

# BianQue: Balancing the Questioning and Suggestion Ability of Health LLMs with Multi-turn Health Conversations Polished by ChatGPT

Yirong Chen<sup>1</sup>, Zhenyu Wang<sup>1</sup>, Huimin Zheng<sup>1</sup>, Xiaofen Xing<sup>1</sup>, Zhipei Xu<sup>1</sup>, Kai Fang<sup>1</sup>,  
Sihang Li<sup>1</sup>, Junhong Wang<sup>1</sup>, Qi Liu<sup>2</sup>, Xiangmin Xu<sup>2,3\*</sup>

<sup>1</sup>School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China

<sup>2</sup>School of Future Technology, South China University of Technology, Guangzhou, China

<sup>3</sup>Pazhou Lab, Guangzhou, China

{eeyirongchen, 202030242072, 202210182130}@mail.scut.edu.cn, xfxing@scut.edu.cn,  
{eexzp, 202030240160, 202030241044, eewjh}@mail.scut.edu.cn, {drliuqi, xmxu}@scut.edu.cn

## Abstract

Large language models (LLMs) have performed well in providing general and extensive health suggestions in single-turn conversations, exemplified by systems such as ChatGPT, ChatGLM, ChatDoctor, DoctorGLM, and etc. However, the limited information provided by users during single turn results in inadequate personalization and targeting of the generated suggestions, which requires users to independently select the useful part. It is mainly caused by the missing ability to engage in multi-turn questioning. In real-world medical consultations, doctors usually employ a series of iterative inquiries to comprehend the patient’s condition thoroughly, enabling them to provide effective and personalized suggestions subsequently, which can be defined as chain of questioning (CoQ) for LLMs. To improve the CoQ of LLMs, we propose BianQue, a ChatGLM-based LLM finetuned with the self-constructed health conversation dataset BianQueCorpus that is consist of multiple turns of questioning and health suggestions polished by ChatGPT. Experimental results demonstrate that the proposed BianQue can simultaneously balance the capabilities of both questioning and health suggestions, which will help promote the research and application of LLMs in the field of proactive health.<sup>1</sup>

users are amazed by the excellent suggestion ability of LLMs.

However, LLMs are deficient in “questioning” which is an important way to proactively understand users needs in medical, psychological, educational and other application scenarios. When we engage in healthcare conversations with these LLMs (ChatGPT<sup>2</sup>, ChatGLM<sup>3</sup>, SparkDesk<sup>4</sup>), they do not yet possess the ability to conduct multiple rounds of questioning, as presented in Appendix B. The above LLMs generally provide reasonable and universal suggestions based on the single-turn instruction provided by users. However, in the real world, doctors often need to conduct multiple turns of questioning with patients in order to provide targeted advice, as shown in Figure 1. During the user consultation, the doctor raised different questions in the first 9 turns of conversations to understand the specific situation of the baby. The above multi-turn questioning process can be defined as **Chain of Questioning (CoQ)**. We found that the current LLMs lack CoQ capabilities because LLMs lack training data for multiple rounds of questioning during the instruction fine-tuning stage and RLHF stage. When researchers construct instructions and answers, on the one hand, they ignore multiple rounds of conversation history, and on the other hand, answers are usually suggestions rather than questions.

## 1 Introduction

Recently, Large language models (LLMs), e.g. ChatGPT (OpenAI, 2022), LLaMA (Touvron et al., 2023), ChatGLM (Zeng et al., 2023), have been extensively applied in various fields. Through high-quality instruction fine-tuning and reinforcement learning based on human feedback (RLHF) (Ouyang et al., 2022), LLMs already have possessed stunning language comprehension, generation, and knowledge reasoning abilities. Overall,

At present, research on LLMs in the health field mainly focuses on evaluating the performance of existing models, constructing suitable datasets, and fine-tuning instructions. Singhal et al. (2022) proposed a medical Q&A benchmark MultiMedQA for evaluating the clinical knowledge QA abilities of LLMs. Li et al. (2023) constructed a real doctor-patient dialogue dataset HealthCareMagic-100k, and used it to fine-tune the ChatDoctor based on

\*Corresponding author. Email: xmxu@scut.edu.cn

<sup>1</sup><https://github.com/scutcyr/BianQue>

<sup>2</sup><https://chat.openai.com>

<sup>3</sup><https://chatglm.cn>

<sup>4</sup><https://xinghuo.xfyun.cn>

病人：孩子九个月，感冒引起的支气管炎和肺炎，现在住院输了两天的液，没有一点效果，现在气喘加重，咳嗽，和鼻塞也没减轻，请问我现在应该采取什么样的治疗？

医生：宝宝咳嗽几天了？

Start of CoQ

病人：一个礼拜了。

医生：宝宝咳嗽是一声声还是一顿顿的？

病人：一顿顿的。

医生：宝宝咳嗽是白天严重还是夜间严重？

病人：一早一晚。

医生：咳嗽痰多吗？

病人：嗯。

医生：有没有发热？

病人：刚开始就是因为发烧，现在没有了。

医生：有没有给宝宝行胸片检查和化验检查？

病人：血项高，拍片是支气管炎和肺炎。

医生：现在给宝宝应用什么药物？具体药名知道吗？

病人：头孢。

医生：还有吗？

End of CoQ

病人：平喘的，和化痰的。

医生：有没有雾化治疗？

病人：有，一天两次，今天加到四次了。

医生：可以，一般肺炎疗程10~14天

病人：还要继续输液吗？

医生：治疗早期3天咳嗽有可能加重，药物完全起效，达到平衡的血药浓度需要2~3天。继续输液巩固治疗，加强雾化治疗。注意多喝水，勤拍背。

病人：现在继续吗？

医生：继续输液巩固治疗，加强雾化治疗。治疗7~10天复查胸片检查。

Patient: My child is nine months old and has suffered from bronchitis and pneumonia caused by a cold. He has been hospitalized for two days and have received fluid transfusions without any effect. His asthma has worsened, and cough and nasal congestion have not alleviated. May I ask what treatment he should take now?

Doctor: How many days has the baby been coughing?

Patient: It's been a week now.

Doctor: Is the baby coughing loudly or repeatedly?

Patient: Repeatedly.

Doctor: Is the baby cough serious during the day or at night?

Patient: Morning and night.

Doctor: Does the baby have much phlegm when coughing?

Patient: Yes, the baby has much phlegm when coughing.

Doctor: Does the baby have fever?

Patient: It started with a fever, but now it's gone.

Doctor: Have you given your baby chest X-ray examination and laboratory examination?

Patient: The blood items are high, and the results of radiography are bronchitis and pneumonia.

Doctor: What medicine does the baby currently take? Do you know the specific drug name?

Patient: Cephalosporin.

Doctor: Anything else?

Patient: Antiasthmatic drugs and expectorant drugs.

Doctor: Is there any atomization treatment?

Patient: Yes, twice a day. Today it has increased to four times.

Doctor: Okay, the general course of treatment for pneumonia is 10-14 days

Patient: Does the baby still need to continue infusion?

Doctor: Cough may worsen in the early 3 days of treatment, and the drug takes effect completely. It takes 2-3 days to reach a balanced blood concentration. Continue infusion consolidation treatment and strengthen atomization treatment. Pay attention to having your baby drink plenty of water and pat his back frequently.

Patient: Continue now?

Doctor: Continue infusion consolidation treatment and strengthen atomization treatment. The chest X-ray examination should be conducted again after 7-10 days of treatment.

Figure 1: Example of chain of questioning (CoQ). The sentence in red font presents the doctor's CoQ: a series of questions about cough time, cough sound, sputum status, fever status, examination, medication and treatment.

LLaMA. Similar health LLMs have been released one after another, e.g. BenTsao (本草) (Wang et al., 2023b), ChatGLM-6B-Med (Wang et al., 2023b), DoctorGLM (Xiong et al., 2023), MedAlpaca (Han et al., 2023), ClinicalGPT (Wang et al., 2023a) and etc. These models are basically based on the assumption that "users can clearly describe their problems or situations". Therefore, during the model construction phase, **the questioning ability of the model was not considered**. Although these models have achieved well performance in the field of medical QA, they do not have the ability to ask users questions.

To enhance the questioning ability of LLMs, we constructed a multi-turn health conversation dataset named *BianQueCorpus*, in which the targets consist of balanced proportional questions (46.2%) and suggestions (53.8%), as shown in Figure 2. Mean-

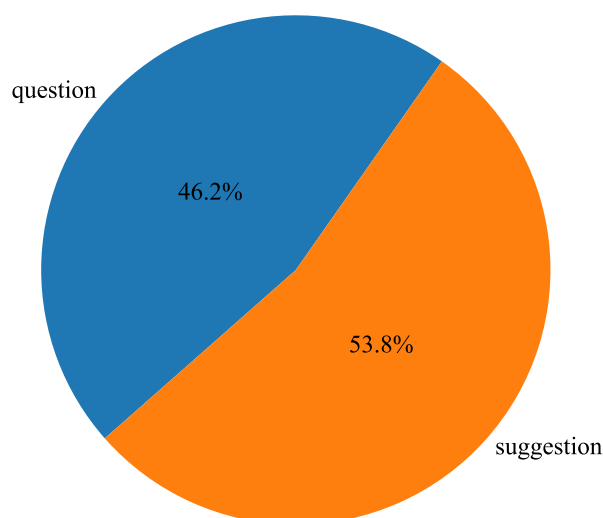


Figure 2: Proportion of questions and suggestions in answers of *BianQueCorpus*.

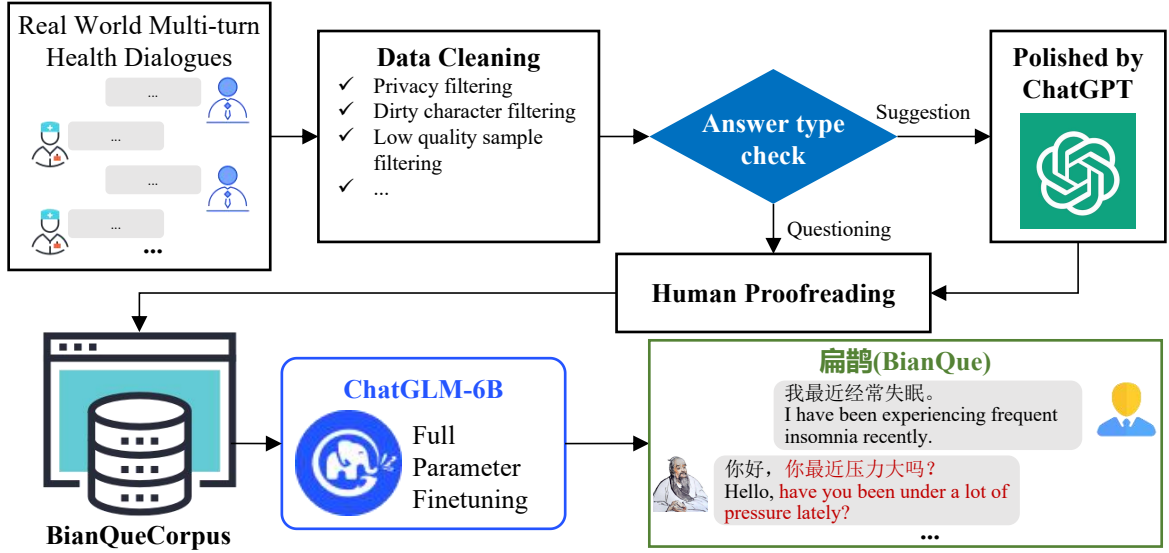


Figure 3: Construction process of *BianQueCorpus* dataset and *BianQue* model.

while, we present *BianQue*, a health LLM that is specifically designed for balancing the questioning and suggestion ability. The results on multi-turn health conversation dataset demonstrate that *BianQue* outperforms existing models and ChatGPT, especially in the ability to conduct chain of questioning (CoQ).

## 2 Methodology

### 2.1 *BianQueCorpus*: Balancing Questioning and Suggestion

In the field of Chinese health conversational AI, there are already some multi-turn conversation datasets, e.g. MedDialog-CN (He et al., 2020), IMCS-V2 (Chen et al., 2022), CHIP-MDCFNPC (Zhang et al., 2022), MedDG (Zhang et al., 2022). However, these conversations are often crawled from internet consultation platforms, such as 好大夫<sup>5</sup>. These datasets are often mixed with a large amount of noise, such as missing content, missing images, reward information, privacy content, incomplete JSON content, website link, website tips, voice recording, text automatically replied by the system, etc. We first collected real-world multi-turn health conversations through data outsourcing services. Then, we performed a two-stage data optimization process: (i) We constructed a data automatic cleaning process based on regularized expression to improve the quality of existing conversation datasets. (ii) We designed a polishing prompt (see Figure 4) and use ChatGPT to polish

<sup>5</sup><https://www.haodf.com/>

the doctors’ suggestion of multi-turn conversations, because doctors often respond very briefly through internet platforms, lacking detailed analysis and suggestions. The whole construction process of *BianQueCorpus* is presented in Figure 3. We ultimately obtained a multi-turn health conversation dataset consisting of 2,437,190 samples, in which the questions accounted for 46.2% among the doctors’ answers.

```
"messages": [
  {"role": "user",
   "content": "I want you to act as a doctor. Given the Chinese conversation context between doctor and patient, you need to provide empathetic, professional, and reasonable Chinese answer. \n\nMulti-turn Chinese dialog: ""+ multi_context +""\n\nPlease analyze the dialog and provide your answer.\nYour Chinese answer is: ""
  ]}
```

Figure 4: The prompt used for polishing the suggestions of doctors based on real-world multi-turn conversation context.

### 2.2 *BianQue* Model

We chose the ChatGLM-6B (Du et al., 2022; Zeng et al., 2023) as the base LLM architecture to construct the *BianQue*, since it is open source and has excellent Chinese understanding and generation performance. The input of model is defined as:

$$input = u_1^u + ' \setminus n' + u_1^p + \dots + u_N^u + ' \setminus n' + u_N^p$$

where  $u_i^u$  = ‘病人: ’ + utterance $_i^u$ ,  $u_i^p$  = ‘医生: ’ + utterance $_i^p$  ( $i < N$ ),  $u_N^p$  = ‘医生: ’,  $N$  is the number of dialogue turn.

Table 1: Evaluation results.

Dataset	Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	R-1	R-2	R-L	PQA
MedDialog-CN	ChatGLM-6B	7.28	3.72	2.10	1.23	10.86	0.92	7.43	0.20
	DoctorGLM	10.39	5.06	2.94	1.80	13.27	1.04	11.17	0.01
	ChatGPT	7.61	3.90	2.21	1.30	11.11	0.96	7.82	0.28
	BianQue	<b>11.12</b>	<b>6.50</b>	<b>4.42</b>	<b>3.10</b>	<b>15.55</b>	<b>2.15</b>	<b>12.96</b>	<b>0.53</b>
IMCS-V2	ChatGLM-6B	6.83	3.61	2.12	1.30	10.24	1.03	7.26	0.36
	DoctorGLM	8.38	4.22	2.52	1.55	11.87	0.95	9.22	0.06
	ChatGPT	8.46	4.54	2.71	1.70	11.48	1.29	8.97	0.38
	BianQue	<b>14.50</b>	<b>10.16</b>	<b>7.85</b>	<b>6.23</b>	<b>21.73</b>	<b>6.24</b>	<b>19.09</b>	<b>0.70</b>
CHIP-MDCFNPC	ChatGLM-6B	6.22	3.11	1.81	1.10	9.62	0.85	0.67	0.35
	DoctorGLM	8.59	4.33	2.68	1.71	12.05	1.11	9.68	0.05
	ChatGPT	7.52	3.74	2.20	1.36	10.51	0.97	8.03	0.38
	BianQue	<b>13.41</b>	<b>8.49</b>	<b>6.05</b>	<b>4.42</b>	<b>19.00</b>	<b>3.99</b>	<b>16.56</b>	<b>0.57</b>
MedDG	ChatGLM-6B	4.76	2.31	1.34	0.81	7.35	0.56	5.06	0.47
	DoctorGLM	6.87	3.47	2.15	1.35	9.62	0.88	7.61	0.09
	ChatGPT	5.11	2.41	1.38	0.83	7.58	0.50	5.46	0.63
	BianQue	<b>14.86</b>	<b>10.43</b>	<b>8.09</b>	<b>6.37</b>	<b>21.56</b>	<b>6.46</b>	<b>19.56</b>	<b>0.81</b>

### 3 Experiments

#### 3.1 Baselines and Benchmarks

We select ChatGLM-6B<sup>6</sup> (Zeng et al., 2023), ChatGPT (gpt-3.5-turbo) (OpenAI, 2022), and DoctorGLM (Xiong et al., 2023) as the baseline models. Comparative experiments were conducted on test set of MedDialog-CN, IMCS-V2, CHIP-MDCFNPC and MedDG respectively, since they are multi-turn conversation datasets that have both suggestions and questions in targets.

#### 3.2 Implementation details

BianQue is finetuned on the proposed *BianQueCorpus* using the *WarmupDecayLR* learning rate scheduler with *warmup\_steps* = 1000 and *warmup\_max\_lr* =  $5e - 5$ . During the training stage, the maximum input length is set to 1,536, while the maximum target length is set to 512. A batch size of 80 and global training steps of 25,000 are applied. The decoding algorithms of Top-p sampling with  $p = 0.75$  and temperature  $\tau = 0.95$  is applied in the inference stage.

#### 3.3 Results and Analysis

Following the ClinicalGPT (Wang et al., 2023a), We evaluated BianQue and other models with the metrics: BLEU-1/2/3/4 (Papineni et al., 2002) and

ROUGE-1/2/L (Lin, 2004). In addition, we define a new metric to measure the model’s Proactive Questioning ability (PQA):

$$PQA = \frac{2P_q R_q}{P_q + R_q},$$

$$P_q = \frac{Q_{tp}}{Q_{tp} + Q_{t\bar{p}}}, R_q = \frac{Q_{tp}}{Q_{tp} + Q_{\bar{t}p}},$$

where  $Q_{tp}$  is the number of samples with both target and prediction are question,  $Q_{t\bar{p}}$  is the number of samples with question target and suggestion prediction,  $Q_{\bar{t}p}$  is the number of samples with both target and prediction are suggestion.

As shown in Table 1, BianQue demonstrates considerable performance on MedDialog-CN, IMCS-V2, CHIP-MDCFNPC and MedDG, achieving better scores than other models across all metrics.

### 4 Conclusion and Future Work

In this study, we introduced BianQue, a health LLM with balanced questioning and suggestion ability, which is finetuned based on the proposed large-scale multi-turn health conversation dataset *BianQueCorpus*, in which the targets consist of balanced proportional questions (46.2%) and suggestions (53.8%). The empirical results highlight the superior multi-turn questioning ability. Future work requires further focus on the conversion mechanism between questioning and suggestion.

<sup>6</sup><https://github.com/THUDM/ChatGLM-6B>



## Limitations

It must be emphasized that there are potential risks when using generative language models for health conversation. Doctors in the real world are rigorous in diagnosing diseases and providing medication guidance. However, the current state-of-the-art LLMs (e.g. ChatGPT) still cannot guarantee the accuracy of the text they generate. Therefore, it is necessary to set up inspection and error correction mechanisms for the health suggestions generated by LLMs. At the same time, when LLMs learn the ability to proactively question, their usage risk also increases, as the models may ask users some questions related to privacy. For example, when users consult AI about cold-related issues, AI may proactively inquire about their age, gender, and other privacy information. Further privacy protection mechanisms need to be considered in the research and application of LLMs. Overall, the methods proposed in this article are still in the early research stage, and the questioning and suggestion mechanisms are not clear enough. The proposed model is limited to academic research and cannot be used in real-world deployment.

## Ethics Statement

The Bianque model is committed to improving the proactive questioning ability of LLMs, rather than providing very professional medical diagnosis or advice. The multi-turn conversation dataset used in this study is mainly based on the real world doctor-patient conversations, which has gone through a strict data cleansing process to eliminate private information and dirty text content. To this end, we constructed 50 regular expressions and used the *re* package for filtering. We compared the data quality before and after data cleansing, and the excellent rate increased from 82% to 93%. Due to the lack of human feedback during the model finetuning stage, the current version of the model may involve user privacy when asking questions, which is particularly important to note. On the other hand, the health recommendations generated by the model have not undergone rigorous examination and proofreading, and therefore cannot be used as a substitute for real-world doctors. We emphasize that this is an early research-oriented model, rather than a mature and directly applicable model. Therefore, future work needs to combine RLHF to improve the safety level of model generated questions or suggestions. Besides, when

Bianque is applied to downstream scenarios, it is necessary to inform the users in advance that the answers they see are generated by health AI, which are for reference only.

## Acknowledgements

This work was supported by the Science and Technology Project of Guangzhou (202103010002), the Natural Science Foundation of Guangdong Province (2022A1515011588), the National Key R&D Program of China (2022YFB4500600), the Science and Technology Project of Guangdong (2022B0101010003), the National Natural Science Foundation of China under Grant U1801262 and Guangdong Provincial Key Laboratory of Human Digital Twin (2022B1212010004).

## References

- Wei Chen, Zhiwei Li, Hongyi Fang, Qianyu Yao, Cheng Zhong, Jianye Hao, Qi Zhang, Xuanjing Huang, Jiajie Peng, and Zhongyu Wei. 2022. [A benchmark for automatic medical consultation system: frameworks, tasks and datasets](#). *Bioinformatics*, 39(1). Btac817.
- Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. [GLM: General language model pretraining with autoregressive blank infilling](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 320–335, Dublin, Ireland. Association for Computational Linguistics.
- Tianyu Han, Lisa C. Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K. Bresssem. 2023. [Medalpaca – an open-source collection of medical conversational ai models and training data](#).
- Xuehai He, Shu Chen, Zeqian Ju, Xiangyu Dong, Hongchao Fang, Sicheng Wang, Yue Yang, Jiaqi Zeng, Ruisi Zhang, Ruoyu Zhang, Meng Zhou, Penghui Zhu, and Pengtao Xie. 2020. [Meddialog: Two large-scale medical dialogue datasets](#).
- Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, and You Zhang. 2023. [Chatdoctor: A medical chat model fine-tuned on llama model using medical domain knowledge](#).
- Chin-Yew Lin. 2004. [ROUGE: A package for automatic evaluation of summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- OpenAI. 2022. [Introducing chatgpt](#).

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. [Training language models to follow instructions with human feedback](#). In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Nathaneal Scharli, Aakanksha Chowdhery, Philip Mansfield, Blaise Aguerre y Arcas, Dale Webster, Greg S. Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. 2022. [Large language models encode clinical knowledge](#).

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. [Llama: Open and efficient foundation language models](#).

Guangyu Wang, Guoxing Yang, Zongxin Du, Longjun Fan, and Xiaohu Li. 2023a. [Clinicalgpt: Large language models finetuned with diverse medical data and comprehensive evaluation](#).

Haochun Wang, Chi Liu, Nuwa Xi, Zewen Qiang, Sendong Zhao, Bing Qin, and Ting Liu. 2023b. [Huatu: Tuning llama model with chinese medical knowledge](#).

Honglin Xiong, Sheng Wang, Yitao Zhu, Zihao Zhao, Yuxiao Liu, Linlin Huang, Qian Wang, and Dinggang Shen. 2023. [Doctorglm: Fine-tuning your chinese doctor is not a herculean task](#).

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, Weng Lam Tam, Zixuan Ma, Yufei Xue, Jidong Zhai, Wenguang Chen, Zhiyuan Liu, Peng Zhang, Yuxiao Dong, and Jie Tang. 2023. [GLM-130b: An open bilingual pre-trained model](#). In *The Eleventh International Conference on Learning Representations (ICLR)*.

Ningyu Zhang, Mosha Chen, Zhen Bi, Xiaozhuan Liang, Lei Li, Xin Shang, Kangping Yin, Chuanqi Tan, Jian Xu, Fei Huang, Luo Si, Yuan Ni, Guotong Xie, Zhifang Sui, Baobao Chang, Hui Zong, Zheng Yuan,

Linfeng Li, Jun Yan, Hongying Zan, Kunli Zhang, Buzhou Tang, and Qingcai Chen. 2022. [CBLUE: A Chinese biomedical language understanding evaluation benchmark](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7888–7915, Dublin, Ireland. Association for Computational Linguistics.

## A Reproducibility Checklist

- **Model and Data:** The BianQue model and BianQueCorpus will be released upon decision of the paper.
- **System Hardware:** BianQue is trained on a Ubuntu 20.04.6 LTS server that has 2 CPUs called "Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz", 8 NVIDIA A800-SXM4-80GB GPUs, and 1,024GB memory.
- **Driver Version:** The version of Nvidia driver is "525.105.17". CUDA=11.6, and Cudnn=8.4.0.27.
- **Package version:** python=3.8.16, torch<sup>7</sup>=1.13.1+cu116, transformers<sup>8</sup>=4.28.0, deepspeed<sup>9</sup>=0.9.3, datasets=2.11.0 and jieba=0.42.1 is recommended. Other dependent packages and versions will be released in our open source repository.
- **Model Parameters:** BianQue has 6.2B parameters with 28 layers and *max\_sequence\_length* of 2,048. During the inference phase, the model requires at least 14GB of GPU memory.
- **Training Time:** BianQue is trained with global steps of 25,000 and *torch\_dtype* of "float16" on 8 NVIDIA A800-SXM4-80GB GPUs. The training time is about 66 hours.

## B Sample Conversations of LLMs

The following are examples of health conversation testing in ChatGPT (Figure 5), ChatGLM (Figure 6), and SparkDesk (Figure 7). These are three common Chinese LLMs, but none of them have CoQ capabilities. The above LLMs generally provide reasonable and universal suggestions based on the single-turn instruction provided by users.

<sup>7</sup><https://pytorch.org/get-started/previous-versions>

<sup>8</sup><https://github.com/huggingface/transformers>

<sup>9</sup><https://github.com/microsoft/DeepSpeed>

However, in the real world, doctors often need to conduct multiple turns of questioning with patients in order to provide targeted advice.

## **C Sample Conversations of BianQue**

Figure 8 shows an example of the BianQue model proactively asking questions.



孩子九个月，感冒引起的支气管炎和肺炎，现在住院输了两天液，没有一点效果，现在气喘加重，咳嗽，和鼻塞也没减轻，请问我现在应该采取什么样的治疗？

My child is nine months old and has suffered from bronchitis and pneumonia caused by a cold. He has been hospitalized for two days and have received fluid transfusions without any effect. His asthma has worsened, and cough and nasal congestion have not alleviated. May I ask what treatment he should take now?



我不是医生，但我可以给你一些建议。对于九个月大的孩子，如果感冒引起了支气管炎和肺炎，你应该尽快寻求医疗帮助。住院治疗是一个正确的决定，因为医生可以监测孩子的病情并提供适当的治疗。

在住院期间，医生可能会采取以下治疗措施：

1. 继续输液：输液可以保持孩子的水分和电解质平衡，并帮助孩子恢复体力。
2. 给予抗生素：如果肺炎是由细菌感染引起的，医生可能会给孩子开具抗生素来对抗感染。
3. 给予支气管扩张药物：气喘和支气管炎常常伴随支气管痉挛，医生可能会开具支气管扩张药物来缓解孩子的气喘症状。
4. 给予退烧药：如果孩子有发热，医生可能会建议使用适当的退烧药来控制体温。

此外，保持孩子的环境温暖、湿润，可以使用加湿器来增加空气湿度，有助于缓解鼻塞症状。

最重要的是，请及时与医生沟通并遵循他们的治疗建议。他们会根据孩子的具体情况制定最适合的治疗计划。祝愿你的孩子早日康复！

I'm not a doctor, but I can give you some advice. For a nine month old child, if a cold causes bronchitis and pneumonia, you should seek medical help as soon as possible. Hospitalization is the right decision because doctors can monitor a child's condition and provide appropriate treatment.

During hospitalization, doctors may take the following treatment measures:

1. Continue infusion: Infusion can maintain a child's water and electrolyte balance and help them recover their physical strength.
2. Administer antibiotics: If pneumonia is caused by a bacterial infection, doctors may prescribe antibiotics to the child to combat the infection.
3. Give Bronchiectasis drugs: Asthma and bronchitis are often accompanied by Bronchospasm, and doctors may prescribe Bronchiectasis drugs to alleviate the child's asthma symptoms.
4. Give Antipyretic: If the child has fever, the doctor may suggest using appropriate Antipyretic to control the temperature.

In addition, keeping the child's environment warm and moist can be achieved by using a humidifier to increase air humidity, which can help alleviate nasal congestion symptoms.

Most importantly, please communicate with the doctor in a timely manner and follow their treatment recommendations. They will develop the most suitable treatment plan based on the child's specific situation. Wishing your child a speedy recovery!

Figure 5: A case of a user confiding to ChatGPT.



孩子九个月，感冒引起的支气管炎和肺炎，现在住院输了两天的液，没有一点效果，现在气喘加重，咳嗽，和鼻塞也没减轻，请问我现在应该采取什么样的治疗？

My child is nine months old and has suffered from bronchitis and pneumonia caused by a cold. He has been hospitalized for two days and have received fluid transfusions without any effect. His asthma has worsened, and cough and nasal congestion have not alleviated. May I ask what treatment he should take now?



首先，建议立即咨询专业医生，因为小孩病情变化很快，需要得到专业治疗。根据您提供的信息，孩子患有感冒引起的支气管炎和肺炎，目前正在住院治疗。在这种情况下，医生通常会采用抗生素治疗、支气管扩张剂、氧气吸入等措施。如果已经输了两天的液，但病情没有好转，建议采取以下措施：

1. 与医生沟通：尽快与医生沟通，告知孩子病情没有好转。医生可能会更换抗生素，调整液体输注量和速度，或者增加其他治疗措施。
  2. 保持呼吸道通畅：让孩子保持半坐位或竖抱位，以减少呼吸道阻塞。可以使用雾化器或氧气吸入，帮助孩子呼吸。
  3. 给孩子喝水：多给孩子喝水，以帮助保持身体水分和稀释痰液。
  4. 注意营养：生病期间，孩子需要足够的营养。如果孩子食欲不振，可以尝试给孩子吃些易于消化的食物，如米汤、稀饭、面条等。
- 最后，请注意保持孩子室内空气流通，避免孩子接触烟雾和其他刺激性气体。同时，随时关注孩子病情变化，及时向医生反馈。

Firstly, it is recommended to consult a professional doctor immediately as the child's condition changes rapidly and requires professional treatment.

According to the information you provided, the child has bronchitis and pneumonia caused by a cold and is currently hospitalized. In this case, doctors usually use antibiotics, Bronchiectasis, oxygen inhalation and other measures. If the condition has not improved after two days of fluid infusion, it is recommended to take the following measures:

1. Communicate with the doctor: Communicate with the doctor as soon as possible to inform the child that the condition has not improved. Doctors may replace antibiotics, adjust fluid infusion volume and speed, or add other treatment measures.
2. Maintain unobstructed respiratory tract: Keep the child in a semi seated or upright position to reduce respiratory obstruction. Atomizers or oxygen inhalation can be used to help children breathe.
3. Drink water for children: Drink more water to help maintain body moisture and dilute phlegm.
4. Pay attention to nutrition: During illness, children need sufficient nutrition. If a child has a loss of appetite, they can try eating easily digestible foods such as rice soup, porridge, noodles, etc.

Finally, please pay attention to maintaining indoor air circulation for children to avoid exposure to smoke and other irritating gases. At the same time, keep an eye on changes in the child's condition and provide timely feedback to the doctor.

Figure 6: A case of a user confiding to ChatGLM.

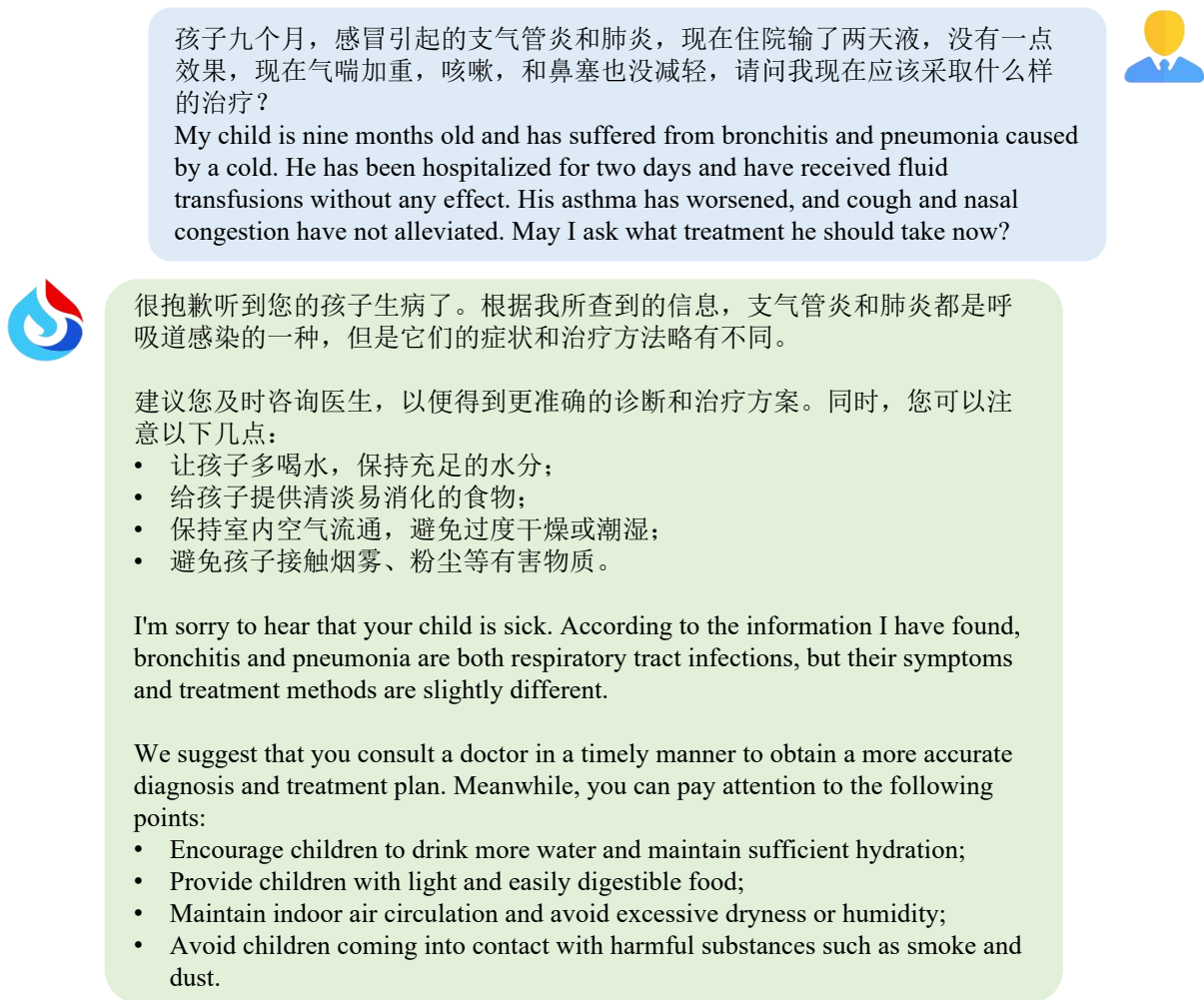


Figure 7: A case of a user confiding to SparkDesk.

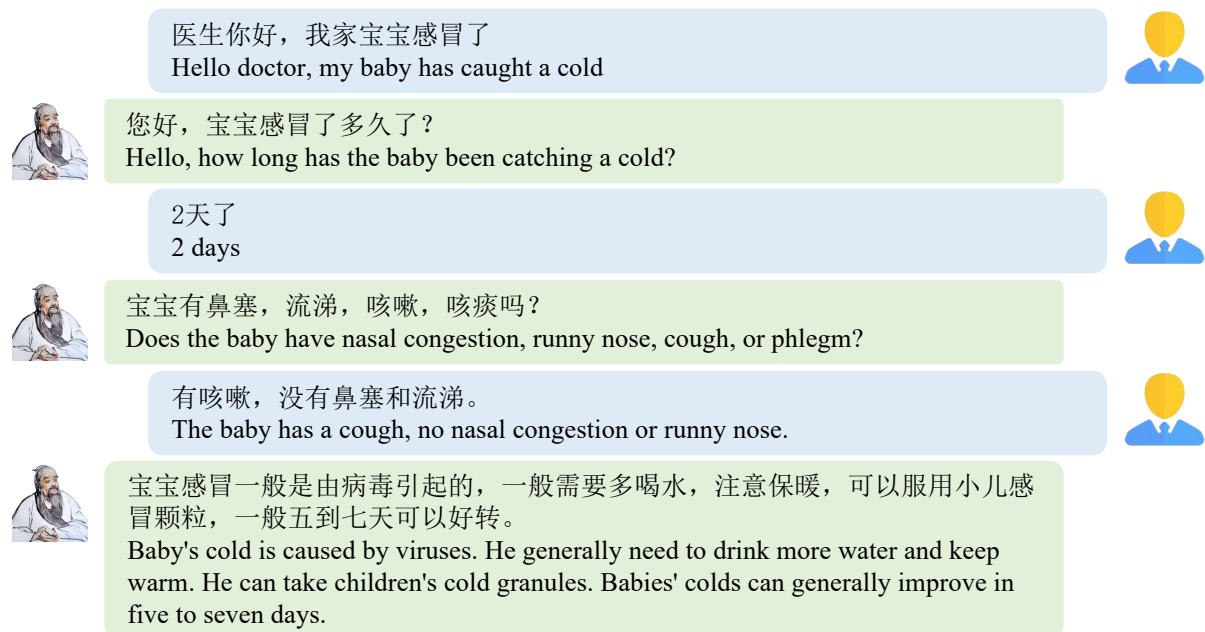


Figure 8: A case of a user confiding to BianQue.