

---

# IVYGPT: INTERACTIVE CHINESE PATHWAY LANGUAGE MODEL IN MEDICAL DOMAIN

---

A PREPRINT

**Rongsheng Wang**  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
p2213046@mpu.edu.mo

**Yaofei Duan**  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
p2213964@mpu.edu.mo

**ChanTong Lam**  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
ctlam@mpu.edu.mo

**Jiexi Chen**  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
jaychan0302@gmail.com

**Jiangsheng Xu**  
Opera inc  
China  
blueshell@whu.edu.cn

**Haoming Chen**  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
p2213065@mpu.edu.mo

**Xiaohong Liu**  
John Hopcroft Center  
Shanghai Jiao Tong University  
Shanghai, China  
xiaohongliu@sjtu.edu.cn

**Patrick Cheong-Iao Pang**  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
mail@patrickpang.net

**Tao Tan\***  
Faculty of Applied Sciences  
Macao Polytechnic University  
Macao, China  
taotanj@gmail.com

July 21, 2023

## ABSTRACT

General large language models (LLMs) such as ChatGPT have shown remarkable success. However, such LLMs have not been widely adopted for medical purposes, due to poor accuracy and inability to provide medical advice. We propose IvyGPT, an LLM based on LLaMA that is trained and fine-tuned with high-quality medical question-answer (QA) instances and Reinforcement Learning from Human Feedback (RLHF). After supervised fine-tuning, IvyGPT has good multi-turn conversation capabilities, but it cannot perform like a doctor in other aspects, such as comprehensive diagnosis. Through RLHF, IvyGPT can output richer diagnosis and treatment answers that are closer to human. In the training, we used QLoRA to train 33 billion parameters on a small number of NVIDIA A100 (80GB) GPUs. Experimental results show that IvyGPT has outperformed other medical GPT models.

**Keywords** Large language models · Medical · Reinforcement Learning

## 1 Introduction

Large language models (LLMs) are an important research direction in the field of artificial intelligence and have seen rapid growth after the introduction of ChatGPT in 2022. These large language models are trained based on large amounts of text data and are capable of generating human-like text, answering questions, and performing other natural language processing (NLP)-related tasks with high accuracy, showing excellent performance in a wide range of tasks. A key development in this area is the use of the Transformer architecture [2] and the underlying attention mechanism [3], which greatly improves the ability of language models to handle long-range dependencies in natural language text. Another important development is the use of pre-training, using LLMs that have been trained on large datasets and then fine-tuned for specific tasks, allowing the models to perform better in domain-specific Q&A, text generation tasks, further enabling LLMs to provide more accurate support for specialized domains or specific industries.

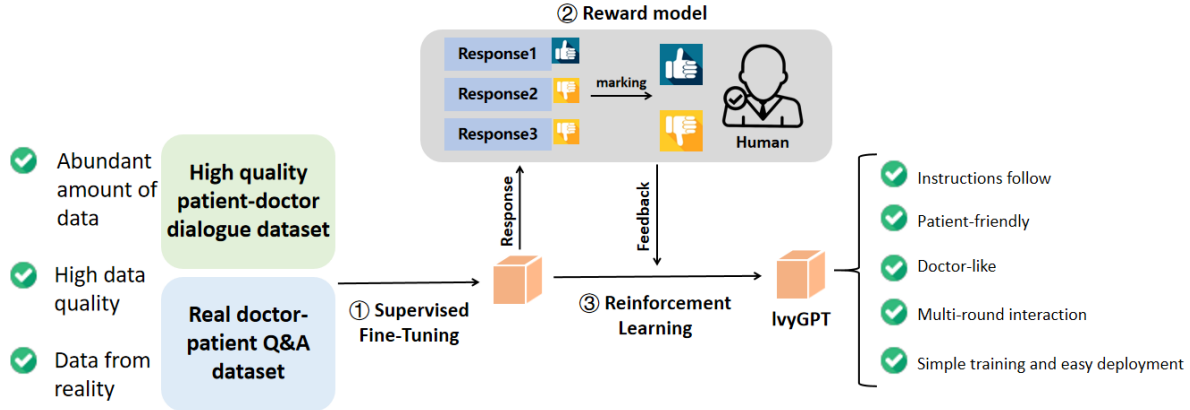


Figure 1: Process strategy for IvyGPT training

LLMs such as ChatGPT and GPT-4 are general-purpose models that are not trained to be used specifically for medical problems and are not designed to solve clinical tasks. Therefore, the use of LLMs in the medical domain may lead to incorrect or misleading medical information, and such incorrect or incomplete information may have harmful consequences for patient care. In addition, due to the diversity of global medical approaches (e.g., Chinese and Western medicine), LLMs for use in the medical domain need to be adapted to local medical practices and treatment methods. Currently, common LLMs are typically trained in English, which limits their ability to understand and respond semantically to other languages and affects the ability of models to reason in populations that do not speak English as their first language.

LLMs for use in the medical field have many advantages. They can provide diagnostic advice, medication recommendations, interpretation of medical reports, health care recommendations, and more. By providing basic information about patients through online chat, LLMs can perform online medical services based on the information provided. Through NLP technology, LLMs automatically sift and generate information about disease diagnosis, treatment, etc. From massive medical data to automate the processing of patient questions and needs. And the processing speed of large language models is so fast that they can respond to patients' needs quickly, ensuring that patients can get medical advice on time and effectively reducing the problem of patient congestion in hospitals [11].

Several LLMs for the medical domain have been proposed for the Chinese language, including HuaTuo [4], ChatMed [5], ZhenNong-TCM [6], and MedicalGPT (zh) [7]. However, these models have shortcomings such as small parameter sizes (like 6B or 7B) and the lack of high-quality training data conforming to real doctor-patient scenarios. These limitations restrict their generalization abilities in medicine. To address these issues, we propose IvyGPT for the medical settings in China and our contributions in this paper include:

1. We propose a comprehensive training method for medical LLMs which comprises three parts: supervised training, reward model training, and reinforcement learning using the QLoRA method to train large models with 33B parameters on devices with low computing power.
2. We contribute a high-quality dataset containing verified and realistic scenarios of doctor-patient conversations.
3. We evaluate the IvyGPT against other LLMs in the medical domain.

## 2 Methodology

Our approach focuses on integrating high-quality data with human feedback to enhance the quality of responses in medical consultations. This is achieved through a two-stage training strategy: Supervised Fine-tuning (SFT) utilizing mixed data and Reinforcement Learning (RL) derived from human feedback, as illustrated in Fig. 1.

### 2.1 Supervised Fine-Tuning

LLMs are pre-trained on large-scale unlabeled data, so they have strong language understanding and generation capabilities. However, for medical question-and-answer tasks, models may require more guidance and tuning to achieve optimal performance. With supervised fine-tuning, the generalization ability of the model can be combined with the requirements of a specific task, thus improving task performance. However, fine-tuning large language models with full parameters is extremely resource-intensive. Hu et al. [10] proposed a method, LoRA, which would freeze the

weights of pre-trained models and inject trainable rank decomposition matrices into each layer of the Transformer architecture, thus greatly reducing the number of trainable parameters required for downstream tasks. The QLoRA [9] method performs better on some fine-tuning tasks of large language models with larger parameters, and it will do a 4-bit quantization of the base model to ensure that the model can be fine-tuned for LLMs with very large parameters in a low memory footprint.

## 2.2 Reinforcement Learning from Human Feedback

**Reward Model** We train a reward model to align with the characteristics of doctors and LLMs. We use real instructions and conversations as training data, sampling multiple responses from our fine-tuned model. For multi-turn conversations, we provide the dialogue history to align our model’s response generation. These responses are then scored by human, considering informativeness, coherence, adherence to human preferences, and factual accuracy based on given real doctors’ diagnoses. The scoring LLM evaluates each response and assigns a score. We use this paired response data to train the reward model, using the fine-tuned model as its backbone for better generalization.

**Reinforcement Learning** Throughout the training process, in order to ensure that the model’s output is closer to human-preferred answers, we employ a reinforcement learning strategy to incentivize the model to produce better results. By defining a reward function, the system can learn to take appropriate actions during the dialogue process to maximize the expected reward. This allows the dialogue system to better understand user intent and provide accurate assistance. In the reinforcement learning process, we input the top k answers returned by the model for the same question into the reward model for scoring and obtain a score value. To prevent the model trained through reinforcement learning from deviating too far from the model in the SFT (Supervised Fine-Tuning) stage, we introduce constraint during training to ensure that the model achieves better results without veering off track. Without this constraint, the model may lean towards answers with extremely high scores, which may not necessarily be good answers.

## 3 Experiments

### 3.1 Training Details

Our model was implemented in PyTorch using the Accelerate, PEFT, and transformers packages, with LLaMA-33B [13] as the base model. For training, we loaded the model onto 4 NVIDIA A100 80GB GPUs and performed fine-tuning. In the supervised fine-tuning process, we set the learning rate, batch size, and maximum context length to  $5e-5$ , 16, and 1024, respectively. We trained all models for 3 epochs and saved the weights that performed the best on the validation set. During the reinforcement learning process, the model required larger memory usage. Therefore, we set the learning rate, batch size, and maximum context length to  $5e-5$ , 8, and 256, respectively, and trained for 2 epochs. The model trained through reinforcement learning demonstrated improved understanding and generation of responses in various conversational scenarios, accurately following different instructions.

### 3.2 Dataset

First we used a dataset released from HuaTuoGPT [8], and leveraged ChatGPT to verify grammatical errors and remove common sense errors. Second, we obtained a large number of real doctor-patient conversations from public websites and added them to the dataset. Finally, our dataset contains 307,038 sets of Q&As.

As shown in Fig. 2 (a), we used a smaller amount of data to complete the training process of the model, which effectively reduces the training time and resource consumption. In Fig. 2 (b), a comparison of the length of the model’s response generated for the query is shown. IvyGPT generates an average word count of 271.05, which is higher than the current models in the Chinese medical domain. Usually, a higher word count of responses indicates that the model can cover richer information in the response.

### 3.3 Evaluation Results

We used the word2vec-based method of computing cosine similarity to evaluate semantic similarity of AI answers and human answers. We used the trained 64-dimensional word2vec model to embed answers. We compared IvyGPT with four proposed baseline models: HuaTuo, ShenNong, ChatMed, and MedicalGPT for analysis. We used 100 query pairs to evaluate the semantic similarity between ChatGPT and the Chinese medical domain model, respectively. The results are shown in Table 1, and our model obtains the highest semantic similarity. This implies that our model is better able to capture and express semantic relatedness, resulting in more accurate and consistent language generation.

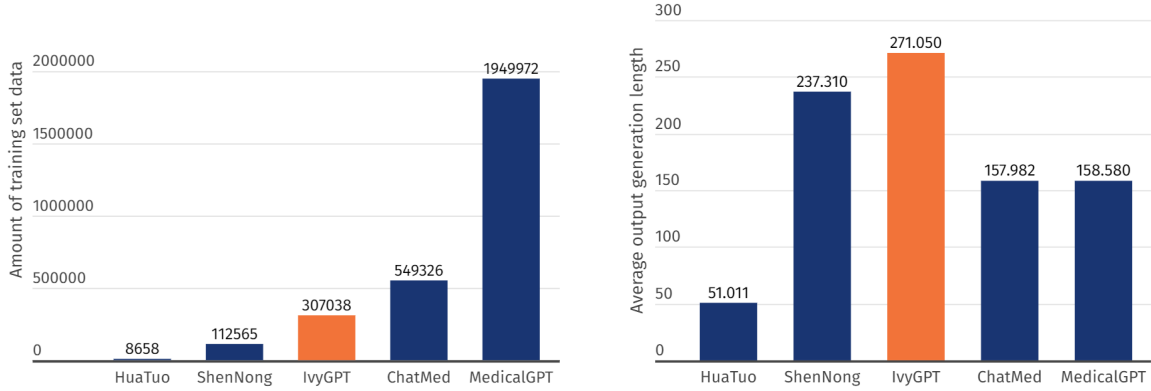


Figure 2: (a) The total amount of data in the training set, (b) The average number of words in the generated responses

Table 1: Semantic similarity comparison with real doctor

Model	Score
ShenNong [6]	77.71
HuaTuo [4]	71.20
ChatMed [5]	84.51
MedicalGPT [7]	83.73
ChatGPT	89.13
<b>IvyGPT (ours)</b>	<b>93.58</b>

We compare the output word count of the model with different training methods. As shown in the Fig. 3 (a), the model with reinforcement learning has richer answers to users' questions. Fig. 3 (b) shows that QLoRA has a shorter training time and is more efficient compared to LoRA.

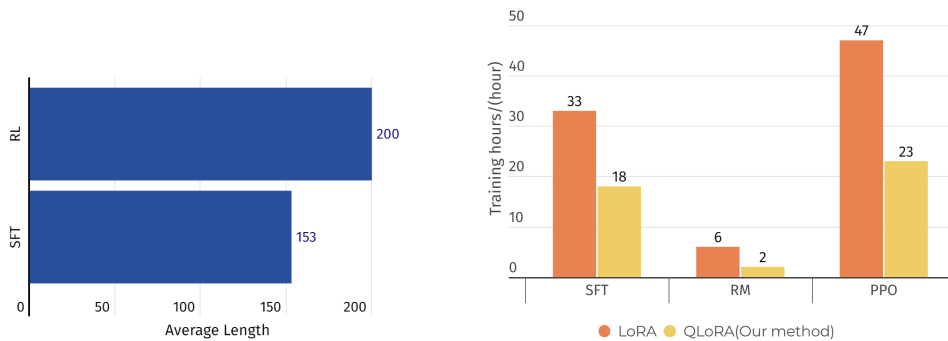


Figure 3: (a) The average word count of responses produced by the model after reinforcement learning. (b) The training duration under different fine-tuning methods

## 4 Discussion

We propose IvyGPT, a medical large-scale language model trained through supervised fine-tuning using high-quality medical question-and-answer instances and reinforcement learning based on human feedback. We demonstrate through experiments that IvyGPT generates responses with more reliable medical knowledge. However, there are still limitations in the application of IvyGPT in the medical field. Firstly, the difficulty in obtaining medical data limits the breadth and depth of the model's knowledge. Additionally, the model's ability to personalize and understand context is limited, leading to insufficient consideration of individual patient circumstances. Lastly, ethical and accountability issues need to be carefully addressed to ensure accuracy, reliability, and to avoid the spread of bias and misinformation. Therefore, while IvyGPT shows potential in the medical field, further improvements and validation are necessary.

There is still much work to be explored in the future, including:

1. Combining large language models with other medical expert systems and knowledge bases can provide more comprehensive and accurate medical diagnostics and recommendations. By integrating multiple models and systems, their strengths can be fully utilized to improve overall performance and reliability.
2. Explore various model compression techniques, such as pruning, quantization, distillation, etc., to reduce the size and computational requirements of models. This will make large language models easier to deploy and use, while lowering hardware requirements.
3. Conducting thorough evaluations and trials in real-world medical settings to assess the model’s performance, effectiveness, and impact on patient outcomes.

## 5 Conclusion

We propose IvyGPT, a medical LLM trained through supervised fine-tuning using high-quality medical question-and-answer instances and RLHF. With QLoRA, IvyGPT is able to load and train a model with 33 billion parameters, even with limited computational resources. The results show that the responses generated by IvyGPT have a higher similarity to the Q&A in real doctor-patient scenarios, indicating its potential in promoting self-service healthcare and supporting healthcare professionals.

## References

- [1] OpenAI, “ChatGPT,” <https://chat.openai.com>, 2022.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2019.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [4] H. Wang, C. Liu, N. Xi, Z. Qiang, S. Zhao, B. Qin, and T. Liu, “Huatuot: Tuning llama model with chinese medical knowledge,” 2023.
- [5] W. Zhu and X. Wang, “Chatmed: A chinese medical large language model,” <https://github.com/michael-wzhu/ChatMed>, 2023.
- [6] W. Zhu and X. Wang, “Shennong-tcm-llm,” <https://github.com/michael-wzhu/ShenNong-TCM-LLM>, 2023.
- [7] Y. M. Y. W. Y. W. Hongcheng Liu, Yusheng Liao, “Medicalgpt-zh: Large Language Model for Chinese Medical Dialogue,” <https://github.com/MediaBrain-SJTU/MedicalGPT-zh>, 2023.
- [8] H. Zhang, J. Chen, F. Jiang, F. Yu, Z. Chen, J. Li, G. Chen, X. Wu, Z. Zhang, Q. Xiao, X. Wan, B. Wang, and H. Li, “Huatuogpt, towards taming language model to be a doctor,” 2023.
- [9] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “Qlora: Efficient finetuning of quantized llms,” *arXiv preprint arXiv:2305.14314*, 2023.
- [10] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” *arXiv preprint arXiv:2106.09685*, 2021.
- [11] J. Li, P. C.-I. Pang, Y. Xiao, and D. Wong, “Changes in doctor–patient relationships in china during covid-19: A text mining analysis,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 20, p. 13446, 2022.
- [12] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, “Training language models to follow instructions with human feedback,” 2022.
- [13] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, “Llama: Open and efficient foundation language models,” 2023.