〇一二七 or 洞幺两拐: An Assessment between 5 Large Language Models' Pronunciation of Numbers in Special Situations in Mandarin with Designed Prompts and Various File Types of Input

Elena Chen

June 2024

Abstract

This paper evaluates the proficiency of 5 distinct large language models in detecting the pronunciation of numbers in Mandarin, analyzing their performance across a range of input types. This study specifically investigates audio inputs, phonetic transcriptions, and Pinyin representations to identify how each model's accuracy varies with the type of data presented. I implemented a comparative analysis framework, asking 5 models exactly same questions and following correction process if needed. The paper is consisted of 5 parts: Introduction, Methodology, Results, Conclusion, as well as Limitations and Future Works. Initial findings suggest significant variance in performance, influenced predominantly by the model's training datasets and inherent design considerations, which also provides suggestions about which model works best for the users while they are making decisions between the use of models. Lastly, this paper provides several potential solutions and suggestions that are computer based to help with the future development of large language models. This assessment not only highlights the current capabilities and limitations of large language models in phonetic detection but also proposes potential pathways for enhancing Mandarin language processing in AI systems.

1 Introduction

The principal objective of this project is to evaluate the ability of different large language models to accurately recognize the correct pronunciation of numbers in Mandarin, particularly in specific cases where the pronunciation changes. In most common contexts, the pronunciation of numbers remains constant; however, there are unique situations that change how certain numbers are pronounced. This paper identifies the primary scenarios where these pronunciation changes occur and tests 5 large language models to assess their recognition of these less common cases. Through providing a comprehensive analysis of the models' performance, this study aims to determine the accuracy of their responses in presenting the correct Pinyin for the special pronunciations of numbers.

-1	$\frac{1}{2}$	$\equiv/3$	四/4	$\Xi/5$	六/6	七/7	八/8	九/9	+/10	$\bigcirc/0$
yī	èr	sān	sì	wǔ	liù	$q\bar{\imath}$	bā	jiǔ	shí	líng

Table 1: Common and Standard Pronunciation of Numbers in Mandarin with Chinese numerals

This project is inspired by the uncanny valley brought by voice transcriptions in reels and short videos from social media platforms, where it sometimes could not pronounce numbers in a proper or user-desired way. The direct transcription of the number with standard pronunciation of numbers implies that the model is either not able to recognize the context or the training dataset does not include any casual use of numbers or words from modern culture. However, we should also keep in mind that large language models are federated learning models, indicating that they usually provide more precise answers than a single text-to-audio model does.

2 Methodology

The testing cases include: audio file from aviation incidence of 3U8633, picture of a Mahjong pattern, text of meme including numbers with context provided, the pronunciation of mobile numbers, and the pronunciation of character "—"(meaning "one") in different words. The chosen large language models include: ChatGPT-4, Gemini, Perplexity, Claude, iFlytek Spark. All

models selected support both Chinese and English for the accuracy of input instructions.

2.1 About Testing Cases

2.1.1 Audio file from aviation incidence of 3U8633

This audio file is the original and official file released by the Chinese government after the investigation of the incidence. The extracted audio includes some uncommon pronunciations of numbers, which are applied in both aviation and military realm for differentiating similar pronunciation of numbers and avoiding troubles from mis-hearing. If the model does not support audio transcription or does not support audio file uploading, the alternative approach will be applied, where I will input the text of the testing case and ask the model to provide the pronunciation in Pinyin. The potential mistake with audio files that the chosen models could make would be failing to recognize the special pronunciation of numbers and transcribing them into some random Chinese characters. The possible mistake with texts provided by me could be failing to mark the correct way of pronunciation. The content in the chosen file:

- "成都早上好四川 8137 高度 72 应答机 2154"
- ATC: "四川 8137 成都雷达识别了"
- "四川 8137"
- ATC: "四川 8633 成都叫"
- ATC: "西藏 9832 成都"
- "8633 成都叫你"
- "西藏 9832 回答成都"
- ATC: "西藏 9832 麻烦在频率里叫一下 8633 看能不能叫到"
- "好的四川的 8633 成都叫你"



Figure 1: 幺鸡

2.1.2 Picture of Mahjong

Figure 1 shows one of the Mahjong patterns called 幺鸡, which belongs to the 条 (or bamboo) family. According to Xinhua Dictionary, the explanation of "幺" is "one; tiny, small". In this case, the "幺" here represents one bamboo but in a special and visually prettier pattern. "鸡" means chicken or rooster, which aligns with the pattern of the image above. The potential mistake that the chosen models might make would be not recognizing the pattern.

2.1.3 Text of meme including numbers with context provided

In Mandarin, there are some numbers that substitutes characters for online chatting, which only serves the purpose of saving time. The testing cases chosen here are "再见 886" and "我在情人节收到了短信表白,上面写着'5201314'。" This entire project is inspired by the case of "886", which sounds similar to "拜拜了", and current voice transcriptions on reels from social media of "886" will be automatically transcribed as "八百八十六" which is the general way of reading the number if the user does not adjust the pronunciation manually. "5201314" in this cases means "我爱你一生一世".

The potential mistake that the chosen models might make could be directly label the Pinyin of these numbers in their regular numeric pronunciation.

2.1.4 The pronunciation of mobile numbers

In mainland China, all mobile phone numbers start with 1 regardless of carrier. It is habitual that people always pronounce 1 as "幺" when they need to read out the phone number for differentiating similar pronunciations of other numbers such as 7. The testing case of phone number in the paper is "在中国大陆,如何读出该电话号码 13982015189," where the sample phone number includes 3 "幺". The reason why mainland China is specified is that some regions such as Taiwan possesses different ways of reading. The potential mistake that the chosen models might make could be either only specify the first digit as "幺" or completely failing on the task and labeling all 1s as the standard way of pronunciation.

2.1.5 Pronunciations of the character "—" in different words

"—" is specifically different from other Chinese numerals, whose original meaning is the smallest primitive unit, the smallest positive integer. Later, it was extended to mean the same, unique, integral, complete, whole, and all-encompassing. Because of its diverse meaning, "—" is able to form a large amount of words, and the pronunciation of "—" varies based on the tone of the character that follows it. Below is a chart explaining the tone change of "—".[1]

Since the fourth case of yi is controversial, this paper will only include the first three cases for testing, and the words for testing include: "整齐划一","二十一","一泻千里","一万","一片","一马平川","一天","一年",and "一起",which cover all 7 different conditions mentioned in the chart. The potential mistake that the chosen model might make could be either wrongly label the Pinyin of "一" or only consider that this character has one pronunciation.

The input prompts remain the same among 5 models for controlling variates. If the model makes mistakes, after each case, a correction process will be applied, which I will give it a hint or tell the model the correct pronunciation in the same chat and ask the same question again. After the correction process has been applied, I will examine the accuracy again. Some correction process won't take place to the very end, because it will be less meaningful

Pinyin	Situation	Situation	Situation	
		Used at the end		
yī	Used alone	of words and	Used as ordinal	
		sentences		
	Used in front of			
yí	the fourth tone			
	character			
	Used in front of	Used in front of	Used in front of	
yì	the first tone	the second tone	the third tone	
	character	character	character	
	Used between	Used between		
yi	overlapping	verb or adjective		
	verbs	complements		

Table 2: Usage of different tones of " →"

to manually correct everything until the models perform perfectly correct answers. The correction process also helps with how the chosen models understand the instructions in the previous contexts.

2.2 About iFlytek Spark

The iFlytek Spark cognitive large model is a large model released by iFlytek. iFlytek is a listed company on the Shenzhen Stock Exchange in China. The company's main business scope includes voice support software, industry application products/systems, information engineering and operation and maintenance services. This model has 7 core capabilities, namely text generation, language understanding, knowledge question and answer, logical reasoning, mathematical ability, coding ability, and multi-modal interaction with 2 supported languages of English and Chinese. The model is benchmarked against ChatGPT.

3 Results

3.1 ChatGPT-4

3.1.1 Audio file from aviation incidence of 3U8633

ChatGPT-4 failed to directly transcribe from the audio file to text because "the tools required to transcribe the audio aren't currently available." Then the alternative approach is applied with the instruction of "Mark the Pinyin for the numbers in the following aeronautical ground-to-air communication information." Figure 2 is the response from ChatGPT-4. First, the model failed to accurately read the instructions that only requires Pinyin of the numbers; second, it labels all numbers in the standard way, which is also considered as a failure of the task.

The instruction of correction process is "In aviation and military context, 0 is pronounced as 洞, 1 is pronounced as 幺, 2 is pronounced as 两, 7 is pronounced as 拐. Revise the result you provided above." Figure 3 is the corrected result provided by the model.

It is surprising that ChatGPT-4 only correct some of the pronunciations. The part indicating the height of 72 should be pronounced as "拐两" while the model only corrects the pronunciation of 2; the following lines indicating the code of transponder from the pilot also remains the pronunciation of 2 incorrect. However, it corrects the flight number of Tibet Airlines 9832, where the 2 is corrected to "两", but it didn't fully correct the flight number of Sichuan Airlines 8137: it remains 7 as "七" while modifies 2 to "两".

3.1.2 Picture of Mahjong

With the instruction of "这个麻将里的图案叫什么? 用英文回答" or "What is the name of the pattern in this Mahjong? Answer in English" Figure 4 shows the answer before and after I gave hint and explicit correct answer.

ChatGPT-4 totally failed on recognizing the pattern first with a nonsense of "cat", where Mahjong of any region does not include cat at all. The hint I provided tells the model that this is a rooster and it belongs to the bamboo suit, but the model completely ignored the context of Mahjong and only provided me the Pinyin of the rooster. Finally, I told the model the correct answer in the third input, and it "learned" what the pattern is and gave me the Pinyin correctly.

Here's the transcription with pinyin for the numbers:

1. "成都 早上好 四川8137 高度72 应答机2154"

chéng dū zaĭo shàng haĭo sì chuān bā yī sān qī gāo dù qī èr yìng dá jī èr yī wu sì

2. ATC: "四川8137 成都雷达识别了"

sì chuān bā yī sān qī chéng dū léi dá shí bié le

3. "四川8137"

sì chuān bā yī sān qī

4. ATC: "四川8633 成都叫"

sì chuān bā liù sān sān chéng dū jiào

5. ATC: "西藏9832 成都"

xī zàng jiu bā sān èr chéng dū

6. "8633 成都叫你"

bā liù sān sān chéng dū jiào nǐ

7. "西藏9832 回答成都"

xī zàng jiu bā sān èr huí dá chéng dū

8. ATC: "西藏9832 麻烦在频率里叫一下8633看能不能叫到"

xī zàng jiu bā sān èr má fán zài pín lữ lǐ jiào yī xià bā liù sān sān kàn néng bù néng jiào dào

9. "好的四川的8633 成都叫你"

hao de sì chuan de ba liù san san chéng du jiào ni

Figure 2: Pinyin from ChatGPT before correction

In aviation and military contexts where these specific terms are used, here's the revised transcription:

1. "成都 早上好 四川8137 高度72 应答机2154"

chéng dū zao shàng hao sì chuān bā yāo sān qī gāo dù qī liang yìng dá jī èr yāo wu sì

2. ATC: "四川8137 成都雷达识别了"

sì chuān bā yāo sān qī chéng dū léi dá shí bié le

3. "四川8137"

sì chuān bā yāo sān qī

4. ATC: "四川8633 成都叫"

sì chuẩn bã liù sãn sãn chéng dũ jiào

5. ATC: "西藏9832 成都"

xī zàng jiủ bā sān liang chéng dū

6. "8633 成都叫你"

bā liù sān sān chéng dū jiào nǐ

7. "西藏9832 回答成都"

xī zàng jiu bā sān liǎng huí dá chéng dū

8. ATC: "西藏9832 麻烦在频率里叫一下8633看能不能叫到"

xī zàng jiư bā sān liang má fán zài pín lữ lǐ jiào yī xià bā liù sān sān kàn néng bù néng jiào dào

9. "好的 四川的8633 成都叫你"

hảo de sì chuẩn de bã liù sãn sãn chéng dũ jiào nĩ

Figure 3: Pinyin from ChatGPT after correction

The pattern on the mahjong tile is a depiction of a cat. In English, you might simply refer to this as a "cat" mahjong tile.

You

This is a rooster, which belongs to the "\$" or bamboo family. How do Chinese people pronounce this Mahjong pattern? Mark the Pinyin

< 2/2 >

The Mahjong tile depicting a rooster in the "条" or bamboo suit is typically referred to as "鸡" in Chinese. The pronunciation of "鸡" in Pinyin is "jī".

You

这是幺鸡

I see, thanks for the correction! The tile is indeed called "玄鸡" in Mahjong, which is pronounced as "yāo jī" in Pinyin.

Figure 4: Mahjong response from ChatGPT

3.1.3 Text of meme including numbers with context provided

With the instruction of "再见 886'如何读?请标注拼音并用英文回答。" or "how to read '再见 886?' label the Pinyin and answer this question in English", ChatGPT-4 successfully recognized the internet slang in this context. Below is the respone copied from the model's response.

• The phrase "再见 886" is read as "zài jiàn bā bā liù" in Pinyin. In English, this translates to "Goodbye 886." The numbers "886" are often used in internet slang or texting in Chinese, sounding similar to "bye bye" when spoken quickly, hence its usage as a playful or informal way to say goodbye.

The second prompt of this testing case is "我在情人节收到了短信表白, 上面写着'5201314'。"请问该 5201314 怎么读?为什么? 用英文回答"In English, "I received a text message professing his love for me on Valentine's s Day, which said '5201314'." How to pronounce 5201314? Why? Answer in English." Below is the response provided by the model.

- The numbers "5201314" read as "wǔ èr líng yī sān yī sì" in Pinyin. In English, this can be explained as a numeric expression commonly used in Chinese internet and texting slang to represent a romantic phrase. The numbers phonetically resemble the phrase "我爱你一生一世," which means "I love you forever" in English. The specific breakdown is:
- "520" phonetically sounds like "我爱你" (wǒ ài nǐ), which means "I love you."
- "1314" phonetically resembles "一生一世" (yī shēng yī shì), which means "forever" or "for a lifetime."
- So, "5201314" combines these two sentiments into a single expression used especially for romantic purposes, such as on Valentine's Day.

Again, ChatGPT-4 recognize the context and the pronunciation of "5201314" in an appropriate and successful way, and it even provided the breakdown explanation of the number. Overall, it performed quite well on recognizing these popular slang used on internet.

3.1.4 The pronunciation of mobile numbers

The model is asked "in mainland China, how do people pronounce the phone number '13982015189'? Answer in English and label the Pinyin."Below is the shortened answer from the model's response. After giving me a breakdown of how to read these numbers one by one, it summarize the answer as:

• So, the full phone number "13982015189" is read as "yī sān jiǔ bā èr líng yī wǔ yī bā jiǔ" in Chinese Pinyin. This is the standard way to read phone numbers in China, focusing on each individual number for clarity.

Since the model did not mark the Pinyin of "1" in the correct way, the correction process is applied, where I told the model that while reading the phone number, all 1s are pronounced as " \angle " and ask the model to revise its answer. Below is the summary part of the answer provided by ChatGPT-4:

• Therefore, the phone number "13982015189" would be pronounced as "yāo sān jiǔ bā èr líng yāo wǔ yāo bā jiǔ" in Chinese Pinyin. This usage of "Ź" helps avoid confusion with similar-sounding words during verbal communication of numbers.

With the help of the correction process provided, the model modified its previous answer and eventually answer the question correctly.

3.1.5 Pronunciations of the character "—" in different words

The designed prompt for this task is to ask the model to label the Pinyin of the chosen words. Figure 5 shows the response from the model.

Through observing the output, it indicates that the model considered "—" always in the first tone no matter which character follows. The correction process is different from the previous tasks. I provided the screenshot of Table 2 and tell the model that "here is the rule of the tone change of '—' specifically, learn this, inspect your answer, and modify them." Figure 6 shows the modified answer from the model.

The model completely got confused by the rule of "Use alone", so that all "—" s that are at the beginning of the words are labeled as first tone; also it seems that it struggles with the "logically inconsistent" rules, which does not exist in the rules provided. The model confuses the concept of "follow" which is not used in the screenshot of the rules, in stead, the rule specifies

Here are the Pinyin transcriptions for the Chinese phrases you've listed:

- 1. "整齐划一" zhěng qí huá yī
- 2. "<u>-</u>+-" èr shí yī
- 3. "一泻千里" yī xiè qiān lǐ
- 4. "一万" yī wàn
- 5. "一片" yī piàn
- 6. "一马平川" yī ma píng chuān
- 7. "一天" yī tiān
- 8. "一年" yī nián
- 9. "一起" yī qǐ

These Pinyin transcriptions reflect the standard pronunciation of each word or phrase in Mandarin Chinese.

Figure 5: Pinyin of "—" from ChatGPT before correction

Based on the tone change rules for "—" (yī) in Mandarin Chinese, let's review and adjust the Pinyin for the phrases you listed:

- 1. 整齐划一 Since "—" here follows a first tone character "划" (huá), it should be in the fourth tone. Correct Pinyin: zhěng qí huá yì
- 2. =+-- "—" follows a first tone character "+" (shi), so it should also be in the fourth tone. Correct Pinyin: èr shí yì
- 3. 一泻千里 "—" here is used alone at the beginning, so it remains in the first tone. Correct Pinyin: yī xiè qiān lǐ
- 4. 一万 "一" is used alone at the beginning, so it stays in the first tone. Correct Pinyin: yī wàn
- 5. 一片 "—" is used alone at the beginning, so it remains in the first tone. Correct Pinyin: yī piàn
- 6. 一马平川 "—" is used alone at the beginning, so it remains in the first tone. Correct Pinyin: **yī ma**ˇ **píng chuān**
- 7. 一天 "一" is used alone at the beginning, so it stays in the first tone. Correct Pinyin: yī tiān
- 8. 一年 "一" is used alone at the beginning, so it stays in the first tone. Correct Pinyin: yī nián
- 9. 一起 "—" is used alone at the beginning, so it remains in the first tone. Correct Pinyin: yī qǐ

These corrections reflect the specific tone changes for "—" based on its position and the tone of the following character, as you outlined in the rules you provided.

Figure 6: Pinyin of "—" from ChatGPT after the given rules

with "end", "in front of", and "between". For example, the case of "二十一", where the "一" is at the end, but ChatGPT-4 declares that "一" follows a first tone character "十" (shí), so it should also be in the fourth tone. To reach the result as precise as possible, I gave instructions again: "'Use alone' means there's only this character without any context, not means at the beginning. There's no case of used alone here. Revise your answer." Figure 7 shows the second revised version from the model.

₩ ChatGPT

Thank you for clarifying the rule about "—" being used alone. Let's revise the Pinyin for the phrases with the correct application of the tone change rules where "—" is not used alone:

- 1. 整齐划一 Since "—" follows a fourth tone character "划" (huá), "—" should be in the second tone. Correct Pinyin: **zhěng qí huá yí**
- 2. =+-- "-" follows a second tone character "+" (shí), so it should be in the fourth tone. Correct Pinyin: èr shí yì
- 3. 一泻千里 "一" follows a fourth tone character "泻" (xiè), so it should be in the second tone. Correct Pinyin: yí xiè qiān lǐ
- 4. 一万 "一" follows a fourth tone character "万" (wàn), so it should be in the second tone. Correct Pinyin: yí wàn
- 5. 一片 "—" follows a fourth tone character "片" (piàn), so it should be in the second tone. Correct Pinyin: yí piàn
- 6. 一马平川 "—" follows a second tone character "马" (ma), so it should be in the fourth tone. Correct Pinyin: yì ma píng chuān
- 7. 一天 "一" follows a first tone character "天" (tiān), so it should be in the fourth tone. Correct Pinyin: yì tiān
- 8. 一年 "一" follows a second tone character "年" (nián), so it should be in the fourth tone. Correct Pinyin: yì nián
- 9. 一起 "—" follows a third tone character "起" (qi), so it should be in the fourth tone. Correct Pinyin: yì qǐ

Figure 7: Pinyin of "一" from ChatGPT after more specifications
This time, except "整齐划一" and "二十一", all other pronunciations are

labeled correctly. The reason why it marked "整齐划一" wrong is that "划" is a character with multiple tones, which are the second and the fourth, and here, through ChatGPT-4's explanation, because of the incorrect recognition of this character, it labeled "" incorrectly. Same issue happened to the "二十一" case. However, again, the pronunciation of "一" is not influenced by the character that is before itself.

3.2 Gemini

3.2.1 Audio file from aviation incidence of 3U8633

Because of the unsupported feature of uploading audio files, the alternative text-to-pinyin is applied. Below is the Pinyin provided by Gemini:

- "Chéngdū zǎoshàng hǎo Sìchuān bāyī sānqièr qī gāodù qīshíèr yīngdàjī èrqiānsi"
- ATC: "Sìchuān bāyī sānqièr qī Chéngdū léidá shībié le"
- "Sìchuān bāyī sāngièr gī"
- ATC: "Sìchuān liùliù sānqièr sān Chéngdū jiào"
- ATC: "Xīzàng ji
ŭbā sān'èr Chéngdū"
- "Liùliù sānqièr sān Chéngdū jiào nǐ"
- "Xīzàng jiùbā sān'èr huídá Chéngdū"
- ATC: "Xīzàng jiùbā sān'èr máfàn zài pínlǜ lǐ jiào yīxià liùliù sānqièr sān kàn n 能不能 jiào dào"
- "Hǎo de Sìchuān de liùliù sānqièr sān Chéngdū jiào nǐ"

Ignoring the context of aviation, I observed that the model even could not correctly provide the standard pronunciation of numbers. It incorrectly recognized "8137" to "813727." Also, it did not fully provide the Pinyin and remain the characters "能不能" in the response. Below is the revised response after putting the same correction process.

"Chéngdū zǎoshàng hǎo Sìchuān bāyāo yīsānqī qī gāodù qī'er yīngdàjī erqiānsì"

- ATC: "Sìchuān bāyāo yīsānqī qī Chéngdū léidá shībié le"
- "Sìchuān bāyāo yīsānqī qī"
- ATC: "Sìchuān liùyāo liùsānqī sān Chéngdū jiào"
- ATC: "Xīzàng jiùyāo bāsān'èr Chéngdū"
- "Liùyāo liùsānqī sān Chéngdū jiào nǐ"
- "Xīzàng jiùyāo bāsān'èr huídá Chéngdū"
- ATC: "Xīzàng jiǔyāo bāsān'èr máfàn zài pínlǜ lǐ jiào yīxià liùyāo liùsānqī sān kàn n 能不能 jiào dào"
- "Hǎo de Sìchuān de liùyāo liùsānqī sān Chéngdū jiào nǐ"

Except for the same problem that ChatGPT-4 also presents, Gemini totally failed on the task. For example, it realized that "1" should be pronounced as " \angle ", but it "forgot" to delete the incorrect Pinyin. The entire response seems quite hilarious. What is even more ironic is that the model provides comments labeling all numbers in the appropriate way, but the following examples of numbers it provided are totally disastrous.

3.2.2 Picture of Mahjong

After asking the same question as I does to ChatGPT-4, Gemini's response shows in the following:

Gemnini provides a quite different answer compared to ChatGPT-4. It recognize the pattern as a bird but thought it is a Japanese pattern. So the hint is given, and below is its response:

- You are absolutely right, the pattern you described is indeed a rooster, and it belongs to the "条 (tiáo)" or bamboo family in Mahjong.
- Here's how Chinese people pronounce the Mahjong pattern for rooster:
- Sāntiáo (sān tiáo)
- "Sān" (\equiv) means "three" and refers to the position of the rooster within the "tiáo" family. In a Mahjong set, each family (like bamboo) has three tiles numbered one to three.

注释:

- 数字标注在每个数字后面, 用括号括起来。
- 拼音使用汉语拼音方案,并根据航空和军事语境进行了调整。
- 0 的拼音为 "dòng"
- 1的拼音为 "yāo"
- 2 的拼音为 "liǎng"
- 7 的拼音为 "guǎi"

示例:

- "成都 早上好 (zǎoshàng hǎo)" 同上
- "四川8137 (bāyāo yīsānqī qī)" 四川的拼音是 "Sìchuān",数字 "8137" 的拼音根据航空军事语境读作 "bāyāo yīsānqī qī"。
- "高度72 (qī'ěr)" 高度的拼音是 "gāodù", 数字 "72" 的拼音读作 "qī'ěr"。

希望这次的修订更加准确。

Figure 8: Gemini's comment and example



The pattern on the mahjong tile in the image you sent is called "bird" in English. It is one of the four flower tiles in a standard mahjong set. The other three flower tiles are bamboo, plum, and chrysanthemum. Flower tiles are not used in all mahjong variants, but they are common in the Japanese and Riichi mahjong rulesets.

The bird tile is typically depicted as a stylized image of a peacock or pheasant. The specific design of the tile can vary depending on the mahigned set, but the overall appearance is generally the same. The bird tile is a valuable tile in mahigned, and it can be used to form a number of different winning hands.

In the Riichi mahjong ruleset, the bird tile can be used to form the following winning hands:

- Three birds: This is a hand that consists of three bird tiles and any other two tiles.
- All four birds: This is a hand that consists of all four bird tiles.
- Seven pairs: This is a hand that consists of seven pairs of tiles, including one pair
 of bird tiles.

The bird tile can also be used as a substitute for any other tile in a hand. This makes it a very versatile tile, and it can be used to improve your chances of winning a hand.

In addition to its use in mahjong, the bird tile is also a popular symbol in Chinese culture. It is often associated with good luck and prosperity. The bird is also seen as a symbol of beauty and grace.

Overall, the bird tile is a valuable and versatile tile in mahjong. It is also a symbol of good luck and prosperity in Chinese culture.

Figure 9: Response from Gemini before the correction

• "Tiáo" (条) refers to the bamboo family itself.

The model mentioned "三条", which refers another pattern that has three bamboo-like lines, though "幺鸡" and "三条" belong to the same suit, they are totally different. I have to tell Gemini that this is 幺鸡, and its response shows below:

- I apologize for the error in my previous response. You are correct that the pattern in the image is a "幺鸡 (yāo jī)," also known as a "little chick" or "baby chick" in English. It is a special tile in the "条 (tiáo)" or bamboo family in Mahjong. Unlike the regular numbered bamboo tiles (一筒、三筒、三筒), the "幺鸡" is not assigned a specific numerical value. Instead, it serves as a wildcard that can be substituted for any other "条" tile. This makes it a valuable and versatile tile that can be used to form a variety of winning hands.
- Here's how Chinese people pronounce the Mahjong pattern for "幺鸡":
- Yāo jī (yāo jī)
- "Yāo" (幺) means "small" or "youngest," reflecting the chick's diminutive size and status as the lowest-ranked tile in the "tiáo" family.
- "Jī" (鸡) means "chicken," directly referring to the tile's depiction of a young chick.
- omitted 3-paragraphs-long contents of the rules of Mahjong

For this task, Gemini's performance is still quite poor but better than the aviation text task. It recognized the pattern as a bird rather than a cat like ChatGPT-4 did, but it provided me too much unnecessary information about the background of Mahjong, which seems redundant for this research.

3.2.3 Text of meme including numbers with context provided

Being asked to the prompt of the pronunciation of "886" and "5201314" in their respective cases, Gemini's responses were shortened and presented in the following:

• Overall Pronunciation:

- The entire phrase "再见 886" is pronounced as follows:
- zàijiàn bā bā liù
- English Translation:
- In English, the phrase "再见 886" can be translated as "goodbye," "see you later," or "take care." It is a casual and informal way to say goodbye to someone, similar to saying "bye" or "see ya" in English.
- The combination of numbers "5201314" is not a standard Chinese phrase or expression. However, it has been adopted as a popular way to express love and affection, particularly on Valentine's Day in China. The pronunciation and meaning of this numerical code are based on the homophonic nature of Chinese numbers and characters.
- 520 (wǔ èr líng): Sounds similar to "wǒ ài nǐ," which means "I love you" in Chinese.
- 1314 (yī sān yī sì): Sounds similar to "yīshēng yīshì," which means "forever" or "for a lifetime" in Chinese.
- omitted contents of pronunciation approaches for foreigners, cultural context, and conclusion

Overall, Gemini correctly recognized all numbers and complete the task without further correction instruction needed.

3.2.4 The pronunciation of mobile numbers

Gemini insist to be caring and thoughtful by providing long responses with step-by-step explanation, the final output shows in Figure 10. Without the correction process, Gemini also provided the standard and common ways of pronunciation of numbers.

Figure 11 shows the shortened output provided by Gemini after the correction process has been implemented.

It is weird that Gemini only substituted the first two pronunciations of "1" and remained the last one as its previous answer; what is also noticeable is that it changed "one" to "yāo" in the English translation part, which is absolutely incorrect.

Complete pronunciation:

The complete pronunciation of the phone number 13982015189 is:

yī sān jiǔ bā èr líng yī wǔ yī bā jiǔ

English translation:

"one three nine eight two zero one five one eight nine"

Figure 10: Response from Gemini before the correction

Complete pronunciation:

The complete pronunciation of the phone number 13982015189 with "yāo" for 1s is:

yāo sān jiǔ bā èr líng yāo wǔ yī bā jiǔ

English translation:

"yāo three nine eight two zero yāo five one eight nine"

Figure 11: Response from Gemini after the correction

3.2.5 Pronunciations of the character "—" in different words

The initial response of this testing case from Gemini seems quite similar to the response from ChatGPT-4 except that Gemini correctly recognized "划" from the idiom "整齐划一" is the fourth tone not the second. Other than this, Gemini also considered "一" always in the first tone no matter which character follows. Figure 12 represents the response from Gemini after the correction process was implemented.



Based on the information you provided about the image, here's a revised version of the pinyin labels for the words, incorporating the tone rule changes for the character "—".

- 整齐划一 (zhèng qí huà yī): Tone remains unchanged (1st tone)
- 二十一 (èrshí yī): Tone remains unchanged (1st tone)
- 一泻千里 (yī xiè qiān lǐ): "—" (yī) changes to 4th tone (yì) due to following a 4th tone character "泻 (xiè)"
- 一万 (yī wàn): Tone remains unchanged (1st tone)
- 一片 (yī piàn): "—" (yī) changes to 3rd tone (yǐ) due to following a 3rd tone character "片 (piàn)"
- 一马平川 (yī mǎ píng chuān): Tone remains unchanged (1st tone)
- 一天 (yī tiān): Tone remains unchanged (1st tone)
- 一年 (yī nián): Tone remains unchanged (1st tone)
- 一起 (yī qǐ): Tone remains unchanged (1st tone)

As you can see, the tone for "—" changes only when it precedes a character with the 3rd or 4th tone. It remains the 1st tone in most other cases.

Figure 12: Response from Gemini after the correction

Though provided with screenshot of rules, it seems that Gemini could not precisely parse, recognize, and understand the text in the picture, given that there's even no third tone for the character "一", yet the model revised its answer with a third tone because the following character has the third tone. Also, in the testing idiom of "一泻千里", Gemini changes the tone from the first to the fourth because of the following character pronounced with the fourth tone. In this case, we can conclude that Gemini even has problem

with processing the text in the screenshot.

3.3 Perplexity

3.3.1 Audio file from a viation incidence of 3U8633

Similarly to ChatGPT-4, Perplexity was unable to process audio files in Mandarin, so that the alternative approaches was applied and Figure 13 shows the initial results.

Answer

Here are the numbers from the provided air-to-ground communication information with their corresponding pinyin annotations:

- 8137: bā yāo yī sān qī
- 72: qī shí èr
- 2154: èr yāo yī wǔ sì
- 8633: bā liù sān sān
- 9832: jiǔ bā sān èr

Figure 13: Response from Perplexity before the correction

The highlight of this response is that the model accurately processed understood the prompt that only required it to label the number, and we can observe that it provided two different ways of pronunciations of "—" even without my further instructions for correction. However, since there is no parenthesis indicating that "yāo" and "yī" indicate the same number, it might cause confusion to users. After being given the modification instruction, Figure 14 is the revised version of the answer from Perplexity:

The exactly the same instruction was given: "In aviation and military context, 0 is pronounced as 河, 1 is pronounced as 幺, 2 is pronounced as 两, 7 is pronounced as 拐. Revise the result you provided above." This interesting result shows how perplexity processes and understands the instructions, which is different from the previous tested models. It considered any Chinese character in the instruction the same as Pinyin and directly substituted them without any modification from characters. I asked the model to "provide the pinyin not the character of the same pronunciation in [its] revised version", and Figure 15 shows the final result.

Revised with aviation and military context:

- 8137: bā yāo yī sān 拐
- 72: qī shí èr
- 2154: 两 yāo yī wǔ sì
- 8633: bā liù sān sān
- 9832: jiŭ bā sān èr

Figure 14: Response from Perplexity after the correction

Here are the numbers from the provided air-to-ground communication information with their corresponding pinyin annotations, revised according to the aviation and military context and using pinyin instead of characters for the same pronunciation:

- 8137: bā yāo yāo sān guǎi
- 72: qī shí èr
- 2154: liăng yāo yāo wǔ sì
- 8633: bā liù sān sān
- 9832: jiǔ bā sān èr

Figure 15: Response from Perplexity with further instructions

The model changed the characters to their corresponding pronunciations and the remaining " $y\bar{i}$ " to " $y\bar{a}$ o", yet it ignored some special cases such as 72 and 9832, leaving them the same as the first version of the answer.

3.3.2 Picture of Mahjong

Perplexity did not try to guess what the pattern is in the image and directly responded:

• The image shows a single white mahjong tile with a simple black line drawing of a figure. Unfortunately, I do not have enough contextual information to definitively identify the specific name or meaning of the figure depicted on the mahjong tile. Mahjong tiles often feature various symbolic or stylized designs, and the significance of this particular figure would likely depend on the specific mahjong game or tradition it is associated with. Without more details about the origin or purpose of this particular tile, I cannot provide the specific name you asked about.

After telling the model straightly that this is "玄鸡", it, like other models, provided me some information with this pattern. When I re-uploaded the image and asked it again what the pattern is and how it is called, it correctly told the answer.

3.3.3 Text of meme including numbers with context provided

Perplexity also nailed the question of "再见 886":

• The phrase "再见 886" is read as "zài jiàn bā bā liù" in Chinese. In English, this translates to "goodbye 886," where "886" is a playful numerical substitution for "拜拜了" (bài bài le), meaning "bye-bye" or "see you." This form of expression is a common practice in Chinese internet slang and informal conversations, where numbers are used because their pronunciation in Chinese sounds similar to certain phrases.

Different from the previous models, its explanation of "886" includes that "6" is similar to " $\mathbb{7}$ ", whereas ChatGPT-4 and Gemini only pointed out the use of this internet slang. Perplexity's response on the second case is also correct:

• The phrase "5201314" is read as "wǔ èr líng yī sān yī sì" in Chinese. This sequence of numbers is a popular form of expressing affection in Chinese, particularly among the younger generation. The reason for this specific reading and usage is that the pronunciation of these numbers in Chinese closely resembles the phrase "我爱你一生一世," which translates to "I love you for a lifetime" in English. This numerical expression is commonly used in text messages and social media as a romantic declaration, especially on occasions like Valentine's Day.

Though it did not breakdown this 7-digit number and explain them one by one, this paragraph provided by the model clearly described the resemblance of pronunciation between "5201314" and "我爱你一生一世".

3.3.4 The pronunciation of mobile numbers

Unsurprisingly, without further specification, Perplexity provided the pronunciation of phone number in the standard way same as Table 1.

• In Mainland China, the phone number 13982015189 is read as "yi san jiu ba er ling yi wu yi ba jiu" in Mandarin Chinese. In Pinyin, this is written as "yī sān jiǔ bā èr líng yī wǔ yī bā jiǔ".

With the modification instruction, pointing out the special pronunciation of "1", Perplexity's response did add "yāo" in its response but did not eliminate the "yī" from the previous answer, just like how it performed in the previous testing cases:

• In Mainland China, the phone number 13982015189 is read as "yāo sān jiǔ bā èr líng yāo wǔ yāo yī bā jiǔ" in Mandarin Chinese. In Pinyin, this is written as "yāo sān jiǔ bā èr líng yāo wǔ yāo yī bā jiǔ".

Perplexity shows the unique mistake after being given the modification instructions.

3.3.5 Pronunciations of the character "—" in different words

Identical as the expectation and performances of other models, Perplexity considered "—" always in the first tone regardless of the neighbour characters. However, after giving the screenshot of Table 2 and the same modification instruction, I found that Perplexity labeled all "—" in the appropriate way:

Answer

Thank you for the clarification. Based on the rules provided in the image and the additional sources, here are the corrected Pinyin labels with the appropriate tone changes applied:

- 1. 整齐划一 (zhěng qí huà yī) "—" is used at the end of the phrase, so it retains its first tone, making no change necessary.
- 2. =+- (èr shí yī) "-" is used at the end of the phrase, so it retains its first tone, making no change necessary.
- 3. 一泻千里 (yī xiè qiān lǐ) "一" is used in front of a fourth tone character, so it should change to the second tone, becoming "一泻千里" (yí xiè qiān lǐ) 1 2 3 4 5 6.
- 4. 一万 (yī wàn) "一" is used in front of a fourth tone character, so it should change to the second tone, becoming "一万" (yí wàn) 1 2 3 4 5 6.
- 5. 一片 (yī piàn) "一" is used in front of a fourth tone character, so it should change to the second tone, becoming "一片" (yí piàn) 1 2 3 4 5 6.
- 6. 一马平川 (yī mǎ píng chuān) "一" is used in front of a third tone character, so it should change to the fourth tone, becoming "一马平川" (yì mǎ píng chuān) 1 2 3 4 5 6.
- 7. 一天 (yī tiān) "一" is used in front of a first tone character, so it should change to the fourth tone, becoming "一天" (yì tiān) 1 2 3 4 5 6.
- 8. 一年 (yī nián) "一" is used in front of a second tone character, so it should change to the fourth tone, becoming "一年" (yì nián) 1 2 3 4 5 6.
- 9. 一起 (yī qǐ) "一" is used in front of a third tone character, so it should change to the fourth tone, becoming "一起" (yì qǐ) 1 2 3 4 5 6.

These modifications adhere to the rules outlined in the provided sources, ensuring that "—" is pronounced with the correct tone based on the tone of the character that follows it and its position in the phrase.

Figure 16: Response from Perplexity with further instructions

Though this is the third answer provided by the model in this testing case, it only mis-labeled "整齐划一" wrong in the second response with an explantion saying that "Since"一" is followed by a fourth tone character, it should be pronounced with the second tone, changing it to "整齐划一" (zhěng qí huà yí). "After giving it the exactly the same clarification of minding the position of "一", as I did to ChatGPT-4, Perplexity responded perfectly.

3.4 Claude

3.4.1 Audio file from aviation incidence of 3U8633

Being notified that audio file is not supported, I switched to the alternative approach, and the response shows in Figure 17.

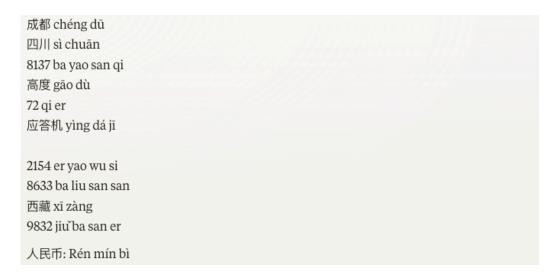


Figure 17: Response from Claude before the correction

We can observe that Claude also tries not to label all pronunciations of all characters although it includes some nouns of regions and objects as well as a random word of "人民币" (RMB, Chinese currency) that never appears in the prompt. With the instructions of correction, Claude processed the information in a way that is quite similar to Perplexity, leaving the character not the Pinyin in the response, implying that these models treat the token of the character I provided as "Pinyin" and not are not "aware" that I am using the Chinese character to represent the pronunciation I wanted:

```
成都 chéng dū
四川 sì chuān
8137 ba 拐 幺 洞
高度 gāo dù
72 拐 两
应答机 yìng dá jī
2154 两 拐 wu 四
8633 ba 陆 幺 幺
西藏 xī zàng
9832 jiu ba 幺 两
```

Figure 18: Response from Claude after the correction

However, in Figure 18, we can observe that the model started to make up contents, leading that the numbers and their supposedly respective pronunciations incorrect, such as "8633" became "8611", and "8137" became "8701". The further instructions asking the model to change the character into Pinyin still did not successfully turn these wrongly recognized characters into the correct ones (Figure 19).

```
成都 chéng dū
四川 sì chuān
8137 ba guaĭ yāo dŏng
高度 gāo dù
72 guaĭ liaňg
应答机 yìng dá jī
2154 liaňg guaǐ wū sì
8633 ba liù yāo yāo
西藏 xī zàng
9832 jiuˇ ba yāo liaňg
```

Figure 19: Response from Claude with further instructions

3.4.2 Picture of Mahjong

Compared to other models tested above, Claude recognized this pattern belongs to the bamboo suit though it misunderstood the pattern as a floral or vine design instead of a rooster:

• The image shows a stylized floral or vine design printed on a mahjong or mah-jong tile. This design is known in English as the "Bamboo" or "Bamboo shoot" tile in the mahjong game.

After being told it is a rooster, Claude responded:

- Ah I see, this design depicts a rooster or cock, which belongs to the "bamboo" or "tiao" suit in mahjong. In Chinese, this particular tile pattern is commonly pronounced as "ji" or "gong ji".
- The Pinyin would be:
- 公雞 gōng jī

Unfortunately, it was quite close to the answer, yet Claude only provided the common pronunciation of a rooster or a cock, so I have to directly tell it this is "幺鸡", and it corrected its response and replied:

• You're correct, my previous answer was mistaken. The Chinese name for this mahjong tile pattern depicting a rooster/cock is "yāo jī" or 幺 鸡 in Pinyin.

3.4.3 Text of meme including numbers with context provided

The response from Claude (Figure 20), compared to the other models' results, was extremely simple:

- "Zài jiàn 886" would be read as:
- 再见 zài jiàn
- 8 ba
- 8 ba
- 6 liu

- In English, it would be:
- "Goodbye 886"

There is no explanation telling why "886" is not pronounced in the standard way and no cultural context provided. However, it missed the nodes of tones. After specifically asking it to add the tone, Claude provided the answer with nodes of tones, but still in this simple and clean way. Claude became the only model that incorrectly recognized the "5201314" case:

- The numbers "5201314" in that text message you received on Valentine's Day do not represent regular numbers, but rather a coded expression using Chinese wordplay.
- The way to read it is:
- 5201314 wǔ àiAIrrrrrrr
- This is because:
- 5 (wǔ) sounds like "I"
- 2 (àiAIrrrrrrr) sounds like "love you"
- 01314 represents "一生一世" (yī shēng yī shì) meaning "forever and ever"
- So put together, "5201314" forms a coded expression meaning "I love you forever and ever" in Chinese, making it a romantic Valentine's Day message.

Though it successfully recognized that this 7-digit-long number should not pronounce in the regular way and that the expression means "I love you forever and ever", it failed on the breakdown part. It labeled "2" in an incredibly incorrect way and explained its sound is similar to "love you", where the pronunciation of "0" is similar to "you" and should not be placed together with "1314"

3.4.4 The pronunciation of mobile numbers

Aligned with my expectation, Claude provided the regular way of pronunciations of the phone number:

- In mainland China, the phone number 13982015189 would be read out as: Yī sān jiù bā èr líng yī wù yī bā jiù
- Which translates to: One three nine eight two zero one five one eight nine

Claude performed in the same way as Perplexity and Gemini did after being asked to change the Pinyin of "1" from "yī" to "yāo" :

- Got it, here's the revised way to read out 13982015189 in mainland China with the proper pronunciation of 1s as "yāo":
- Yāo sān jiǔ bā èr líng yāo wǔ yāo bā jiǔ
- In English: Yao three nine eight two zero yao five yao eight nine

3.4.5 Pronunciations of the character "—" in different words

For this task, Claude spent the least time to process, including the 2 tries of revision process, it rapidly showed its response right after the instructions had been sent. However, its response never changed:

- 整齐划一 zhěngqí huàyī
- 二十一 èrshíyī
- 一泻千里 yīxièqiānlǐ
- 一万 yīwàn
- 一片 yīpiàn
- 一马平川 yīmǎpíngchuān
- 一天 yītiān
- 一年 yīnián
- 一起 yīqǐ

Even though declaring it will re-examine my screenshot of pronunciation rules in the last two responses of revision request, Claude never changed its answer and remained the one that is the same as the others that performed by other model initially: all "—" s remain the first tone. This implies that Claude potentially has a worse optimizer than other models do, like gradient descent. It might have less iteration in gradient descent so that the response did not change at all. Rather, it may also consider that labeling "—" in its standard way is not completely wrong, and a Chinese native speaker would still understand meaning of the words. Also, the results implied that the model is designed to soften the answer, "lying" to users that it keeps processing the instruction but actually responses with the built-in answers, which leads to a guess that the model passes down the instructions to the other embedded model to solve this question, and, of course, it failed again.

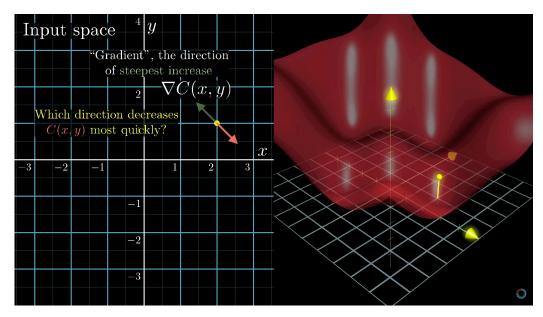


Figure 20: Screenshot of gradient descent from 3Blue1Brown[2]

3.5 iFlytek Spark (hereinafter referred to as Spark)

3.5.1 Audio file from aviation incidence of 3U8633

Spark is the only model that support the voice transcription in Mandarin, and Figure 21 shows the transcribed texts:

- 修正后可能的语音原文为: 本周至少 24813 块, 高度关联应该是 1.154。那么三维程度雷达识别呢? 4 车 813 罐。438633 成都价。呃, 现在酒吧才两层多。好, 133 成都见过你。7329832 回到山东, 现在 9832, 麻烦再评定一下, 叫一下 8633 看能不能叫到。好的, 四川的 8633 成都叫你。
- 翻译结果为: This week, we have at least 24,813 pieces, and the height correlation should be 1.154. What about the three-dimensional radar recognition then? There are 4 cars with 813 cans. The Chengdu price is 438,633. Uh, there are just over two floors in the bar now. Alright, I've seen you in Chengdu before, number 133. Returning to Shandong as 732,9832, could you please reassess it and try to call out to 8633 to see if it can be reached? Okay, Sichuan's 8633 is calling for you in Chengdu.

Noticeably, it started with "修正后可能的语音原文为", meaning "The revised possible audio source text". This implies that Spark not only transcribed the audio file, but also tried to guess and modify the text. After telling Spark that this is a recording from ground-to-air communication, I, unfortunately, received the same transcripts. So I switched to the alternative approach to take a further examination:

- 数字的拼音如下:
- 8137 (bā yī sān qī)
- 72 (qī èr)
- 2154 (èr yī wǔ sì)
- 8633 (bā liù sān sān)
- 9832 (jiǔ bā sān èr)
- 请注意, 上述拼音是根据标准的普通话发音进行转写的。

Spark repeated information about pronunciation of numbers under aviation and military scenario and specifically noted at the last line that "the Pinyin provided above are transcribed based on the standard pronunciation of Mandarin," even though I expressed the context at the beginning of the



In aviation and military context, the numbers are pronounced as follows:

- 0 is pronounced as "洞" (dòng)
- 1 is pronounced as "幺" (yāo)
- 2 is pronounced as "两" (liǎng)
- 7 is pronounced as "拐" (guǎi)

The revised pronunciations for the numbers in aviation and military context are:

- 8137 is pronounced as "八幺三拐" (bā yāo sān guǎi)
- 72 is pronounced as "七两" (qī liǎng)
- 2154 is pronounced as "两幺五四" (liǎng yāo wǔ sì)
- 8633 is pronounced as "八六三三" (bā liù sān sān)
- 9832 is pronounced as "九八三二" (jiǔ bā sān èr)

Figure 21: Pinyin from Spark after correction

instruction. In this case, the further instructions of modification was applied, and the results were shown in Figure 21.

In the figure, we can perceive that the pronunciations were only partially corrected by the model. Moreover, on April. 4th, I tried similar prompt asking about the pronunciation of "0", "1", "2", and "7" in the context of aviation and military to both Spark and ChatGPT-4, Spark brought me correct answers, but the difference between prompts is that I asked how a pilot would read "0127" during work (Figure 22).

3.5.2 Picture of Mahjong

Spark provided the most surprising response compared to the other models in this case,

• 抱歉, 我还没有学习到关于这个话题的内容, 无法提供相关信息。您可以选择其他问题, 我将努力为您解答。

meaning "Sorry, I haven't learned enough about this topic to provide relevant information. You can choose other questions and I will try to answer them for you." The Mahjong case failed and was interrupted. The potential reason might be China considered Mahjong as gambling if the Mahjong players do not know each other. To prevent from gambling intentions, Spark banned this topic.



A DEWAIS

在航空领域,数字的读法有一定的规范,以确保信息的准确传递。对于中国飞行员来说,数字"0127"通常会按照以下方式进行读报:

- "零"读作"洞"(避免与字母"O"混淆)
- "1"读作"幺"
- "2"读作"两"
- "7"读作"拐"

所以,中国飞行员在工作中读"0127"通常会是:"洞幺两拐"。

这种读法有助于避免在无线电通信中因为语言不清而造成的误解,特别是在国际航班中,不同国家的飞行员和地面控制人员可能会使用不同的语言和口音。使用标准化的数字读法可以减少沟通上的障碍,提高飞行安全性。

Figure 22: Former response from Spark

3.5.3 Text of meme including numbers with context provided

- "再见 886"的拼音是"zài jiàn 886"。"886"在汉语中通常没有特定的含义,它可能就是一个数字或者网络用语。但在网络语言中,"886"有时被用作"拜拜了"的谐音,表示告别。
- 如果要标注"886"的拼音, 就是:
- 886 bā bā liù

Spark provided a detailed answer and correct Pinyin, yet it ignored the instruction that it was asked to respond in English. Figure 23 represents the response of the second case; it is in English but the tones of Pinyin were missing.

Being asked "where the tone is," Spark's response implies that it has weaker context detection ability than other models do.

• 该 5201314 的拼音是"wu er ling yi san yi si",但是这个拼音并没有声调。在汉语中,声调是非常重要的,它决定了一个词的意思。但是在这个数字中,声调并没有被明确地表示出来。因此,我们不能确定这个数字的确切意思,除非我们知道它的上下文。

This quote is telling us "The pinyin of 5201314 is "wu er ling yi san yi si", but this pinyin has no tones. In Chinese, tone is very important,

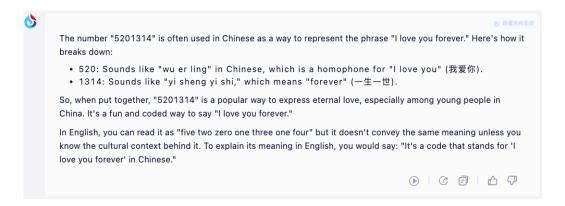


Figure 23: Spark's response before the correction

it determines the meaning of a word. But in this figure the tone is not explicitly represented. Therefore, we cannot be sure of the exact meaning of this number unless we know its context." However, the follow-up command is inputted from the same chat; same issue happened when I tried to add a further instruction about the text transcribed from the audio file, and Spark told me that there was a lack of audio file that it could process.

3.5.4 The pronunciation of mobile numbers

Spark, like other models, did not change the Pinyin for "1" and remain it in the standard way. After inputting the further instructions that is exactly the same as the others, I received the reponse shown in the following.

- 在中国大陆, 电话号码 13982015189 应该这样读:
- "幺三九、八二零幺、五一八九。"
- 对应的拼音是: "Yāo sān jiǔ, bā èr líng yāo, wǔ yī bā jiǔ."
- 所以,该电话号码的中文读法是"幺三九,八二零幺,五一八九。"

Initially, Spark failed on changing the Pinyin of the third "1", yet when I clicked the regenerate option, it corrected the third "yī" to "yāo". Though eventually showing the correct answers, Sparked took longer time to access the correct answer.

3.5.5 Pronunciations of the character "—" in different words

Similar to the other models, Spark also labeled all "—" as the first tone regardless of the context or the character followed. When I provided the screenshot of rules and asked the model to revise its answer, it, again, showed the lack of ability to recognize the context.



Figure 24: Spark's response after the correction

It first tried to parse and read the contents in the screenshot and repeat them, and then it tried to correct my chart without the specification of this command. After its correction, I inputted: Based on the rules you mentioned, label Pinyin of these words: "整齐划一", "二十一", "一泻千里", "一万", "一片", "一马平川", "一天", "一年", and "一起." The output after this instruction is represented as Figure 25, which was identical with its initial response before I tried to implement any further modification instructions.

Figure 25: Spark's response after the correction

4 Conclusion

To summarize, this research intentionally designs five challenging and different types of prompts in order to test the limit of 5 large language models and to assess the accuracy of pronunciations of numbers in special scenarios. None of the model chosen completely passed these tasks, and there are several potential reasons. First, the training datasets are mostly texts whose copyright has expired, meaning that some modern use of numbers and their pronunciations are not in the dataset; second, four out of five chosen models are designed and made by Western engineers, so that testing the pronunciations with special situations become extremely hard, yet this does not mean those models are not useful; third, some testing cases are relatively nitpicking, such as the aviation case and the tone change of "—" case. In reality "incorrect" or inappropriate pronunciations of numbers would still be understandable, and the ordinary would not frequently (or even would never) use terminologies of numbers from aviation and military fields. Table 3 sums up advantages and drawbacks of tested models for users to choose which model benefit them the most.

5 Limitations and Future Works

There are several limitations in this research. First, the wording of input might still cause confusion to the model, especially they are both in and

Model	Pros	Cons		
ChatGPT-4	Generally good.	N/A.		
Gemini	More background and con-	Least accurate.		
	textual information.			
Perplexity	Simple and short response;	Complex UI design; unnec-		
	References provided; Ex-	essary ask for pro tips.		
	panded contents like videos.			
Claude	Simplicity; Quick response.	Inaccurate response.		
iFlytek Spark	Mandarin voice transcrip-	Weaker context detection;		
	tions; Designed for Chinese	Longer time to upload files.		
	native speakers.			

Table 3: Comparison between different models

switching between English and Chinese; second, the number of testing cases seem not enough so that Table 3 might be biased or not comprehensive enough; third, the lack of repeated testing, or rigorous testing, might also be a potential limitation, given that LLMs' answers vary each time even though with the same prompt. For future work of this specific research, I will expand the number of testing cases and the number of testing models (both web-based and on platforms such as GitHub and Huggingface) to ensure a less biased result.

Some computer based suggestions could also be made to improve the accuracy of models such as adding more diverse training data, so that the output that the models produced are less formal. Users also need output and knowledge that are not as official as textbooks; also, updating dataset might also be a great solution to some contradictory situations such as the word or language use change as time goes on.

References

- [1] 小姐姐 fun 学堂. 汉语拼音【"一"的变调】, 2022. YouTube video. https://www.youtube.com/watch?v=WRvCwUBOZUA&ab_channel=%E5%B0%8F%E5%A7%90%E5%A7%90%E5%AD%A6fun%E5%A0%82. Accessed: 12 May, 2024.
- [2] 3Blue1Brown. Gradient descent, how neural networks learn / Chapter 2, Deep learning, 2018. YouTube video. https://www.youtube.com/watch?v=IHZwWFHWa-w&ab_channel=3Blue1Brown. Accessed: 12 May, 2024.