

Testing of applying LLMs to Quan Song Shi

Prompts Engineering

I used three prompts in total. I fed them to LLMs sequentially. Below are the prompts that I used:

1rst Prompt:

Names in ancient China can take various forms. Including 名 (given name), 字 (styled name), 姓 (Last name, or surname), and 號 (pseudonym). It is unsure which one will appear. There are some general rules to learn about names that appear in the communication poem: 字, 別號, 姓+字, or 姓+名.

you can learn the general rule from 'ifla_names_of_persons_chinese_names_2020' pdf file and common Chinese names in the Song dynasty from the following webpage. (

<https://www.gutenberg.org/cache/epub/25196/pg25196-images.html>

) Please read both to ensure you have enough knowledge about the task I will give you.

Output "ready" when you have learned the rules and common names.

2nd Prompt:

I have a list of poetry titles in classical Chinese (see 'cases_62' csv file). Your task is to extract names from each case. If you can't identify any names in a case, just leave 'N/A'.

Generate in this format(Fill in the blanks by yourself):

```
|case | extracted_name|  
|--|--||
```

'case' column should be the original titles I provide, and 'extracted_name' should be the output.

3rd Prompt:

Here's the correct answer (see 'true_answer_62' csv file) for the extraction. Please attach this correct answer in the output. Convert 'x' to 'N/A' to match the format. Lastly, calculate the accuracy rate (if 'extracted_name' match with 'true_answer', then correct; vice versa)

Generate in this format(Fill in the blanks by yourself):

```
|case | extracted_name| true_answer | accuracy_rate ||  
|--|--|--|--||
```

Under the 'accuracy_rate' column, calculate an overall accuracy rate. Note that you should have just one total accuracy rate, which is put at the top, aligned with the line of 伯和見過. You can only claim one case is accurate if 'extracted_name' and 'true_answer' match each

other entirely. Partial matching does not count. For example, 宇文文 and 宇文 are not the same, thus not accurate.

Outputs


1. Claude3.5: 87.30 % accuracy rate. (see the output [here](#))
2. ChatGPT: 24.59 % accuracy rate. (see the output [here](#))

Findings

Initial trial: (Prompt 2 and 3)

I didn't provide any background information about the rules and common names initially. In other words, I only gave the 2nd and 3rd prompts. Claude 3.5 already had an accuracy rate of 87.30%, while ChatGPT-4.0 had 0%.

Claude3.5 first trial:

 You can only claim one case is accurate if 'extracted_name' and 'true_answer' are entirely matching with each other. Partial matching does not count. For example, 宇文文 and 宇文 are not the same, thus not accurate. Please revise your output

I apologize for the misunderstanding. You're absolutely right, and I'll revise the output based on exact matches only. Here's the corrected table with the updated accuracy rate:

case	extracted_name	true_answer	accuracy_rate
伯和見過	伯和	伯和	87.30%
別宇文文	宇文文	宇文	
君章見過	君章	君章	
寄歐陽公	歐陽公	歐陽公	
挽何見山	何見山	何見山	
有懷安國	安國	安國	
襄陽懷古	N/A	N/A	
見放翁退	放翁	放翁	
見蔡儒效	蔡儒效	蔡儒效	
見袁蒙齋	袁蒙齋	袁蒙齋	
送丁惠安	丁惠安	丁惠安	
送仙居令	N/A	N/A	
送歐陽辯	歐陽辯	歐陽辯	
送潘華陽	潘華陽	潘華陽	
送鞠彭山	鞠彭山	鞠彭山	

ChatGPT4o first trial:

ChatGPT 4o >

Extracted Names Batch 3

↓

↗

#	case	extracted_name			
1	呈歐陽伯威				
2	呈水心先生				
3	和南仲見寄				
4	和叔奇見寄				
5	和君時見寄				
6	和堯夫見寄				

The third batch of records has been processed and displayed. Let me know when you'd like me to continue with the next batch. [↩](#)

You didn't extract any names from three batch. How can I teach you to be able to extract them?

📄

✎

🧠

To teach me how to extract names accurately from your data, you could provide specific patterns or rules to help me identify the names within the titles. For example, if you can define how a name is typically structured in classical Chinese poetry titles (like indicating that 字 appears in a certain context or location), I can use that information to improve the extraction process.

Here are a few ways you could teach me:

- Provide Examples:** Give me a few concrete examples of titles along with the names that should be extracted from each. This way, I can learn the patterns you're using.
- Rules/Patterns:** Share specific rules, such as:
 - If a title contains “字” or “別號”, the name is typically preceded or followed by this character.
 - If a name in classical Chinese often follows a particular structure, such as a family name plus 字 or 別號, specify how that appears.
- Position in the Title:** Are names usually at the beginning, middle,

↓

 and of the title? Knowing this can help narrow down potential name candidates

3. **Position in the Title:** Are names usually at the beginning, middle, or end of the title? Knowing this can help narrow down potential name candidates.

With these details, I can refine the approach and extract the correct names from each title. Let me know how you'd like to proceed!

for the first two cases, 伯和 and 宇文 are name

🧠

📄 Memory updated

Thanks for providing those examples! Now that I know “伯和” and “宇文” are names, I can apply a similar approach to the rest of the cases by looking for similar structures. I'll go ahead and use this pattern to try extracting names in the next batch. Let's proceed and refine the extraction process.

Let me re-process the first two batches with this insight and then move on to the next batch.

Reprocessed Names for Batches 1 and 2

↓

↗

#	case	extracted_name			
15	錢塘懷古				
16	題仙居閣				
17	題昌國院				
18	仁壽堂古梅				
19	代壽德興尹				
20	別宇文季蒙	宇文			

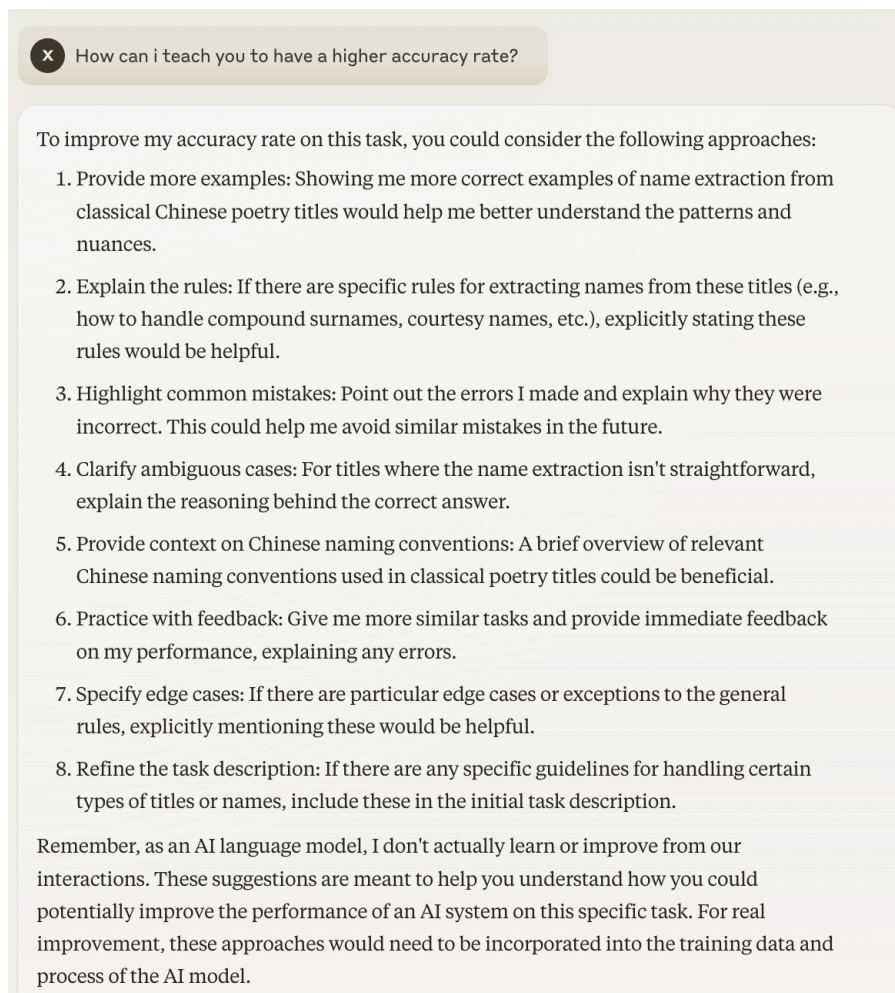
I've reprocessed the first two batches using the new insight, and the results are now displayed. Would you like me to continue processing the next batch with this approach? If you have additional patterns

↓

 es to guide the extraction, feel free to share! [↩](#)

It seems that Claude 3.5 has already done a decent job without any background knowledge, whereas ChatGPT-4.0 struggled. To fix this, I provided the first two cases as training examples to ChatGPT-4.0, but it still failed to apply those examples to other cases.

Revised trail: (Prompt 1, 2 and 3)



I asked how to teach both ChatGPT-4.0 and Claude 3.5 to generate more accurate results, and both suggested providing examples and explaining general rules.

On the second trial, I fed some general rules and common names (Hundred Family Surnames) from the Song dynasty before asking the LLMs to conduct the task. As a result, ChatGPT-4.0 improved its accuracy rate from 0% to 24.59%, while Claude 3.5 generated the same output (no changes).

Conclusion:

It takes less effort to communicate with Claude 3.5 (87.30%) than with ChatGPT-4.0 (24.59%). Providing background knowledge didn't affect Claude 3.5's performance, but it improved ChatGPT-4.0's, although its accuracy is still not acceptable. Better prompts are needed for more accurate extraction.