and cross-conversation details.

AI agents will have a profound impact as they are more flexible and reliable, addressing the errors that ChatGPT might make. Like many science fiction

scenarios, robots will eventually work alongside humans, either autonomously or as assistants in performing tasks.

3) Real-World AI Agent Experiment

To illustrate AI Agent's capabilities, let's look at a real-world experiment involving AI agents.

In 2023, a significant experiment on AI agents captured public attention. (<https://hai.stanford.edu/news/computational-agents-exhibit-believable-humanlike-behavior>)

Stanford University and Google collaborated on this experiment, creating 25 unique AI characters, each with its own background story and expertise, and had them interact in a simulated small town.

These AI characters, built using generative AI similar to ChatGPT, possessed a high degree of autonomy and planning ability.

Conversations between characters were recorded in their memory devices, allowing researchers to observe and analyze these interactions. The findings revealed that these characters could engage in natural language conversations and form complex social relationships.

In the daily life of the small town, these AI characters not only participated in everyday conversations but also took part in various town affairs, such as elections and festivals. One AI character, Sam, decided to run for mayor, sparking discussions among other residents. The research team was surprised because this campaign was not pre-programmed—it was the result of AI analyzing the town's needs and expectations.

Another AI character, Isabella, was assigned by the research team to organize a Valentine's Day party. Although she was only instructed to invite her boyfriend, she ended up inviting other AI residents and planned activities based on each person's preferences and personalities. This event showcased AI's organizational and coordination abilities.

These AI residents were not just lines of code—they demonstrated the ability to respond, plan, and even dream. The experiment is ongoing, and as technology advances, our expectations of machines continue to grow.

I'm curious to see when they might start arguing or even fighting.

4) Bill Gates' Perspective

Complementing these advancements, Bill Gates offers a perspective on the future impact of AI assistants.

At the end of 2023, Bill Gates shared his views on AI: "AI assistants will take care of everything. Will humans still need education?"

(<https://udn.com/news/story/6811/7581146>)

He predicts that within the next five years, smart assistants will function like a second brain for humans, capable of handling various daily tasks, big and small. These assistants will not be limited to basic conversations or content generation but will operate across applications, learning and predicting users' preferences, and actively executing tasks.

Bill Gates is concerned that as AI becomes more widespread, humans may lose the motivation to learn.

2. Multi-Modal LLM

Building on the idea of advanced AI capabilities, another exciting development is multi-modal large language models.

Multi-modal technology involves combining and analyzing different types of data such as text, images, and sounds, to improve the accuracy of information processing and decision-making.

Multi-modal AI integrates data from various sensory modalities, enabling it to understand complex user needs and environments more comprehensively, thereby achieving more accurate judgments and more efficient interactions.

Currently, tools like ChatGPT primarily process text data and do not inherently support multi-modal input. However, ChatGPT can now leverage OpenAI products like DALL-E and CLIP to handle multi-modal data.

As AI tasks become increasingly complex, AI agents will also require multi-modal technology, allowing them to span multiple domains, handle complex interactions, and make informed decisions.

For AI agents, multi-modal technology not only enhances interaction

capabilities but also improves effectiveness and adaptability in real-world applications. Multi-modal systems provide rich information that transcends a single sensory modality, crucial for interpreting complex or ambiguous user intentions.

For example, in medical assistance, a multi-modal AI agent can more accurately assess the severity and urgency of a patient's condition by analyzing their tone of voice, facial expressions, and textual descriptions.

Another example is in retail, where a customer seeking a specific color of clothing can have a multi-modal AI agent automatically capture the color and their body size, then immediately analyze in-store inventory and respond in natural language. This enhances the customer experience and boosts sales efficiency.

After the developer conference in November 2023, OpenAI began releasing several multi-modal products that integrate text, images, video, and audio data. The goal is to create AI systems capable of comprehensively understanding and generating cross-media content.

OpenAI's products demonstrate its leadership in multi-modal AI technology:

-DALL·E: A revolutionary AI tool that generates detailed images from text descriptions.

-CLIP: Learns to understand images by viewing large numbers of pictures and related text descriptions. By analyzing both images and associated text, it can better understand image content. For instance, if you need to find all the images of "Mars rovers" on the internet, you can show CLIP some pictures and descriptions of Mars rovers, and it will quickly and accurately find all relevant images, even recognizing new ones it hasn't seen before through learned descriptions.

-Whisper: An efficient speech recognition system capable of recognizing and translating speech input in multiple languages.

-Sora: Generates videos from text, though it is not yet available for public testing.

In Addition, two noteworthy technologies emerging from Google and Meta:

-Google's Gemini: A multi-modal generative AI that understands and

generates text, images, and other data types. Its key innovation is integrating information across modalities, enhancing the model's ability to process complex queries and produce relevant outputs. For example, it handles queries with both text and image inputs, delivering consistent results—a challenge for earlier models.

-Meta's Ferret: An advanced multi-modal AI that processes text, images, video, and sound simultaneously. Ferret's main innovation is integrating data from various sources for a holistic understanding of user queries. For example, it analyzes video content, audio tracks, and text descriptions to provide accurate content identification and classification.

Multi-modal technology can be further extended and combined with edge AI, enhancing the functionality of AI agents.

In May 2024, OpenAI and Google respectively launched the multi-modal ChatGPT-4o and Gemini 1.5, making significant breakthroughs in the field of multi-modal AI, greatly enhancing AI's capabilities in processing and understanding multi-modal data.

3. Artificial General Intelligence

Lastly, let's explore the concept of Artificial General Intelligence (AGI) and its potential to reshape our understanding of AI.

1) OpenAI's Vision

On OpenAI's official website (www.openai.com), it states: "Our vision for the future of AGI: Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity."

This statement highlights two key points:

- AGI is defined as an AI system that is generally smarter than humans.
- AGI can perform intellectual tasks at or beyond human levels across a wide range of domains, significantly differing from current applied AI, like ChatGPT.
- OpenAI's mission is to ensure that AGI benefits all of humanity.

When AGI was first proposed, many people considered it impossible. But as technology advanced, belief in its potential grew.

Let's explore some recent developments.

2) OpenAI's Announced Sora :

2023 marked a pivotal year for text-to-video generation technology. Most tools, like Stable Video Diffusion and Pika, could only produce short clips lasting 3 to 4 seconds, with the longest being no more than 10 seconds.

When OpenAI launched a new product called "Sora," it immediately gained widespread attention. Sora can create realistic and imaginative scenes based on text prompts, producing videos up to a minute long with high picture quality, fully aligned with the user's instructions. OpenAI's demonstrations of Sora-generated videos were astonishing.

For example, with the prompt, "A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage," Sora could generate a 1-minute video.

(<https://openai.com/research/video-generation-models-as-world-simulators>)

To achieve such effects, AI must understand the human physical world and simulate the motion of objects within it. Fully mimicking the physical laws and causal relationships of the real world remains a challenge.

However, compared to other AI tools that generate videos from text, Sora's progress is remarkable. Some even predict that OpenAI could achieve AGI within seven months.

But predicting the arrival of AGI is misleading, as the transition from current AI technology to true AGI is not a sudden leap, but a process divided into several stages.

3) Stages of AGI Development:

AI Agent Stage:

AI exhibits highly specialized efficiency in specific domains or tasks, capable of automating tasks ranging from simple to moderately complex, but lacks cross-domain generality.

Enhanced Learning Stage:

AI begins to learn and adapt across multiple tasks and contexts, learning from experience and applying acquired knowledge in new environments.

Cross-Domain General Stage:

AI starts handling problems in different domains without needing retraining for new fields, showing human-like adaptability in solving both concrete and abstract problems.

Autonomous Learning and Innovation Stage:

AI not only learns new skills and knowledge but also proposes new ideas and solutions, capable of developing and progressing without human guidance.

Beyond Human Stage:

The final stage of AGI, where it surpasses human intelligence and cognitive abilities, solving complex problems, creating new knowledge, understanding, and creating art.

While the precise progress of OpenAI's AGI development remains unknown, it is clear that we are steadily advancing toward a future where AGI is a reality.

Some estimate we are at least in stages 1) or 2).

Some believe that OpenAI is cautiously releasing AGI in stages, allowing society to gradually adapt to its impact.

When will AGI be achieved?

Actually, "Beware, AGI is all around you!"

Appendix

A. List of Directive Prompts

To find out other useful "Direct Prompts" are available, we can directly ask ChatGPT.

I read there are 'Direct Prompts' that can make ChatGPT's responses more efficient.

Examples include: 'Pause and think' and 'Pause and think step-by-step before answering.'

Please list ALL 'Direct Prompts' you know in a table.

The leftmost column should be numbers, the middle should be the 'Direct Prompts', and on the right should be the explanation.



Here is a table listing several "Direct Prompts" that can make ChatGPT's responses more efficient:

#	Direct Prompt	Explanation
1	Pause and think	Encourages the model to take a moment before generating a response, potentially improving accuracy.
2	Pause and think step-by-step before answering	Prompts the model to break down its thought process into steps before responding.

3	Be brief	Instructs the model to provide a concise response.
4	Explain your reasoning	Directs the model to provide the rationale behind its response.
5	Give me a summary	Asks for a brief overview of the topic instead of a detailed explanation.
6	Answer in bullet points	Requests the model to organize its response in a list format.
7	Provide examples	Instructs the model to include examples in its response.
8	Assume the reader is a beginner	Directs the model to simplify its language and concepts for a novice audience.
9	Analyze pros and cons	Asks the model to evaluate both the advantages and disadvantages of a topic.
10	Consider alternative perspectives	Encourages the model to explore different viewpoints on the issue.
11	Focus on key points	Instructs the model to highlight only the most important aspects of the topic.
12	Clarify complex terms	Requests the model to define or explain difficult or technical terms used in its response.
13	Compare and contrast	Prompts the model to identify similarities and differences between two or more subjects.
14	Predict future implications	Asks the model to forecast the potential consequences or outcomes of a situation.
15	Suggest improvements	Instructs the model to provide recommendations for betterment.
16	Explore in detail	Encourages the model to provide a thorough and comprehensive analysis of the topic.
17	Answer with a focus on practicality	Directs the model to emphasize practical aspects and applications in its response.

18	Rephrase in simpler terms	Asks the model to simplify its language for better understanding.
19	List key takeaways	Instructs the model to summarize the most important points or lessons from the discussion.
20	Stay objective	Directs the model to provide an unbiased and neutral response.
21	Discuss historical context	Prompts the model to include relevant historical background in its response.
22	Offer a step-by-step guide	Instructs the model to provide a detailed process or instructions on how to achieve something.
23	What if scenarios	Asks the model to explore hypothetical situations and their possible outcomes.
24	Explain with an analogy	Requests the model to clarify a concept using a relatable comparison.
25	Argue both sides	Instructs the model to present arguments for both sides of an issue. 

Figure A1: List of Directive Prompts

B. Output Formats

The following are various output formats that ChatGPT can generate.

1. Please list the top 5 cities in the world with their names and their population figures, presented in bullet point format.

Please list the top 5 cities in the world with their names and their population figures, presented in bullet point format.



Here are the top 5 cities in the world by population:

- Tokyo, Japan: 37.4 million
- Delhi, India: 32.9 million
- Shanghai, China: 29.2 million
- São Paulo, Brazil: 22.6 million
- Mexico City, Mexico: 22.2 million

These figures are based on the most recent estimates for metropolitan areas.

Figure B1: Bullet Point Format

2. Please provide information on the NBA championship teams for 2020 and 2021 in table format, including winning team name, opponent name, date, and venue.

Please provide information on the NBA championship teams for 2020 and 2021 in table format, including the winning team name, opponent team name, date, and venue.



Here is the information on the NBA championship teams for 2020 and 2021 in table format:

Year	Winning Team	Opponent Team	Date	Venue
2020	Los Angeles Lakers	Miami Heat	October 11, 2020	ESPN Wide World of Sports Complex, Orlando, Florida
2021	Milwaukee Bucks	Phoenix Suns	July 20, 2021	Fiserv Forum, Milwaukee, Wisconsin

These details cover the key information about the NBA championships for those years.

Figure B2: Table Format

3. Please provide information on NBA championship teams for 2020 in scorecard format, including the winning team name, opponent name,

winning date, and score.

Please provide information on NBA championship teams for 2020 in scorecard format, including the winning team name, opponent name, winning date, and score.



2020 NBA Finals Scorecard

- Winning Team Name: Los Angeles Lakers
- Opponent Team Name: Miami Heat
- Winning Date: October 11, 2020
- Score: Lakers 106:93 Heat
- Series Result: Lakers 4:2 Heat

Figure B3: Scorecard Format

4. Provide information on NBA championship teams for 2020 and 2021 in CSV format, including the winning team name, opponent name, winning date, and score.

Provide information on NBA championship teams for 2020 and 2021 in CSV format, including the winning team name, opponent name, winning date, and score.



Here is the information in CSV format:

CSV	Copy code
<pre>Year,Winning Team,Opponent Team,Winning Date,Score,Series Result 2020,Los Angeles Lakers,Miami Heat,October 11, 2020,Lakers 106:93 Heat,Lakers 4:2 Heat 2021,Milwaukee Bucks,Phoenix Suns,July 20, 2021,Bucks 105:98 Suns,Bucks 4:2 Suns</pre>	

Figure B4: CSV Format

5. Please plot the interest rates in the United States over the past 12 months in a bar chart.

Please plot the interest rates in the United States over the past 12 months in a bar chart.



Figure B5: Bar Chart Format

6. Please plot the interest rates in the United States over the past 12 months in a line chart.

Please plot the interest rates in the United States over the past 12 months in a line chart.

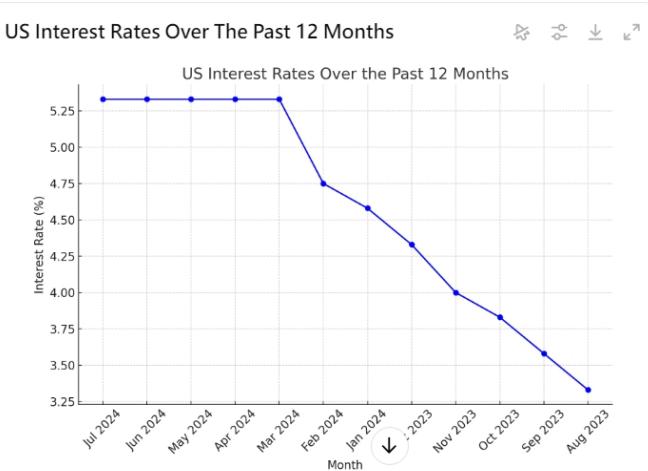


Figure B6: Line Chart Format

7. Please write a simple poem for New Year's greetings presented in HTML format:

Please write a short poem for New Year's greetings presented in HTML format:



Here's an even more concise version of the New Year's greeting poem in HTML format:

```
html
<html>
<body>
    <center>
        <h3>Happy New Year!</h3>
        <p>New dreams rise,<br> Old ones fade.<br> Joy and peace,<br> For the year ahead!</p>
    </center>
</body>
</html>
```

Figure B7: HTML Code

8. Provide information on NBA championship teams for 2020 in JSON format, including the winning team name, opponent name, winning date, score, and series result.

Provide information on NBA championship teams for 2020 in JSON format, including the winning team name, opponent name, winning date, score, and series result.



Here's the information in JSON format:

```
json
```

```
{  
    "year": 2020,  
    "winning_team": "Los Angeles Lakers",  
    "opponent_team": "Miami Heat",  
    "winning_date": "October 11, 2020",  
    "score": "Lakers 106:93 Heat",  
    "series_result": "Lakers 4:2 Heat"  
}
```

```
Copy code
```

Figure B8: JSON Format

9. Please write a Python program that performs the following: "Input two integers X and Y, calculate and return the result of X raised to the power of Y."



Here's a simple Python program that performs the task of calculating X raised to the power of Y :

```
python Copy code

# Function to calculate X raised to the power of Y
def power(x, y):
    return x ** y

# Input two integers X and Y
x = int(input("Enter the base (X): "))
y = int(input("Enter the exponent (Y): "))

# Calculate and display the result
result = power(x, y)
print(f"{x} raised to the power of {y} is: {result}")
```

Figure B9: Python Program

10. Please provide a very simple overview of Markdown's main features, including only headers and lists. Format the response as Markdown.



Markdown Overview

Markdown is a lightweight markup language that you can use to add formatting elements to plaintext text documents. Here's a quick overview of its main features:

Headers

You can create headers by adding `#` symbols before your text. The number of `#` symbols determines the level of the header.

Example:

```
markdown Copy code

# This is a First Level Header
## This is a Second Level Header
### This is a Third Level Header
```

Figure B10-1: Markdown Format- Headers

Lists

Markdown supports ordered and unordered lists.

Unordered Lists

Use `-, `*, or `+` to create unordered lists.

Example:

```
markdown
Copy code
- Item 1
- Item 2
- Item 3
```

Figure B10-2: Markdown Format -Lists

C. Types of Emotional Styles

Here are different emotional styles that can be applied to enhance communication through ChatGPT.

1. Attitude

- Approachable / Unemotional
- Comedic / Sober
- Conversational / Serious
- Flowery / Concise
- Funny / Witty
- Inspirational / Mundane
- Serious / Lighthearted
- Sincere / Sarcastic
- Whimsical / Practical

2. Communication Style

- Adaptable / Rigid
- Confident / Reserved
- Consistent / Spontaneous
- Direct / Indirect
- Engaging / Detached
- Empathetic / Aloof
- Inquisitive / Skeptical
- Persuasive / Neutral
- Respectful / Challenging
- Understanding / Unyielding

3. Register/Terminology/Expression style

- Comfortable / Formalized
- Formal / Informal
- Professional / Casual

4. Tone

- Encouraging / Critical
- Friendly / Assertive
- Humble / Passionate
- Optimistic / Pessimistic
- Respectful / Playful
- Supportive / Detrimental

Afterword

Reflecting on this creative journey, from the initial spark of inspiration to the completion of the manuscript, it has been challenging yet fulfilling. In this book, we have explored the development of generative AI, understood the essence of Prompt Engineering, and delved into the endless possibilities of custom GPTs. This is not just a technical guide but an adventure that intertwines the mind and technology.

During the process of writing this book, we have felt the limitless possibilities brought by the rapid development of technology. From the initial ChatGPT to various generative AIs like Copilot, Gemini, Claude, and others, they have changed the way we interact with the AI world.

The theme of the book, "Prompt Engineering" is actually a professional technique used by AI engineers to fine-tune large language models, usually seen only in academic papers. However, in a world filled with AI, it is too important. My original intention in writing this book was to teach the intricate "Prompt Engineering" to readers without a technical background, enabling everyone to learn and apply it in daily life.

Through this book, we hope to transform these seemingly complex techniques into tools that everyone can master, allowing more people to unleash their creativity and potential in this new era.

Writing this book required more effort than we thought. The technical aspects were manageable, but designing examples was exhausting and required creativity, especially since ChatGPT's responses varied each time, showcasing the love-hate nature of generative AI.

The key message we hope everyone takes away is: do not rely solely on prompt collections. Building a solid foundation will lead to more sustainable development.

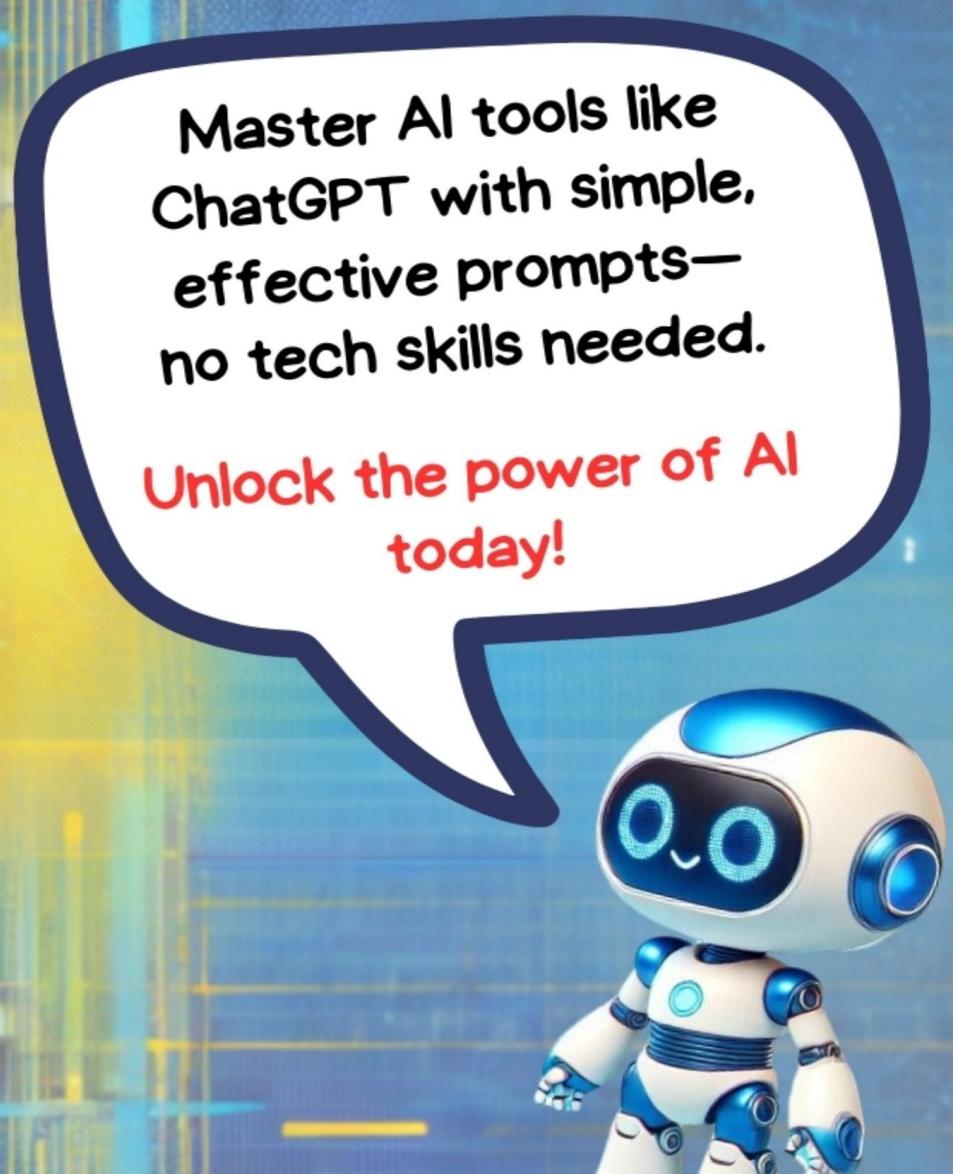
Every chapter and every paragraph in this book contain my passion and respect for this field. Throughout the writing process, we faced many challenges and difficulties.

Sometimes inspiration ran dry, and sometimes resources were scarce. But whenever we felt lost, we would think of those whose lives were changed by AI. Their stories gave me endless motivation and inspiration, making me realize the power of technology lies in changing each of our lives.

The future is unknown but full of hope. As AI technology continues to advance, we will encounter more opportunities and challenges. How to seize opportunities and face challenges in this unpredictable era, becoming leaders in technology and creativity, is a question we all need to ponder. This book is my exploration and answer to this question.

Finally, we want to thank every reader who accompanied me on this journey. We hope this book inspires and helps you. Whether you are a tech enthusiast or a beginner, may we all find our place and shine in this AI era.

Thank you all, and may the future be bright.



Master AI tools like ChatGPT with simple, effective prompts—no tech skills needed.

Unlock the power of AI today!