

# Prompt Engineering for Applied Linguistics: Elements, Examples, Techniques, and Strategies

Zhiqing Lin<sup>1</sup>

<sup>1</sup> School of English Studies, Shanghai International Studies University, Shanghai, China

Correspondence: Zhiqing Lin, School of English Studies, Shanghai International Studies University, 1550 Wenxiang Road, Shanghai, 201620, China. E-mail: LinZQ2023@163.com

Received: July 12, 2024

Accepted: August 10, 2024

Online Published: August 12, 2024

doi: 10.5539/elt.v17n9p14

URL: <https://doi.org/10.5539/elt.v17n9p14>

## Abstract

Generative artificial intelligence, represented by large language models (LLMs), has the potential to revolutionize applied linguistics, such as language teaching, language learning, and language testing. However, how to write an effective prompt (i.e., prompt engineering) remains underexplored in applied linguistics. This study aims to elaborate on the important elements of prompt engineering, including persona, audience, contexts, instruction, and output specification. Examples were used to demonstrate how these elements can form an effective prompt in applied linguistics contexts. Besides, this study also delineates several important prompting strategies to handle more complex tasks, such as iterative prompting and few-shot prompting. Most importantly, it provides some practical tips to mitigate the potential shortcomings of LLMs, including data privacy, potential bias, explainability, and hallucinations, from the perspective of prompt engineering. This study highlights the potential applications of LLMs in applied linguistics by prompt engineering and methods in prompt engineering to navigate the potential pitfalls of LLMs, fostering the application of LLMs in applied linguistics effectively and responsibly.

**Keywords:** generative artificial intelligence, large language models, prompt engineering, applied linguistics

## 1. Introduction

The rapid development of artificial intelligence (AI), in particular the emergence of Large Language Models (LLMs) represented by ChatGPT, has brought great potential to the field of applied linguistics. To date, researchers have delved into the applications of LLMs in language testing (Lin & Chen, 2024), language learning (Su et al., 2023), and language teaching (Imran & Almusharraf, 2023). Moreover, many journals also call for papers in special issues on the applications of LLMs in language learning and teaching (e.g., Han, 2024) because people take great interest in how LLMs are capable of assisting and revolutionizing language teaching and language learning (Kohnke et al., 2023).

Prompt engineering is a fundamental skill that humans need when they interact with LLMs, which should be given due attention. Moreover, grasping the technique of prompt engineering is a learned skill that requires expertise and practice (Oppenlaender et al., 2023). The necessity of mastering the principle of prompt engineering lies in several aspects. (1) Prompt engineering can increase the LLMs' performance in generating more personalized material. (2) It can enhance the efficiency of interaction between humans and LLMs, increasing user experience. (3) It can reduce the biases and errors while outputting the materials, which can increase the content quality for the users.

Several studies have been conducted to investigate the effect of prompt engineering (e.g., Henrickson & Meroño-Peñuela, 2023) and demonstrate the importance of prompt engineering in improving the model performance in practical use (Heston & Khun, 2023; Wang et al., 2024). However, most of them are restricted to the field of medical education (Grabb, 2023; Heston & Khun, 2023) and computer science (Shah, 2024; Wang et al., 2024). Although there are several studies on prompt engineering related to applied linguistics (e.g., Giray, 2023; Lee et al., 2023), they were limited in providing a systematic overview of prompt engineering. Most importantly, they did not provide a transparent and reasonable framework for writing and iteratively improving the prompt and did not provide some possible solutions to mitigate potential drawbacks of LLMs, such as

explainability and hallucinations. Therefore, how to write an effective prompt to enhance efficiency and navigate through potential pitfalls of LLMs might still be elusive for some of the stakeholders in applied linguistics.

Therefore, this study seeks to provide comprehensive guidance on mastering prompt engineering techniques, tips, and best practices to achieve optimal outcomes with LLMs. This study has important pedagogical and practical implications. First, language learners benefit from knowing how to effectively interact with LLMs following the methods demonstrated in this study in the course of learning English. Second, for language teaching, English teachers can also follow the prompt engineering guidance given by the study to generate teaching material or even a syllabus, which can greatly shorten their lesson preparation time and largely enrich teaching activities. Third, researchers in applied linguistics can benefit from mastering the skills of prompt engineering on how to use the LLMs to increase their efficiency in conducting research.

This study was guided by the following questions.

- (1) What are the basic elements of an effective prompt?
- (2) How can these elements form an effective prompt?
- (3) What are the most important techniques in prompt engineering?
- (4) What are the prompting engineering techniques to deal with the potential challenges of LLMs in practice?

## 2. Important Elements in Prompt Engineering

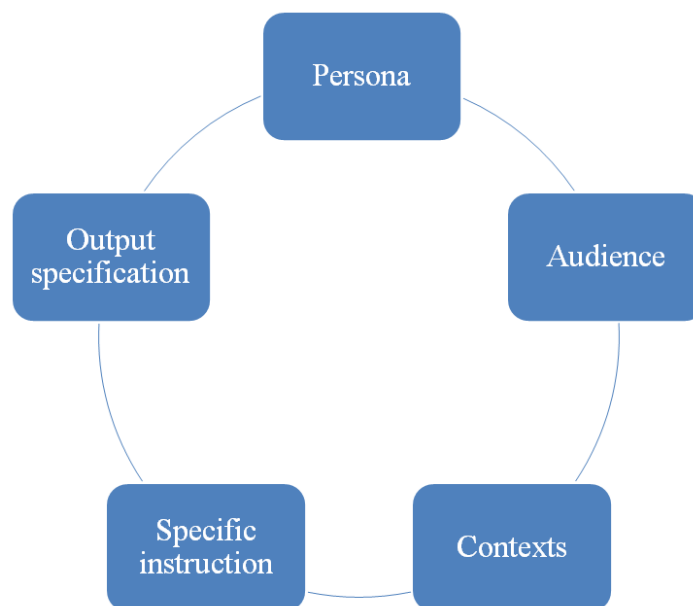


Figure 1. The elements in the prompt

The elements during the prompt engineering proposed by different scholars might not be the same sometimes (Giray, 2023; Lee et al., 2023). However, the most important elements during prompt engineering usually include five elements, as shown in Figure 1, including persona, audience, contexts, specific instruction, and output specifications. Some of them are required, and some of them might be optional. This suggests that the element of a prompt might be task-dependent and task-specific. However, specific instruction is usually required since it is about what LLMs should do in practice.

### 2.1 *Persona (Role)*

In prompt engineering, a persona is defined as the role LLMs are informed to play (White et al., 2023). The concept of persona is usually utilized when conducting prompts to tailor them to the needs and preferences of users. The types of persona depend on the demands of different disciplines. In applied linguistics, the persona LLMs can play is multifaceted according to diverse identities. For students, the persona can be a virtual teacher, an oral English partner, a writing assistant, or a dictionary, among others. For language teachers, the persona can

be a teaching assistant, a test developer, a material developer and an essay rater, etc. For researchers, the persona LLMs can play includes a thesis interpreter, research designer, and so on.

For example, the following prompts are examples of the persona elements.

"I would like you to act as my classmate who can help me enhance my oral English ability"

"I would like you to act as a test developer."

"Provided you are a research designer, you need to generate a research plan for me."

### *2.2 Audience*

Audience refers to the intended audience of the output LLMs generate (Marvin et al., 2024). This element is crucial since it determines whether the content LLMs generate is suitable for the targeted readers. In applied linguistics, this element also determines whether LLMs can serve educational purposes since different audiences might have different needs, and the content generated by LLMs should also be geared to the specific audience group. Concretely speaking, the audience in applied linguistics can be students at different levels, such as English as a foreign language (EFL) learners at kindergarten, EFL young learners, and EFL learners at the university. These learners have different language abilities, different knowledge backgrounds and different psychological characteristics, which suggest that content generated by LLMs should be tailored to their corresponding characteristics. Therefore, this element is very helpful to ensure that the content generated by LLMs is suitable to the intended audience. In prompt engineering, specifying the audience is also very easy. For example, "Suppose you are going to develop a grammatical test for college EFL students whose language ability is about C1". This can inform the LLMs of the audience of the grammatical test it developed.

### *2.3 Contexts*

In prompt engineering, setting a specific context for LLMs is crucial in that it can provide detailed backgrounds for LLMs to understand the task they are expected to do (Shin et al., 2023). Therefore, adding information about contexts in the prompt can enhance the content quality. For instance, the following prompts are examples of context elements of a prompt.

"In this scenario, an English teacher at an international high school where students come from various cultural backgrounds, like English, Singapore, Korean, and Japanese backgrounds. She wants to design language learning tasks that cater to the individual background, catering to cultural differences."

"In this scenario, a researcher in applied linguistics needs to investigate whether students' learning motivation and their emotions will make an impact on their writing proficiency. Please generate a scientific research plan."

### *2.4 Specific Instruction*

Providing specific instructions to LLMs is the most important element in prompt engineering, which also suggests that vague expressions should be avoided in prompt engineering (Gruda, 2024). The more detailed the prompt is, the more likely the generated content can meet our desires. This requirement is quite intuitive since machines need precise information to fulfill the task assigned by humans. Therefore, a general rule for this element (i.e., specific instruction) is to be specific on the task descriptions as much as possible when writing a prompt for LLMs.

However, it was not easy to provide precise instruction to LLMs sometimes because the definition of human language and the definition in the LLMs might not be the same. Therefore, providing the operational definition to LLMs is also an effective method for providing specific task descriptions. This method can help human describe their desired task or outcome more precisely.

### *2.5 Output Specification*

Output specification is the description of the desired output, which can specify the style, tone, and format of output. The parameters in this element are essential when users want to tailor the content to their special needs. Additionally, the format of the output can be diverse, such as in a table format or in a different file format.

In applied linguistics, the potential output specification might cover the lexical and syntactical complexity of the material, such as B1, B2, and C1 based on the CEFR level. Moreover, in terms of the output of English dialects, it can be formulated in American, British, Australian, or Indian English. Besides, this element can also involve requesting the output format. This is very helpful whenever the user (for example, English teachers) wants to use LLMs to create the content and save the content in PPT for later presentation. For example, the following prompts can reveal different styles.

- (1) The output material should be around the B1 level, and the difficulty level of all the words in the material should not be beyond the B1 level.
- (2) Please ensure the output content should include at least three attributive clauses and one adverbial sentence.
- (3) Please generate the content in a table form/an EXCEL file, which can be directly downloaded.

### 3. Putting All the Elements Together

#### 3.1 Examples of Prompts for Potential Applications

After understanding the most essential element of a prompt, the next step is to integrate all the elements to form an effective prompt. This section shows what an effective prompt is and how the element might change the output generated by LLMs.

Table 1. Potential application scenarios in applied linguistics

Domains	Scenarios
Language learning	(1) Listening practitioners
	(2) Speaking partners
	(3) Reading helpers
	(4) Writing Evaluator
Language teaching	(1) Teaching plan writer
	(2) Teaching material writers
	(3) Quiz developer
Researchers	(1) Academic writing assistant
	(2) Reading assistant
	(3) Language proofreader

Table 1 shows the potential applications of LLMs in applied linguistics. Three domains were taken as examples, including language learning, language teaching, and research in applied linguistics. These three domains were taken as examples since they are the most commonly seen subareas in applied linguistics in practice. It is acknowledged that there are also other application scenarios in applied linguistics, such as forensic linguistics, psycholinguistics, etc. Due to the space limitations, they are not discussed in this study.

##### 3.1.1 Language Learning

The following are the prompts that are designed by the author according to elements mentioned in previous sections for language learning. When it comes to multimodal data (e.g., voice), some plugins, such as voice control for ChatGPT, can facilitate the interaction between humans and machines.

###### (1) Listening Practitioners

Suppose you are an English broadcaster, you need to generate two items of each listening subskill I give you, and you should broadcast both the listening material and items for me. The listening subskills include listening for gist, listening for specific information, and listening for detailed understanding. Each item should be broadcast twice. Meanwhile, the item difficulty should be around the B2 level according to the CEFR standard.

###### (2) Speaking Partners

I am a college student, and I am going to take part in the oral IELTS test one week later. Now, suppose you are an English tutor, and you need to practice oral English with me. The oral tasks include three conversations, and you need to have conversations with me on the following topics, namely, talking about my study experience, introducing one of my favorite books, and sharing one interesting film. Moreover, after the conversation, you need to evaluate my performance and give me some comments on how to improve my oral English.

###### (3) Reading Helper

Suppose you are a college English teacher. I have finished a CET-4 reading test, but I have some puzzles about some items. I will show you the reading material and also the items, and then you need to explain why this option is correct and also point out why the other options are wrong. Furthermore, you need to translate the whole passage into Chinese to help me answer the material better.

#### (4) Writing Evaluator

At present, I need to write an English essay about whether we should study English any longer under the influence of artificial intelligence. I have no idea about how to write the essay. Suppose you are an English teacher teaching writing. First, you need to help me conceive the structure of the essay and give me some inspiration before writing. Second, while writing, you need to help me translate some sentences and provide me with more synonym words. Finally, after writing, you need to evaluate it from different perspectives and polish my essay accordingly.

#### 3.1.2 Language Teaching

The following are the prompts that are designed by the author according to elements mentioned in previous sections for language teaching.

##### (1) Teaching Plan Writer

I am an English teacher in a senior high school, and I need to prepare a lesson on "traveling across borders". Before teaching, you need to help generate a teaching plan about the teaching objectives and the whole teaching process. While teaching, you need to help me design several class activities to activate the students' interest and improve their class engagement.

##### (2) Teaching Material Writer

I am a college English teacher, and I need to deliver a writing class on how to write an argumentative essay. However, my students are very confused with the notion of an argument. Could you help me define what an argumentative essay is and then give some examples to help students distinguish an argument from a fact? After that, you need to provide some exemplary essays to illustrate how to write an argumentative essay on the topic "Is online learning as effective as traditional classroom learning?"

##### (3) Quiz Developer

I would like you to act as an English teacher who is going to develop some items for students. The purpose of the items is to serve as a quiz for students during the class. Please develop ten true or false items for students based on the reading materials I give you.

#### 3.1.3 Researchers

##### (1) Academic Writing Assistant

I would like you to act as a professor in applied linguistics. I have finished researching students' self-assessments. Now, I am going to write an academic paper. However, I do not know what I should write in the discussion section. Please provide me with 20 perspectives on how to write the discussion.

##### (2) Reading Assistant

I would like you to act as a reading assistant who can help me understand the paper I will send you later. Please read the paper I will give you carefully in the next turn, and then I will ask you some questions about the paper.

##### (3) Language Proofreader

Suppose you are an expert in academic writing, and your task is to help me improve my writing quality. Make sure not to change the meaning of the sentence while you polish the sentence.

#### 3.2 Detailed Example of ChatGPT as a Vocabulary Teaching Assistant

Section 3.1 shows the prompt designed based on the theoretical framework discussed in the previous section. This section shows a more concrete demonstration of the application of prompt engineering in language teaching. The prompt designed and used in this case is as follows.

"Assume that you are an English teacher, and your task is to teach English vocabulary to EFL learners in China. Your task is to figure out how to help students memorize the new words quickly. So, how should you teach the word "chrysanthemum" and figure out effective methods, such as the morphological analysis method and association mnemonic technique, for students to memorize?"

The first sentence is about setting a persona for GPT-4. The second sentence is about providing a context for GPT-4. The third sentence is about a specific task. This is the most essential element in the prompt."

The following content is a plan generated by GPT-4.

### (1) Introduction to the Word

Show a Picture: Start by showing a picture of a chrysanthemum to the class.

Contextual Use: Use the word in a sentence, such as "The garden was full of beautiful chrysanthemums."

### (2) Morphological Analysis Method

Break Down the Word: Explain the structure of the word by breaking it down into its roots and affixes.

Root: "chrys" (comes from the Greek word "chrysos," meaning gold)

Suffix: "anthemum" (comes from the Greek word "anthemon," meaning flower)

Explain the Meaning: Combine the meanings of the parts to explain the whole word: "gold flower."

Practice Pronunciation: Break it into syllables and practice pronunciation: chrys-an-the-mum.

### (3) Association Mnemonic Technique

Find Similar Sounding Words: Identify words that sound similar in English or the students' native language.

chrys: sounds like "Chris" (a common name)

an: sounds like "an"

them: sounds like "them"

um: sounds like "um"

Create a Story: Form a simple, memorable story using these words.

Example: "Chris and Ann found them in the garden, um, they were beautiful chrysanthemums."

Visualize the Story: Have students draw a picture of Chris and Ann in a garden full of chrysanthemums, and label the drawing with "chrysanthemum".

### (4) Repetition and Practice

Flashcards: Create flashcards with the word "chrysanthemum" on one side and a picture of the flower on the other.

Writing Exercise: Have students write the word multiple times and use it in sentences.

Speaking Practice: Encourage students to say the word out loud in sentences.

### (5) Engaging Activities

Drawing: Students draw the story scene and label it with the word.

Memory Games: Organize games like word matching or crosswords that include "chrysanthemum".

Role Play: Have students act out the story of Chris and Ann finding chrysanthemums.

### (6) Cultural Connections

Discuss Significance: Talk about the importance of chrysanthemums in Chinese culture, such as their use in festivals and traditional medicine.

Comparisons: Compare chrysanthemums to other flowers familiar to the students.

### (7) Review and Reinforcement

Quizzes: Regularly quiz students on the word in different contexts.

Peer Teaching: Have students teach the word to each other in pairs or small groups.

Real-Life Use: Encourage students to use the word when they see chrysanthemums in real life or in pictures.

It can be seen that the GPT-4 can generate a teaching plan for the new word "chrysanthemum". Interestingly, GPT-4 begins by introducing the new word in a real situation, which sounds like a communicative teaching approach. Then GPT-4 shows how to use the morphological analysis method to break down the word, which indicates how to understand the word. Unexpectedly, GPT-4 can show how the meaning of "chrysanthemum" derives from the meaning carried by the suffix and roots. This looks like the approach in cognitive linguistics for lexicography.

It is noteworthy that the example generated by GPT-4 for the same task can also be personalized by changing how the essential elements introduced above are specified. For instance, the audience element can also be specified as children in kindergarten. This can help the content be more tailored to the targeted users. Besides,

the output specification element in the prompt has not been defined. It is also possible to define the length and format of the desired content. It is also possible to ask LLMs to divide the word "chrysanthemum" in other ways. For example, GPT-4 can also divide the word "chrysanthemum" with the syllable that humans might be more accustomed to. Here is another solution LLMs provide:

"Chrysan: Think of "cry" (as it sounds similar and relates to mourning).

the: Acts as a connector, pointing to the next part of the word.

mums: Link this to the "mums" or mothers who might be attending a funeral, mourning, or remembering a loved one."

Therefore, writing a prompt for LLMs is the process of continuous adjustment and iterative improvement. This highlighted the strategies for prompt engineering, which will be introduced in the next section.

#### **4. Important Techniques for Prompt Engineering**

The most important elements and how these important elements can be formed together have been discussed. However, in practice, things cannot be done simply by putting all the elements together. There are also other important techniques in prompt engineering, which can help users in applied linguistics better achieve their goals via LLMs. Therefore, this section delineates the most commonly used techniques in prompt engineering after users form their prompt using the elements introduced above.

##### *4.1 Iterative Prompting*

Iterative prompting refers to a technique where multiple rounds of prompts are used to refine and improve the output generated by a language model (Shi et al., 2023). As far as the author's knowledge, iterative prompting might be the most important technique in prompt engineering since it usually takes several turns of iterative adjustment to obtain the desired output from LLMs.

Iterative prompting is commonly achieved in four steps, i.e., (1) writing an initial prompt, (2) analysis of output, (3) modifications, and (4) evaluations. The purpose of utilizing iterative prompting is to help ensure that the final output is aligned with the specific needs and expectations of the task at hand, making it an essential technique in prompt engineering for generating high-quality responses from language models. It is also noteworthy that iterative prompting also entails domain knowledge to evaluate the model performance, which can provide directions for iterative prompting.

##### *4.2 Negative Prompting*

Negative prompting is defined as the technique to clarify what the AI model should avoid generating or focusing on its response (Gu et al., 2023). Specifically, negative prompting aims to inform the LLMs not to generate unexpected information, including wrong information, irrelevant information, etc. By including explicit instructions about what not to include, negative prompting is beneficial to ensure that the model's responses align more with users' expectations and requirements.

For example, you are asking the model to generate a text illustrating the importance of technology applications in language learning, especially extending the role of technology in English listening and writing parts. However, you want to avoid any content of specific brands or commercial products to keep the discussion neutral and academic. Therefore, you might include a negative prompt such as "do not generate any information related to commercial purposes". As a whole, employing negative prompts helps guide the model to produce content that adheres to specific constraints, making the output more useful and appropriate for the given context.

##### *4.3 Few (Zero) Shot Prompting*

In prompt engineering, few-shot and zero-shot are two important techniques (Chen et al., 2023). The difference between them lies in whether specific examples are provided when LLMs are conducting task performances. Few-shot prompt involves giving the language model a few examples of the task you want it to perform (Robert et al., 2022), while a zero-shot prompt involves giving the language model a task without providing any specific examples (Yong et al., 2023). In applied linguistics, these two techniques are widely used, such as grammar correction, text classification, language translation, and so on. More importantly, both few-shot and zero-shot prompts capitalize on the extensive pre-training of large language models, making them powerful tools for performing complex language tasks with minimal input.

In the few-shot prompting, at present, one of the easiest methods is to provide an example for LLMs. It can be done by screenshot or uploading a file to LLMs, which can be easily done in GPT-4. By instructing LLMs with

examples, LLMs might better understand the intent of users and might generate content that is more tailored to users' needs.

#### *4.4 Chain of Thought Prompting*

Chain-of-thought prompting is a technique used in prompt engineering to enhance the reasoning capabilities of language models (Wei et al., 2022; Yu et al., 2023). This technique involves segmenting a complex question into a series of steps, which allows the model to generate a more coherent and accurate response by reasoning through each step sequentially. The prominent aspects of the chain of thought prompting include step-by-step reasoning, accuracy enhancement, and applicability.

In prompt engineering, the way to implement this technique is to use the phrase "Let's think step by step". This prompt has been shown as an effective method to improve the LLMs' performance, especially when the task is related to reasoning (Zhang et al., 2022). The possible reason is that LLMs might get lost on how to solve multiple tasks at a time. This casts important light on the prompt engineering strategies. As for applied linguistics, users can also add a sentence like "Let's think step by step" in the prompt. Alternatively, they can also try to break down the task into several tasks when they find LLMs cannot handle a complicated task that consists of several subtasks.

### **5. Strategies to Combat Possible Challenges in Prompt Engineering**

After understanding the important techniques in prompt engineering, it is also essential to discuss how to overcome the possible challenges of LLMs in practice via prompt engineering, such as potential bias, data privacy, explainability, and hallucinations (Kaddour, 2023).

#### *5.1 Potential Bias*

In prompt engineering, potential bias is a major factor influencing the output quality of LLMs, which also causes researchers great concerns (Pack & Maloney, 2024). The bias can be embodied in bias in different genders, regions, languages, ideologies, etc. (Dwivedi et al., 2023).

Potential biases might come from training data and model training. Additionally, the user biases mean that the person designing the prompt may unconsciously introduce their own biases based on their background, beliefs, or preferences. Moreover, it would be influenced by the users' literacies, cultural backgrounds, and languages. Meanwhile, representation bias occurs when certain groups or viewpoints are underrepresented or misrepresented in the training data. For instance, Bender et al. (2021) revealed that LLMs like GPT-3, when prompted with gender-neutral phrases, would often produce gender-skewed outputs.

Therefore, it is essential to take some measures to mitigate the aforementioned biases. From the perspective of prompt engineering, negative prompting might be an effective method for users in applied linguistics to mitigate potential biases. For example, it could be: "Please ensure that the content should not be biased towards different genders, cultural backgrounds, nations, etc., and should conform to the ethical principles in education". Another possible way to provide more specific guidelines about the meaning of ethical principles and bias is to upload files about these topics to LLMs (e.g., Gallegos et al., 2024), asking LLMs to learn the content related to these topics. These kinds of prompting engineering methods can lower the challenge of potential bias of the output LLMs generate.

#### *5.2 Data Privacy*

Data privacy in prompt engineering is a crucial aspect, especially given the increasing use of AI systems in handling private information (Wang et al., 2023). The problem of data privacy of LLMs might be more related to the underlying technology of LLMs, such as data transition and data storage. One comprehensive review of methods to tackle data privacy was conducted by Edemacu and Wu (2024).

However, these possible solutions might be challenging for users in applied linguistics. This study focuses on prompt engineering. Therefore, methods for mitigating data privacy concerns are summarized from the perspective of prompt engineering. The first tip is to anonymize the data when we prompt with LLMs. This method can lower the possibility of leaking important information. The second method is to split the content that one would like to upload into LLMs, which may lower the possibility of data privacy concerns. The third method is local processing. This means that users can also use the LLMs that are deployed on their own computers. This might be a more practical method compared with the previous two methods, which can greatly solve the problem related to data privacy. However, it might also involve more technical issues. The fourth method is to use the temporary chat. Take GPT-4 as an example. Users can chat with LLMs using the temporary chat options, in



which the data cannot be used to train the model, and data can only be stored for no more than 30 days, according to the statement of OpenAI.

### *5.3 Explainability*

Explainability in prompt engineering refers to the ability to understand, interpret, and articulate the rationale of the responses generated by language models based on the given prompt (Doshi-Velez & Kim, 2017; Du et al., 2019).

The importance of increasing explainability lies in two aspects. That is, explainability can establish proper trust by elucidating the reasoning mechanism, ensuring accountability, and improving the performance of AI systems (Strobelt et al., 2019). Based on previous studies, some strategies to improve explainability were given. For instance, clarifying the prompt design and avoiding overly complex or vague prompts can help generate more predictable and explainable responses.

In prompt engineering, studies have shown that asking LLMs to explain is an effective method to increase its explainability (Chiang & Lee 2023). This can help users better understand why LLMs generate the content. For example, the prompt can be written like this: "Please explain why you think this sentence is negative". Users in applied linguistics also ask LLMs to explain the rationale why LLMs made that decision. This can increase the transparency of LLMs, and most importantly, users can validate whether the process or the reasons LLMs provide can be justified.

### *5.4 Hallucination*

Hallucination refers to the phenomenon where an AI model creates information that seems believable but is completely invented and not grounded in the input data or any actual facts (Tonmoy et al., 2024). The potential causes of hallucinations might be related to the model training and overgeneralization (Ji et al., 2023). To avoid hallucinations, Varshney et al. (2023) put forward a method. That is, before proceeding to sentence generation, the approach initially utilizes the model's output values to detect potential hallucinations, confirm their accuracy, and then address any identified hallucinations. Moreover, Tonmoy et al. (2024) conducted a review that summarized some extant approaches to decrease hallucinations, such as retrieval-augmented generation (RAG) (Lewis et al., 2021; Gao et al., 2023) and Chain of Natural Language Inference (Lei et al., 2023).

Regarding prompt engineering, one practical method for researchers in applied linguistics to tackle hallucination is to ask LLMs to provide the link to the references. This can provide a way for researchers to triangulate the content LLMs generate. Another effective method is negative prompting, asking LLMs not to generate irrelevant information and requiring LLMs to check whether the content is consistent with the existing data.

### *5.5 Output Variability*

Output variability refers to the content LLMs generate based on a given prompt might vary across different times or situations (Lo, 2023). The possible reason is that the output of LLMs is based on random sampling. However, this might pose a challenge to the replicability of the content generated by LLMs, which might make it hard for others to replicate the study conducted based on LLMs (Yan et al., 2023).

The variability of the output LLMs can be controlled by the parameter called "temperature". Models with different temperatures generate diverse responses. Specifically, a higher temperature setting increases randomness, leading to more varied outputs, while a lower setting makes outputs more deterministic. In prompt engineering, users can not directly set up the parameter for "temperature" in the model like GPT-4o. One easy and practical method is to use the OpenAI Playground, which can provide users with temperature options. Another possible solution is to use Python, which can provide more flexibility in setting the temperature parameter. Readers who are interested in how to interact with LLMs with Python can refer to the book written by Alammam and Grootendorst (2024).

## **6. Conclusion**

This study aims to delineate the basic elements of prompt engineering and how these elements can form an effective prompt. The prompt for the potential applications of LLMs in applied linguistics was also shown. This study also demonstrated several important prompting strategies, such as negative prompting, iterative prompting, and few-shot prompting. These important components in prompt engineering play a key role in manipulating the output and can serve as a guideline or framework for researchers in applied linguistics who would like to leverage LLMs. As far as the author's knowledge, this is the first paper that investigates how to write an effective prompt for the applied linguistics domain.

Besides, this study can also contribute to figuring out potential prompting engineering methods to navigate potential pitfalls and challenges of LLMs in practice, such as potential bias, explainability, data privacy, output variability, and hallucinations. Understanding how to combat these challenges by prompting engineering can help researchers in applied linguistics better harness the LLMs in a proper and responsible way.

However, this study also has some limitations. Prompt engineering is not a magic. It cannot solve all potential pitfalls and problems. Human review is also very important, especially when LLMs are used in high-stake circumstances. Besides, prompt engineering also entails some domain knowledge to write an effective prompt, which cannot be exhaustively listed and discussed in the study. Moreover, this study discusses potential strategies to mitigate the potential pitfalls of LLMs, most of which were from the perspective of prompt engineering. Other possible technical strategies have not been discussed.

Despite these limitations, this study has important practical use and implications for applied linguistics. First, this study can serve as a learning tutorial for researchers in applied linguistics who would like to leverage LLMs to conduct interdisciplinary studies between generative AI and applied linguistics. Second, this study can serve as a tutorial for language teachers in applied linguistics, which can enhance their digital literacy in prompt engineering. Third, this study can also benefit EFL learners who would like to use LLMs in learning English, such as obtaining written feedback and inquiring about a specific expression in English writing.

## References

- Alammar, J., & Grootendorst, M. (2024). *Hands-on large language models*: O'Reilly Media, Inc.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Paper presented at the FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. <https://doi.org/10.1145/3442188.3445922>
- Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2023). *Unleashing the potential of prompt engineering in Large Language Models: A comprehensive review*. <https://doi.org/10.48550/arXiv.2310.14735>
- Chiang, C. H., & Lee, H. Y. (2023). *Can large language models be an alternative to human evaluations?* <https://doi.org/10.18653/v1/2023.acl-long.870>
- Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. <https://doi.org/10.48550/arXiv.1702.08608>
- Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, 63(1), 68-77. <https://doi.org/10.1145/3359786>
- Dwivedi, S., Ghosh, S., & Dwivedi, S. (2023). Breaking the bias: Gender fairness in LLMs using prompt engineering and In-context learning. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 15(4). <https://doi.org/10.21659/rupkatha.v15n4.10>
- Edemacu, K., & Wu, X. (2024). *Privacy preserving prompt engineering: A survey*. <https://doi.org/10.48550/arXiv.2404.06001>
- Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Derroncourt, F., ... Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, 1-79. [https://doi.org/10.1162/coli\\_a\\_00524](https://doi.org/10.1162/coli_a_00524)
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... Wang, H. (2023). *Retrieval-augmented generation for large language models: A survey*. <https://doi.org/10.48550/arXiv.2312.10997>
- Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*. <https://doi.org/10.1007/s10439-023-03272-4>
- Grabb, D. (2023). The impact of prompt engineering in large language model performance: A psychiatric example. *Journal of Medical Artificial Intelligence*, 6. <https://doi.org/10.21037/jmai-23-71>
- Gruda, D. (2024). Three ways ChatGPT helps me in my academic writing. *Nature*. <https://doi.org/10.1038/d41586-024-01042-3>
- Gu, J., Han, Z., Chen, S., Beirami, A., He, B., Zhang, G., ... Torr, P. (2023). *A systematic survey of prompt engineering on vision-language foundation models*. <https://doi.org/10.48550/arXiv.2307.12980>
- Han, Z. (2024). ChatGPT in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition*, 46(2), 301-306. <https://doi.org/10.1017/S0272263124000111>

- Henrickson, L., & Meroño-Peñuela, A. (2023). Prompting meaning: A hermeneutic approach to optimising prompt engineering with ChatGPT. *AI & Society*. <https://doi.org/10.1007/s00146-023-01752-8>
- Heston, T., & Khun, C. (2023). Prompt engineering in medical education. *International Medical Education*, 2(3), 198-205. <https://doi.org/10.3390/ime2030019>
- Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4). <https://doi.org/10.30935/cedtech/13605>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... Pung, F. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*. <https://doi.org/10.1145/3571730>
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). *Challenges and applications of large language models*. <https://doi.org/10.26434/chemrxiv-2023-2307>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, 29, 11483-11515. <https://doi.org/10.1007/s10639-023-12249-8>
- Lei, D., Li, Y., Hu, M., Wang, M., Yun, V., Ching, E., & Kamal, E. (2023). *Chain of natural language inference for reducing Large Language Model ungrounded hallucinations*. <https://doi.org/10.48550/arXiv.2310.03951>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2021). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. <https://doi.org/10.26434/chemrxiv-2021-11401>
- Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, 103344. <https://doi.org/10.1016/j.system.2024.103344>
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4). <https://doi.org/10.1016/j.acalib.2023.102720>
- Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2023). Prompt Engineering in Large Language Models. in *International Conference on Data Intelligence and Cognitive Informatics*, Springer Nature Singapore, Singapore. [https://doi.org/10.1007/978-981-99-7962-2\\_30](https://doi.org/10.1007/978-981-99-7962-2_30)
- Oppenlaender, J., Linder, R., & Silvennoinen, J. (2023). *Prompting AI art: An investigation into the creative skill of prompt engineering*. <https://doi.org/10.26434/chemrxiv-2023-2303>
- Pack, A., & Maloney, J. (2024). Using artificial intelligence in TESOL: Some ethical and pedagogical considerations. *TESOL Quarterly*, 58(2), 1007-1018. <https://doi.org/10.1002/tesq.3320>
- Robert L. Logan IV, Ivana Balažević, Eric Wallace, Fabio Petroni, Sameer Singh, & Sebastian Riedel. (2021). *Cutting down on prompts and parameters: simple few-shot learning with language models*. <https://doi.org/10.18653/v1/2022.findings-acl.222>
- Shah, C. (2024). *From prompt engineering to prompt science with human in the loop*. <https://doi.org/10.48550/arXiv.2401.04122>
- Shi, Y., Ren, P., Wang, J., Han, B., ValizadehAslani, T., Agbavor, F., ... Liang, H. (2023). Leveraging GPT-4 for food effect summarization to enhance product-specific guidance development via iterative prompting. *Journal of biomedical informatics*, 148, 104533. <https://doi.org/10.1016/j.jbi.2023.104533>
- Shin, J., Tang, C., Mohati T., Nayebi, M., Wang, S., & Hemmati, H. (2023). *Prompt engineering or fine tuning: An empirical assessment of large language models in automated software engineering tasks*. <https://doi.org/10.48550/arXiv.2310.10508>
- Strobelt, H., Gehrmann, S., Behrisch, M., Perer, A., Pfister, H., & Rush, A. M. (2019). SEQ2SEQ-VIS : A visual debugging tool for sequence-to-sequence models. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 353-363. <https://doi.org/10.1109/TVCG.2018.2865044>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57. <https://doi.org/10.1016/j.asw.2023.100752>

- Tonmoy, S. M., Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). *A comprehensive survey of hallucination mitigation techniques in large language models*. <https://doi.org/10.48550/arXiv.2401.01313>
- Varshney, N., Yao, W., Zhang, H., Chen, J., & Yu, D. (2023). *A stitch in time saves nine: Detecting and mitigating hallucinations of LLMs by validating low-confidence generation*. <https://doi.org/10.48550/arXiv.2307.03987>
- Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., Li, Q., & Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *Npj Digital Medicine*, 7(1). <https://doi.org/10.1038/s41746-024-01029-4>
- Wang, T., Zhou, N., & Chen, Z. (2024). *Enhancing computer programming education with LLMs: A study on effective prompt engineering for Python code generation*. <https://doi.org/10.48550/arXiv.2407.05437>
- Wang, Y., Pan, Y., Yan, M., Su, Z., & Luan, T. H. (2023). A Survey on ChatGPT: AI-Generated contents, challenges, and solutions. *IEEE Open Journal of the Computer Society*, 4, 280-302. <https://doi.org/10.1109/OJCS.2023.3300321>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., ... Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. <https://doi.org/10.48550/arXiv.2302.11382>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. <https://doi.org/10.1111/bjet.13370>
- Yong, G., Jeon, K., Gil, D., & Lee, G. (2023). Prompt engineering for zero-shot and few-shot defect detection and classification using a visual-language pretrained model. *Computer-Aided Civil and Infrastructure Engineering*, 38(11), 1536-1554. <https://doi.org/10.1111/mice.12954>
- Yu, Z., He, L., Wu, Z., Dai, X., & Chen, J. (2023). *Towards better chain-of-thought prompting strategies: A survey*. <https://doi.org/10.48550/arXiv:2310.04959>
- Zhang, Z., Zhang, A., Li, M., & Smola, A. (2022). *Automatic chain of thought prompting in Large Language Models*. <https://doi.org/10.48550/arXiv:2210.03493>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).