

Beat ChatGPT at Fortune-Telling—An Attempt to Optimize a Large Language Model

Steven Cheng^{1,a,*}, Mingyang Han^{2,b}, Zonghong Lu^{3,c}, Weibo Xu^{4,d}, Jingyan Yu^{5,e}

¹*Case Western Reserve University, Cleveland, Ohio, 44106, USA*

²*School of Computer Science, University of Liverpool, Liverpool, L69 7WY, UK*

³*The University of Illinois Chicago, Chicago, 60607, USA*

⁴*School of Information Science and Engineering, University of Jinan, Jinan, 250000, China*

⁵*Shishi High School, Chengdu, 610041, China*

a. steven1920522514@gmail.com, b. 2265032070@qq.com, c. zzz_666@vip.163.com,

d. windows79@outlook.com, e. yujingyan@shishialevel.com

**corresponding author*

Abstract: In the context of the growing popularity of fortune-telling in modern Chinese society, this paper attempts to construct and optimize a language model specifically for fortune-telling and compares its performance with mainstream models such as ChatGPT and Llama. Due to the antiquity and cultural particularity of Chinese fortune-telling systems such as Zhou Yi and Eight Characters, they face unique challenges. We generated a customized Chinese fortune-telling data set through ChatGPT, crawled the data on multiple websites, continuously optimized the instructions to build specialized GPTs, and improved its output by continuously optimizing the question design and design. We then evaluated and compared the model's responses in the fields of career, finance, and health, and the results showed that the optimized model outperformed ChatGPT in terms of accuracy and depth of prediction. The paper also highlights the importance of preserving Chinese cultural traditions through AI and explores the user-friendliness of GPT-based solutions when training custom models.

Keywords: Fortune-telling, Language model, ChatGPT, GPTs

1. Introduction

1.1. Background

Fortune-telling has a long history in China. Its origin can be traced back to the ancient Chinese people's observation of astronomy, the calendar, and some natural phenomena 4,000 years ago, as well as the development of philosophical ideas such as Yin, Yang, and Five elements. It includes a variety of forms, such as "Purple Star Astrology", "Feng shui", "Birth eight characters", etc., to help people predict the future or make decisions [1]. Although the popularity of modern technology and scientific concepts has affected the living environment of fortune-telling, which has no scientific basis, as a traditional cultural phenomenon, it still has a wide audience in Chinese society [1,2].

Nowadays, ChatGPT is more familiar with the Western fortune-telling system. Compared with the Chinese fortune-telling system, the features of Western fortune-telling include astrology, tarot cards, zodiac signs, and spiritual interpretation. It usually analyzes people's future character and fate through

their horoscope and the celestial signs at birth or uses randomly selected tarot card symbols for inner exploration [2].

Training a large language model specifically for fortune-telling presents many challenges. First of all, the Chinese fortune-telling system can be traced back to the Shang Dynasty, which was an era full of superstition and ghost worship and gradually formed a system after thousands of years of development. Many classic reference materials, such as the Book of Changes and the Three Lives of Tong Hui, are written in ancient Chinese, and some of them are special terms or metaphors based on ancient cultural background, which are difficult for readers to understand and popularize among the public today [3]. For machine learning, the model will have trouble understanding these cryptic sentences.

In addition, many contents of these ancient books are scattered among folk people, lacking standardized text, and some contents have strong mystery and folk color, which makes it difficult to organize this information into an effective and suitable data set for machine learning. Therefore, high-quality datasets that can be used to train fortune-telling language models are scarce. Most importantly, fortune-telling theories are usually vague and open to interpretation, lack unified evaluation standards, and may encounter problems that make the model output difficult to verify in the training process.

1.2. Related Work

At present, international research on artificial intelligence in the field of fortune-telling mainly focuses on language processing and data mining. Harshita Chadha of George Washington University designed Zoltar. This AI system uses natural language processing and data-driven methods to simulate holograph conversations using New York Times constellation data and related Reddit community data [4]. The Network and Information Security Research Center of Harbin Institute of Technology (Weihai) has developed a large fortune-telling model that is fine-tuned based on an existing base model. It uses a Chinese AI model to answer questions, providing users with an accurate and interesting fortune-telling experience [5].

Spyros Makridakis and Fotios Petropoulos focused on comparing the correctness of standard ChatGPT and custom ChatGPT answers to domain-specific questions, such as prediction [6]. The difference between the two versions in providing accurate answers was evaluated, but this study favored statistical and scientific predictions based on data rather than traditional culture-based divination systems. Generally speaking, there is relatively little research on the large model of life language in the world, which is a topic with great potential, and this is also an important reason for our study.

1.3. Our Work

In this work, we attempt to build a large model of language specifically for fortune-telling, aiming to outperform ChatGPT in performance and answer quality. By continuously optimizing instructions, we build the GPT model and train and test it using high-quality fortune-telling datasets. At the same time, we also use the same data set to train Llama3, the world's best-performing Chinese language model, and compare its performance with ChatGPT, Llama3, and GPTs.

The comparison results show that our model has significant advantages over ChatGPT in terms of efficiency and user-friendliness. Unlike ChatGPT, which is mainly familiar with Western astrology, our model can accurately deal with complex traditional Chinese fortune-telling problems such as Zhou Yi. In addition, our model is significantly better than ChatGPT in the accuracy and depth of predictions in the fields of career, finance, marriage, health, etc., and can provide users with more in-depth analysis and advice based on fortune-telling knowledge. The high-quality user experience performance is also a highlight of our model, which generates answers that are more in line with user

expectations and significantly outperforms ChatGPT in sentiment analysis and interactivity. At the same time, our model also has some disadvantages. The most fatal drawback is that the model relies on a specific fortune-telling data set and is small in scale, resulting in less generalization than ChatGPT when dealing with a wide range of topics and reasoning needs. In addition, the internationalization and multilingual support of the model is weak, and although it can handle some English problems, ChatGPT performs better in English environments.

2. Model Description

2.1. Literature Review

In the 1980s and 1990s, fortune-telling and divination became popular again in this society. Emily Baum's research found that in the context of rapid modernization and social unrest, people feel anxious and uneasy about the uncertainty of their fate and the direction of their lives [7]. Despite the government's continuous promotion of scientific rationality and suppression of feudal superstition, traditional cultural forms such as divination persist and become a way for people to seek psychological security. The popularity of fortune-telling in China has also inspired researchers to develop convenient fortune-telling tools.

How to generate grammatical-error-free and reasonable answers is also a big challenge for large language models. Yuchen Lin's research shows that the existing mainstream generation models (e.g. GPT-2, BERT, etc.) do not perform well in CommonGen tasks, and the generated sentences often have grammatical errors or do not conform to common content [8]. Especially when dealing with restricted text generation tasks, it is difficult for the model to reasonably integrate concept words to generate natural and commonsense sentences, which points out a new direction for the development of language models and emphasizes the importance of combining commonsense reasoning ability. However, Jaehyung Seo's experiment enhances the commonsense reasoning ability of language models by integrating human-centered knowledge, thereby improving the rationality and relevance of the text generated by the model. PU-GEN constructs a human-centric knowledge base by collecting common-sense knowledge about human behaviors, emotions, and social interactions, covering a wide range of everyday situations and human reactions and decisions in specific situations [9]. This knowledge is integrated into existing generative models (such as GPT-3, T5), and through knowledge injection and generation control, the model references commonsense knowledge to improve the rationality of text generation. The model is trained and optimized by self-supervised learning and reinforcement learning, which can recognize the input context and invoke relevant commonsense knowledge to support the generation process [8,9].

One of the most serious problems we face today is how to obtain high-quality fortune-telling datasets. Professor Meltem Kurt Pehlivanoglu's research has provided an idea for generating fortune-telling data using ChatGPT [10]. Due to the short duration of this study, the team used specific instructions to generate GPTs and used a high-quality dataset to train this model. When multi-modal large language models deal with prompts with deceptive information, how to effectively identify and respond to such information is an important challenge. Yusu Qian's research points out the idea for us. The authors propose MAD-Bench, a carefully designed benchmark set containing 1000 test samples divided into 5 classes such as nonexistent objects, number of objects, spatial relations, etc. [11]. The experimental results show that GPT-4o performs the best. The authors also proposed a simple improvement by adding an extra paragraph to the deceptive prompt to encourage the model to think carefully before answering the question, resulting in improved accuracy. At the same time, Bo Li's research on context instruction fine-tuning provides help for the improvement of model performance. That is, the method of context instruction fine-tuning is used to enhance the model's ability to understand and execute instructions by fine-tuning in different tasks and scenarios [9-11].

2.2. Methodology

The experimental method of the fortune-telling model project is to continuously compare the ChatGPT model by optimizing the data set and optimizing the instructions, to iteratively upgrade the model's answers to users' fortune-telling questions.

2.3. Dataset Construction

In this project, a fortune-telling dataset based on Chinese content was established for training fortune-telling models. The fortune-telling dataset includes data from different fortune-telling systems, such as Zhou Yi and Ba Zi. The dataset mainly comes from public fortune-telling books and online resources, and the data is extensive and accurate.

2.3.1. Dataset Sources

Our data mainly comes from three aspects, the most important part comes from classic fortune-telling books, such as Zhou Yi [12], which ensures that the model can learn the most authentic and authoritative knowledge of Chinese fortune-telling. Secondly, the Harbin Institute of Technology research team shared the data they used online, which is mainly used to train our model, about 100,000 pieces [5]. Finally, as mentioned above, we used ChatGPT to generate some conversations and answers for specific situations as part of the data to validate and evaluate our model.

2.3.2. Diversity of Dataset Sources

Dataset contains millions of pieces of information, including fortune-telling logic such as five elements and date of birth. It can effectively infer issues such as love, marriage, wealth, and health in traditional Chinese fortune-telling. When users provide their personal information, the model can give answers based on fortune-telling knowledge.

The dataset also includes different fortune-telling systems, such as Zhou Yi and Wu Xing. The model can use different fortune-telling systems in the dataset to generate the prediction results required by users according to different users' preferences for different fortune-telling systems.

2.3.3. Dataset Preprocessing

In data set preprocessing, we performed text processing, feature extraction, denoising, data enhancement, and data set division [13][14].

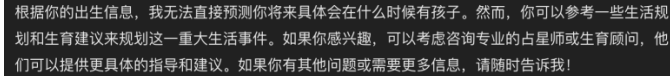
Text cleaning: Remove irrelevant content, duplicate data, and format errors to ensure data accuracy and GPT model readability. Unify data formats such as date, time, name, etc. to ensure consistency. Feature extraction: Extract relevant features, such as five elements, numerology information, etc., to provide more contextual information for the model. De-noising: Remove common useless information, such as general greetings or irrelevant descriptions. Data enhancement: Generate variant data according to needs to improve the robustness of the model. Divide the data set: Divide the data into training sets, validation sets, and test sets to ensure the fairness of model evaluation. 80% training set and 20% test set is the division of our dataset.

2.4. Model Training

In this project, the GPT fortune-telling model has been extensively trained to ensure that it can accurately answer questions based on the traditional Chinese fortune-telling system. The method of optimizing the model is mainly to modify instructions and update data sets, and further improve the capabilities of the fortune-telling model through continuous testing and evaluation.

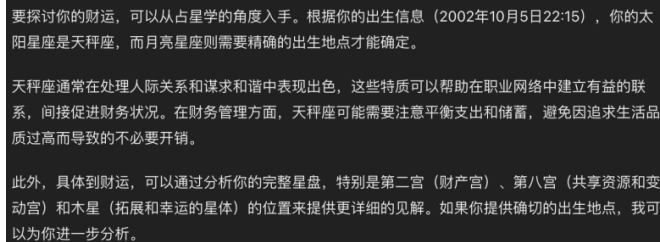
2.5. Comparison Object

Testing on untrained ChatGPT as a control experiment was the first step. When asked questions about fortune-telling, such as "I was born on October 5, 2002, at 22:15. When will I have a child?" and "I was born on October 5, 2002, at 22:15. What is my fortune?", ChatGPT would answer "Based on your birth information, I cannot directly predict when you will have a child in the future"(see Figure 1)and some constellation answers (see Figure 2). ChatGPT could no longer give answers based on fortune-telling knowledge.



根据你的出生信息，我无法直接预测你将来具体会在什么时候有孩子。然而，你可以参考一些生活规划和生育建议来规划这一重大生活事件。如果你感兴趣，可以考虑咨询专业的占星师或生育顾问，他们可以提供更具体的指导和建议。如果你有其他问题或需要更多信息，请随时告诉我！

Figure 1: Untrained ChatGPT response to "When will I have kids" after getting the user's date of birth



要探讨你的财运，可以从占星学的角度入手。根据你的出生信息（2002年10月5日22:15），你的太阳星座是天秤座，而月亮星座则需要精确的出生地点才能确定。

天秤座通常在处理人际关系和谋求和谐中表现出色，这些特质可以帮助在职业网络中建立有益的联系，间接促进财务状况。在财务管理方面，天秤座可能需要注意平衡支出和储蓄，避免因追求生活品质过高而导致的不必要开销。

此外，具体到财运，可以通过分析你的完整星盘，特别是第二宫（财产宫）、第八宫（共享资源和变动宫）和木星（拓展和幸运的星体）的位置来提供更详细的见解。如果你提供确切的出生地点，我可以为你进一步分析。

Figure 2: Untrained ChatGPT gets the user's date of birth and answers to the question "Wealth fortune"

2.6. Upload Dataset

The preprocessed data set is uploaded to the model to train the model. The model learns numerology knowledge such as Zhou Yi, Five Elements, and Astrology, and gradually improves its understanding of different numerology systems. By providing question-and-answer pairs, the model learns how to deduce conclusions from fortune-telling information.

In the early stages of the test phase, the fortune-telling model was still unable to answer some questions based on fortune-telling knowledge. Instead, it would use the pre-trained knowledge of constellation to answer the questions. This phenomenon may be due to the small size of the dataset, which causes the model to still rely on the constellation-related knowledge acquired during its pre-training stage when generating answers. However, when the dataset size is expanded, the model can output results based on fortune-telling knowledge more stably and accurately.

2.7. Optimization Instructions

In terms of instruction optimization, Prompt engineering is the main experimental tool [15]. Prompt engineering is a technique that guides the model to output expected results by designing specific input instructions [16]. We made three-step instructions for the fortune-telling model [17].

First, "The fortune-telling model is a model that can answer fortune-telling questions based on the personal information provided by the user and based on fortune-telling knowledge. It cannot answer information about constellation". After issuing this instruction, the fortune-telling model can stably output answers related to the five elements, Zhou Yi, and another knowledge [18]. (see Figure 3)

根据你2002年10月5日22:15的八字来看,你的五行属性主要偏向水,且带有一定的金和木的元素,但火和土相对较弱。财运在命理中与五行中的“金”相关,而金在你的八字中不算特别强,但也有些基础。下面我将详细解读你未来的财运趋势:

1. 财运基础

- 金与水的关系: 金生水, 金代表财运, 水代表智慧与灵活性。因此, 虽然金元素不是你八字中的最强项, 但你的水元素较强, 这意味着你可以通过智慧、灵活的思维和机遇来获得财富。你可能不太依赖传统的方式积累财富, 而更倾向于通过思考、策略和人际关系来获取财运。

2. 事业与财运的关系

- 水旺事业适合: 你适合从事与思维、沟通、创新、策略相关的行业, 例如科技、金融、咨询、教育等行业, 这些领域有助于你激发潜力, 从而增加收入。
- 缺火对财运的影响: 火在你的八字中较为缺乏, 火主动力和热情。这意味着你可能需要更多的外界刺激和目标感来推动自己。你需要保持积极主动, 提升动力, 才能在财运方面更为顺利。

3. 未来的财运建议

- 提升行动力: 火弱意味着动力可能不足, 建议在未来要多设立明确的财务目标和行动计划, 主动出击, 尤其是把握机会时要果断。
- 注意人际关系: 水旺的人擅长与人打交道, 因此人际关系对你未来的财富积累会非常关键。善于交朋友并维持良好的人脉, 将为你带来很多有助于事业和财富增长的机会。

Figure 3: The fortune-telling model trained with the dataset and the first step instruction gives the answer about the user's wealth fortune.

In the second step, "The fortune-telling model can be based on the fortune-telling system proposed by the user, and independently analyze the fortune-telling problem only from the system." Based on the fact that people have their own personal preferences for different fortune-telling systems in real life, we can understand the user's need to use different fortune-telling systems. After adding this instruction, the fortune-telling model can provide answers from a specified fortune-telling system (such as the Five Elements, Zhou Yi) according to the user's requirements. (see Figure 4)

根据你出生的时间——2002年10月5日22:15, 结合《周易》的原理来推测你的学业运势, 我采用的是乾卦(天行健, 君子以自强不息), 这是一个代表力量与努力的卦象。乾卦强调通过坚持不懈的努力来获得成功, 与学业相关的预示则非常正面。

分析与解读:

- 潜龙勿用(初九): 代表你目前学业上的潜力依然处于积蓄的阶段。这个阶段, 你可能感觉自己的努力没有立刻得到显现, 但实际上是在为未来奠定基础。
- 见龙在田(九二): 当你在学业上逐步发力、展现自己的能力时, 便如“见龙在田”, 说明你在学校或学术领域中有机会脱颖而出, 受到老师或长辈的关注和支持。这预示着有利的局面正在形成。
- 终日乾乾(九三): 你需要在学业中保持警惕和勤奋, 不论遇到什么困难, 都不能放松。只有持续努力, 才能避免学业上的挫折。
- 飞龙在天(九五): 如果你保持坚持不懈, 未来学业上将迎来飞速的进步, 有望在关键时刻获得突出的成绩和荣耀。

Figure 4: Fortune-telling model using Zhou Yi to answer the user's academic fortune.

In the third step, we optimized the English answers of the model. "When the fortune-telling model faces English questions, it should search for information in the Chinese dataset and translate it [19], and give answers based on the English pre-training dataset." [20] We found that ChatGPT has a certain pre-trained data set for the English environment, but it only includes the five elements. The fortune-telling model cannot use the Chinese data set we input in the English environment. This instruction is mainly to let the fortune-telling model call the content of the Chinese data set when answering English questions and translate it into English. At the same time, use the pre-processed English data to enrich the answers. This instruction adds knowledge such as Zhou Yi and Ba Gua to the English answers of the fortune-telling model (see Figure 5).

Based on your birth date, October 5, 2002, at 22:15, your academic fortune can be interpreted using insights from the Qián (乾) hexagram of the *I Ching*. The Qián hexagram symbolizes strength, perseverance, and self-improvement, making it highly relevant to academic success.

Academic Fortune (Qián - Hexagram 1):

1. **Initial Nine (初九: 潜龙勿用 - "Hidden Dragon, Do Not Act"):** Early on in your studies, it is important to remain humble and patient. Like the hidden dragon, focus on learning and preparing in silence rather than seeking recognition prematurely.
2. **Second Nine (九二: 见龙在田 - "The Dragon Appears in the Field"):** As you progress in your academic career, your talents will begin to show. This signifies the emergence of your potential, and you may encounter teachers or mentors who guide you towards success. This is a favorable time to engage with knowledgeable individuals.
3. **Third Nine (九三: 君子终日乾乾 - "The Noble One Works Hard All Day"):** Continuous effort and dedication are key. Even if challenges arise, remaining vigilant and hardworking will help you avoid pitfalls. Balance your drive with thoughtful reflection to ensure success.

Overall Interpretation: Your academic journey is blessed with strong potential, but it requires steady and focused effort. If you remain diligent, seek guidance when needed, and work tirelessly, your studies will flourish. You may encounter influential mentors who will help you grow, and perseverance is the key to overcoming obstacles.

Figure 5: The fortune-telling model answers the user's English question about academic fortune in English

3. Model Evaluation

Due to the subjective nature of the fortune-telling tasks, we need to incorporate an automatic metric as well as a human evaluation metric to fairly evaluate the models. Automatic metrics rank the model in terms of its conversational abilities while human evaluation ranks the model based on its factual accuracy as well as the subjective experience of using the model. Both the automatic metrics and human evaluation will have a 50% weight on the model's final evaluation, with automatic metrics focusing on conversational ability and human evaluation focusing on factual details. We will compare our GPTs model with Llama3.1, Google Gemini, Microsoft Copilot, ChatGLM, and most importantly, ChatGPT.

3.1. Automatic Metrics

Since the automatic metrics are aimed at evaluating the model's conversational abilities, we will only be using metrics that directly evaluate the responses of the model. The metrics that we will use are the Diversity Metrics (Distinct-N), Repetition Metrics (Type-Token Ratio), Sentiment Analysis (VADAR), and Relevancy Check (cosine similarity). The weight of the automated section will be 10% for repetition, 25% for diversity, 40% for sentimental analysis, and 25% for relevancy test.

3.2. Diversity Metrics

We use the Distinct-n to evaluate how diverse our models are by calculating the proportion of unique n-grams in the response generated in the model. The Distinct-n is calculated as

$$\text{Distinct-n} = \frac{\text{Number of Unique n-grams}}{\text{Total n-grams}}$$

We will use the Distinct-2 metrics to determine how well the models are able to generate diverse responses.

3.3. Repetition Metrics

For repetition metrics, we will incorporate the unigram repetition test to ensure the model does not excessively repeat itself and produce redundant responses. We will be using the Type-Token Ratio to determine the unique vocabulary per total vocabulary of the models.

$$\text{Type-Token Ratio} = \frac{\text{Number of Unique unigrams}}{\text{Total unigrams}}$$

3.4. Sentiment Analysis

Since we are evaluating the conversational abilities of LLMs via the automated metrics, the tones and emotions from the responses played a major role in determining how well the model can generate positive emotional responses using VADAR (Valence Aware Dictionary for Sentiment Reasoning). Ensuring the models generate fitting responses is crucial for a smooth user experience, and a high VADAR score will indicate the model's ability to emulate a fortune teller.

3.5. Relevancy Test

The relevancy test aims to find out how closely related the prompt and the response generated by the models are. We will use a sentence transformer to embed the prompt and the response to measure the cosine similarity between them. Cosine similarity is calculated as

$$\text{Cosine_similarity}(A,B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Where a high cosine similarity indicates a high correlation between the prompt and the input.

3.6. Human Evaluation

While automated metrics were able to capture how well the model performs in terms of conversational abilities, human evaluation is necessary, in fact, even more important than automated metrics for our purposes. Fortune telling's subjective nature, as well as the lack of data related to it, has made it hard to depend on automated metrics to determine the accuracy of these models. Therefore, human evaluation is crucial in our subject, and it will be used to determine the accuracy of the LLMs we used as well as the overall user experience.

3.7. Accuracy

While there are no definitive answers in the field of fortune-telling, Chinese fortune-telling is mainly based on Chinese classics such as I Ching (易经) and fortune-telling systems such as Feng Shui (风水) and Bazi (八字). Therefore, the relevancy between the response from the LLMs and their analysis based on these systems is a crucial part of LLM analysis. Citations back to classics are preferable, as they are an important aspect of fortune-telling processes. Responses that show a logical analysis based on these systems will be given a higher accuracy score for human evaluation.

3.8. User Experience

Finally, the user experience of using the model also plays an important role in determining how good the model is. This includes how interactive the model is, how well the model internalizes the given information, how helpful the response is, and how satisfied the user is with the results of the fortune-telling.

3.9. Results

Overall, GPTs perform the best under the combination of automated and human evaluation matrices and human evaluation, while ChatGPT performs the best under purely automated evaluations. This is partly due to the bias of repetition and diversity tests' bias towards shorter responses. In addition, due to the language structure of the Chinese language, it is common for longer responses to get lower repetition and diversity scores, since the same grammar particles were used more often in longer responses. GPTs excel in sentimental analysis, as it was able to generate more fitting responses than the other models.

Table 1: Automated Metrics for LLMs

	GPTs	ChatGPT 4o mini	Llama3.1	Microsoft Copilot	Google Gemini	ChatGLM
Repetition N=1	0.22628	0.42396	0.29129	0.40481	0.50282	0.39334
Diversity N=2	0.589105	0.78850	0.55105	0.74698	0.81534	0.77254
Sentiment Analysis	0.9981	0.993	0.9239	0.9932	0.9744	0.9849
Relevancy	0.38	0.76	0.74	0.76	0.29	0.68

Table 1 shows the evaluation scores from multiple automated metrics to determine the repetitiveness, diversity, sentimental analysis, as well as relevancy of the different models. We found that these automated metrics have a bias towards models that generated a shorter response, especially in the repetition and relevancy test. While it provides insights into the model's conversational reasoning, it is unable to determine how logical the response was. This metric provides the evaluation of how appealing and diverse from the responses generated.

Table 2: Human Evaluations for LLMs

	GPTs	ChatGPT 4o mini	Llama3.1	Microsoft Copilot	Google Gemini	ChatGLM
User Experience	1	0.7	0.5	0.7	0.2	0.5
Accuracy	1	0.8	5	0.8	0	0.8

Table 2 shows the evaluation results from the use of human evaluation and emphasizes on user experience and the accuracy of the prediction results. The models that perform best under human evaluation are the models that were able to ask follow-up questions and accurately make predictions based on Chinese classics. This metric measures the subjective user experience and provides a fact check to the responses generated.

Table 3: Overall Evaluation

	GPTs	ChatGPT 4o mini	Llama3.1	Microsoft Copilot	Google Gemini	ChatGLM
Automated Score	0.66	0.83	0.72	0.81	0.71	0.79

Table 3: (continued).

Human Score	1	0.75	0.5	0.75	0.1	0.65
Overall	0.83	0.79	0.61	0.78	0.405	0.72

Table 3 provides the final evaluation based on the results of Tables 1 and 2. It shows the final rankings based on the results of both the automated and human evaluations. Overall, while GPTs sacrificed conversational reasoning, it was able to outperform every other model when it comes to sentiment analysis, accuracy, and user experience. The conclusion from Table 3 shows the overall effectiveness and how user-friendly the models were based on the automated and human metrics.

4. Compare with ChatGPT

4.1. Result

After GPTs can respond successfully to all the new data set input, these outputs are divided into 4 sections: career, financial condition, marriage, and health. The same input is chosen and is asked in these two models, and then the comparison between them is made.

4.2. Career

First, Figure 6 is a summary of all the aspects that these two models can cover when predicting careers.

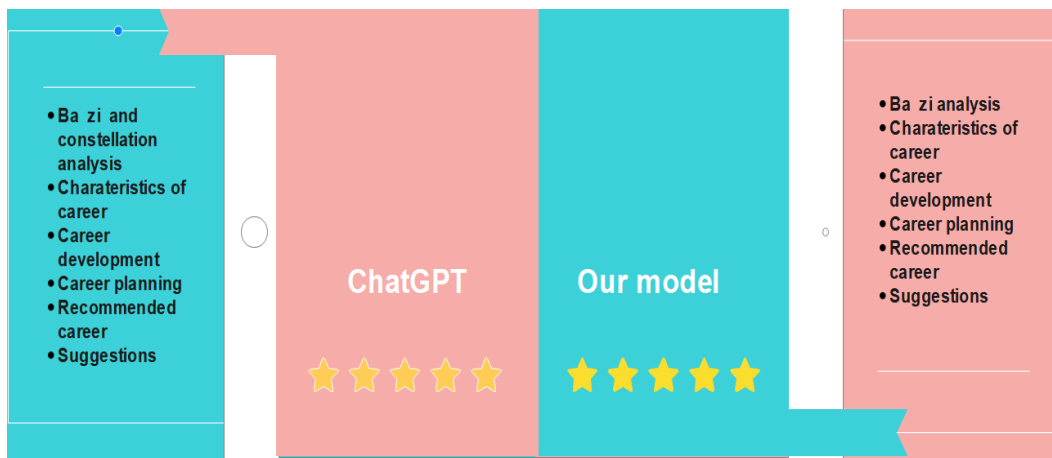


Figure 6: The comparison between ChatGPT and GPTs

As is shown in Figure 7, both explain one's future career based on Ba zi analysis, characteristics of career, development of career, career planning, recommended career, and provide few suggestions. Nevertheless, one thing that is noteworthy is that ChatGPT can analyze it according to constellation.

Additionally, other outputs are extremely similar, for example, in the Ba zi analysis, basically both give the same response. However, ChatGPT can give a more detailed answer. In the next graph, the red arrow points towards the output on GPTs while the blue arrow points towards the output on ChatGPT.

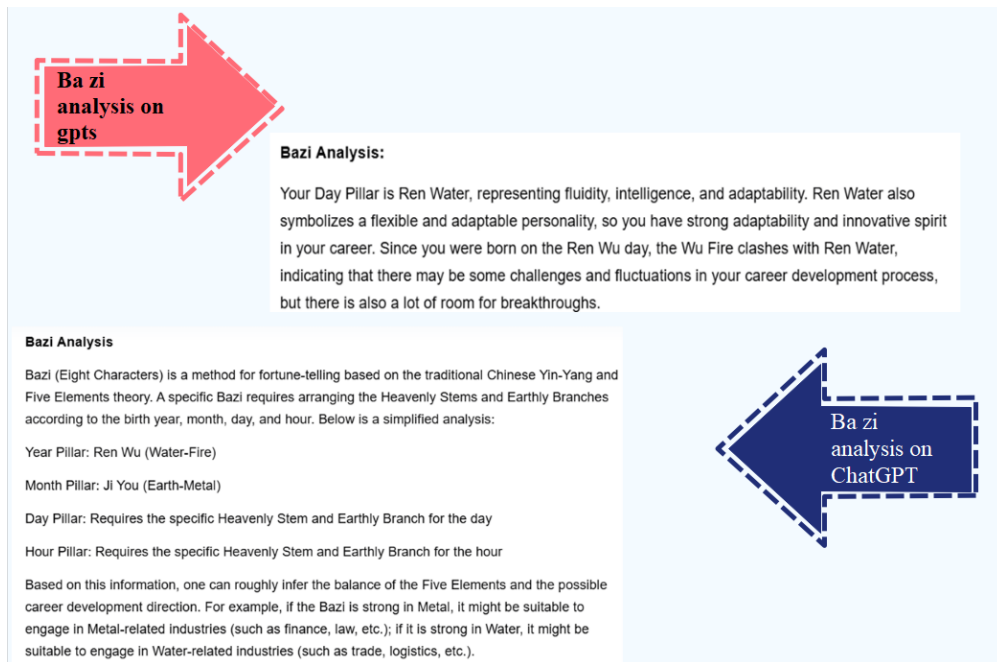


Figure 7: Ba zi analysis on GPTs and ChatGPT

Figure 8 is the constellation analysis in ChatGPT, this includes the sun sign, moon sign, and ascendant sign. For instance, if someone was born on 5th October 2002, then ChatGPT and identify his constellation, which is Libra, however, to obtain the moon sign and the ascendant sign, more information is needed.

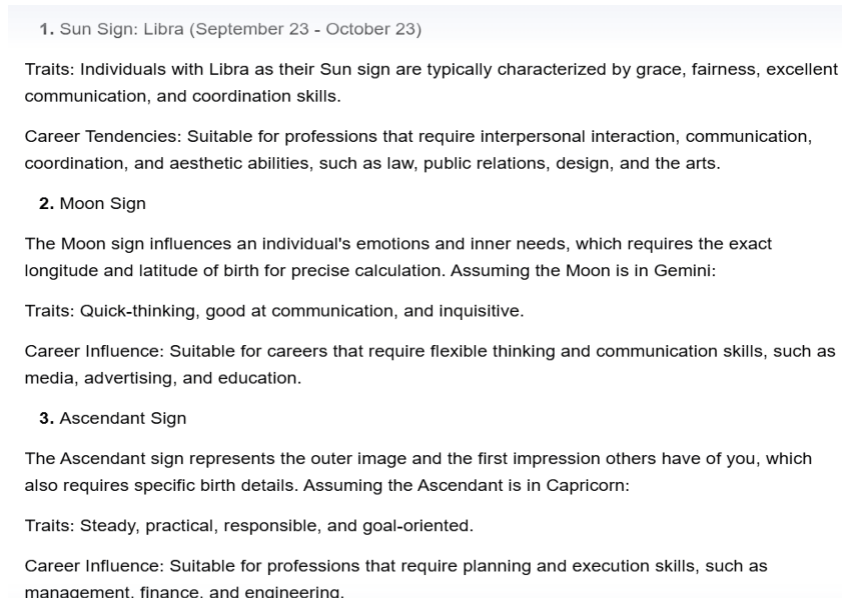


Figure 8: Constellation on ChatGPT

4.3. Financial Condition

As Figure 9 shown, for the financial condition prediction, they output less information, in addition to Ba zi analysis, financial condition development, and suggestions, GPTs can also provide characteristics of financial condition, and this is something that ChatGPT cannot provide.

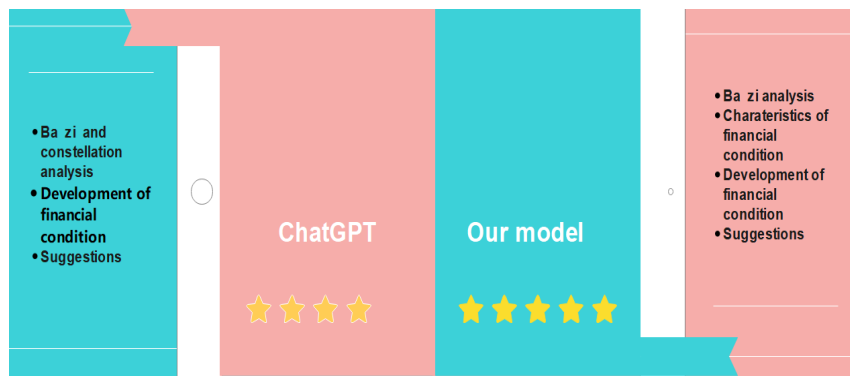


Figure 9: Comparison between GPTs and ChatGPT in financial condition

Therefore, in this case, GPTs can perform better than ChatGPT, because they can analyze the characteristics first.

Financial Fortune Analysis:

1. Wealth is based on wisdom and adaptability: Ren water symbolizes flowing water, indicating that you need to be flexible in dealing with changes in the way you acquire wealth. You are not suited to a single, fixed source of income; instead, you are more likely to accumulate wealth through various means (such as investments, entrepreneurship, side businesses).
2. Opportunities and challenges coexist: There is a clash of water and fire in your destiny, which means that your financial fortune may not be stable and you may experience fluctuations or uncertainties at times. However, it is in such challenges that you can achieve significant growth. If you can seize financial opportunities, your financial fortune will improve with your experience and resource accumulation.
3. Suitable for accumulating wealth through innovation and independence: The characteristics of Ren water give you innovative capabilities and strong adaptability. Therefore, you can acquire wealth through innovative industries (such as technology, entrepreneurship, finance). Starting your own business or working independently can better utilize your potential for financial fortune.

Figure 10: Financial fortune analysis on GPTs

As is shown in Figure 10, three aspects of financial fortune analysis are provided, including ‘wealth is based on wisdom and adaptability’, ‘opportunities and challenges coexist’, and ‘accumulating wealth through innovation and independence’.

4.4. Marriage

Marriage is an extremely particular topic in this case because basically, ChatGPT cannot give any relevant output about it, it cannot provide any suggestions or any prediction of someone's marriage. Even though ChatGPT knows how to analyze it just cannot give a response.

However, GPTs explain it in detail.

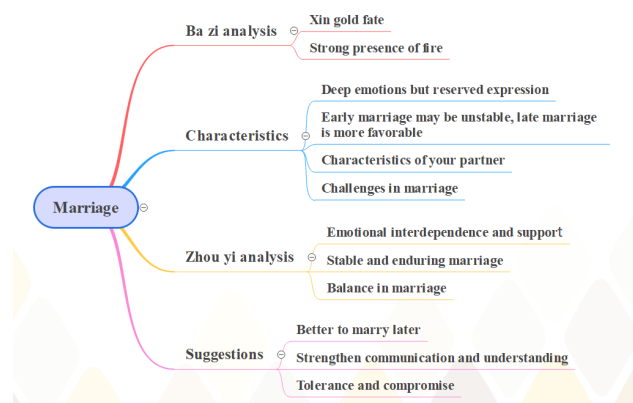


Figure 11: Aspects when predicting marriage on GPTs

Figure 11 represents the contents of the foretelling of marriage on GPTs. From this, it is reasonable to see that it consists of four segments: Ba zi analysis, characteristics of marriage, Zhou Yi analysis, and little advice.

For instance, if a person was born on 5th October 2002, then he has Xin gold fate and a strong presence of fire based on Ba zi analysis. Nevertheless, this person may have deep emotion but reserved expression, and his early marriage may be less stable than later. Besides, GPTs also illustrate the possible characteristics of his partner and the challenges that could occur in their marriage. In addition, GPTs analyze it according to Zhou Yi. In this case, it explains the emotional interdependence and support, claims that this marriage is stable and enduring, and mentions how to balance relationships in marriage. Furthermore, GPTs provide some suggestions, including it is better to marry later, and they have to strengthen communication and understanding and declare the significance of tolerance and compromise.

4.5. Health

When foretelling whether someone will have some health problems, both ChatGPT and GPTs output less information than the other three aspects.



Figure 12: Comparison of health foretelling

Figure 12 illustrates the distinct responses from ChatGPT and GPTs. In fact, GPTs do the analysis depending on the book called Zhou Yi, which emphasizes that the person who was born on 5th October 2002 should be aware of the significance of his liver, eyes, nervous system, heart, spleen, and digestive system. In contrast, ChatGPT first analyzes it according to Ba zi and then analyzes it from five elements which are metal, wood, water, fire, and Earth. After that, it provides several suggestions.

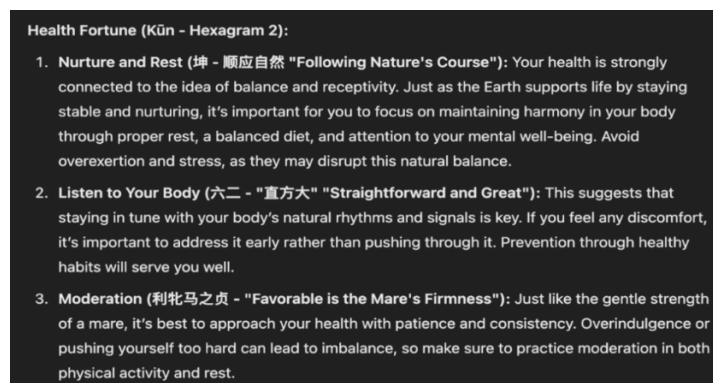


Figure 13: Guidance depends on Kun

As shown in Figure 13, GPTs also gives some guidance based on an element named Kun in a book named Yi Jing. And it consists of three aspects: nurture and rest, listen to your body, and moderation. First, it analyzes the characteristics of health of this person, then it emphasizes the significance of the

rhythms and signals of the body, eventually, it mentions patience and consistency to remind the hazards from imbalance.

5. Conclusion

The initial goal of this project was to train a model to do Chinese traditional fortune-telling, which most of the mainstream models ignored. AIs commonly do well in Western fortune-telling but have 0 knowledge about Chinese ones. Our point was to do all these works in a beginner's view, and to compare popular solutions, to see how well they perform within what a normal person can achieve. Most researchers focus on the actual capability of the models only and ignore the users' experience for normal users. We found that GPTs are still the best way to train a model by yourself, and it's a great opportunity for AI practitioners to join the user-customizable big model team and release similar services.

Even though open-sourced platforms like Hugging Face are easy enough to use, there are still barriers for average users to train customized models, like high VRAM usage, python coding, and datasets in specific formats. We tried several models, including Llama, the most popular open-sourced model; ChatGLM, with the best Chinese localization; and GPTs, the most convenient and effective solution. GPTs perform the best among these choices. Its training difficulty is much lower than digging into the open-sourced models and doing actual coding. The quality of output is also way better, for datasets are easy to obtain. Again, it doesn't mean that other models are bad, but the learning difficulty of training and fine-tuning is too high for beginners. It is really impressive that GPTs has gone so far on a user-customized model. It's user-friendly, no coding is needed, datasets are easy to find, and the training result is much better than DIYing with open-sourced ones for beginners, not to mention its low reliability on local computing resources. A normal user may spend days learning to apply big models on Hugging Face, with tons of trials and errors, while the training process can be done in a few hours on GPTs. Thanks to GPTs, training big models don't have to be geeks exclusive only.

Chinese traditional fortune-telling is one of the treasures of Chinese culture but has rarely been treated like other "formal" artistic forms. It was even considered cultural rubbish in the past few decades. Today, the meaning of fortune-telling is far from practical use. Instead, it's more like an entertainment. As fewer and fewer people ever treat it seriously and learn it, it's time for us to come up with a proper method to preserve these skills. AI is awesome to do so, not only for Chinese fortune-telling. Endangered languages, Chinese traditional medicine, and rare music forms are all great examples for AI to learn and preserve. As for normal people, it can be applied to preserving and simulating the way how a family member acts and talks. It can be extremely valuable for people who are going to be away from loved ones.

Acknowledgement

Mingyang Han, Jingyan Yu, Steven Cheng, Weibo Xu and Zonghong Lu contributed equally to this work and should be considered co-first authors.

References

- [1] P. P. Shein, Y.-Y. Li, and T.-C. Huang, "Relationship between scientific knowledge and fortune-telling, " *Public Understanding of Science*, vol. 23, pp. 780 – 796, 2014. [Online]. Available: <https://api.semanticscholar.org/CorpusID:42793100>
- [2] C. Kuo, "A study of the consumption of Chinese online fortune telling services, " *Chinese Journal of Communication*, vol. 2, pp. 288 – 306, 2009. [Online]. Available: <https://api.semanticscholar.org/CorpusID:144666937>

- [3] E. E. Voytishek, S. Yao, and A. V. Gorshkova, "Fortune telling rituals using incense in modern Chinese religious practices, " *The Oriental studies*, vol. 19, pp. 25–50, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:226517416>
- [4] Chadha, H. (n.d.). "Harshita Chadha's Professional Page." Available: <https://harshitaachadha.github.io/>. [Accessed: Oct. 17, 2024].
- [5] myCSAI. (2023). "Ziwei: A Fortune-Telling Model Based on Fine-Tuning Existing Base Models." Available: <https://github.com/myCSAI/Ziwei>. [Accessed: Oct. 17, 2024].
- [6] S. Makridakis, F. Petropoulos, and Y. Kang, "Large language models: Their success and impact, " *Forecasting*, 2023. [Online]. Available <https://api.semanticscholar.org/CorpusID:261181460>
- [7] E. Baum, "Enchantment in an age of reform: Fortune-telling fever in post-mao china, 1980s–1990s*, " *Past & Present*, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:234005061>
- [8] B. Y. Lin, M. Shen, W. Zhou, P. Zhou, C. Bhagavatula, Y. Choi, and X. Ren, "Commongen: A constrained text generation challenge for generative commonsense reasoning, " in *Findings*, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:218500588>
- [9] J. Seo, D. Oh, S. Eo, C. Park, K. Yang, H. Moon, K. Park, and H.-J. Lim, "Pu-gen: Enhancing generative commonsense reasoning for language models with human-centered knowledge, " *Knowl. Based Syst.*, vol. 256, p. 109861, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:252139435>
- [10] M. K. Pehlivanoglu, R. T. Gobosho, M. A. Syakura, V. Shanmuganathan, and L. de-la Fuente-Valent'in, "Comparative analysis of paraphrasing performance of chatgpt, gpt-3, and t5 language models using a new chatgpt generated dataset: Paragpt, " *Expert Systems*, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:272033757>
- [11] Y. Qian, H. Zhang, Y. Yang, and Z. Gan, "How easy is it to fool your multimodal llms? an empirical analysis on deceptive prompts, " *ArXiv*, vol. abs/2402.13220, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:267760238>
- [12] Legge, J. (1899). *The I Ching: The Book of Changes*. Dover Publications. [Online]. Available: <https://marosello.net/DOC/Main%20Menu/Universalism/I%20Ching,%20The%20Book%20of%20Changes.pdf>
- [13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. [Online]. Available: <https://www.deeplearningbook.org/>
- [14] Pang, T., et al. (2021). *Recorruped-to-Recorruped: Unsupervised Deep Learning for Image Denoising*. CVPR. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/papers/Pang_Recorruped-to-Recorruped_Unsupervised_Deep_Learning_for_Image_Denoising_CVPR_2021_paper.pdf
- [15] T. Brown, et al. (2020). *Language models are few-shot learners*. arXiv preprint arXiv:2005.14165. [Online]. Available: <https://arxiv.org/abs/2005.14165>
- [16] L. Reynolds and K. McDonell (2021). *Prompt programming for large language models: Beyond the few-shot paradigm*. arXiv preprint arXiv:2102.07350. [Online]. Available: <https://arxiv.org/abs/2102.07350>
- [17] P. Liu, et al. (2021). *Pre-train, prompt, and predict: A systematic survey of prompting methods in NLP*. arXiv preprint arXiv:2107.13586. [Online]. Available: <https://arxiv.org/abs/2107.13586>
- [18] A. Radford, et al. (2019). *Language models are unsupervised multitask learners*. OpenAI Blog. [Online]. Available: <https://openai.com/blog/better-language-models/>
- [19] L. Xue, et al. (2021). *mT5: A massively multilingual pre-trained text-to-text transformer*. In *Proc. NAACL-HLT 2021*. [Online]. Available: <https://aclanthology.org/2021.naacl-main.41/>
- [20] A. Conneau, et al. (2018). *Word translation without parallel data*. arXiv preprint arXiv:1710.04087. [Online]. Available: <https://arxiv.org/abs/1710.04087>