

Homework 5: Let's Play LLM

Part 1: Designing a Task for LLM and Explore the Capability of LLM (70%)

Task Description (10%)

- I want to solve the Chinese literature exams. Analyzing whether LLM have the ability to understand underlying meaning of articles.
- I will provide GPT with a Chinese article and a multiple-choice question with four options. analyze the article and choose the correct option, explaining how the four options were analyzed to select the answer.
- In my opinion it can not know the meaning behind the article without any hint. Therefore, I want to compare GPT answer the question with hint and without hint.

Motivation (10%)

- Due to the high controversy surrounding Chinese literature exams over the years, I'm curious whether LLMs have the ability to understand texts, such as understanding the emotions the author wants to express or uncovering hidden meanings in the text, to address this ongoing controversy. By examining GPT's comprehension of a Chinese article from the General Scholastic Ability Test, we can gain insights into its capabilities and limitations in interpreting complex literary texts.
- I want to explore whether LLMs, such as GPT, can accurately interpret and analyze Chinese literary texts. Specifically, I am interested in their ability to understand the emotions the author wants to convey and to uncover hidden meanings within the text. This exploration will help determine if LLMs can effectively address the controversies surrounding Chinese literature exams and provide reliable interpretations.

Describe your Initial Attempt (15%)

- **Input:** Enter the article and four options without giving any hints, and ask them to analyze the options and make a choice.

中國湖南省江永地區，一種神祕的菱形文字——女書——悄悄流傳超過百年。它是一種使用當地方言，須以「唱誦」方式表達與理解的表音文字。女人用它來自傳訴情、互通心跡、結交姊妹、祝禱祈願、傳說紀聞.....。

江永是典型的農業社會，女孩從小「不得亂步出娘房」，因此未嫁的女兒稱作「樓中女」。未嫁的女兒在娘樓做女紅時，邊唱歌、邊學認女書、寫女書、織女書字，進入正統教育體系之外的知識系統，且因熱愛女書而結為姊妹。當有姊妹出嫁，每個姑娘從婚前幾天便開始哭嫁，餽贈女書留念，感懷女性身不由己的共同命運，娘樓是她們建立封閉又親密的女

性空間，但事實上又是「無常」的象徵。

在以哭嫁為主題的女書作品中，常常描述父系社會婚姻將女人從原生土壤連根剝起的滋味，顯示哭嫁不僅是少女面對未知命運的恐懼，更是女人剪斷血緣、地緣與人緣臍帶時的哀號。但也由於有女書，因而婚後她們並不像其他文盲婦女喪失向外通訊、結盟的能力。姊妹們的通信中苦多樂少，但總是盡心相勸、彼此扶持。

她們透過女書，在社會現實與磨難中掙扎、拉鋸、妥協，尋找生存之道。甚至在書寫過程中，執筆者也不止於代書，悲憫勸解的同時，常將自己的身世寫了進去，形成風格特殊的「參與寫作」。

關於女書的起源至今仍是個謎，江永女性發明了世界上獨一無二的女書，但因以土話為音，加上地方阻隔，僅局限於江永一帶流傳。也由於女性死後女書多作為殉葬而被焚毀，以及隨著教育的普及、傳統習俗、社會情境的改變等因素，女書逐漸式微。然而，女書的研究與應用方興未艾，尤其是在消費時代裡，女書反成為一種時尚符號，常出現在文創設計中。（改寫自鄭至慧、劉斐玟等相關研究報告）

依據上文，關於女書的說明，不適當的是：

- (A)女書是婦女可掌握自我話語權的發聲筒，也是姊妹相濡以沫情誼的表現
- (B)以女書書寫他人故事，常在他人生命中照見自己，而融入個人身世感懷

(C)女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計

(D)藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公

分析並解釋四個選項並選出答案

- **Output:**

| 從所提供的段落中，我們可以分析以下選項：

(A) 女書是婦女可掌握自我話語權的發聲筒，也是姊妹相濡以沫情誼的表現

這個選項與原文相符。女書被用作婦女間的溝通工具，透過女書，未嫁女子和已嫁女子能夠保持聯繫，表達彼此的支持和情誼。

(B) 以女書書寫他人故事，常在他人的生命照見自己，而融入個人身世感懷

根據原文，女書常用於寫作過程中融入個人的身世與感懷，不只是單純的代書，而是包含了寫作者的情感和經歷，與這個選項相符。

(C) 女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計

原文提到，女書雖然歷史悠久，但隨著時間的推移和社會的變遷，逐漸式微。然而，它在當代被重新發現，並用於文創設計。這說明了女書不僅是一種歷史的記錄，也成為現代文化的一部分。

(D) 藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公

這個選項暗示了女書具有表達抗議或批評社會不公的功能。然而，根據原文，女書主要是作為一種溝通和表達情感的工具，並沒有直接提到它被用來公開批評社會或面對現實困境。

根據這些分析，(D) 選項是不適當的，因為它賦予女書一個在原文中未被提及的社會抗議功能。其他選項都與原文中描述的女書的用途和影響相吻合。

- **Analysis:** GPT ultimately chose D, but the actually answer is C. And GPT said, "根據原文，女書主要是作為一種溝通和表達情感的工具，並沒有直接提到它被用來公開批評社會或面對現實困境." However, according to the original text, "每個姑娘從婚前幾天便開始哭嫁，餽贈女書留念，感懷女性身不由己的共同命運，娘樓是她們建立封閉又親密的女性空間，但事實上又是「無常」的象徵," it is evident that GPT's description conflicts with the original text. GPT did not interpret the underlying meaning hidden in the original text.

Describe Improvements (35%)

● Input:

1. Enter the article and four options with giving hints, and ask them to analyze the options and make a choice.

中國湖南省江永地區，一種神祕的菱形文字——女書——悄悄流傳超過百年。它是一種使用當地方言，須以「唱誦」方式表達與理解的表音文字。女人用它來自傳訴情、互通心跡、結交姊妹、祝禱祈願、傳說紀聞……。

江永是典型的農業社會，女孩從小「不得亂步出娘房」，因此未嫁的女兒稱作「樓中女」。未嫁的女兒在娘樓做女紅時，邊唱歌、邊學認女書、寫女書、織女書字，進入正統教育體系之外的知識系統，且因熱愛女書而結為姊妹。當有姊妹出嫁，每個姑娘從婚前幾天便開始哭嫁，餽贈女書留念，感懷女性身不由己的共同命運，娘樓是她們建立封閉又親密的女

性空間，但事實上又是「無常」的象徵。

在以哭嫁為主題的女書作品中，常常描述父系社會婚姻將女人從原生土壤連根剝起的滋味，顯示哭嫁不僅是少女面對未知命運的恐懼，更是女人剪斷血緣、地緣與人緣臍帶時的哀號。但也由於有女書，因而婚後她們並不像其他文盲婦女喪失向外通訊、結盟的能力。姊妹們的通信中苦多樂少，但總是盡心相勸、彼此扶持。

她們透過女書，在社會現實與磨難中掙扎、拉鋸、妥協，尋找生存之道。甚至在書寫過程中，執筆者也不止於代書，悲憫勸解的同時，常將自己的身世寫了進去，形成風格特殊的「參與寫作」。

關於女書的起源至今仍是個謎，江永女性發明了世界上獨一無二的女書，但因以土話為音，加上地方阻隔，僅局限於江永一帶流傳。也由於女性死後女書多作為殉葬而被焚毀，以及隨著教育的普及、傳統習俗、社會情境的改變等因素，女書逐漸式微。然而，女書的研究與應用方興未艾，尤其是在消費時代裡，女書反成為一種時尚符號，常出現在文創設計中。（改寫自鄭至慧、劉斐玟等相關研究報告）

依據上文，關於女書的說明，不適當的是：

- (A)女書是婦女可掌握自我話語權的發聲筒，也是姊妹相濡以沫情誼的表現
- (B)以女書書寫他人故事，常在他人生命中照見自己，而融入個人身世感懷
- (C)女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計
- (D)藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公

"也由於女性死後女書多作為殉葬而被焚毀，以及隨著教育的普及、傳統習俗、社會情境的改變等因素，女書逐漸式微"

根據這段描述，尤其是"女書逐漸式微"來判斷答案

2. Because GPT tends to lean towards option D being correct, we can narrow down the choices for it. In the final description, we will remove option D to help it.

(D)藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公

其中D選項是適當的

3. Directly tell GPT that the answer is incorrect without providing any other hints.



You

這個回答是錯誤的。

● Output:

Method1:

根據所提供的段落，特別是「女書逐漸式微」這部分描述，我們可以重新分析選項：

(A) 女書是婦女可掌握自我話語權的發聲筒，也是姊妹相濡以沫情誼的表現
這個選項與原文相符。女書確實是婦女之間交流情感和保持聯繫的工具。

(B) 以女書書寫他人故事，常在他人的生命照見自己，而融入個人身世感懷
這個選項與原文相符。女書的書寫者常將自己的身世感懷融入所寫的內容中。

(C) 女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計
這個選項與原文部分相符。雖然女書被現代的江永女性用於文創設計，但「女書逐漸式微」的描述表示女書在當地的傳承和使用實際上正在減少。

(D) 藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公
這個選項並不完全適當。雖然女書被用來表達情感和訴說困境，但原文並未提及它被用來公開批評社會不公。

因此，根據「女書逐漸式微」的描述，答案應該是：

(C) 女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計

這個選項中的「而今江永女性為復甦女書多投身文創設計」與「女書逐漸式微」的說法存在矛盾，因此不適當。

Method2:

ChatGPT

好的，讓我逐一解釋每個選項：

(A) 女書是婦女可掌握自我話語權的發聲筒，也是姊妹相濡以沫情誼的表現：根據文中描述，女書被用來自傳情感、互通心跡，並且女性在女書中建立了姊妹情誼，因此這個描述是合適的。

(B) 以女書書寫他人故事，常在他生命中照見自己，而融入個人身世感懷：文中提到女性在女書中表達自己的情感和身世感懷，同時也可以通過他人的故事反映出自己的情感，因此這個描述也是合適的。

(C) 女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計：文中沒有提到江永女性為復甦女書而投身文創設計的情況，這個描述在原文中並未提及，因此不適當。

(D) 藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公：根據文中描述，女性使用女書來表達情感、傳達情況，因此可以用來面對現實困境和悲訴社會的不公，這個描述是合適的。

—

Method3:

ChatGPT

你是對的，我抱歉，請容許我重新檢查。

(A) 女書是婦女可掌握自我話語權的發聲筒，也是姊妹相濡以沫情誼的表現 - 正確。文中提到女書是婦女表達自我、傳達情感、結交姊妹、寄託祈願等方式，因此這個選項是適當的。

(B) 以女書書寫他人故事，常在他生命中照見自己，而融入個人身世感懷 - 這個說法在文中沒有提及，因此不適當。

(C) 女書是江永婦女的生命刻痕，而今江永女性為復甦女書多投身文創設計 - 正確。文中提到女書是江永婦女的一個重要文化遺產，而現今的江永女性則將其用於文創設計，使女書在當代獲得新生。

(D) 藉由女書音聲與表意的交織，女性藉以面對現實困境、悲訴社會的不公 - 正確。文中提到女書是婦女用來表達情感、面對困境、結交姊妹等的工具，因此這個選項是合適的。

所以，不適當的選項是(B)。

Analysis:

Method1: By providing key sentences corresponding to option C as hints, we can see that GPT is able to grasp the main points of the article and answer the correct option. Therefore, organizing the key points of the article before asking GPT to answer the question can improve the quality of its response.

Method2: After eliminating one option for GPT, although it chose the correct option, it is uncertain whether it increased the accuracy due to the elimination or if it was a random guess that led to the correct answer.

Method3: Directly telling GPT that its answer is incorrect, but it chose the wrong option (B), so this method does not seem to be effective.

Providing key sentences related to the correct option (Method 1) significantly improves GPT's response quality. Removing one option (Method 2) is simple but its effectiveness is uncertain. Directly informing GPT of an incorrect answer (Method 3) seems ineffective. Thus, the most effective method is providing key sentences as hints, helping GPT grasp the main points of the article and improve accuracy.

Among the three methods, providing key sentences as hints (Method 1) is the most effective in helping GPT accurately understand the article and make the correct choice.

Part 2: Problem Answering (30%)

Question 1

1. Besides providing a more detailed description of the question in the prompt, what other methods can be employed to improve the accuracy of the responses?

Other methods, such as summarizing the key points and highlighting the main content of the question, can also improve accuracy. For instance, by providing keywords or summaries of important sentences, you can help focus the analysis on the most relevant information. This approach ensures that the response is based on a clear understanding of the critical aspects of the text.

Additionally, highlighting key phrases or sentences within the article can guide the analysis towards the most pertinent parts, thereby enhancing the likelihood of selecting the correct option. This technique, combined with detailed descriptions, can significantly improve the overall quality and accuracy of the responses.

2. Please provide an example demonstrating the impact of implementing the aforementioned methods on a prompt.

情緒訊號系統有個特色，就是它總是「開啟」的。情緒開始時，幾乎會立刻出現訊號。例如傷心時，眉毛內緣會向上拉，聲音會變得較輕柔低沉。最立即的情緒訊號是臉部表情。人們常把生氣、害怕、嫌惡、哀傷和輕蔑視為負面情緒，驚訝和愉快是正面情緒。這簡單的二分法有兩個問題，一是忽略所謂負面情緒並不總是令人不舒服，如許多人看了一場哀傷的電影後，喜歡大哭一場。二是被視為正向情緒的快樂，也可能很殘忍，如嘲笑別人。唯有檢視各種情緒事件的特性，才能分辨使人愉快還是不愉快。

聲音是另一種情緒訊號系統，與臉部表情有許多有趣的差別。臉部是可以觀察的，聲音則是時斷時續的，可以憑意志完全關閉。想隱藏表情，可能是人常以電話取代面對面溝通的原因。電子郵件甚至不需要說和聽，所以不會從聲音流露情緒。

身體動作也是情緒訊號：輕鬆時會出現放鬆身體的姿勢，輕蔑的動作是從上往下看對方，驚訝則是把注意力固定在產生情緒的對象。身體動作雖然像臉部和聲音的情緒訊號一樣是不由自主的，但對大多數人而言，控制身體的動作，比完全不露出臉部和聲音的情緒訊號更為容易。（改寫自保羅·艾克曼《心理學家的面相術》）

依據上文，關於情緒訊號的敘述，最適當的是：

- (A) 正向情緒會使人露出笑容，總能帶給他人愉快
- (B) 臉部表情比身體動作情緒訊號，更易憑意志控制
- (C) 人們常以電話溝通，可能是因為聲音比表情容易隱藏情緒
- (D) 情緒訊號系統總是開啟的，通常從身體動作最先出現訊號



依據上文，最適當的敘述是：

(D) 情緒訊號系統總是開啟的，通常從身體動作最先出現訊號。

請根據"聲音則是時斷時續的，可以憑意志完全關閉。想隱藏表情，可能是人常以電話取代面對面溝通的原因。"來決定答案



根據 "聲音則是時斷時續的，可以憑意志完全關閉。想隱藏表情，可能是人常以電話取代面對面溝通的原因。" 這段文字，最適當的答案是：

(C) 人們常以電話溝通，可能是因為聲音比表情容易隱藏情緒。

🔊 📄 ↺ 🔍 ↻

The correct answer to this question is (C). This observation aligns with the approach discussed in PART 1, where summarizing important sentences can significantly enhance the accuracy of responses. By providing a focused summary of the key points and highlighting crucial content within the text, it becomes easier for GPT to identify and analyze the most relevant information. This method helps in narrowing down the options and directs attention to the essential parts of the article, leading to a more accurate selection of the correct answer. This demonstrates that organizing and summarizing key sentences is an effective strategy for improving response quality and accuracy.

Question 2

- Question :

How to enhance the efficiency of Large Language Models (LLMs) while reducing computational and memory requirements
- Motivation :

Despite their impressive performance in language-related tasks, the substantial computational and memory demands of LLMs limit their widespread application in resource-constrained environments; this drives researchers to explore more efficient model compression and system optimization techniques.
- Challenges :
 - 1.How to compress LLMs without significantly degrading performance?
 - 2.How to evaluate the practical effectiveness of various compression

techniques?

3. Can system-level optimization methods effectively complement model compression techniques?

- Solution :

1. Implementing techniques like structured pruning and quantization, which selectively reduce model parameters and precision, can effectively compress models while maintaining acceptable performance levels. Knowledge distillation, which transfers knowledge from a large model (teacher) to a smaller one (student), also helps retain performance.

2. Use standardized benchmarks and evaluation metrics such as model weight memory, runtime memory consumption, inference token rate, and perplexity. Experimental analysis on models like LLaMA-7B with different compression techniques provides practical insights and performance comparisons.

3. Yes, system-level optimization methods like Paged Attention, Flash Attention, and Fused Operations can significantly enhance the runtime efficiency of LLMs. These techniques optimize memory management and computation, thereby complementing model compression methods to achieve better overall performance.

- Reference: <https://arxiv.org/html/2402.01799v1>