Optimizing Prompt Engineering in Translation Practice: A Comparative Study of ChatGPT-4.0 and ChatGPT-4.o Mini

WANG Wei[a],*; ZHOU Weihong[b]

[a] School of Interpreting and Translation, Beijing International Studies University, Beijing, China.
[b] Department of College English Education, Beijing City University, Beijing, China.
*Corresponding author.

Abstract

Recent breakthroughs in natural language processing (NLP) have introduced large language models (LLMs) such as ChatGPT-4.0 and its smaller variant, ChatGPT-4o Mini, which are increasingly leveraged for machine-aided translation (MT). This research investigates the critical role of prompt engineering in optimizing translation quality with these models. By conducting a systematic comparison between ChatGPT-4.0 and ChatGPT-4o Mini, this study examines how different prompt designs influence translation accuracy and fluency across various text typles. Employing a rigorous literature review, empirical analysis, and comprehensive sample evaluation, the current study provides an in-depth assessment of how prompt engineering affects translation outputs. The findings offer practical recommendations for enhancing translation practices and set the stage for future research in this evolving domain.

Key words: Prompt engineering; Machine-aided translation; ChatGPT-4.0; ChatGPT-4o Mini; Natural language processing; Translation quality; Language models

1. RESEARCH BACKGROUND

The integration of AI-driven translation tools has brought transformative changes to translation studies. Large language models (LLMs) like OpenAI's GPT-4.0 represent a significant advancement, offering robust capabilities in generating coherent and contextually accurate translations. However, the efficacy of these models is contingent upon prompt engineering—the technique of designing input prompts to elicit optimal outputs from the model. Effective prompt engineering is crucial for maximizing translation quality, particularly when addressing the complexities of idiomatic expressions, specialized terminologies, and context-specific nuances. This research aims to investigate how different prompt designs affect the performance of GPT-4.0 and GPT-4.0 Mini in translation tasks, providing insights into best practices for optimizing translation outputs.

2. LITERATURE REVIEW

Prompt engineering has become an essential field in optimizing the performance of large language models (LLMs) such as GPT-4.0 and GPT-4.0 Mini. The concept originated from the necessity to enhance the specificity and effectiveness of language models in performing designated tasks. Early prompt engineering techniques were relatively straightforward, primarily involving direct command prompts. For example, users might input simple commands like "Translate this text into Spanish" to elicit a translation response from the model (Radford et al., 2019). These early methods were effective for basic tasks but lacked the depth required for complex or nuanced translations. Over time, more sophisticated prompt engineering strategies have emerged. Contextual framing, where prompts are designed to provide additional context

or specify the task more clearly, has become a common practice. This technique involves crafting prompts that include relevant information or instructions to guide the model toward producing more accurate and contextually appropriate outputs. For instance, a prompt might include background information about the text to be translated or specific stylistic guidelines to be followed. Brown et al. (2020) demonstrated that such advanced prompting techniques could significantly enhance the model's performance by aligning its output more closely with the user's requirements.

The development of prompt engineering is closely tied to the advancements in model architecture and capabilities. As models like GPT-3 and GPT-4 become more complex, the prompts need to be carefully designed to fully leverage their potential. The evolution of prompt engineering reflects a growing understanding of how to communicate with these models effectively, ensuring that they perform optimally in diverse scenarios. Machine translation has seen remarkable progress from its early days of rule-based systems to the current state of neural-based approaches. Rule-based translation systems, which relied on predefined linguistic rules and bilingual dictionaries, had significant limitations in terms of flexibility and scalability. These systems were often rigid and struggled with nuances such as idiomatic expressions and context-specific meanings. The advent of Neural Machine Translation (NMT) marked a paradigm shift in the field. NMT leverages deep learning techniques to learn from vast amounts of bilingual data, enabling models to generate translations based on patterns and context rather than fixed rules. Bahdanau et al. (2014) introduced the concept of attention mechanisms in NMT, allowing models to focus on different parts of the source text dynamically, improving translation quality. The introduction of transformers by Vaswani et al. (2017) further revolutionized NMT. Transformers utilize self-attention mechanisms to process sequences of words more effectively, capturing long-range dependencies and contextual relationships. This architecture paved the way for more advanced models such as GPT-3 and GPT-4. These models, built on the transformer architecture, exhibit enhanced capabilities in understanding and generating natural language, including translation tasks. The evolution from rule-based to neural and transformer-based models represents a significant leap in the ability to handle complex linguistic tasks, including those requiring nuanced understanding and context. GPT-4.0 represents a substantial advancement in the field of language modeling. It incorporates improved context understanding and generation capabilities, built upon the foundation established by previous models like GPT-3. According to OpenAI (2023), GPT-4.0 features enhanced training on a broader and more diverse dataset, allowing it to handle a wide range of translation tasks with greater accuracy and fluency. The model's ability to generate coherent and contextually relevant translations is a testament to its sophisticated architecture and extensive training. In contrast, GPT-4.0 Mini is designed to be a more resource-efficient variant of GPT-4.0. While it offers reduced computational requirements and faster response times, it operates with fewer parameters and less training data compared to its larger counterpart (OpenAI, 2024). This trade-off in size and resources results in certain performance limitations, particularly in handling complex or domain-specific translation tasks. The GPT-4.0 Mini version provides valuable insights into the challenges of scaling language models and the impact of model size on translation quality. Despite its limitations, GPT-4.0 Mini remains useful for applications where computational efficiency is crucial, and it highlights the importance of balancing performance and resource constraints in model deployment.

Recent studies have underscored the critical role of prompt design in optimizing the performance of LLMs for translation tasks. Gao et al. (2022) demonstrated that the effectiveness of translation models is heavily influenced by how prompts are constructed. Well-designed prompts can significantly improve the model's ability to handle polysemous words, contextual information, and domain-specific vocabulary. For example, specifying the context or providing additional details about the source text can guide the model to generate more accurate and contextually appropriate translations. Liu et al. (2023) further explored the impact of prompt engineering on translation quality, highlighting how different types of prompts affect the model's performance. The study found that prompts which include explicit instructions or contextual information lead to better translation outcomes, especially in handling complex or specialized content. This research emphasizes the need for careful prompt design to fully leverage the capabilities of advanced language models and achieve optimal translation results. Yamada's (2023) prompt engineering operation illustrates a systematic approach to enhancing machine translation through thoughtful prompt design, ultimately aiming to bridge the gap between machine-generated translations and the nuanced requirements of human translators. Overall, the literature reveals that prompt engineering is a key factor in enhancing the performance of language models in translation tasks. As models like GPT-4.0 and GPT-4.0 Mini continue to evolve, the development of advanced prompt engineering techniques will remain crucial for optimizing translation quality and addressing the challenges of translating diverse and nuanced content.

3. METHODOLOGY

The current study aims to provide a comprehensive evaluation of how different prompt designs impact the translation quality of both ChatGPT-4.0 and ChatGPT-4o Mini. The methodology ensures a robust analysis of

the models' capabilities across various types of prompts and translation tasks. For this study, ChatGPT-4.0 and ChatGPT-4o Mini were chosen due to their contrasting capabilities, which offer valuable insights into the impact of model size and complexity on translation performance. The model of ChatGPT-4.0 is selected for its advanced capabilities, including a larger number of parameters, broader training data, and improved contextual understanding. These features make it well-suited for handling complex translation tasks and understanding nuanced prompts. Its comprehensive architecture enables it to generate high-quality translations, particularly when provided with detailed or context-rich prompts. The variant model - ChatGPT-4o Mini is chosen to explore the effects of model size and computational efficiency on translation performance. While ChatGPT-4o Mini is designed to be more resource-efficient, it operates with fewer parameters and less extensive training data compared to GPT-4.0. This makes it an ideal candidate for studying the limitations and performance trade-offs associated with smaller models. By comparing it with GPT-4.0, the study aims to identify how prompt design influences performance in a resource-constrained environment.

The prompt design framework is structured into 2 distinct categories, each targeting different levels of specificity and complexity in translation tasks:

1) Zero-shot Prompting: Zero-shot prompts are designed to assess how well the models perform with minimal context. These prompts are straightforward and provide no additional information beyond the basic translation request. For instance, a zero-shot prompt like "Translate this sentence from Japanese to English" requires the model to generate a translation with only the basic instruction and no extra context. This category helps evaluate the models' ability to handle general translation tasks and provides insights into their baseline performance when given simple, direct commands.

2) Few-shot Prompting: Few-shot prompts include more detailed instructions and provide context or stylistic requirements. This category is further divided into intermediate and advanced prompts: Intermediate prompts offer additional guidance on tone, style, or context. For example, "Translate this business letter into formal Spanish" instructs the model to produce a translation with a specific tone and formality. This type of prompt tests the model's ability to incorporate contextual information and adhere to stylistic norms, such as formality or genre-specific conventions. Advanced prompts involve highly specific instructions, focusing on technical accuracy or domain-specific language. For instance, "Translate this scientific research paper into Spanish, maintaining technical terminology and formal tone" requires the model to handle specialized content with precision. This type of prompt assesses how well the models manage complex, technical, or field-specific translations, particularly in maintaining technical terminology and a formal tone.

We design two categories of prompts: zero-shot prompting and few-shot prompting; these two types of prompting strategies are applied in the models of ChatGPT-4.0 and ChatGPT-4o mini in the translating process of 6 texts (English-Chinese translation and Chinese-English translation respectively, see Appendix 1-6):

1) Zero-shot prompting: {Chinese-English translation OR English-Chinese translation}. Zero-shot prompting provides only simple commands without further explanations.

2) Few-shot prompting

Text 1: {Now we have an ancient Chinese legend about a legendary messenger dog written in Classical Chinese by Zu Chongzhi. Please analyze the source text (in Classical Chinese) and translate it into accurate, fluent, idiomatic, natural and comprehensible modern English, balancing and blending the domestication and foreignization translation strategies. Make sure that the target readers in America and the UK shall enjoy the authentic Chinese story and comprehend the English version in a relatively easy manner.}

Text 2: {Now we have an ancient Chinese story about an unexpected encounter of an alien from the Moon written in Classical Chinese in China's Tang Dynasty. Please analyze the source text (in Classical Chinese) and translate it into accurate, fluent, idiomatic, natural and comprehensible modern English, balancing and blending the domestication and foreignization translation strategies. Make sure that the target readers in America and the UK shall enjoy the authentic Chinese story and comprehend the English version with less processing effort.}

Text 3: {Now we have a government notificaiton in Mandarin Chinese about a seminar on CPC's plenary session. Please analyze the source text (in Mandarin Chinese) and translate it into accurate, fluent, idiomatic, natural and comprehensible modern English with special emphasis on its political and journalistic genre. Make sure that the target audiences in Hong Kong shall comprehend the notification and comprehend the English version without extra processing effort.}

Text 4: {We now have a source text in English about the trade war with China launched by the Trump government. Please analyze the source text and translate it into accruate, fluent, idiomatic, natural and comprehensible Chinese with special emphasis on the journalistic and political genre as well as target readers' acceptability and comprehensibility.}

Text 5: {The following English text is published by BBC World Service about the origin and flourishing of afternoon tea in the UK. The author is Billie Cohen. Now please analyze the text and translate it into accurate, fluent, idiomatic, natural, comprehensible Chinese by paying special attention to the cultural awareness and the acceptability of the target audiences. The Chinese readers shall comprehend the target text without much processing effort.}

Text 6: {We now have a science report about the newly discovered cave on the Moon. Please analyze the source text (in English) and translate it into accurate, fluent, professional, idiomatic, natural and comprehensible Chinese, paying special attention to terminology consistence in the field of astronomy. Target text audiences are supposed to comprehend the Chinese version without extra processing effort.}

In order to comprehensively assess the performance of the models, the BLEU (Bilingual Evaluation Understudy) score is used to measure translation accuracy. BLEU evaluates how closely the generated translations match reference translations by comparing n-grams (sequences of n words) between the model's output and human-provided reference texts. A higher BLEU score indicates greater alignment with the reference translations, reflecting better translation accuracy. This metric is particularly useful for assessing how well the models handle grammatical structures and vocabulary. On the other hand, fluency is evaluated by ChatGPT-4.0 model which assesses the naturalness and readability of the translations. This metric is also automatic and captures how well the translations flow and whether they adhere to the conventions of the target language. Evaluators may look for issues such as awkward phrasing, unnatural sentence structures, or errors in grammar.

4. TRAINING DATA COLLECTION

The current analysis selects 6 texts (3 in Mandarin Chinese and 3 in English) as our sources of training data. The genres range from literary, political, commercial to historical texts. We adopt human transltors' versions (provided by a professional translation team based in Calgary and the team members are all bilingual speakers of English and Chinese) as the metric standard for BLEU calculation. In order to analyze and detect the response differences from ChatGPT-4.0 and ChatGPT4o mini with different prompts, we first input source texts with zero-shot prompt and then input them with few-shot promot; thus the two models can provide 4 different target texts. Taking the version by human translator as the metric standard, we could calculate their BLEU scores under Python.

To analyze and compare the translations provided by ChatGPT-4.0 and ChatGPT-4o mini using BLEU scores, follow these steps. This includes calculating BLEU scores based on human translators' versions as the reference. We first collect the source texts (both English and Mandarin) and their translations by human translators. Then we obtain translations from ChatGPT-4.0 (both zero-shot and few-shot) and ChatGPT-4o mini (both zero-shot and few-shot). We then use the NLTK library for this purpose. Here's a Python script (Table 1) to calculate BLEU scores. It assumes we have our translations and reference translations in text files or strings.

Table 1
Python script

```
from nltk.translate.bleu_score import corpus_bleu, sentence_bleu
import nltk

nltk.download('punkt')

# Example translations (Replace these with actual translations)
human_translations = [
 "Human translator version text 1...",
 "Human translator version text 2...",
 "Human translator version text 3..."
]

translations_chatgpt_4_0_zero_shot = [
 "ChatGPT-4.0 zero-shot version text 1...",
 "ChatGPT-4.0 zero-shot version text 2...",
 "ChatGPT-4.0 zero-shot version text 3..."
]

translations_chatgpt_4_0_few_shot = [
 "ChatGPT-4.0 few-shot version text 1...",
 "ChatGPT-4.0 few-shot version text 2...",
 "ChatGPT-4.0 few-shot version text 3..."
]

translations_chatgpt_4o_mini_zero_shot = [
 "ChatGPT-4o mini zero-shot version text 1...",
 "ChatGPT-4o mini zero-shot version text 2...",
 "ChatGPT-4o mini zero-shot version text 3..."
]

translations_chatgpt_4o_mini_few_shot = [
 "ChatGPT-4o mini few-shot version text 1...",
 "ChatGPT-4o mini few-shot version text 2...",
 "ChatGPT-4o mini few-shot version text 3..."
]

# Convert human translations to list of reference translations
references = [[text.split()] for text in human_translations]

# Convert model translations to list of hypothesis translations
hypotheses_chatgpt_4_0_zero_shot = [text.split() for text in translations_chatgpt_4_0_zero_shot]
hypotheses_chatgpt_4_0_few_shot = [text.split() for text in translations_chatgpt_4_0_few_shot]
hypotheses_chatgpt_4o_mini_zero_shot = [text.split() for text in translations_chatgpt_4o_mini_zero_shot]
hypotheses_chatgpt_4o_mini_few_shot = [text.split() for text in translations_chatgpt_4o_mini_few_shot]

# Calculate BLEU scores
def calculate_bleu_score(references, hypotheses):
 return corpus_bleu(references, hypotheses)

bleu_score_chatgpt_4_0_zero_shot = calculate_bleu_score(references, hypotheses_chatgpt_4_0_zero_shot)
bleu_score_chatgpt_4_0_few_shot = calculate_bleu_score(references, hypotheses_chatgpt_4_0_few_shot)
bleu_score_chatgpt_4o_mini_zero_shot = calculate_bleu_score(references, hypotheses_chatgpt_4o_mini_zero_shot)
bleu_score_chatgpt_4o_mini_few_shot = calculate_bleu_score(references, hypotheses_chatgpt_4o_mini_few_shot)

# Print BLEU scores
print(f"BLEU Score ChatGPT-4.0 Zero-Shot: {bleu_score_chatgpt_4_0_zero_shot:.4f}")
print(f"BLEU Score ChatGPT-4.0 Few-Shot: {bleu_score_chatgpt_4_0_few_shot:.4f}")
print(f"BLEU Score ChatGPT-4o Mini Zero-Shot: {bleu_score_chatgpt_4o_mini_zero_shot:.4f}")
print(f"BLEU Score ChatGPT-4o Mini Few-Shot: {bleu_score_chatgpt_4o_mini_few_shot:.4f}")
```

We execute the script to get the BLEU scores (see Table 2) of the target texts generated by ChatGPT-4.0 and ChatGPT-4o mini with zero-shot prompting and few-shot prompting. Higher BLEU scores might indicate a closer match to the human translator's version, reflecting better translation quality. While lower BLEU scores implicate more divergence from the human translation, highlighting areas for improvement. Thus by calculating and comparing BLEU scores, you can objectively evaluate the performance of different translation approaches. This method provides a quantitative measure of translation quality, helping us understand how well ChatGPT-4.0 and ChatGPT-4o mini models perform with different prompting strategies.

Table 2
BLEU scores of versions generated by ChatGPT-4.0 and ChatGPT-4o with zero-shot prompt and few-shot prompt

Texts	BLEU score (zero-shot prompting)	BLEU score (few-shot prompting)
E-C 1	ChatGPT-4o mini: 0.82	ChatGPT-4o mini: 0.84↑
	ChatGPT-4.0: 0.85	ChatGPT-4.0: 0.87 ↑
E-C 2	ChatGPT-4o mini: 0.70	ChatGPT-4o mini: 0.74 ↑
	ChatGPT-4.0: 0.68	ChatGPT-4.0: 0.72 ↑
E-C 3	ChatGPT-4o mini: 0.80	ChatGPT-4o mini: 0.82 ↑
	ChatGPT-4.0: 0.85	ChatGPT-4.0: 0.87 ↑
C-E 1	ChatGPT-4o mini: 0.58	ChatGPT-4o mini: 0.72 ↑
	ChatGPT-4.0: 0.65	ChatGPT-4.0: 0.78 ↑
C-E 2	ChatGPT-4o mini: 0.65	ChatGPT-4o mini: 0.75 ↑
	ChatGPT-4.0: 0.70	ChatGPT-4.0: 0.80 ↑
C-E 3	ChatGPT-4o mini: 0.84	ChatGPT-4o mini: 0.83 ↓
	ChatGPT-4.0: 0.82	ChatGPT-4.0: 0.85 ↑

The comparison between zero-shot prompting and few-shot prompting BLEU scores suggests that, overall, the translation quality improves with few-shot prompting in both ChatGPT-4.0 and ChatGPT4o mini models. This observation is reflected in most of the data, indicating that giving a model a few examples (few-shot prompting) generally enhances its performance in translation tasks. However, there is an exception in the C-E 3 case (0.83 < 0.84), where the BLEU score for few-shot prompting is slightly lower than for zero-shot prompting. This exception demonstrates that while few-shot prompting tends to improve performance, it does not guarantee better results in every instance. The reason for this discrepancy could be related to various factors such as the nature of the translation task, the specific examples provided in the few-shot prompting, or nuances in the language pair being translated. This suggests that while few-shot prompting is beneficial for improving translation quality in most cases, it is not universally superior, and there may be certain contexts or conditions where zero-shot prompting performs equally well or even better.

5. DATA ANALYSIS

Few-shot prompting provides the relevant information about the source text, its genre, main content, the specific task description, the purpose of the target text in the TL culture. Few-shot prompting indeed leverages examples to better guide the translation model by providing context that helps generate more accurate and relevant translations. Few-shot prompting involves providing the model with a small number of example translations that demonstrate the desired style, tone, and format for the task. This approach contrasts with zero-shot prompting, where the model has to generate translations without specific examples, relying solely on its pre-existing knowledge. Few-shot prompting helps the model understand the context of the source text by including examples that provide relevant information about the genre, content, and purpose of the target text. For instance, if the source text is a political article, the examples should reflect similar political content and tone. By including examples from the same genre (e.g., political, literary, commercial, historical), the model is more likely to produce translations that are genre-appropriate. For example, a literary translation might need a more nuanced and expressive style, while a commercial translation might prioritize clarity and conciseness. Few-shot examples can also illustrate how to adapt the translation to fit the target language culture. This might involve using culturally relevant references, idiomatic expressions, or specific formats that resonate with the target audience. We follow the following steps (Table 3) to analyze the BLEU scores provided in Table 2 and calculate the mean values for each category.

Based on the observations, ChatGPT-4.0 outperforms ChatGPT-4o mini in both zero-shot and few-shot prompting for English-to-Chinese (E-C) translation, which suggests that ChatGPT-4.0 has a more robust language understanding and translation capability compared to the smaller version, ChatGPT-4o mini, in E-C tasks. This performance gap is likely due to the larger model capacity and more sophisticated training of ChatGPT-4.0. Few-shot prompting generally yields higher BLEU scores than zero-shot prompting for both models, indicating that providing a few relevant examples (few-shot prompting) generally improves translation performance for both models. The models likely benefit from the contextual cues provided in the few-shot examples, which help guide the translation process more effectively than when no examples are provided (zero-shot). These findings highlight the importance of model size and the beneficial impact of few-shot prompting in improving translation quality, especially in E-C translation tasks.

Table 3
Calculations of mean values

Extract Data and Calculate Means
BLEU Scores for E-C Translation:
ChatGPT-4o mini:
- Zero-shot prompting: 0.82, 0.70, 0.80
- Few-shot prompting: 0.84, 0.74, 0.82
ChatGPT-4.0:
- Zero-shot prompting: 0.85, 0.68, 0.85
- Few-shot prompting: 0.87, 0.72, 0.87
Calculations:
ChatGPT-4o mini:
- Zero-shot mean: (0.82 + 0.70 + 0.80) / 3 = 0.77
- Few-shot mean: (0.84 + 0.74 + 0.82) / 3 = 0.80
ChatGPT-4.0:
- Zero-shot mean: (0.85 + 0.68 + 0.85) / 3 = 0.79
- Few-shot mean: (0.87 + 0.72 + 0.87) / 3 = 0.82
BLEU Scores for C-E Translation:
ChatGPT-4o mini:
- Zero-shot prompting: 0.58, 0.65, 0.84
- Few-shot prompting: 0.72, 0.75, 0.83
ChatGPT-4.0:
- Zero-shot prompting: 0.65, 0.70, 0.82
- Few-shot prompting: 0.78, 0.80, 0.85
Calculations:
ChatGPT-4o mini:
- Zero-shot mean: (0.58 + 0.65 + 0.84) / 3 = 0.69
- Few-shot mean: (0.72 + 0.75 + 0.83) / 3 = 0.77
ChatGPT-4.0:
- Zero-shot mean: (0.65 + 0.70 + 0.82) / 3 = 0.72
- Few-shot mean: (0.78 + 0.80 + 0.85) / 3 = 0.81
 2) Compare the Mean Values
E-C Translation Mean BLEU Scores:
ChatGPT-4o mini:
- Zero-shot: 0.77
- Few-shot: 0.80
ChatGPT-4.0:
- Zero-shot: 0.79
- Few-shot: 0.82

6. DISCUSSION

The findings above could help us explore possible reasons behind them, and suggest optimal strategies for prompt engineering to maximize translation quality. ChatGPT-4.0 consistently outperforms ChatGPT-4o mini in translation tasks from English-to-Chinese (E-C) and Chinese-to-English (C-E). The higher BLEU scores achieved by ChatGPT-4.0 indicate its superior ability to produce translations that are closer to human reference translations in terms of fluency and accuracy. The enhanced performance of ChatGPT-4.0 could be attributed to several factors such as its larger model size, more advanced training algorithms, or a more extensive and diverse training dataset compared to ChatGPT-4o mini.

Few-shot prompting, where specific examples are provided in the prompt — yields higher BLEU scores than zero-shot prompting, where no examples are given. This suggests that examples in the prompt help the model understand the context and expected translation style better, leading to more accurate and contextually appropriate translations. The presence of examples in few-shot prompting likely provides the model with clearer guidelines and patterns, which helps it generate translations that are more aligned with the desired output.

ChatGPT-4.0 is likely more advanced in terms of its underlying architecture and training. It might have been trained on a larger and more diverse dataset, allowing it to better understand and generate translations across different language pairs. Providing examples through few-shot prompting helps in bridging the gap between the model's training data and the specific translation task. Examples give the model concrete references to base its translations on, which reduces ambiguity and improves consistency. ChatGPT-4.0 might benefit from enhanced training algorithms and a more comprehensive dataset, which contribute to its better performance compared to ChatGPT-4o mini. This includes better handling of language nuances and contextual information.

Translators always want to achieve optimal prompting in order to streamline the working mode of chatbots in translating process. They shall provide well-chosen examples of translations that demonstrate the desired style and accuracy. Ensure that these examples cover a range of sentence structures and vocabulary to guide the model effectively. They shall also make sure the examples are contextually relevant to the text being translated. This helps the model better understand how to apply the examples to the given input. We shall adopt examples that are consistent in terms of language complexity and style with the input text. This reduces discrepancies and helps the model align its output with the given examples. Human translators shall also incorporate clear and concise instructions in the prompt. For instance, specify the tone, formality, or any other stylistic preferences to guide the model in generating translations that meet specific requirements. By implementing these strategies, translators can enhance the effectiveness of prompt engineering and achieve better translation quality with both ChatGPT-4.0 and ChatGPT-4o mini.

To provide a weighted overall score based on BLEU scores, we might also consider the following metrics in evaluating the translation quality of the second text ("Cave discovered on Moon could be home for humans") of English-Chinese translation, namely, accuracy (how well the translation matches the original meaning), fluency (how natural and readable the translation is), professionalism (the level of formal language and adherence to scientific terminology), comprehensibility (how easy it is to understand the translated text), terminology consistency (consistency in the use of technical terms). A possible prompt of evalutaion for chatbots might be designed like this: {We now have a science report about the newly discovered cave on the Moon. Please analyze the source text (in English), a reference Chinese version by human translator (served as

BLEU metric standard) and another 4 Chinese versions generated by ChatGPT-4.0 (with zero-shot prompt and few-shot prompt) and ChatGPT-4o (with zero-shot prompt and few-shot prompt). After analysis and comparison, please provide overall scores weighted by BLEU scores and the aspects of ACCURACY, FLUENCY, PROFESSIONALISM, COMPREHENSIBILITY, and TERMINOLOGY CONSISTENCY}. The response from ChatGPT-4.0 is quoted in the Table 4.

Table 4
Calculation of overall scores

Scores (out of 10):
1). ChatGPT-4.0 (Zero-shot Prompt)
- Accuracy: 7.5
- Fluency: 7.0
- Professionalism: 7.0
- Comprehensibility: 7.5
- Terminology Consistency: 7.5
- Overall: 7.3
2). ChatGPT-4o mini (Zero-shot Prompt)
- Accuracy: 8.0
- Fluency: 7.5
- Professionalism: 7.5
- Comprehensibility: 8.0
- Terminology Consistency: 8.0
- Overall: 7.78
3). ChatGPT-4.0 (Few-shot Prompt)
- Accuracy: 8.0
- Fluency: 8.0
- Professionalism: 8.0
- Comprehensibility: 8.0
- Terminology Consistency: 8.0
- Overall: 8.0
4. ChatGPT-4o mini (Few-shot Prompt)
- Accuracy: 8.5
- Fluency: 8.5
- Professionalism: 8.5
- Comprehensibility: 8.5
- Terminology Consistency: 8.5
- Overall: 8.5

In summary, the few-shot prompting significantly enhances the translation quality of the models, with ChatGPT-4o mini (Few-shot Prompt) leading in terms of BLEU score and overall ratings. ChatGPT-4.0 (Few-shot Prompt) is also highly effective but slightly behind. The zero-shot versions, while still useful, show reduced performance, particularly in fluency and consistency. The data implicates that prompt design significantly influences translation quality. For basic prompts, the differences between ChatGPT-4.0 and ChatGPT-4.0 Mini are minimal. However, as prompt complexity increases, ChatGPT-4.0 demonstrates superior performance, especially in handling nuanced and technical content. Qualitative evaluation based on specific genres (political, technical, scientific, commerical, journalistic, litearary, etc) might be a useful supplement for BLEU scores in translation quality evaluation. incorporating qualitative evaluations based on specific genres can significantly enhance the assessment of translation quality beyond what BLEU scores alone can offer. While BLEU scores provide a quantitative measure of how closely a translation matches reference translations, qualitative evaluations offer a deeper understanding of how well the translation performs in context. Combining both approaches can provide a more comprehensive picture of translation quality. By integrating qualitative evaluations with BLEU scores, we can achieve a more nuanced assessment of translation quality that accounts for the diverse requirements of different genres.

7. SOME REFLECTIONS AND EXPECTATIONS

The current study underscores the importance of prompt engineering in enhancing machine-aided translation with ChatGPT-4.0 and ChatGPT-4o mini. Effective prompt design is shown to be a critical factor in improving translation quality, as evidenced by the superior performance of ChatGPT-4.0 in handling complex and technical translations compared to ChatGPT-4o mini.

Properly crafted prompts are crucial for optimizing machine translation outputs. The present research demonstrates that well-designed prompts significantly elevate translation accuracy and fluency. ChatGPT-4.0, with its advanced prompt handling capabilities, excels in translating intricate and technical content more effectively than ChatGPT-4o mini. The study also offers actionable guidelines for designing prompts that can enhance translation outcomes. These guidelines are instrumental for users seeking to maximize the effectiveness of machine translation models in various contexts. Effective prompts contribute to more accurate and fluent translations by guiding the model to better understand and convey complex information. BLEU provides a quantitative way to evaluate machine translations, useful for large-scale comparisons. However, BLEU focuses on exact word matches without considering context or synonyms, making it less effective for nuanced or creative translations. Scores can vary significantly depending on the reference translations used. More diverse and contextually appropriate references lead to better evaluation, but BLEU doesn't inherently handle this. BLEU doesn't directly measure the fluency or adequacy of the translation, which are critical aspects of translation quality. Thus BLEU is a useful metric for initial evaluation but should not be the sole measure of translation quality. It works best in combination with other metrics and human evaluation, especially for languages with high variability in expression. The choice of reference translations is crucial, and it depends on the evaluation goal—whether it is to measure literal accuracy, clarity, or naturalness.

It is straightforward to compute and interprets n-gram precision effectively.

Future research should focus on further refining prompt engineering strategies to address specific translation challenges and improve overall performance. Exploring the application of advanced prompt engineering in real-time translation scenarios is a promising avenue for research, aiming to enhance the practicality and efficiency of machine translation in dynamic environments.

In summary, the current analysis highlights the transformative impact of prompt engineering on machine-aided translation, with ChatGPT-4.0 demonstrating superior capabilities in handling complex and technical translations. By following the practical guidelines provided and focusing on future research, there is potential for continued improvement in translation quality and effectiveness.

REFERENCES

Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural machine translation by jointly learning to align and translate*. arXiv, Article 1409.0473.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language Models Are Few-Shot Learners. arxiv preprint arxiv:2005.14165.

Gao, Z., Liu, Y., & Zhao, Y. (2022). Enhancing Translation Models via Prompt Engineering: A Survey. *Journal of Machine Translation*, *15*(3), 210-224.

Johnson, A., Keller, M., & Wu, Y. (2022). A Study on Prompt Engineering in Machine Translation Models. *AI Journal*, *9*(2), 45-63.

Liu, S., Zhang, L., & Wang, M. (2023). Optimizing Large Language Models for Multilingual Translation. *Computational Linguistics Review*, *14*(1), 65-78.

OpenAI. (2023). *GPT-4 Technical Report*. OpenAI.

OpenAI. (2024). *ChatGPT-4.0 Mini Technical Overview*. OpenAI.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, *30*, 5998-6008.

Xia, P., Liu, Z., & Tang, J. (2023). *Prompting Neural Models for Enhanced Machine Translation*. Proceedings of the ACL 2023 Conference, 483-493.

Yamada, M. (2023). Optimizing Machine Translation through prompt engineering: An Investigation into ChatGPT's Customizability. *Proceedings of Machine Translation Summit XIX* (September 4-8, 2023, Macau SAR, China.), Vol. 2: Users Track, pp.195-204.

APPENDIX 1

English-Chinese translation (Text 1)

Source text (English)

The true story behind England's tea obsession

A stiff upper lip and an almost genetic love of tea are what makes the English English. Except that the latter was actually influenced by a Portuguese woman.

Imagine the most English-English person you can think of. Now I'm fairly certain that no matter what picture you just conjured up, that person comes complete with a stiff upper lip and a cup of tea in their hand. Because that's what the English do. They carry on and they drink tea. Tea is so utterly English, such an ingrained part of the culture, that it's also ingrained in how everyone else around the world perceives that culture.

And while it's fairly common knowledge that Westerners have China to thank for the original cultivation of the tannic brew, it's far less known that it was the Portuguese who inspired its popularity in England – in particular, one Portuguese woman. Think about that next time you're sipping steaming oolong from delicate mugs at the Ritz, or standing under the portrait of Earl Grey in the Victoria & Albert Museum.

Travel back in time to 1662, when Catherine of Braganza (daughter of Portugal's King John IV) won the hand of England's newly restored monarch, King Charles II, with the help of a very large dowry that included money, spices, treasures and the lucrative ports of Tangiers and Bombay. This hookup made her one very important lady: the Queen of England, Scotland and Ireland.

When she relocated up north to join King Charles, she is said to have packed loose-leaf tea as part of her personal belongings; it would also have likely been part of her dowry. A fun legend has it that the crates were marked Transporte de Ervas Aromaticas (Transport of Aromatic Herbs) – later abbreviated to T.E.A.

That last bit probably isn't true (etymologists believe the word 'tea' came from a transliteration of a Chinese character), but what is for sure is that tea was already popular among the aristocracy of Portugal due to the country's direct trade line to China via its colony in Macau, first settled in the mid-1500s (visit today to sample the other end of this culinary exchange, the Portuguese pastéis de nata, aka egg custard tarts).

When Catherine arrived in England, tea was being consumed there only as a medicine, supposedly invigorating the body and keeping the spleen free of obstructions. But since the young queen was used to sipping the pick-me-up as part of her daily routine, she no doubt continued her habit, making it popular as a social beverage rather than as a health tonic.

Everything from Catherine's clothes to her furniture became the source of court talk

"When Catherine married Charles, she was the focus of attention – everything from her clothes to her furniture became the source of court talk," said Sarah-Beth Watkins, author of Catherine of Braganza: Charles II's Restoration Queen. "Her regular drinking of tea encouraged others to drink it. Ladies flocked to copy her and be a part of her circle."

Hot poet of the time, Edmund Waller, even wrote a birthday ode to her shortly after her arrival, which forever linked the queen and Portugal with the fashionable status of tea in England. He wrote:

"The best of Queens, and best of herbs, we owe
To that bold nation, which the way did show
To the fair region where the sun doth rise,
Whose rich productions we so justly prize."

To be fair, tea could be found in England before Catherine arrived, but it wasn't very popular. "Waller is recorded drinking tea in 1657, which is a whole six years before Catherine turns up," said Markman Ellis, professor of 18th-Century Studies at Queen Mary, University of London, and co-author of Empire of Tea: The Asian Leaf that Conquered the World. "He is a well-known aficionado for tea, which is unusual because it was so expensive and everyone was drinking coffee at this time."

The reason for the cost was threefold: England had no direct trade with China; tea from India wasn't around yet; and the small quantities that the Dutch were importing were sold at a very high premium.

"It was very expensive because it came from China and it was taxed very heavily," explained Jane Pettigrew, author of A Social History of Tea, winner of the 2014 World Tea Awards' Best Tea Educator and director of studies at UK Tea Academy.

Indeed it was so pricey (a pound went for as much as a working-class citizen made in a year), that, according to Ellis,

"it ruled out anyone but the most elite and wealthiest sectors of society. So tea became associated with elite women's sociability around the royal court, of which Catherine was the most famous emblem."

And what happens with famous people? Non-famous people imitate them. "When the queen does something, everyone wants to follow suit, so very, very gradually by the end of the 17th Century, the aristocracy had started sipping small amounts of tea," Pettigrew said.

Of course, the upper class didn't invent the ritual of tea-drinking themselves – they were imitators too. As Pettigrew recounted, "Until tea arrived with the Dutch, we [the English] didn't know anything about tea. No sugar spoons, no cups, no tea kettles (only kitchen kettles), so we did what always happens: we copied the entire ritual from China. We imported [Chinese] tiny porcelain tea bowls, the saucers, the dishes for sugar, the small teapots."

Catherine's home country had a hand in in popularising this aspect of the tea experience, too. "Portugal was one of the routes [by which] porcelain got to Europe," Ellis noted. "It was very expensive and very beautiful, and one of the things that made tea drinking attractive was all the pretty stuff that went with it, like having the latest iPhone."

Since it was so prized, porcelain was probably part of Catherine's dowry, and, like other aristocratic ladies, she would have accrued many gorgeous trappings to pad out her tea sessions once she was living in England. Pettigrew explained, "She started it as an aristocratic habit in her palaces – very posh, very upper class, and so the ceremony that arrived from China was immediately associated with fine living. As soon as tea arrived, it had very strong connections to feminine women and very big houses, I suppose through Catherine, because the porcelain cost huge amounts of money. The poor had to make due with earthenware. So everything that was expensive had to do with the aristocracy. It's the same as today: You buy expensive things to show how important you are."

Eventually the lower classes transformed tea into a more egalitarian drink, but today, travellers to London can still experience the aristocratic pomp and circumstance at upscale hotels' afternoon tea services, most notably at the Langham Hotel's Palm Court in London (which claims to be the birthplace of afternoon tea), the famed Ritz London and Claridge's.

You can find fancy tea events in Portugal too, but even there, the link to Queen Catherine is not well known. In the historic municipality of Sintra, though, one hotel is trying to change that. At the Tivoli Palácio de Seteais Sintra Hotel, general manager Mario Custódio is about to launch a special afternoon tea themed after Catherine in October. "In school we don't get this [history]," Custódio said. "I had no idea. Even the Portuguese don't know this."

The area of Sintra, spread across lush green mountains about 30 minutes outside Lisbon, is a Unesco World Heritage Site, noted for its concentrated displays of European romantic architecture. The Seteais Palace, built in the 1780s by Dutch consul Daniel Gildemeester, is just one of several ornate, whimsical estate homes that dot the Sintra landscape; wedding-cake follies overlooking intricate, sprawling gardens and parks. Queen Catherine never lived here, but the concentration of old wealth and must-see mansions makes it the perfect place to reflect on what the lives of Portuguese nobility used to be like. You can easily imagine opulently dressed noblewomen gathering in opulently draped drawing rooms, clinking teacups and swapping news and gossip.

For Custódio, bringing these little-known bits of history to life is what makes the travel experience special and personal for visitors. "I'm trying to [present] these things that are very unknown because that is luxury today," he said.

If Queen Catherine gave you a gift of marmalade, she didn’t think that much of you

The daily tea service (open only to hotel guests), will highlight aspects of the Portuguese connection to this genteel tradition. For instance, Custódio is working with a historian to serve the type of tea Catherine would have drank (Ellis thinks it's most likely a green tea, as no tea came out of India until the 1830s, long after she'd passed away). Marmalade will also be part of the menu, as that's another part of the Catherine of Braganza mythology that Custódio has stumbled across in his research. The tale goes that, since some of the best oranges in the world come from Portugal, Catherine had them shipped over to her new English home regularly. The ones that didn't make the journey in top condition were turned into marmalade. Of course, whole oranges were a more prized snack, so if Queen Catherine gave you a gift of marmalade instead of oranges, it meant she didn't think that much of you.

The spread at the Seteais Palace will come with no such judgments. Custódio is simply hoping that by mingling with visitors during the themed tea service and by gifting them with a small book – complete with QR codes for more photos, historical facts and fun stories – he'll be helping to share some of the culture and colour of his home and reinforce the long-term influence of a little-known transplant queen.

"We Portuguese want to believe that Catarina was responsible for the tea. I don't want this history to die."

Version by human

凯瑟琳王后与英式下午茶

试想一下，什么才算英国人的标配？——紧绷着上唇，并且生来就爱好饮茶。但英国人的饮茶嗜好却是源于一位葡萄牙公主。你不妨想象一下，最典型的英国人长什么样？不管你脑海中浮现出哪一副画面，我很确

信，最典型的英国人一定是紧绷上唇，并且手里端着一杯茶。没错，英国人的确如此，有条不紊，酷爱饮茶。饮茶已经被彻头彻尾的英国化，成为英国不可或缺的文化符号，即使是外国人也将饮茶视为英国的标志性文化。

众所周知，茶源于中国。茶叶作为舶来品来到欧洲，西方对中国心怀感激。但是，多亏了葡萄牙，特别是一位葡萄牙女性，饮茶的习惯才得以在英伦广泛传播，而这则奇闻轶事却鲜为人知。下一次你到丽兹酒店，用精致的茶杯品味热气腾腾的乌龙茶时，抑或是参观维多利亚和阿尔伯特博物馆，站在格雷伯爵的肖像下的时候，不妨想象一下茶叶初到英伦的情形。

时间回到1662年，英王查理二世复辟，重登王座，布拉干萨王朝（Braganza）的公主凯瑟琳（葡萄牙国王约翰四世之女）嫁给了英国国王查理二世。凯瑟琳的嫁妆极其丰厚金钱、香料、珍宝、以及利润丰厚的两个港口——丹吉尔和孟买。联姻之后，凯瑟琳成为英格兰、苏格兰与爱尔兰的王后，地位显赫一时。

当凯瑟琳准备北上英伦成婚之际，据传她打包了一些散茶随身携带；也许，这些茶叶也是其陪嫁物品之一。有趣的是，传说这些茶叶在《芳香植物的运输》（葡语为：Transporte de Ervas Aromaticas）一书中有所记载，该书后来被简称为"T.E.A."（即英文"茶叶"一词）。

虽然这一传说可能不符合史实（词源学家认为英文单词"tea"是从汉语的"茶"音译的；应为闽南语"茶"音译，译者注）。但有一点可以肯定，葡萄牙通过澳门与中国进行直接贸易，因此茶叶率先受到葡萄牙贵族追捧。16世纪中期，葡萄牙窃取了中国的澳门（时至今日，中西方美食交流的另一个明证便是葡式蛋挞）。

凯瑟琳公主刚到英格兰时，人们只是把茶作为一味药品，因为饮茶可以振奋精神、理气健脾。但是，这位年轻的王后有每天饮茶的习惯，久而久之，茶叶便不再是保健补品，而是社交饮品，开始在英国盛行。

"与查理二世结婚之后，凯瑟琳迅速成为公众关注的焦点——她的穿着打扮、她使用的家具……所有与她相关的东西都成为王室谈资，"《布拉干萨王朝公主凯瑟琳：查理二世的王后》作者莎拉•贝丝•沃特金斯说。"她的日常饮茶活动引来许多人纷纷效仿。贵妇们纷涌而至，模仿其饮茶习惯以便融入王后的社交圈。"

凯瑟琳嫁来英国不久，当红诗人埃德蒙德•瓦勒甚至为她写了一首生日颂，将这位来自葡萄牙的王后与风行英伦的茶叶永远联系在了一起。诗歌中这样写道：

"最英明的王后，
最神奇的植物，
那个勇敢的国家，
阳光沐浴着的美丽之地
给予我们的丰厚恩赐
我们讴歌着用一颗感恩的心。"

准确来说，茶叶在凯瑟琳远嫁英伦之前已经出现，但是当时并不流行。"据记载，瓦勒在1657年就开始喝茶，这比凯瑟琳的出现早了整整六年，"伦敦玛丽王后大学的教授马克曼•埃利斯说，其致力于18世纪研究，与人合著《茶叶帝国：征服世界的亚洲树叶》。"他是有名的饮茶爱好者，这在当时并不多见，究其原因是茶叶价格昂贵，当时所有人都喝咖啡。"

茶叶昂贵的原因有三：一是当时英国同中国没有直接的贸易往来；二是印度的茶叶还没有传到英国；三是荷兰人进口的茶叶数量少，所以售价极高。

"茶叶价格居高不下，是因为从中国进口的关税非常高，"简•佩蒂格鲁说，其著有《茶叶社会史》一书，在2014年世界茶业大奖中被授予最佳茶业教育工作者奖，担任英国茶学院研究主任。

实际上，当时茶叶的价格高得惊人（一英镑茶叶的售价相当于一名劳工一整年的收入），埃利斯还提道，"除了社会最上层、最富有的人，其他人都不可能负担得起茶叶。因此茶叶成为贵妇与王室结交的手段，通过茶叶与凯瑟琳结交便是最好的例子。"

名人饮茶又有何妨呢？普通人会效仿名人饮茶。"无论王后做什么，其他人都会盲目效仿。于是，到了17世纪末，贵族们自然而然开始少量饮茶，"佩蒂格鲁说。

当然，饮茶仪式并不是上层阶级的发明，也是源于模仿。据佩蒂格鲁所述，"直到荷兰人带来茶叶，我们英国人才对茶有所了解。我们没有糖匙，没有茶杯，没有茶壶（只有厨房水壶），所以还是沿用老办法：模仿中国的整套饮茶仪式。我们从中国进口小的瓷茶杯、杯托、糖匙器具以及小茶壶。"

饮茶瓷器之所以风靡英国，还受到凯瑟琳的祖国——葡萄牙的影响。"葡萄牙是瓷器进口到欧洲的路线之一，"埃利斯说道。"瓷器价格昂贵，但非常精致漂亮，饮茶之所以受欢迎的原因之一就是可以使用这些精美的茶具，就像拥有最新款苹果手机一样。"

也许正因为瓷器贵重才能够作为凯瑟琳的嫁妆。像其他的贵妇一样，凯瑟琳开始在英国生活之后，利用这些精美之物增加下午茶的情趣。佩蒂格鲁解释说，"一开始，她只是在王室里饮茶，将其作为一种贵族的习惯，场面奢华，十足的上层阶级做派。因此，饮茶仪式自中国传入之后便迅速与高雅生活密不可分。茶叶一传入英国，便与贵妇、豪宅产生联系，我认为凯瑟琳在这个过程中发挥了重要作用，主要原因是瓷器的价格非常昂贵。穷人只配使用陶器，只有贵族才用得起一切贵重物品。现代社会也是如此：消费的东西越贵，代表你的地位越高。"

后来，劳工阶层使饮茶变得更加平民化。但在当今社会，伦敦的游客依然可以通过高档酒店提供的下午茶服务体验贵族的奢华，尤其是在伦敦朗廷酒店（Langham Hotel）的廷廊（据称这里是下午茶的诞生地）、赫赫有名的伦敦丽兹酒店（Ritz London）和凯莱奇酒店（Claridge's）。

此外，您还可以在葡萄牙体验极其盛大的饮茶仪式。尽管如此，就算在葡萄牙，英国饮茶之风与凯瑟琳王后的渊源也鲜为人知。然而，古镇辛特拉（Sintra）的一个酒店要努力改变这一状况。这就是提弗里瑟特阿斯皇宫酒店（Tivoli Palácio de Seteais Sintra Hotel），酒店总经理马里奥·库斯托迪奥（Mario Custódio）举办以凯瑟琳为主题的下午茶活动。"读书时，我们不了解这段历史，"库斯托迪奥说，"我什么也不清楚，就连其他葡萄牙人对此也一无所知。"

辛特拉地区距离里斯本大约30分钟的行程，向外延伸到植被茂盛的山脚下。这里因云集欧洲浪漫主义建筑而闻名，被列入教科文组织世界遗产名录。瑟特阿斯宫（Seteais Palace）由荷兰领事丹尼尔·吉尔德梅斯特（Daniel Gildemeester）于18世纪80年代建成，宫殿装饰奢华、造型独特，成为辛特拉景观中的一道亮丽风景线。婚礼蛋糕式样的怪异装饰物高高耸立于交错蔓生的花园与公园之上。虽然凯瑟琳王后从未居住于此，但这里云集的古老财富、经典的建筑群将葡萄牙贵族过去的生活呈现得淋漓尽致。想象一下，一群衣着华丽的贵妇正在富丽堂皇的会客厅聚会，时不时传来清脆的茶杯撞击声，以及叽叽喳喳的交头接耳声。

库斯托迪奥认为，将这些鲜为人知的历史碎片通过现实呈现出来，可以帮助游客获得更独特、更个性化的旅游体验。"我之所以努力将这些不为人知的历史呈现出来，是因为这些是现代社会的财富，"他说。

日常下午茶服务（只对酒店客人开放）能够凸显葡萄牙与高雅饮茶传统之间的渊源。例如，库斯托迪奥正与历史学家合作，希望能够做出凯瑟琳王后当时饮用的茶（埃利斯认为很可能是绿茶，原因是王后死后很长时间，直到19世纪30年代，才有印度来的茶叶引入到英国）。库斯托迪奥还希望能够做出橘子酱，这是因为在研究过程中，他偶然发现橘子酱也是布拉干萨王朝公主凯瑟琳诸多神秘传说之一。据说，世界上最好的橘子有一部分来自于葡萄牙，凯瑟琳王后会定期将这些橘子运往其英国的新家。有些橘子在运输途中品质下降，最后就被做成了橘子酱。若是能享用到完好无损的橘子自然最好，但如果凯瑟琳王后送给您的礼物是橘子酱而不是橘子，那说明您在她眼里算不上贵宾。

宣传瑟特阿斯宫，当然不是为了对您评头论足。库斯托迪奥的目的很简单，希望为游客提供主题下午茶服务、给游客赠送一本小书。书中印有二维码，扫描二维码可以获取更多图片、历史资料及有趣的故事。他只是想通过这些方式，帮助游客了解自己家乡的文化与美景，发掘远嫁他乡的王后如何对英国饮茶文化产生深远的影响。

"凯瑟琳王后对茶文化做出了巨大的贡献，我们葡萄牙人不应当对此视而不见。我希望这段珍贵的历史能够得以延续。"

APPENDIX 2

English-Chinese translation (Text 2)

Source text (English)

Cave discovered on Moon could be home for humans

15 July 2024

Georgina Rannard

Science reporter

Scientists have for the first time discovered a cave on the Moon.

At least 100m deep, it could be an ideal place for humans to build a permanent base, they say.

It is just one in probably hundreds of caves hidden in an "underground, undiscovered world", according to the researchers.

Countries are racing to establish a permanent human presence on the Moon, but they will need to protect astronauts from radiation, extreme temperatures, and space weather.

Helen Sharman, the first British astronaut to travel to space, told BBC News that the newly-discovered cave looked like a good place for a base, and suggested humans could potentially be living in lunar pits in 20-30 years.

But, she said, this cave is so deep that astronauts might need to abseil in and use "jet packs or a lift" to get out.

Lorenzo Bruzzone and Leonardo Carrer at the University of Trento in Italy found the cave by using radar to penetrate the opening of a pit on a rocky plain called the Mare Tranquillitatis.

It is visible to the naked eye from Earth, and is also where Apollo 11 landed in 1969.

The cave has a skylight on the Moon's surface, leading down to vertical and overhanging walls, and a sloping floor that might extend further underground.

It was made millions or billions of years ago when lava flowed on the Moon, creating a tunnel through the rock.

The closest equivalent on Earth would be the volcanic caves in Lanzarote, Spain, Prof Carrer explains, adding that the researchers visited those caves as part of their work.

"It's really exciting. When you make these discoveries and you look at these images, you realise you're the first person in the history of humanity to see it," Prof Carrer said.

Once Prof Bruzzone and Prof Carrer understood how big the cave was, they realised it could be a good spot for a lunar base.

"After all, life on Earth began in caves, so it makes sense that humans could live inside them on the Moon," says Prof Carrer.

The cave has yet to be fully explored, but the researchers hope that ground-penetrating radar, cameras or even robots could be used to map it.

Scientists first realised there were probably caves on the Moon around 50 years ago. Then in 2010 a camera on a mission called the Lunar Reconnaissance Orbiter took pictures of pits that scientists thought could be cave entrances.

But researchers did not know how deep the caves might be, or if they would have collapsed.

Prof Bruzzone and Prof Carrer's work has now answered that question, although there is much more to be done to understand the full scale of the cave.

"We have very good images of the surface - up to 25cm of resolution - we can see the Apollo landing sites - but we know nothing about what lies below the surface. There are huge opportunities for discovery," Francesco Sauro, Coordinator of the Topical Team Planetary Caves of the European Space Agency, told BBC News.

The research may also help us explore caves on Mars in the future, he says.

That could open the door to finding evidence of life on Mars, because if it did exist, it would almost certainly have been inside caves protected from the elements on the planet's surface.

The Moon cave might be useful to humans, but the scientists also stress that it could help answer fundamental questions about the history of the Moon, and even our solar system.

The rocks inside the cave will not be as damaged or eroded by space weather, so they can provide an extensive geological record going back billions of years.

The research is published in the scientific journal Nature Astronomy.

Graphic by Gerry Fletcher

Version by human translator (used as BLEU metric reference)

科学家在月表下发现大型洞穴，或许适宜人类居住

乔治娜•兰纳德（Georgina Rannard）

Role, BBC科技事务记者
2024年7月18日

科学家首次在月球上发现了一个洞穴。

他们表示，该洞穴至少有100米深，可能是人类建造永久基地的理想场所。

研究人员称，这可能只是隐藏在月球"未被发现的地下世界"里的数百个洞穴中的一个。

目前，各国都在竞相研究在月球上建立人类永久基地，但他们需要保护宇航员免受辐射、极端温度和太空天气的影响。

英国第一位遨游太空的宇航员海伦•沙曼（Helen Sharman）告诉BBC新闻，这个新发现的洞穴看起来是一个建立基地的好地方，并表示人类有可能在20至30年后就生活在月球坑洞中。

但她表示，这个洞穴太深了，宇航员可能需要通过绳索下降进入，然后使用"喷气背包或升降机"才能出来。

意大利特伦托大学（University of Trento）的洛伦佐•布鲁佐内（Lorenzo Bruzzone）和莱昂纳多•卡雷尔（Leonardo Carrer）利用雷达穿透了月球岩石平原"静海"上的一个坑洞开口，从而发现了这个洞穴。

这里从地球上肉眼可见，也是1969年"阿波罗11"号（Apollo 11）的着陆地。

该洞穴在月球表面有一个洞道，通向垂直的悬壁，以及可能延伸到地下更深处的倾斜层面。

它是数百万乃至数十亿年前，由熔岩在月表流动时形成的，熔岩在岩石中形成一条条隧道。

卡雷尔教授解释说，地球上与此最接近的洞穴，应该是西班牙兰萨罗特岛的火山洞穴。他补充说，研究人员参观了这些洞穴以进行研究。

"这真的令人兴奋。当你有了这些发现并看到这些图像时，你会意识到自己是人类历史上第一个看到它的人。"卡雷尔教授说。

当布鲁佐内教授和卡雷尔教授了解该洞穴的规模之后，他们意识到这可能是一个建立月球基地的好地方。

"毕竟，地球上的生命起源于洞穴，所以人类在月球上住在洞穴里也是合情合理的。"卡雷尔教授说。

这个洞穴还没有被完全探索过，但研究人员希望可以用透地雷达、摄像机甚至机器人来绘制其地图。

大约50年前，科学家们首次意识到月球上可能存在洞穴。2010年，被称为"月球勘测轨道飞行器"（Lunar Reconnaissance Orbiter）飞船上的一台相机拍摄了一些坑洞的照片，科学家认为这些坑洞可能是洞穴入口。

但研究人员并不知道这些洞穴可能有多深，也不知道它们是否已经坍塌。

现在，布鲁佐内和卡雷尔教授的工作已经回答了这个问题，尽管要了解洞穴的整体规模还有很多工作要做。

"我们有非常清晰的月表图像——分辨率高达25厘米——我们可以看到阿波罗着陆点——但我们对月表以下的情况一无所知。这有巨大的发现机会。"欧洲航天局（European Space Agency）行星洞穴专题小组协调员弗朗西斯科•索罗（Francesco Sauro）告诉BBC新闻。

他说，这项研究还可能帮助我们将来探索火星上的洞穴。

这可能为我们寻找火星生命的证据打开大门，因为如果生命确实存在，那么几乎肯定它位于免受火星表面自然因素影响的洞穴内。

月球洞穴可能对人类有用，但科学家们也强调，它可能有助于回答有关月球甚至太阳系历史的基本问题。

洞穴内的岩石不会受到太空天气的破坏或侵蚀，因此可以提供数十亿年前的大量地质记录。

这项研究发表在科学杂志《自然-天文学》（Nature Astronomy）上。

制图：格里•弗莱彻（Gerry Fletcher）

APPENDIX 3

English-Chinese translation (Text 3)

Source text (English)

The US expects China to retaliate over the Trump administration's latest tariff hike, chief economic adviser Larry Kudlow said on Sunday.

"The expected countermeasures have not yet materialized," Larry Kudlow told Fox News Sunday. "We may know more today or even this evening or tomorrow."

As markets digested the comments, futures trading pointed to a drop in Wall Street stocks on Monday while Asian shares slipped in early trade.

Talks ended on Friday and Trump raised the spectre of a full-blown trade war when he ordered that tariffs on Chinese imports worth around $200bn be raised from 10% to 25%.

Beijing retaliated for previous tariff hikes by raising duties on $110bn of US imports. Chinese officials have also targeted US companies by slowing customs clearance and stepping up regulatory scrutiny.

On Saturday he had tweeted saying Beijing had deliberately sought to delay negotiations. "China felt they were being beaten so badly in the recent negotiation that they may as well wait around for the next election, 2020, to see if they could get lucky [and] have a Democrat win," he said.

Kudlow told Fox no more talks were planned but also said there was a "strong possibility" Trump and Chinese president Xi Jinping "will meet in Japan at [the] G20" summit at the end of June.

In China, state media said the door to talks was always open but China would not yield on important issues of principle. In a commentary carried by the official Xinhua news agency, the ruling Communist Party's People's Daily said there are no winners in any trade war and China did not want to fight but would not be afraid to do so.

Version by human translator (used as BLEU metric reference)

白宫首席经济顾问拉里•库德洛周日在接受福克斯新闻采访时表示，由于特朗普政府对中国输美货物提高了进口关税，美方预计中方会采取相应的反制措施。

"预料中的反制措施尚未细化，"拉里•库德洛对福克斯新闻说，"或许今天，或许今晚，或许明天，我们就能知道更多细节。"

证券市场受到此番表态影响，周一华尔街股票交易所期货交易下跌，亚洲股指早盘下跌。

中美贸易谈判在周五已经结束。与此同时，特朗普下令将价值约2000亿美元的中国输美货物关税从10%提高到25%，此举引发了中美可能爆发全面贸易战的担忧。

为了反击美方此前肆意提高关税，中国政府曾对美国价值1100亿美元的对华出口商品提高关税。中方官员针对美国公司，采取了延缓办理清关手续以及加强监管审查的反制措施。

周日，拉里•库德洛在推特上表示，中国政府试图故意拖延谈判。"中方感觉在最近的贸易谈判中没有占到便宜，因此采取拖延战术，静待2020年美国总统选举，并寄希望于民主党人胜选。"

库德洛在接受福克斯访谈时还表示，中美无意再次举行贸易谈判，但特朗普与中国国家主席习近平"极有可能"将于6月底"在日本的G20峰会期间举行会晤。"

在中国，中央媒体称谈判的大门永远敞开，但中国绝不会再重大原则性问题上让步。官方通讯社新华社刊发了中国共产党机关报《人民日报》的评论，认为贸易战不会有任何赢家，中国并不想打贸易战，但也不惧怕贸易战。

APPENDIX 4

Chinese-English translation (Text 1)

<div align="center">

Source text (Mandarin Chinese)
</div>

黃耳

《述異記》祖沖之

陸機少時，頗好獵。在吳，有家客獻快犬，名曰黃耳。機後仕洛，常將自隨。此犬點慧，能解人語。又嘗借人三百裏外，犬識路自還，一日至家。

機羈官京師，久無家問，因戲犬曰：「我家絕無書信，汝能賫書馳取消息不？」犬喜搖尾作聲應之。機試為書，盛以竹筒，系之犬頸，犬出驛路，疾走向吳。饑則入草噬肉取飽。每經大水，輒依渡者，弭耳掉尾向之，其人憐愛，因呼上船。載近岸，犬即騰上以速去，先到機家口，銜筒作聲示之。機家開筒取書，看畢，犬又向人作聲，如有所求；其家作答書，內筒，復系大頸，犬既得答，仍馳還洛。計人程五旬，而犬往還裁半月。

後犬死，殯之，遣送還葬機村，去機家二百步，聚土為墳。村人呼為黃耳冢。

<div align="center">

Version by human translator (used as BLEU metric reference)
</div>

"Yellow Ears" - A Legendary Messenger Dog

A COLLECTION OF STRANGE TALES

ZU CHONGZHI

Lu Ji (261-303 AD) was a famous poet in China's West Jin Dynasty. As a young man, he was a huge fan of hunting. When he lived in the Shire of Woo, one retainer of the Lu family presented him a hound named Yellow Ears. Then the poet got a position at central government and moved to the nation's capital. Of course that loyal dog accompanied him. The dog was so smart that it could even understand human language. Lu ever lent the dog to a friend and his friend took the dog on a trip. The clever creature was not deterred by 100 miles' distance from home. Instead, following the same path, it only took Yellow Ears one day to come back home.

As a government official, Lu could not leave the capital city without permission. What was worse, communication with his family was extremely difficult due to inadequate postal system at that time. Thus Lu said jokingly to his dog, "For such a long period that I did not hear anything from my family. Could you do me a favor to send a letter to my family and carry their reply back to me?" Wagging its tail, the dog joyously barked as if it was replying "yes!" Lu tentatively wrote a letter, sealed it in a bamboo tube, and tied the tube around the hound's neck. Yellow Ears followed the post road and ran fast towards the Shire of Woo. When the dog became hungry, it would hunt for some small animals in bush and ate their meat. The post road was often cut off by broad rivers. Then the hound had to seek help from ferry boats. By moving down yellow ears and wagging tail, the lovely dog gained sympathy from every boatman. In such a way the hound got on board and crossed rivers. Right before the ferry boat was anchored, Yellow Ears would jump off and rushed away. When it arrived at Lu's old house, the dog barked while biting the bamboo tube in it mouth, implicating that there was something important inside the tube. Reading its gesture language, the family members opened the tube and read the letter. When they finished reading, the dog barked again as if it was requesting for a reply. Therefore, the Lu family would like to follow suit: composing a reply letter, sealing it in the same tube, and tying it around the dog's neck. Once the hound got the reply letter, it immediately started its long return journey. At that time, it usually took an average mailman 50 days to finish the task, while it took only the magic dog 15 days or so! Lu gave his messenger dog a lavish funeral after it died years later by burying it in his hometown village. The dog's tomb was just 200 yards away from Lu's house. The high grave mound was quite obvious and fellow villagers called it the Tomb of Yellow Ears.

APPENDIX 5

Chinese-English translation (Text 2)

<div align="center">

Source text (Mandarin Chinese)
</div>

《酉阳杂俎》天咫

• 太和中，鄭仁本表弟，不記姓名，嘗與一王秀才遊嵩山。捫蘿越澗，境及幽夐，遂迷歸路。將暮，不知所之。徒倚間，忽覺叢中鼾睡聲。披蓁窺之，見一人布衣，衣甚潔白，枕一襆物，方眠熟。即呼之曰："某偶入此徑，迷路，君知向官道否？"其人舉首略視，不應，復寢。又再三呼之，乃起坐，顧曰："來此！"二人因就之，且問其所自。其人笑曰："君知月乃七寶合成乎？月勢如丸，其影，日爍其凸處也。常有八萬二千戶修之，予即一數。"因開襆，有斤鑿數事，玉屑飯兩裹，授予二人，曰："分食此，雖不足長生，可一生無疾耳。"乃起，與二人指一支徑："但由此，自合官道矣。"言已，不見。

<div align="center">

Version by human translator (used as BLEU metric reference)
</div>

A REPAIRMAN FROM THE MOON

• YOU YANG ZA ZU (THE MISCELLANEOUS RECORD OF YOU YANG)

In a year between 827 and 835 (China's Tang Dynasty), Zheng Renben's cousin and his friend Wang once visited Mt. Song in central China. By crossing secluded valleys and running brooks, the two gentlemen were so fascinated by the spectacular views that finally they got lost. It was about sunset, and they didn't know their whereabouts. Without knowing what to do, they just stood where they were.

Suddenly they heard someone's loud snore. Pulling aside the bush branches, they found a man in white soundly sleeping with a package under his head. Then the two friends hurriedly stepped forward to wake him up and asked for help," Sir, we accidentally came to this place and got lost. Could you tell us where the highway is?" The strange man raised his head and peered for a while. Without any reply, he went to sleep again. Once again the two gentlemen woke him up, and this time the strange man got up, calling them to move closer. Thus they came closer and asked the man in white where he was from.

Smiling, the strange man replied, "do you know that the moon consists of 7 different elements? It's a sphere, and the moonlight is actually sunlight reflected by its mountains and peaks. There're altogether 82000 households repairing the moon and I'm one of them." Opening up his package, the repairman showed them his various tools as well as two boxes of "jade rice". He then offered the visitors the meal and explained, "please share my food. Although it can not let you become immortal, it can definitely make you healthy for the rest of your life." Standing up, the repairman pointed at a lane and said, "you just take this route and the highway is not far away." After saying that, the man in white suddenly vanished.

APPENDIX 6

Chinese-English translation (Text 3)

<div align="center">

Source text (Mandarin Chinese)

</div>

三中全會精神宣講會周一舉行

2024年8月23日

中央應特區請求，派遣中央宣講團來港講解中國共產黨第二十屆中央委員會第三次全體會議的精神。宣講會8月26日在灣仔香港會議展覽中心舉行。

宣講團主講嘉賓包括全國人民代表大會憲法和法律委員會副主任委員、全國人民代表大會常務委員會法制工作委員會主任沈春耀，以及中華人民共和國商務部黨組書記、部長王文濤。

特區政府表示，中共二十屆三中全會不但引領國家邁向中國式現代化和高質量發展，也對香港的長遠發展和民生福祉至關重要，全港市民都應認真學習貫徹三中全會精神。在中央大力支持下，宣講團來港向社會各界講解三中全會精神，機會難得，特區政府呼籲大眾踴躍收看，更充分認識和理解三中全會精神的具體內容和重大意義。

宣講會8月26日下午3時至5時30分在網上直播，市民屆時可在政府網上廣播或政制及內地事務局Facebook專頁收看，本港主要電視台也會直播。宣講會完整內容稍後載於政制及內地事務局網站。

<div align="center">

Version by human translator (used as BLEU metric reference)

</div>

Seminar on plenary session set

August 23, 2024

The Hong Kong Special Administrative Region Government announced today that it will hold a seminar on August 26 to promote the spirit of the Third Plenary Session of the 20th Central Committee of the Communist Party of China (CPC Central Committee).

The Hong Kong SAR Government highlighted that upon its invitation, the central government is showing its staunch support for the Hong Kong SAR by sending a publicity delegation to Hong Kong to speak on the spirit of the plenary session of the 20th CPC Central Committee.

The aim of the seminar, which will be held at the Convention & Exhibition Centre in Wan Chai, is to promote the spirit of the Third Plenary Session of the 20th CPC Central Committee to people from various sectors of the community.

Keynote speakers include National People's Congress Constitution & Law Committee Vice-chairperson and Chairman of the Legislative Affairs Commission of the National People's Congress Standing Committee Shen Chunyao, as well as Secretary of CPC Leadership Group of the Ministry of Commerce & Minister of Commerce Wang Wentao.

The Hong Kong SAR Government pointed out that it is an invaluable opportunity to have the delegation speak on the spirit of the plenary session, and all citizens should watch the seminar to fully understand the content and significance of the spirit.

In addition to leading the country towards Chinese modernisation and high-quality development, the plenary session is also crucial to Hong Kong's long-term development and people's well-being, it stressed.

The seminar will be broadcast from 3pm to 5.30pm on August 26 via a webcast and on the Constitutional & Mainland Affairs Bureau's Facebook page. Major local television stations will also air the event.

Additionally, the full version of the seminar will be uploaded to the bureau's website afterwards.