Chinese Chorales Dataset: A High-Quality Music Dataset for Score Generation

Yongjie Peng[1], Lei Zhang[2], and Zhenyu Wang[1(✉)]

[1] The School of Control and Computer Engineering, North China Electric Power University, Beijing, China
`{yjpeng,zywang}@ncepu.edu.cn`
[2] Beijing National Day School-Longyue Experimental Middle School, Beijing, China

Abstract. For a long time, the JSB Chorales Dataset has served as the benchmark for choral composition generation, with numerous models and algorithms achieving remarkable results on this dataset, which is designed to generate Bach-style choral music. However, when we aim to tackle the task of generating Chinese vocal choral compositions, we encounter a lack of suitable Chinese music datasets for this purpose. The Chinese Chorales Dataset presented in this paper is a high-quality collection of Chinese choral music, comprising 125 Chinese choral songs stored in MusicXML format, divided into 441 musical segments. This dataset has been professionally crafted to meet the needs of Chinese composers seeking to create high-quality choral compositions. We also provide a compressed .npz file version containing pitch, fermata, tempo, and chord information, split into training, validation, and test sets. Additionally, we conducted multiple experiments on this dataset to validate the effectiveness of the information contained within. For access to the dataset and usage details, please visit https://github.com/123654ad/Chinese-Chorales-Dataset/tree/main.

Keywords: music dataset · Chinese chorales · machine learning · benchmark

1 Introduction

Musical score is a way of recording music in symbols. It typically deals with symbolic representations and encodes abstract musical features including beats, chords, pitches, durations, and rich structural information. The relationship between musical score and audio is like that between text and speech; a musical score is a highly symbolic and abstract visual representation that effectively records and communicates musical ideas, while audio contains all the details that we can hear. The symbolic generation of musical scores, i.e., Score Generation, is one of the important subtasks of music generation. [4,10]

There are many compositional tasks included in Score Generation, such as: generating a harmonic accompaniment to a given melody, composing a melody based on chords, melody generation, etc. [5,7,8]. Some of these score-generation

tasks are automated without constraints, and some are conditionally constrained. Still, most of them are aimed at assisting composers to create more novel and creative songs more easily. In order to accomplish these musical score-generation tasks, a high-quality musical score dataset is essential.

In recent years, the JSB Chorales Dataset [1] and the JSB fake Chorales Dataset is often used for melodic harmony tasks. The goal of this task is to generate a harmonic accompaniment to a given melody. Harmony is one of the basic expressive tools of music, which can add richness and expressiveness to music, and the difficulty of this task lies in ensuring the correct relationship between the generated melody and the harmony. In both datasets, each musical score contains four voices, namely soprano, alto, tenor, and bass, and the soprano voice is usually used as the main melody to generate the other three voices, thus obtaining a vocal chorus in the Bach style.

For the Bach-style dataset, many excellent models have been developed in recent years, such as Choir Transformer [20], DeepChoir [19], BachBot [11], Deep-Bach [6], and so on. These models all utilize deep learning techniques to learn the styles and rules of Bach chorales and are able to generate high-quality and varied choral excerpts.

However, there is a shortage of datasets specifically dedicated to authentic Chinese folk music, especially in polyphonic formats. Additionally, there is a lack of datasets for generating choral compositions featuring human voices singing Chinese folk songs.

Creating such a dataset holds significant value, as it:

Preserves Cultural Heritage: By curating an extensive collection of authentic Chinese folk music in a structured dataset, we effectively preserve and document our rich musical heritage.

Fosters Cultural Expression: Music is a profound expression of cultural identity. This dataset enables composers to infuse their works with the unique character of our cultural traditions. It contributes to the celebration and continued relevance of our cultural identity.

Empowers Creative Exploration: Composers gain access to a valuable resource for inspiration and exploration. They can draw from the extensive pool of Chinese folk melodies, making it easier to incorporate these elements into their compositions. This empowers contemporary composers to create music that echoes our cultural history.

In this paper, we introduce the Chinese Chorales Dataset, a high-quality dataset of Chinese chorales containing 125 choral songs from modern China, all the musical scores were produced by manual scoring, and were segmented into a total of 441 musical score fragments. The dataset is stored in two ways, one in MusicXML format, and the other is a .npz compressed file version containing pitch, fermata, beat, and chord information, which is divided into a training set, a validation set, and a test set.

We hope that our dataset can help in future choral song generation, especially in the following two tasks:

Table 1. Existing Datasets for score generation

Datasets	Dataset information				
	Size	Polyphony	Choral Music	Annotation	Music Type (Chinese/Western)
JSB Chorales Dataset [1]	380+	✓	✓	✗	Western
JS Fake Chorales Dataset [14]	500	✓	✓	✗	Western
POP909 [18]	909	✓	✗	✓	Chinese-Western mixture
Opencpop [17]	100	✗	✗	✓	Chinese
Lakh MIDI [15]	170k	✓	✗	✗	Chinese-Western mixture
Chinese Chorales Dataset	**441**	✓	✓	✓	**Chinese**

Task 1: Generate a complete Chinese vocal chorus based on a given soprano melody, to help Chinese composers create better quality chorus more easily.

Task 2: Help composers complete the task of melody generation, generate a whole melody with Chinese national style, and then create a complete song.

This paper has three main contributions:

1. We introduce the Chinese Chorales Dataset, a MusicXML-formatted dataset of Chinese chorales containing a total of 125 chorales divided into 441 musical score segments for storage.
2. We have done several experiments on the Chinese Chorales Dataset to demonstrate the validity of pitch, fermata, beat, and chord information in this dataset.
3. We have curated a Chinese choral music dataset manually transcribed, which addresses the current absence of a readily available Chinese folk song dataset suitable for generating vocal choral compositions. This dataset will facilitate easier creation of Chinese folk songs, contribute to the promotion of Chinese culture, and make Chinese music more accessible to musicians and audiences worldwide.

2   Related Work

In this section, we will discuss in Sect. 2.1 some existing musical score datasets, their respective characteristics, and whether they are able to fulfill the purpose of generating Chinese chorales that we wish to accomplish. After that, we will discuss in Sect. 2.2 what requirements our dataset should fulfil in order to accomplish Task 1 and Task 2, which we hope to do.

2.1   Existing Music Dataset

The currently available datasets for musical score music generation are summarized in Table 1. The first column in Table 1 shows the name of the dataset and the remaining columns indicate some other information about the dataset, such as the size of the dataset, its format, and whether it contains external information.

One of the popular datasets for chorales generation is JSB Chorales, which is available in several versions, with the common version containing a total of 380+ musical score fragments. Each musical score fragment contains four voice parts: soprano, alto, tenor, and bass. JSB fake Chorales [14] provides 500 musical score fragments generated by TonicNet [13]in MIDI format by using models with narrow expertise as a source of high-quality scalable synthesis data.

However, neither dataset provides external labels such as beats, chords, etc., and by providing external labels the controllability of the dataset can be effectively improved [2,16], as well as making it easier to evaluate the performance of the generated models. [9]

POP909 [18] provides a dataset of piano arrangements of multiple versions of 909 popular songs composed by professional musicians, which provides rhythmic, metric, key, and chord annotations for each song. POP909 has now also become one of the well-known datasets for polyphonic music generation, but the 'polyphony' it provides is vocal melody, lead melody, and piano accompaniment, rather than multi-part choral pieces.

Lakh MIDI is one of the most popular datasets in notation format, containing 176,581 songs in MIDI format from a wide range of genres, produced by a PhD at Columbia University [15]. Most of the songs in this dataset have multiple tracks, most of which are consistent with the original audio.

None of the datasets mentioned above are purely Chinese song datasets, and therefore, they do not meet our expectations for the task of creating high-quality Chinese choral songs.

In addition to the above datasets, Opencpop serves as a publicly available high-quality Mandarin singing corpus for singing voice synthesis (SVS). The corpus consists of 100 popular Chinese songs sung by a professional female singer. The audio files are recorded in studio quality at a sampling rate of 44,100 Hz, with corresponding lyrics and musical scores.

Although Opencpop is a pure Chinese pop song dataset, it is not a choral song dataset containing multiple voices.

Hence, in pursuit of our ultimate goal of automating the composition of Chinese chorales, it becomes imperative for us to develop a dedicated chorales dataset comprising exclusively authentic Chinese musical content.

The creation of such a dataset is a fundamental step towards enabling advanced AI and machine learning systems to not only understand the intricacies of traditional Chinese musical composition but also to foster innovation in this realm. By exclusively focusing on Chinese music, we can preserve the essence of Chinese musical heritage and pave the way for the automated generation of new Chinese chorales that respect and perpetuate these deep-rooted traditions.

Moreover, this specialized dataset serves as a crucial bridge between technology and culture, facilitating the seamless integration of traditional Chinese musical elements into contemporary music. It is a testament to the harmonious coexistence of technology and cultural heritage, empowering the next generation

of musicians and composers to create and share music that is rich in Chinese cultural identity on both local and global stages.

From Table 1, our dataset is comparable in terms of data volume to other existing datasets such as Opencpop, JSB Chorales Dataset, and JS Fake Chorales Dataset. However, our dataset stands out as it encompasses "polyphonic" vocal Chinese music scores and includes external annotations. These characteristics, unique to our dataset, enable it to undertake tasks that other datasets cannot accomplish, such as **Task 1**.

2.2   Requirements for the Choral Music Dataset

To fulfill our expectation of realizing a dataset of Chinese choral songs, we first identified the requirements that need to be met for the production of this dataset [12,16], The requirements are as follows.:

Four Voices: to provide valid information for the generation of chorales in Task 1.

Chinese Music: to meet our expectation of creating Chinese chorales.

External Annotation: to ensure the controllability of the generation process and the evaluability of the generation results. [2,16]

3   Dataset Description and External Annotation Methods

Chinese Chorales Dataset is a musical score dataset consisting entirely of Chinese chorales, providing choral musical scores for a total of 125 songs. In the following Sect. 3.1, we will talk about the production process of the dataset in detail, and in Sect. 3.2, we will describe the format of the dataset and the external annotations we give.

3.1   Dataset Production Process

In order to collect a dataset of Chinese songs with contemporary features and ethnic flavors, we have hired professional composers to re-arrange Chinese songs obtained from the Internet into four-part musical scores. These songs are primarily from the last century to the present and revolve around patriotic themes. We will release the list of the song titles list along with the dataset.

We followed the following process to collect and process the dataset:

1. Initially, we selected songs that met our criteria based on their time period and style, ensuring that these songs exhibited relatively consistent historical characteristics and musical styles. We conducted searches on the internet to find and download the sheet music for these songs.
2. After the preliminary selection of musical compositions, we engaged professional composers to re-arrange the chosen songs, creating four-part musical scores in MusicXML format. These scores were stored on local servers and backed up in the cloud.

Table 2. Examples of chord information

Chord	Interval		Root	Duration
Song1	[0, 3, 7]	[0, 3, 7]	C4 C4	1 1

Table 3. Examples of pitch information

Pitch	Soprano, Alto, Tenor, Bass
Song1	[63, 58, 70, 67] [63, 58, 72, 67] [63, 58, 70, 67]
Song2	[76, 67, 71, 0] [76, 67, 74, 0] . . .

3. Furthermore, we assigned a group of individuals to review the composed musical scores. They meticulously assessed and compared each song to ensure that they retained the essence and unique characteristics of the original songs. They also coordinated the styles of various choral pieces, ensuring that all songs in the dataset met our requirements and goals.

3.2 Dataset Format

We saved our dataset in the MusicXML data format, and we used the Music21 library for the processing and saving of datasets in the MusicXML format. [3]

In addition to the musical score fragments saved in MusicXML format, we also provide .npz compressed files containing pitch information, fermata, beat information, and chord information. Each of these files is a dictionary containing the training set, validation set, and test set, where the corresponding value for each song is a list of sequences.

The example pitch sequences are given in Table 2, from which we can get a rough idea of how the pitch sequences are stored in the Chinese Chorales Dataset.

The number of musical score segments contained in the training, validation, and test sets are 309, 88, and 44, respectively. In a file containing pitch information, each sequence is itself a list of time steps, and at each time step, there are four numbers corresponding to one pitch for each voice.

The pitches are encoded as 0–129, which contains 128 pitches, a rest, and a hold, with the time resolution set to 16th notes. In the file containing the fermata information, the fermata is represented by 0, 1, and the beat is encoded as 0–3, which represents no beat, weak beat, medium-heavy beat, and strong beat, respectively. [19]
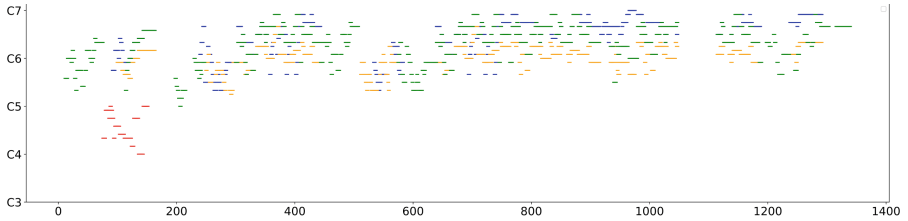
(a) Example a
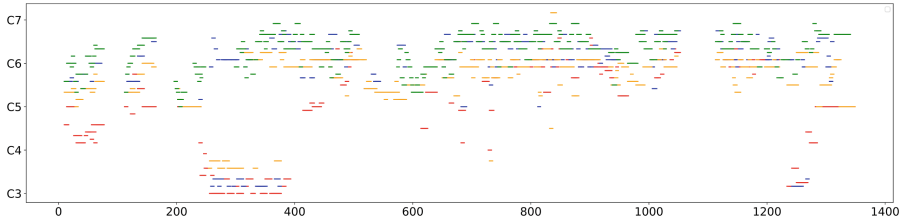
(b) Generation results of example a

Fig. 1. Example generated using the Chinese chorus dataset. In this example, the green part is the main melody, the orange part is the alto, the blue part is the tenor, and the red part is the bass.The generative model uses Deepchoir (full) with sixteenth note resolution. (Color figure online)

As for the information in the chord section, the previous practice tends to encode some chords that are commonly used; the authors of TonicNet encoded chords into fifty categories, namely twelve major chords (one chord for each pitch category in the Western chromatic scale), twelve minor chords, twelve diminished chords, twelve augmented chords and special notation [13], while the authors of DeepChoir coded into twelve categories. [19]

Considering the large number of chord categories contained in this dataset, and to represent the various categories as completely as possible, so that the users of this dataset can encode the chord categories by themselves according to their own needs, this paper adopts the representation of intervals, roots and durations. The duration is expressed in quarter notes, and a duration of 1 means that the chord lasts one-quarter note.

As shown in Table 3, this approach not only provides a complete representation of each chord but also adds the durations of different chords to the model, which provides more complete and rich information. The four columns in Table 3 show the corresponding song, the chord's interval, the chord's duration, and the chord's root note (the lowest note).

(a) Original Musical score Fragment.

(b) Musical score fragment generated with DeepChoir(full)

Fig. 2. Comparison of the musical score fragment generated using DeepChoir(full) with the original fragment

4   Experiments

In this section, we conduct a baseline experiment on music generation using the Chinese Chorales Dataset: choral song generation conditional on melody. We conducted the experiment using several models, including DeepChoir [19], BachBot [11], and DeepBach [6]. These three models have shown very good results on the JSB Chorales Dataset and are also well-known models under this task. Therefore, we decided to test our dataset on these models.

4.1   Model and Parameter Settings

To ensure that the various information in our dataset is reliable, we modified the publicly available code in [19] and conducted several experiments.

We first cut the two-channel structure in the DeepChoir model, leaving only the branch used to input pitch information. We tested the learnability of the pitch information of the dataset by inputting only pitch information to the model in this way, and the experiments ultimately determined that the model was able
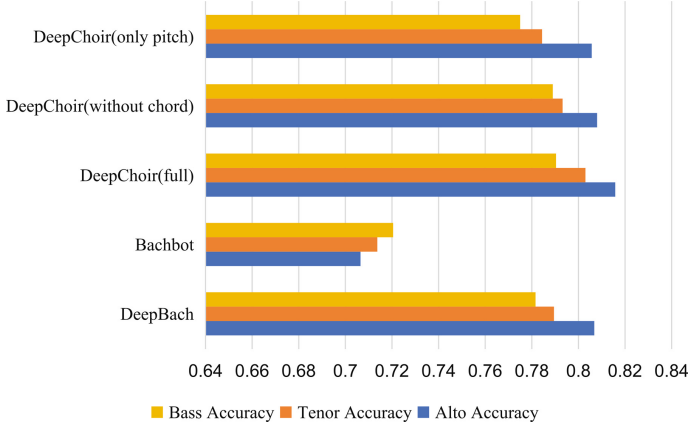
Fig. 3. Comparison of the generation accuracy of each model for the three voices

to learn valid information from the pitch information alone, thus generating chorales with a high degree of accuracy that could be convincing.

Secondly, we tried the effect of adding beat and intensity information to pitch information, and chord information, respectively. It is worth mentioning that during these multiple experiments, we removed the gamma sampling algorithm from our code, as it was found that post-processing methods such as the gamma sampling algorithm mentioned in [8] would have a large impact on the generation of our dataset.

In contrast, we did not make any changes in the DeepBach and Bach-Bot models. All the experiments we conducted were trained with the same parameters.

4.2   Results

Figure 1 shows an example of DeepChoir(full) generation, where the top piano roll is a fragment of the score from the dataset, and the bottom piano roll is the generated score, with the main theme in green, alto in orange, tenor in blue, and bass in red.

The musical score fragment generated with DeepChoir(full) is shown in Fig. 2 along with the original musical score fragment. As can be seen in Fig. 1 and 2, the generated scores capture the basic harmonic relationship between melody and accompaniment and contain consistent rhythmic patterns. While the quality leaves something to be desired, it serves as a baseline to illustrate the usage of our dataset.

Figure 3 shows the generation results of the above three models, and it can be seen that all three models can achieve a high correct rate on the Chinese Chorales Dataset. Meanwhile, on the DeepChoir model, the order of the correct rates of the three voices is also in line with our expectation, i.e., the model containing only pitch information has the lowest correct rate, the one with fermata information
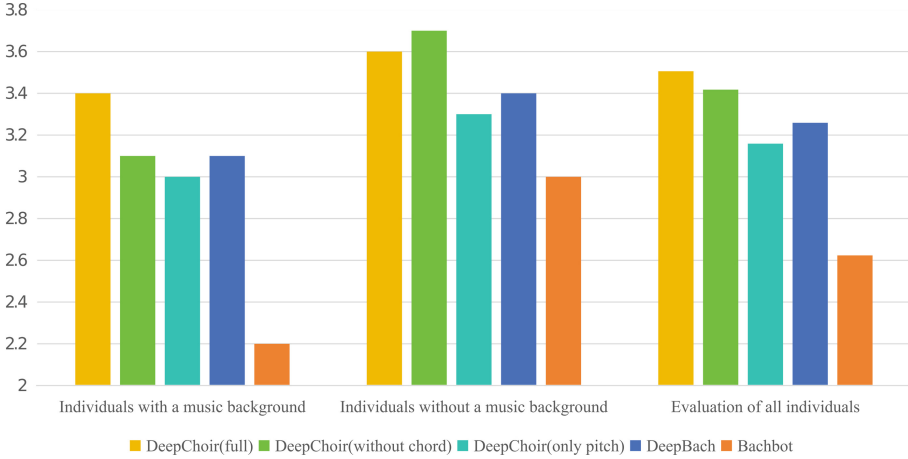
Fig. 4. Subjective evaluation scores of each model

and beat information has a slightly higher correct rate, and the one with chord information has an even higher rate.

In addition to objective evaluation criteria, we also conducted subjective assessments. We invited 34 students to provide feedback on our generated results, with 16 of them possessing a certain level of music background, while the remaining 18 students evaluated the results without any prior musical knowledge. All evaluators assigned scores to the musical scores on a scale of 1–5 based on the perceived quality of the songs.

As shown in Fig. 4, although the evaluations from individuals without a music background deviated slightly from our expectations, the final evaluation results were generally consistent with our expectations. Specifically, the Deepchoir (full) model, which incorporates all four types of information as input, produced the best results. This demonstrates that when all four specified types of information are provided as input, the model can more effectively learn the compositional patterns from the main melody to the various vocal parts, resulting in the creation of chorale compositions that are generally well-received by the audience.

Through the analysis of multiple experiments employing both subjective and objective evaluation criteria, we conclude that the four pieces of information in our dataset are effective and valuable for assisting the model in learning the patterns of choral composition.

5   Conclusion

In this paper, we have produced the Chinese Chorales Dataset, a dataset of Chinese chorales. It contains 125 Chinese chorales, which are divided into 441 parts and stored in MusicXML format. We also provide a compressed file version containing pitch, fermata, beat, and chord information. In order to ensure

the quality of the data, we used manual scoring to create the dataset, and we designed several experiments on the dataset to verify the validity of the information in the dataset. We hope that this dataset can help in the future task of composing Chinese choral music.

References

1. Boulanger-Lewandowski, N., Bengio, Y., Vincent, P.: Modeling temporal dependencies in high-dimensional sequences: application to polyphonic music generation and transcription. arXiv preprint arXiv:1206.6392 (2012)
2. Chen, K., Zhang, W., Dubnov, S., Xia, G., Li, W.: The effect of explicit structure encoding of deep neural networks for symbolic music generation. In: 2019 International Workshop on Multilayer Music Representation and Processing (MMRP), pp. 77–84. IEEE (2019)
3. Cuthbert, M.S., Ariza, C.T.: music21: a toolkit for computer-aided musicology and symbolic music data. In: Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands, 9–13 August 2010. DBLP (2010)
4. Elowsson, A., Friberg, A.: Algorithmic composition of popular music. In: The 12th International Conference on Music Perception and Cognition and The 8th Triennial Conference of the European Society for The Cognitive Sciences of Music, pp. 276–285 (2012)
5. Gardner, J., Simon, I., Manilow, E., Hawthorne, C., Engel, J.: Mt3: multi-task multitrack music transcription. arXiv preprint arXiv:2111.03017 (2021)
6. Hadjeres, G., Pachet, F., Nielsen, F.: DeepBach: a steerable model for Bach chorales generation. In: International Conference on Machine Learning, pp. 1362–1371. PMLR (2017)
7. Hernandez-Olivan, C., Beltran, J.R.: Music composition with deep learning: a review. In: Advances in Speech and Music Technology: Computational Aspects and Applications, pp. 25–50 (2022)
8. Hernandez-Olivan, C., Puyuelo, J.A., Beltran, J.R.: Subjective evaluation of deep learning models for symbolic music composition. arXiv preprint arXiv:2203.14641 (2022)
9. Hernandez-Olivan, C., Zay Pinilla, I., Hernandez-Lopez, C., Beltran, J.R.: A comparison of deep learning methods for timbre analysis in polyphonic automatic music transcription. Electronics **10**(7), 810 (2021)
10. Ji, S., Luo, J., Yang, X.: A comprehensive survey on deep music generation: multi-level representations, algorithms, evaluations, and future directions. arXiv preprint arXiv:2011.06801 (2020)
11. Liang, F.T., Gotham, M., Johnson, M., Shotton, J.: Automatic stylistic composition of Bach chorales with deep LSTM. In: ISMIR, pp. 449–456 (2017)
12. Manilow, E., Wichern, G., Seetharaman, P., Le Roux, J.: Cutting music source separation some Slakh: a dataset to study the impact of training data quality and quantity. In: 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 45–49. IEEE (2019)
13. Peracha, O.: Improving polyphonic music models with feature-rich encoding. arXiv preprint arXiv:1911.11775 (2019)
14. Peracha, O.: JS fake chorales: a synthetic dataset of polyphonic music with human annotation. arXiv preprint arXiv:2107.10388 (2021)

15. Raffel, C.: Learning-based methods for comparing sequences, with applications to audio-to-MIDI alignment and matching. Doctoral dissertation (2016)
16. Su, L.: Attend to chords: improving harmonic analysis of symbolic music using transformer-based models (2021)
17. Wang, Y., et al.: Opencpop: a high-quality open source Chinese popular song corpus for singing voice synthesis. arXiv preprint arXiv:2201.07429 (2022)
18. Wang, Z., et al.: POP909: a pop-song dataset for music arrangement generation. arXiv preprint arXiv:2008.07142 (2020)
19. Wu, S., Li, X., Sun, M.: Chord-conditioned melody harmonization with controllable harmonicity. In: ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5. IEEE (2023)
20. Zhou, J., Zhu, H., Wang, X.: Choir transformer: generating polyphonic music with relative attention on transformer. arXiv preprint arXiv:2308.02531 (2023)