Efficient Knowledge Infusion via KG-LLM Alignment

Anonymous ACL submission

Abstract

To tackle the problem of domain-specific knowledge scarcity within large language models(LLMs), knowledge graph-retrieval-augmented method has been proven to be an effective and efficient technique for knowledge infusion. However, existing approaches face two primary challenges: knowledge mismatch between public available knowledge graphs and the specific domain of the task at hand, and poor information compliance of LLM with knowledge graphs. In this paper, we leverage a small set of labeled samples and a large-scale corpus to efficiently construct domain-specific knowledge graphs by LLM, addressing the issue of knowledge mismatch. Additionally, we propose a three-stage KG-LLM alignment strategy to enhance the LLM's capability to utilize information from knowledge graphs. We conduct experiments with a limited-sample setting on two biomedical question-answering datasets, and the results demonstrate that our approach outperforms existing baselines.

1 Introduction

Recent advancements in large language models(LLMs), such as ChatGPT, have demonstrated impressive capabilities in general-purpose content creation (OpenAI, 2022; Touvron et al., 2023). Nevertheless, their proficiency in domain-specific applications, particularly in the medical field, is notably constrained by insufficient knowledge (Bao et al., 2023; Zhang et al., 2023; Han et al., 2023b). To enhance the domain-specific performance of LLMs, the primary strategies for knowledge infusion include two main approaches: continual pre-training on domain-specific corpora and retrieval-augmented method, which involves integrating external information into the models.

Compared to continual pre-training, the retrieval-augmented approach is gaining popularity in knowledge-intensive scenarios due to its cost efficiency and enhanced traceability (Lewis et al., 2020; Lan et al., 2023). Some retrieval-augmented method involve integrating LLMs with resources directly such as corpora, news articles and tables through supervised fine-tuning (Borgeaud et al., 2022; Hu et al., 2023). However, the knowledge required by the model may be scattered among vast amounts of data, and directly retrieving from raw data instances will introduce noise inevitably, preventing the model from effectively utilizing the information. To mitigate this issue, leveraging structured knowledge, especially knowledge graphs(KGs), is an effective method (Moiseev et al., 2022; Ranade and Joshi, 2023; Wang et al., 2023).

However, the existing KG-retrieval-augmented methods still encounter two principal challenges. The first challenge pertains to knowledge mismatch. While many existing strategies utilize public KGs for knowledge infusion, the knowledge demanded by domain-specific tasks is frequently of a highly specialized nature, leading to a substantial likelihood that the KG might not cover all the requisite information, or might even present gaps. The second challenge involves poor information compliance. The structured format of triples in KGs diverges from the free-flowing format of natural language (Li et al., 2021; Ke et al., 2021) and the target text often includes additional information that is not found in the triples. This disparity can lead to confusion within LLMs, which could result in outputs from the trained model that do not align with the information incorporated from the KG, particularly in scenarios with a scarcity of supervised examples.

In this work, we construct a domain-specific corpus-based knowledge graph efficiently by LLMs and develop a knowledge infusion approach to enhance the ability of LLMs to utilize graph information, enabling them to generate comprehensive, logical, and low-hallucination responses. Firstly, we train a knowledge extraction model based on LLM using a small amount of labeled data. Then, we obtain a domain knowledge graph that resolves knowledge mismatch by performing extraction on unsupervised domain-specific corpora and reducing errors in the results through simple post-processing. Subsequently, we propose a novel three-phase KG-LLM alignment framework to optimize the exploitation of KG content by LLMs. The framework consists of the following stages:

- In the initial pre-learning phase, we synthesize substantial triples-to-text generation task examples derived from the previously mentioned extraction outcomes. We then train a Low-Rank Adapter(LoRA) (Hu et al., 2021), designated as K-LoRA, to assimilate the process of KG infusion and acquire proficiency in the domain-specific linguistic modality.

- The subsequent phase involves supervised fine-tuning. For each question-answer pair in the training set, we retrieve knowledge graph based on the question, concatenate the resultant subgraph into the input and proceed to train an additional LoRA. This process is designed to refine the model's output, tailoring it to the specific demands of the given task.

- The final phase is the reinforcement learning from knowledge graph feedback (RLKGF). In this phase, we extract knowledge triples from the generated responses and compare them with the KG to provide evaluative feedback on the knowledge correctness. This feedback serves as a basis for further fine-tuning the model to achieve more comprehensive, less hallucinatory, and more logical content.

To simulate a realistic context where specialized annotations are scarce, we conduct experiments on limited-sample datasets constructed based on two public biomedical question answering datasets, BioASQ (Nentidis et al., 2022) and CMedQA (Cui and Han, 2020). In summary, our main contributions are as follows:

1) We propose a modular knowledge infusion framework. Building upon the efficiently constructed KG, our approach aligns LLMs with the KG through lightweight parameter adjustment, addressing issues of knowledge mismatch and poor information compliance. Experimental results demonstrate that our method significantly outperforms the baselines.

2) We introduce two innovative strategies, namely "pre-learning" and "RLKGF", aiming at forging a stronger link between KGs and LLMs. In pre-learning, we demonstrate that triples-to-text task can serve as a simple and effective knowledge infusion strategy. In RLKGF, we illustrate that KGs can function as automated evaluators for the knowledge correctness of generated responses.

2 Related Works

Retrieval-augmented LLMs. Retrieval-augmented generation methods (Izacard and Grave, 2021; Lewis et al., 2020; Min et al., 2023; Borgeaud et al., 2022) retrieve relevant information from an external database for the query and enable the LLM to generate results using this information. ChainRAG (Xu et al., 2023) focuses on addressing the problem of incorrect knowledge retrieved by information retrieval systems, which can mislead the LLM or disrupt its reasoning chain through their interaction. While these methods enhance factuality, they also introduce new hallucinations. To address this challenge, WebBrain (Qian et al., 2023) incorporates both specific information and general knowledge, which are intertwined with text snippets and used as references to complete the task.

LLM-augmented KG Construction. AutoKG (Yu et al., 2021) proposes a framework for constructing a KG from unstructured documents using information extraction (IE) and internal semantic alignment. Since the graphs constructed by IE typically suffer from edge sparsity and node redundancy, Wu et al. (2023) have applied contrastive pre-training and node clustering to overcome this issue. Leveraging the capabilities of LLMs, Zhu et al. (2023) designs prompts for various knowledge graph construction tasks. Another line of research has aimed to extract knowledge from LLMs to construct KGs (Bosselut et al., 2019; Hao et al., 2023; West et al., 2022). Additionally, PiVe (Han et al., 2023a) utilizes iterative verification prompts to rectify errors in KGs generated by larger LLMs.

3 Methodology

Figure 1 illustrates the proposed framework, referred to as Enhanced LLM with Knowledge Pre-learning and Feedback (ELPF).

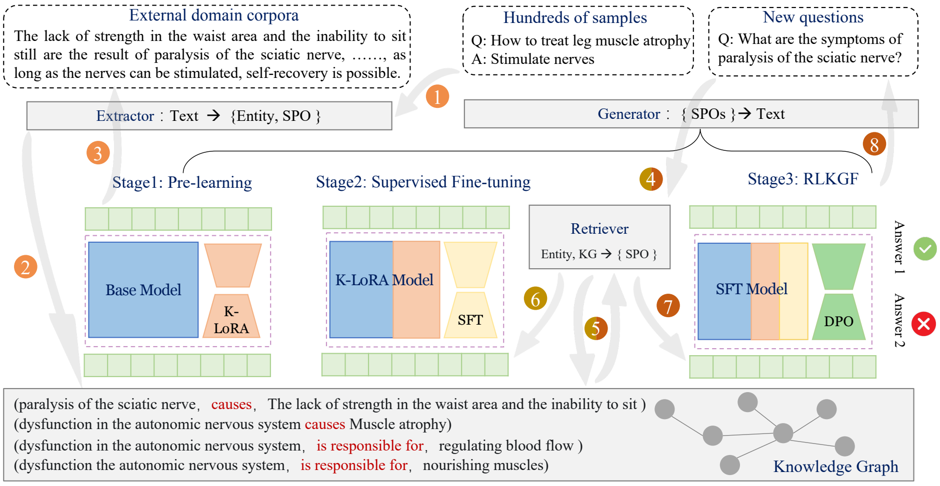

Figure 1: The ELPF framework can be divided into four main stages. 1) **Efficient construction of domain KGs** The process entails labeling a limited set of examples and developing a LLM-based knowledge extraction system to construct a domain KG from corpora efficiently. 2) **Pre-learning with K-LoRA**: Gain an understanding of domain-specific knowledge through LoRA-based triples-to-text generation, referred as K-LoRA. 3) **SFT with KG retrieval**: It involves retrieving subgraphs from the domain-specific KG, modifying the input accordingly and performing supervised fine-tuning. 4) **RLKGF**: The KG acts as an evaluator, providing feedback on knowledge correctness and enabling the model to better align with domain knowledge.

3.1 Efficient construction of domain KGs

For tasks within a specific domain, public available knowledge graphs fail to meet our needs frequently, referred as knowledge mismatch. To counter this issue, a viable solution is to establish a domain-specific knowledge graph utilizing extensive corpora. For one unsupervised document $d \in \mathcal{D}$, the process of KG construction can be formalized as Formula 1.

$$\{\mathcal{S}^a, \mathcal{P}, O^a\} = \mathcal{F}(d) \tag{1}$$

where $\mathcal{F}$ is KG construction system, $\mathcal{S}^a$ is set of subjects, $\mathcal{P}$ is set of defined relationships and O^a is set of objects.

The knowledge triples in the results are organized in the form of Formula 2.

$$\mathcal{T}_d = [< s_1, p_1, o_{11} >, ..., < s_j, p_k, o_{jk} >] \tag{2}$$

where $o_{jk} = o_{jk1}|o_{jk2}|...|o_{jkl}$. These triples are assembled and merged based on the same subject-relation pair. For example, "*Rome*" and "*Florence*" are both cities of *Italy*, so the instance should be represented as "<*Italy, City, Rome|Florence*>".

However, traditional methods of constructing such graphs can be intricate and often depend on substantial manual labor. Here we have designed an efficient KG construction workflow that requires only minimal annotation, leveraging the advanced semantic comprehension capabilities of LLM. We have streamlined the procedure into three stages: "knowledge triples extraction", "error removal", and "entity resolution".

Initially, we examine prevalent standards or seek guidance from domain experts to define a schema, and then we manually annotate a small set of examples($\approx$ 100) to generate training data in the format of "text->knowledge triples". Subsequently, we fine-tune a LLM using LoRA, which is a popular parameter-efficient fine-tuning method. Upon merging the trained LoRA parameters into the base model, we perform inference on extensive corpora to derive knowledge triples. Finally, we employ simple post-processing strategies to minimize errors within the extracted triples:

1. Remove results with incorrect output format, such as triples lacking a subject.

2. Remove results where either the subject or object does not appear in the original text.

3. Remove results where the relationship is not in the defined schema.

4. Remove results where the subject and object are the same.

In the entity resolution phase, we utilize an open-source text embedding tool[1] and set a similarity threshold. If the cosine similarity of two subject nodes surpass this threshold, we regard them as equivalent and subsequently combine their respec-

[1] https://github.com/shibing624/text2vec

tive subgraphs. This merging process contributes to the construction of a comprehensive domain-specific knowledge graph.

We perform a quality assessment on 200 samples of the extracted results from our experimental datasets, where the precision (the ratio of correct triples to the total number of generated triples) surpasses 85%. This suggests that the level of noise within the knowledge graph is within an acceptable range. For additional details and statistical outcomes, please refer to Appendix A.

3.2 Pre-learning with K-LoRA

Given that the triple form of KGs deviates from the natural language, LLMs exhibit limited proficiency in processing it. Moreover, acquiring copious amounts of annotated data in specialized domains frequently poses a challenge. Consequently, even with fine-tuning, it remains challenging to enhance the model's capability to leverage information from KGs. We hypothesize that it might be feasible to devise a method for low-cost, extensive data construction that enables the model to assimilate the task format in advance. Fortunately, by inverting the extraction process described earlier, we can create a "triples-to-text" generation task. With extensive fine-tuning on a multitude of instances, the model can be trained to recognize the information format infused by the KG. Additionally, as the target text is domain-specific, the model is able to acquire the unique linguistic style associated with that domain. To boost the fine-tuning process's efficiency, we continue to utilize LoRA-based SFT. We refer to the LoRA obtained in this step as K-LoRA.

3.3 SFT with KG retrieval

Pre-learning enables LLMs to better comprehend inputs in the triple form. However, it does not directly resolve specific tasks. Consequently, further refinement through fine-tuning with supervised learning examples remains essential. We adhere to the normal procedure of KG-retrieval-augmented methods (Lewis et al., 2020; Pan et al., 2024), which involves retrieving pertinent subgraphs from the previously established domain-specific KG and modifying the input accordingly. The comprehensive input construction is designed to adhere to the following template:

> **[KG]:** $\{g_q\}$
> **[Instruction]:** Refer to the KG and answer the following question: $\{q\}$

An initial observation reveals that the subjects and relations inherent in the subgraphs exhibit a significant correlation with the core purpose of the input query. To leverage this observation, we employ an open-source embedding tool[2] to encode all (s, p) pairs within the knowledge graph. Subsequently, we apply the same embedding tool to encode the input query. This approach facilitates the calculation of similarity scores between the query's embedding and those of the top-k (s, p) pairs. Finally, we retrieve the corresponding objects from the original knowledge graph for each (s, p) pair and reconstruct them into triples. These triples are subsequently integrated with the input to provide subgraph information. To maximize the benefits provided by K-LoRA, it is crucial to ensure that the representation of the subgraph remains consistent with the format used during the pre-learning phase.

3.4 RLKGF

After SFT, the model may still exhibit hallucinations in its responses due to issues such as overfitting. Inspired by the RLHF (Reinforcement Learning with Human Feedback) approach (Ziegler et al., 2020; Ouyang et al., 2022), we hope that the knowledge graph can serve as an automated evaluator, providing feedback on knowledge correctness of the current response, thereby guiding the model towards further optimization.

First, we generate a variety of responses for each query by employing diverse input formats or random seeds. Subsequently, we incorporate the knowledge graph to score and rank these responses. The scoring process entails the utilization of the extraction system described in Section 3.1 to extract triples from these responses, which are then compared with the knowledge graph to ascertain their correctness. The reward is determined by the number of correctly matched knowledge triples. The formula for calculating the reward is represented by Formula 3.

$$reward = \log(r_{spo} + \alpha * r_e) \tag{3}$$

where α is a hyperparameter, r_{spo} represents the number of SPO matches, and r_e represents the number of entity matches. For more details on

[2]https://github.com/shibing624/text2vec

the specific implementation process, please refer to Algorithm 1, where $Jcard$ represents the Jaccard similarity coefficient (Levandowsky and Winter, 1971). Appendix B demonstrates our automatic reward scoring mechanism using a case example.

To facilitate the training process, we utilize the Direct Preference Optimization (DPO) (Rafailov et al., 2023) training strategy, which mitigates sensitivity to reward values, thereby yielding more stable training process. For a comprehensive introduction to the DPO, please refer to Appendix C. This strategy involves creating pairs of samples according to their reward values. It is crucial to discard any pairs where the difference in rewards is not significant(i.e., $reward_{pos} - reward_{neg} \geq thresh$) and handle issues like repetitive generation in positive samples. To assess the extent of duplication within the positive samples, we can determine the ratio of unique clauses to the overall count of clauses following the deduplication process. Should this ratio fall below a predefined threshold, it would indicate the presence of considerable duplication within the sample, which will be dropped then. By utilizing the knowledge graph for automated evaluation, this method eliminates the requirement of manual scoring, thereby reducing labor costs. Another advantage of this approach is that it is not limited by the quantity of supervised samples, which allows for better learning of knowledge correctness.

Algorithm 1 Constructing pairwise samples

Input: Unsupervised questions Q, graph with entities $\mathcal{N}^g$ and SPOs $\{\mathcal{S}^g, \mathcal{P}, O^g\}$

for $q \leftarrow Q$ **do**
  $answers = \mathcal{F}(q)$
  for $answer \leftarrow answers$ **do**
    $\{\mathcal{S}^a, \mathcal{P}, O^a\} = \mathcal{F}_{ie}(answer)$
    $r_{spo} \leftarrow 0, r_e \leftarrow 0$
    for $\{s^g, p, o^g\} \leftarrow \{\mathcal{S}^g, \mathcal{P}, O^g\}$ **do**
      if $Jcard(\{n_s^a, p, n_o^a\}, \{s^g, p, o^g\}) \geq thresh_{sim}$ **then**
        $r_{spo} \leftarrow r_{spo} + 1$
      end if
    end for
    for $n^g \leftarrow \mathcal{N}^g$ **do**
      if $n^a = n^g$ **then**
        $r_e \leftarrow r_e + 1$
      end if
    end for
    $reward = \log(r_{spo} + \alpha * r_e)$
  end for
  for $ans_{pos}, ans_{neg} \leftarrow answers \times answers$ **do**
    if $reward_{pos} - reward_{neg} < thresh$ **then**
      Drop the pairwise sample
    end if
    if ans_{pos} contains a lot of repetitive content **then**
      Drop the pairwise sample
    end if
  end for
end for

Output: pairwise samples $[Ans_{pos}, Ans_{neg}]$

4 Experimental Settings and Results

4.1 Datasets

We select two biomedical question-answering datasets, CMedQA (Cui and Han, 2020) and BioASQ (Nentidis et al., 2022), for evaluating our model because both demand extensive domain-specific knowledge. CMedQA is a comprehensive dataset of Chinese medical questions and answers, consisting of over 10,000 pairs. In contrast, BioASQ is an English biomedical dataset that includes 4,719 question and answer pairs and 57,360 reference passages. To simulate a scenario with limited samples, we randomly choose 500 instances from each dataset for training and designate 1,000 instances from each for testing. For CMedQA, we employ the answer texts from the non-selected QA pairs as corpora to construct a knowledge graph in a weakly supervised manner. Similarly, with BioASQ, we use all the provided reference passages as the domain-specific corpora.

4.2 Evaluation Metric

In our evaluation, we employ multiple metrics, including BLEU (n=4), ROUGE-1, ROUGE-2, and ROUGE-L, to assess the performance of the models. In addition to these automated metrics, we also perform manual evaluations based on five dimensions: fluency, relevance to the question, correctness of the core viewpoint, diversity & completeness, and knowledge hallucination, using reference answers as a benchmark. Since it is challenging to assign an absolute score through manual evaluation, we rank the outputs of models under different settings according to various dimensions. A smaller ranking score indicates better performance, e.g. "1" means the best performance.

4.3 Experimental Settings

During the pre-learning stage, we perform fine-tuning of K-LoRA on both base models. The learning rate and number of epochs in the pre-learning stage are 5e-5 and 3. During the supervised fine-tuning stage, we establish the similarity threshold for subgraph retrieval to 0.9, and select the top-5 subgraphs. For more hyper-parameters and details, please refer to Appendix D.

Model	CMedQA				BioASQ			
	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU
ChatGPT-3.5	18.77	2.80	14.20	1.78	27.18	9.94	21.14	5.93
LLM-base	19.73	3.22	14.62	1.17	13.46	2.77	8.38	1.20
LLM-base-SFT(No-retrieval)	17.90	2.80	14.41	2.43	27.67	11.57	23.09	7.05
LLM-CP-SFT(No-retrieval)	18.31	2.84	14.71	2.56	26.99	11.31	23.55	7.23
LLM-base-SFT(RAG)	17.94	2.88	14.28	2.98	27.19	11.44	22.78	**9.07**
GAP	13.23	1.488	10.23	1.82	26.5	11.31	**24.37**	6.25
ELPF(ours)	**19.83**	**3.86**	**15.44**	**3.46**	**28.55**	**12.70**	24.21	7.79

Table 1: Performance comparison on CMedQA & BioASQ. "CP" indicates "continual pre-trained". We consider continual pre-training as a basic method of domain knowledge infusion, on par with other retrieval-based methods. Consequently, we do not report on the outcomes of hybrid approaches.

Model	CMedQA				BioASQ			
	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU
ELPF(ours)	**19.83**	**3.86**	**15.44**	**3.46**	28.55	**12.70**	**24.21**	**7.79**
w/o K-LoRA&RL	18.55	3.19	14.02	2.86	28.17	11.94	23.47	7.11
w/o K-LoRA	18.62	3.33	15.05	2.90	28.21	11.91	23.41	7.24
w/o RL	19.77	3.85	15.31	3.35	**28.61**	12.31	23.79	7.44
w/o KG retrieval	19.55	3.59	15.28	3.28	28.29	11.91	23.60	7.27

Table 2: Ablation experiment comparison on CMedQA & BioASQ.

4.4 Baselines

Base LLMs: On the CMedQA dataset, we choose ChatGLM2-6B (Zeng et al., 2023) as base model. On the BioASQ dataset, we choose Llama2-chat-7B (Touvron et al., 2023) as base model. Both of the models are initialized with Hugging Face's pre-trained checkpoints[3][4]. Additionally, we opt to utilize the API of ChatGPT-3.5. For the base LLMs, we present the results of querying the model in a zero-shot scenario. Moreover, to compare the difference with basic continual pre-training method, we conduct continual pre-training on the base LLMs using the aforementioned constructed unsupervised corpus. For the settings of hyper-parameters of continual pre-training, please refer to the Appendix D.

No-retrieval Models: We evaluate the performance of base LLMs and continual pre-trained LLMs after LoRA-based SFT with the constructed training set, where the inputs do not contain any retrieval results.

Retrieval-based Models: For KG-level retrieval, we utilize the state-of-the-art KG-to-text method called GAP (Colas et al., 2022) as a baseline. GAP enhances KG-to-text generation by incorporating graph-aware elements into pre-trained language models. For document-level retrieval, we compare our approach with the representative method called RAG (Lewis et al., 2020). RAG ensures that the text retrieval source aligns with the unsupervised corpus used for KG construction. The retrieval method employed here is the same as the subgraph retrieval approach discussed in Section 3.3. We place the retrieval results on the inputs and then perform LoRA-based SFT.

4.5 Main Results

Our results on the CMedQA and BioASQ datasets are shown in Table 1. We observe that the zero-shot querying method achieved ROUGE scores that are close to those obtained through supervised fine-tuning. However, it is worth noting that the zero-shot querying method results in significantly lower BLEU scores on both datasets. These results indicate that the zero-shot querying method does not effectively balance the professionalism and fluency of the generated text. Consequently, this method may not be suitable for generating domain-specific text that meets the desired criteria.

In terms of fine-tuning-based methods, our model shows improvements across various metrics. On the CMedQA dataset, our model achieves a 1.03 ROUGE-L improvement and a 1.03 BLEU improvement compared to the vanilla LoRA-based SFT method. On the BioASQ dataset, we have achieved a 1.12 improvement in ROUGE-L and a 0.74 improvement in BLEU. It is worth noting that our method achieves a significant performance improvement even compared to continual pre-training

[3]https://huggingface.co/THUDM/chatglm2-6b
[4]https://huggingface.co/meta-llama/Llama-2-7b-chat-hf

followed by fine-tuning. These results highlight the effectiveness of our proposed KG collaborative method in enhancing the performance of fine-tuning for LLM. Compared to the GAP method, our approach not only exhibits significant improvements but also offers the advantage of not requiring the full-parameter joint training of a graph encoder with a pre-trained model like GAP. In comparison to RAG, which focuses on document retrieval, our method achieves higher ROUGE scores but lower BLEU scores on the BioASQ dataset. This difference may be attributed to the document retrieval system's ability to recall more extensive information. On the other hand, the process of constructing a KG introduces information loss, which results in ELPF generation relying more on the implicit knowledge of LLM itself when the subgraph is insufficient, leading to lower accuracy. At the same time, document retrieval also introduces more noise, leading to some answers deviating from the original question.

5 Analysis

5.1 Ablation Studies

We conduct several ablation experiments to evaluate the effectiveness of each module. These experiments involve the individual removal of K-LoRA, KG prompt, and RLKGF, as well as the simultaneous removal of both K-LoRA and RLKGF. The results of these experiments, including ROUGE and BLEU scores, can be found in Table 2. Additionally, the manual evaluation results for BioASQ are presented in Figure 2. Here are the key observations from our analysis: (1) Removing K-LoRA leads to the most significant performance drop, reflected in ROUGE, BLEU, and the diversity of knowledge. The main reason is that the format of triples-to-text training samples is similar to the format of the subsequent fine-tuning task, allowing the model to better incorporate the knowledge implied by the input. (2) RLKGF has a less significant impact on ROUGE and BLEU metrics. This is because the reinforcement objective is not focused on replicating the target answer, but rather on incorporating comprehensive, effective, and accurate domain knowledge. It improves the diversity of knowledge, as well as the correctness of viewpoints, and reduces hallucinations, achieving the goal of reinforcement learning. (3) The results of manual evaluation indicate that the ablated models with knowledge integration demonstrate

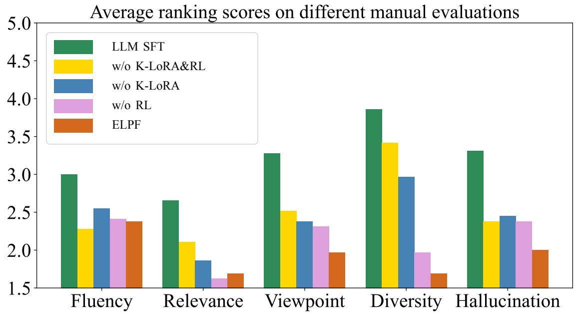

Figure 2: On the BioASQ dataset, different methods are ranked based on five human evaluation dimensions: fluency, relevance to the question, correctness of the core viewpoint, diversity & completeness, and knowledge hallucination. The ranking score represents the manual ranking of the content generated by different models, where a lower ranking score indicates higher quality of the generated content.

improvements over the baseline model that relies solely on fine-tuning, in terms of knowledge correctness (question relevance, viewpoint correctness, and hallucinations) and knowledge diversity. Our ELPF method outperforms others across all dimensions, demonstrating its effectiveness. Appendix E presents a specific case, which allows for a more intuitive understanding of the effectiveness of the answers output by different models.

5.2 In-depth Analysis of K-LoRA

To further analyze the overall impact of K-LoRA on the model, we will examine its effects on domain awareness and the alignment of generated text with the knowledge graph. K-LoRA aims

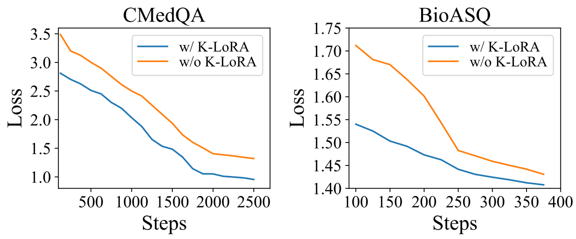

Figure 3: The loss curve of ELPF was compared under the same settings, with and without K-LoRA.

to enable the LLM to pre-learn within the current domain and generate text with the assistance of the knowledge graph in a weakly supervised manner. Regarding domain awareness, as illustrated in Fig.3, we observe a reduction in loss during the model training process and find that the model fused with K-LoRA not only has a smaller initial loss but also converges faster. This indicates that

Question
What could be the cause of menstrual period back pain that lasts for a long time?

Reference
Introduction to dysmenorrhea: Dysmenorrhea refers to the occurrence of lower abdominal pain, lower back pain, lower abdominal bloating, or other discomfort before, during, or after the menstrual period, which can significantly affect daily life and work. According to traditional Chinese medicine, this is often caused by blood stasis. Taking some traditional Chinese medicines that promote blood circulation and remove blood stasis can be used for treatment. Options include Nujin Pill and Motherwort Cream.

Retrieved subgraph
<Dysmenorrhea, clinical manifestations, lower abdominal pain|low back pain|lower abdominal bloating|other discomforts|>
<Dysmenorrhea, etiology, blood stasis|>
<Dysmenorrhea, medication treatment, Nujin Pill|Motherwort Cream|>

ELPF
Menstrual period back pain is caused by **blood stasis**, which can be alleviated by taking Chinese patent medicine Xiaoyao Pill and **Nujin Pill** for regulation. Additionally, taking **Motherwort Cream** in combination can help relieve the symptoms of dysmenorrhea.

ELPF w/o K-LoRA
It is considered to be menstrual irregularities. There are many reasons that can cause menstrual irregularities, such as endocrine disorders, gynecological inflammation, and psychological/environmental factors. It is recommended that you actively go to the hospital for a blood HCG test to determine if it is gynecological inflammation. Treatment should be administered accordingly.

Figure 4: Case study on impact on information compliance of K-LoRA. Information provided by KG in text is highlighted.

the model develops a certain level of awareness of the current domain through the training and weight fusion of K-LoRA. Therefore, K-LoRA serves as a straightforward and efficient approach for domain learning. In terms of the information compliance of the generated text to the knowledge graph, we analyze the text generated with and without K-LoRA, as shown in Figure 4. We notice that although the same knowledge graph information is provided, the original model does not effectively utilize this knowledge graph for generation. On the other hand, the model integrated with K-LoRA relies more on the knowledge graph and generates answers that are closer to the reference answers. This is because the task format of pre-learning and SFT is similar, and K-LoRA enhances the model's ability to adapt to input from the knowledge graph.

5.3 Knowledge Completeness

As our approach depends on information from the knowledge graph, this section explores the impact of the knowledge graph's completeness on our method. The completeness of knowledge can be measured by the size and quality of the knowledge graph. First, we explore the influence of graph size. We offer various sizes of KG, including full (100%), 80%, 60%, 40%, 20%, and 0%. The size control is achieved by randomly removing a certain proportion of nodes from the entire graph. Next, we investigate the impact of graph quality. We construct a set of target data to simulate the upper limit of model performance. The target data consists of triples extracted from the reference answers that correspond to the questions. The results are shown in Table 3. Firstly, we find that reducing the size of the knowledge graph does lead to a decrease in performance, but it is not a purely positive relationship. This is because our knowledge graph contains noise, and the model needs to balance between useful information and noise during the learning process. The model cannot effectively learn when the graph is sparse, resulting in even worse performance compared to not incorporating the graph information. Secondly, we observe that the current results still exhibit a certain gap when compared to the results obtained from the target data. This indicates that there is room for improvement in the quality of the graph constructed by LLM and the subgraph retrieval method. We will address these issues in future work.

	CMedQA		BioASQ	
	Rouge-L	BLEU	Rouge-L	BLEU
0 %	15.04	2.97	23.70	7.23
20%	14.98	3.02	23.59	7.14
40%	15.12	2.95	24.20	7.61
60%	15.26	3.10	24.39	7.70
80%	15.30	3.22	24.39	7.68
100%	15.44	3.46	24.21	7.79
target	16.40	3.56	25.32	8.03

Table 3: The performance comparison on knowledge completeness.

6 Conclusions

In this work, we propose a framework for efficiently infuse domain knowledge into LLMs. By employing efficient construction of domain knowledge graphs and a three-stage KG-LLM alignment process, we address the issues of knowledge mismatch and poor information compliance. Experiments demonstrate that our method significantly improves the quality of text generation and knowledge correctness in limited sample scenarios. We hope our work will provide insight into the challenge of connecting KG with LLM in future exploration.

Limitations

Although ELPF is relatively friendly in terms of sample size and computational resources, this method still has certain limitations. Since the construction of the domain knowledge graph is required in both SFT and RLKGF, the ELPF method is highly dependent on the quality of the graph construction. However, our graph is established based on weak supervision signals, so there are inevitably noises in the results. Insufficient noise handling can affect the effectiveness of the method. Furthermore, because the self-built domain knowledge graph (KG) is incomplete, it is challenging to detect knowledge errors unless they conflict with known knowledge. Additionally, determining the relevance of the knowledge to the query is a vague concept that is difficult to assess. Therefore, to enhance the stability and versatility of reinforcement learning, we have adopted a more conservative reinforcement strategy in RLKGF. This approach somewhat limits the optimization space. However, in actual vertical domain application scenarios, the positive reinforcement or conflict penalty strategies can be adjusted according to the actual situation to achieve better results. Finally, our method focuses on domain-specific text generation. However, due to the limited availability of appropriate public datasets, we only conducted experiments on medical domain texts. This limitation may pose a risk to the generalized ability of our findings in other scenarios.

References

Zhijie Bao, Wei Chen, Shengze Xiao, Kuang Ren, Jiaao Wu, Cheng Zhong, Jiajie Peng, Xuanjing Huang, and Zhongyu Wei. 2023. Disc-medllm: Bridging general large language models and real-world medical consultation.

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, et al. 2022. Improving language models by retrieving from trillions of tokens. In *International conference on machine learning*, pages 2206–2240. PMLR.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4762–4779. Association for Computational Linguistics.

Anthony Colas, Mehrdad Alvandipour, and Daisy Zhe Wang. 2022. GAP: A graph-aware language model framework for knowledge graph-to-text generation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 5755–5769, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Xiongtao Cui and Jungang Han. 2020. Chinese medical question answer matching based on interactive sentence representation learning. volume abs/2011.13573.

Jiuzhou Han, Nigel Collier, Wray Buntine, and Ehsan Shareghi. 2023a. Pive: Prompting with iterative verification improving graph-based generative capability of llms.

Tianyu Han, Lisa C. Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K. Bressem. 2023b. Medalpaca – an open-source collection of medical conversational ai models and training data.

Shibo Hao, Bowen Tan, Kaiwen Tang, Bin Ni, Xiyan Shao, Hengzhe Zhang, Eric P. Xing, and Zhiting Hu. 2023. Bertnet: Harvesting knowledge graphs with arbitrary relations from pretrained language models. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 5000–5015. Association for Computational Linguistics.

Chenxu Hu, Jie Fu, Chenzhuang Du, Simian Luo, Junbo Zhao, and Hang Zhao. 2023. Chatdb: Augmenting llms with databases as their symbolic memory.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models.

Gautier Izacard and Edouard Grave. 2021. Leveraging passage retrieval with generative models for open domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 874–880. Association for Computational Linguistics.

Pei Ke, Haozhe Ji, Yu Ran, Xin Cui, Liwei Wang, Linfeng Song, Xiaoyan Zhu, and Minlie Huang. 2021. JointGT: Graph-text joint representation learning for text generation from knowledge graphs. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2526–2538, Online. Association for Computational Linguistics.

Tian Lan, Deng Cai, Yan Wang, Heyan Huang, and Xian-Ling Mao. 2023. Copy is all you need. In *The Eleventh International Conference on Learning Representations*.

Michael Levandowsky and David K. Winter. 1971. Distance between sets. *Nature*, 234:34–35.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. volume 33, pages 9459–9474.

Junyi Li, Tianyi Tang, Wayne Xin Zhao, Zhicheng Wei, Nicholas Jing Yuan, and Ji-Rong Wen. 2021. Few-shot knowledge graph-to-text generation with pretrained language models. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1558–1568, Online. Association for Computational Linguistics.

Ling Luo, Po-Ting Lai, Chih-Hsuan Wei, Cecilia N Arighi, and Zhiyong Lu. 2022. BioRED: a rich biomedical relation extraction dataset. volume 23, page bbac282.

Sewon Min, Weijia Shi, Mike Lewis, Xilun Chen, Wen-tau Yih, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2023. Nonparametric masked language modeling. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 2097–2118. Association for Computational Linguistics.

Fedor Moiseev, Zhe Dong, Enrique Alfonseca, and Martin Jaggi. 2022. SKILL: Structured knowledge infusion for large language models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1581–1588, Seattle, United States. Association for Computational Linguistics.

Anastasios Nentidis, Georgios Katsimpras, Eirini Vandorou, Anastasia Krithara, Antonio Miranda-Escalada, Luis Gasco, Martin Krallinger, and Georgios Paliouras. 2022. Overview of BioASQ 2022: The tenth BioASQ challenge on large-scale biomedical semantic indexing and question answering. In *Lecture Notes in Computer Science*, pages 337–361. Springer International Publishing.

OpenAI. 2022. Chatgpt: Optimizing language models for dialogue.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. 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 Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback.

Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2024. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–20.

Hongjing Qian, Yutao Zhu, Zhicheng Dou, Haoqi Gu, Xinyu Zhang, Zheng Liu, Ruofei Lai, Zhao Cao, Jian-Yun Nie, and Ji-Rong Wen. 2023. Webbrain: Learning to generate factually correct articles for queries by grounding on large web corpus.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D Manning, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model.

Priyanka Ranade and Anupam Joshi. 2023. Fabula: Intelligence report generation using retrieval-augmented narrative construction. *ArXiv*, abs/2310.13848.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models.

Xintao Wang, Qianwen Yang, Yongting Qiu, Jiaqing Liang, Qianyu He, Zhouhong Gu, Yanghua Xiao, and Wei Wang. 2023. Knowledgpt: Enhancing large language models with retrieval and storage access on knowledge bases.

Peter West, Chandra Bhagavatula, Jack Hessel, Jena D. Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, and Yejin Choi. 2022. Symbolic knowledge distillation: from general language models to commonsense models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 4602–4625. Association for Computational Linguistics.

Siwei Wu, Xiangqing Shen, and Rui Xia. 2023. Commonsense knowledge graph completion via contrastive pretraining and node clustering. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 13977–13989. Association for Computational Linguistics.

Shicheng Xu, Liang Pang, Huawei Shen, Xueqi Cheng, and Tat-seng Chua. 2023. Search-in-the-chain: Towards the accurate, credible and traceable content generation for complex knowledge-intensive tasks.

Seunghak Yu, Tianxing He, and James Glass. 2021. Autokg: Constructing virtual knowledge graphs from unstructured documents for question answering.

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.

Xinlu Zhang, Chenxin Tian, Xianjun Yang, Lichang Chen, Zekun Li, and Linda Ruth Petzold. 2023. Alpacare:instruction-tuned large language models for medical application.

Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2023. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities.

Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2020. Fine-tuning language models from human preferences.

A Weakly Supervised Domain-specific IE System Construction

For the annotation standard of the CMedQA dataset, we referred to the CMeIE v2 dataset[5], which is a large-scale Chinese medical domain relation extraction dataset. For BioASQ, we referred to BioRED (Luo et al., 2022), an English medical relation extraction dataset annotated on the PubMed data source.

The types of relationship defined in the CMedQA dataset are: ["Differential Diagnosis", "Pathological Typing", "Clinical Manifestation", "Adjuvant Therapy", "Pharmacotherapy", "Surgical Treatment", "Etiology", "Synonyms", "Imaging Examination", "Auxiliary Examination", "Department of Consultation", "Complications", "Laboratory Test", "Susceptible Population", "Genetic Factors", "High-risk Factors", "Pathogenesis", "Site of Onset", "Medical History", "Incidence Rate", "Prognosis", "Age of Onset", "Prevention", "Post-treatment Symptoms", "Pathophysiology", "Transmission Route", "Peak Season", "Histological Examination", "Stage", "Radiotherapy", "Screening", "Chemotherapy", "Risk Assessment Factors", "Metastatic Sites", "Prevalence Area", "Mortality Rate"].

The types of relationship defined in the BioASQ dataset are: ["Association", "isa", "Negative_Correlation", "Positive_Correlation"].

For each reference dataset, we only utilized its relational schema and manually annotated 100 samples sampled from unsupervised corpora.

During manual annotation, we assigned two annotators for blind labeling and one quality control personnel for inspection. The final inter-annotator agreement was 0.9, and the accuracy of acceptance was 0.97. During the training, we employed the generative information extraction paradigm and trained a LoRA on top of LLM. The hyperparameter settings were consistent with those in the SFT stage.

Statistical details of the constructed graph are provided in Table 4. The symbol "#" denotes a sign for counting. We performed a quality assessment on 200 samples of the extracted results from experimental datasets and calculated the precision (the ratio of correct triples to the total number of generated triples).

Datasets	#Subjects	#Triples	Precision
CMedQA	25963	220111	0.85
BioASQ	20922	53209	0.89

Table 4: Statistics of the constructed domain KGs.

B Automated Reward Function

In RLKGF, we primarily propose an automated reward scoring mechanism that integrates a Knowledge Graph (KG). Here, we will demonstrate this process through a specific case study as show in Fig.5. For detailed information about the reward calculation, please refer to Algorithm 1.

C Direct Preference Optimiz ation (DPO)

Construct a static pairwise dataset $\mathcal{D} = \{x^i, y^i_\omega, y^i_l\}_{i=1}^N$ according to Section 3.3, where y_ω represents the positive samples and y_l represents the negative samples, and then perform reward modeling. According to DPO, the reward model $r_\phi(x, y)$ is trained using a negative log-likelihood loss as follows:

$$\mathcal{L} = -\mathbb{E}_{(x,y_\omega,y_l)\sim\mathcal{D}}[\log\theta(r_\phi(x, y_\omega) - r_\phi(x, y_l))]$$

where θ is the logistic function. In the context of LMs, the network $r_\phi(x, y)$ is often initialized

[5] https://tianchi.aliyun.com/dataset/95414

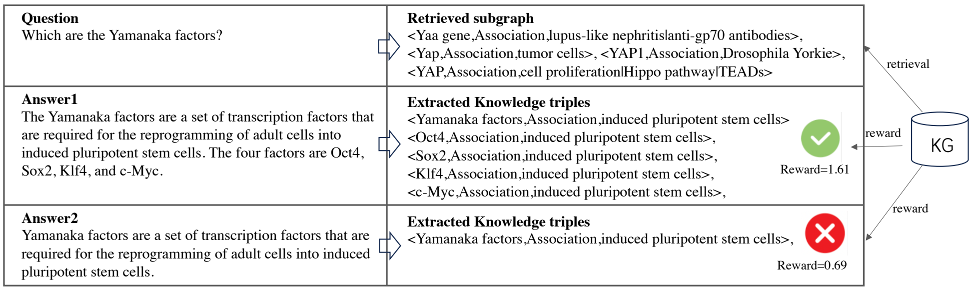

Figure 5: Case study on RLKGF dataset generation.

	CMedQA			BioASQ		
	K-LoRA	SFT	RLKGF	K-LoRA	SFT	RLKGF
LLM	ChatGLM2-6B			Llama2-chat-7B		
batch size	32	32	8	32	32	8
fine-tuning type	LoRA	LoRA	LoRA	LoRA	LoRA	LoRA
epochs	3	20	1	3	3	1
lora rank	8	8	8	8	8	8
lora target	QKV	QKV	QKV	QKVO	QKVO	QKVO
learning rate	$5e^{-5}$	$1e^{-4}$	$1e^{-6}$	$5e^{-5}$	$5e^{-5}$	$1e^{-6}$
max-input-length	512	512	512	512	512	512
max-output-length	512	512	512	512	512	512
KL-div β	-	-	0.4	-	-	0.4
top-p	0.7	0.7	0.7	0.7	0.7	0.7
temperature	0.9	0.9	0.9	0.01	0.01	0.01

Table 5: The parameter settings on CMedQA and BioASQ.

Q

Which are the main causes of fetal echogenic bowel?

Reference

Fetal echogenic bowel is mainly associated to feto-maternal, intramniotic bleeding but in several cases it is linked to cystic fibrosis, cytomegalovirus (CMV), herpes simplex virus and other viral infections and fetal aneuploidy.

KG information:
['<fetal aneuploidy,Association,Cytomegalovirus>', '<fetal aneuploidy,Association,herpes simplex virus>', '<fetal aneuploidy,Association,cystic fibrosis>', '<fetal aneuploidy,Association,feto-maternal>', '<fetal echogenic bowel,Association,fetal aneuploidy>']

Comparisons

ChatGPT 3.5
Fetal echogenic bowel refers to a condition where there is increased brightness or echogenicity of the fetal bowel on ultrasound. It can be associated with several possible causes, including:
1. Maternal Factors: Certain maternal conditions can contribute to fetal echogenic bowel. These may include infections such as cytomegalovirus (CMV), toxoplasmosis, or parvovirus B19. Maternal diabetes and intrauterine growth restriction (IUGR) can also be associated with this condition.
2. Genetic Disorders: Fetal echogenic bowel can be a marker for certain genetic disorders. It is commonly seen in cases of trisomy 21 (Down syndrome), but can also be associated with other chromosomal abnormalities such as trisomy 18 (Edwards syndrome) or trisomy 13 (Patau syndrome).
3. Fetal Infections: In some cases, fetal infections can cause echogenic bowel. These may include intrauterine infections such as CMV, toxoplasmosis, or rubella.
4. Bowel Obstruction or Meconium Peritonitis: Certain conditions that affect the fetal bowel, such as bowel obstruction or meconium peritonitis (inflammation of the lining of the abdominal cavity due to the leakage of meconium), can lead to echogenic bowel.
It is important to note that the presence of fetal echogenic bowel does not always indicate a serious problem, as it can be a benign finding in some cases. However, it should be evaluated further by a healthcare professional to determine the underlying cause and any associated risks or implications for the baby.

LLAMA2 SFT
Fetal echogenic bowel (FEB) is a common sonographic finding in the second trimester of pregnancy. The main causes of fetal echogenic bowel are:
1. Intrahepatic cholestasis of pregnancy (ICP)
2. Fetal bile duct obstruction
3. Fetal bile acid synthesis
4. Fetal bile acid transport
5. Fetal bile acid metabolism

Ours

ELPF
Fetal echogenic bowel (FEB) is a sonographic finding characterized by the presence of echogenic bowel loops in the fetal abdomen. It is a common finding in the first trimester of pregnancy and is usually benign. However, it may be associated with several fetal anomalies, such as cystic fibrosis, herpes simplex virus, and structural malformations.

ELPF w/o K-LoRA&RL
Fetal echogenic bowel (ECHO) is a common sonographic finding in the second trimester of pregnancy. It is characterized by the presence of a hyperechoic bowel wall, which is thought to be due to the accumulation of bile and/or meconium in the bowel lumen.

ELPF w/o K-LoRA
Fetal echogenic bowel is a common sonographic finding in the second trimester of pregnancy. It is associated with intrauterine fetal demise, IUGR, aneuploidy, structural malformations, growth deficiency, and pregnancy complications.

ELPF w/o RL
Fetal echogenic bowel (FEB) is a sonographic finding characterized by the presence of a hyperechoic bowel wall. It is a common sonographic finding in the second trimester of pregnancy and is usually benign. However, it may be associated with several fetal anomalies, such as cystic fibrosis, aneuploidy, and structural malformations.

Figure 6: Performance of various models in one case.

from the SFT model $\pi^{SFT}(y|x)$ with the addition of a linear layer on top of the final transformer layer that produces a single scalar prediction for the reward value. To ensure a reward function with lower variance, prior works normalize the rewards, such that $\mathbb{E}_{(x,y)\sim\mathcal{D}}[r_\phi(x,y)] = 0$ for all x. During the DPO RL phase, use the learned reward function to provide feedback to the language model, with the optimization objective as follows:

$$\mathcal{J} = \max_{\pi_\theta} \mathbb{E}_{x\sim\mathcal{D},y\sim\pi_\theta(y|x)}[(r_\phi(x,y)] \\ -\beta\mathbb{D}_{KL}[\pi_\theta(y|x)||\pi_{ref}(y|x)]$$

where β is a parameter that controls deviation from the baseline reference policy π_{ref}, and constraints on the KL divergence ensure that the reinforced strategy does not deviate too far from the baseline reference strategy (SFT). We also analyzed the impact of the value of the β parameter on the reinforcement training and selected an optimal parameter for subsequent training, as seen in Table 6.

	Rouge-1	Rouge-2	Rouge-L	BLEU
β=0.1	28.1	11.81	23.29	7.2
β=0.2	28.2	11.88	23.36	7.25
β=0.4	28.61	12.27	23.81	7.42

Table 6: In BioASQ, performance comparison of ELPF on different parameters β.

D Implementation Details

We conduct experiments on four A100 80GB GPUs and two V100 32GB GPUs. For details of the parameters used in the experimental training at each stage, please refer to Table 5. As for continual pretraining, we fine-tune full parameter of LLM with batch_size=4, epochs=3, learning_rate=5e-5.

E Case Study

We evaluate the effectiveness of the model through several case studies, as shown in Figure 6. ELPF provided concise and relatively comprehensive answers regarding the characteristics and main causes of fetal intestinal echoes. It mentioned both physiological and pathological situations. ELPF(w/o RLKGF) is close to ELPF in performance. However, the other answers were not as complete. ELPF(w/o K-LoRA&RLKGF) only mentions the physiological condition, while ELPF(w/o K-LoRA) only addresses the pathological factors. Untrained models like ChatGPT-3.5 and Llama2-chat-7B exhibit obvious hallucinations.