Llama-R1-MedRAG: Retrieval Augmented Generation Facilitated Reinforcement Learning for Patient Diagnosis

Thomas He

October 19, 2025

1 Introduction

With the rapid development of Large Language models (LLMs) and Vision-Language models (VLMs) in the recent years, we have seen unforeseen revolution across many industries, changing the lives of millions, particularly the medical domain. In patient diagnosis, such models could help clinicians interpret patient data, clinical notes, and even radiology images and provide feedback and analysis. Although current state-of-the-art VLMs are powerful for general tasks, they often hallucinate and lack the ability to reason among structured and unstructured data. This is problematic in medical applications because diagnosis requires extensive experience with diseases, their symptoms, and how they are shown on radiology images. Thus, the LLM or VLM for medical domain requires high reasoning capability and deep understanding of diseases and symptoms.

2 Dataset

There are several medical domain datasets that have a variety of modalities and diseases. Most formats are Question Answer pairs (QA). PMC-VQA is a QA pairs dataset derived from PubMed Central Open Access.[14] This dataset contains 227,000 Visual question answering samples over 149,000 unique images. An example of one data pair is: given a medical image, a question, four answer choices, the ground truth label is the correct answer. This dataset could potentially be suited for Supervised-Fine-Tuning (SFT) as it is a decent size and allows the model to get exposure to medical questions and medical images as a start.

Another dataset that could be used is the EHRXQA, a multi-modal question answering dataset for electronic health records with Chest X-ray images.[1] This dataset is derived from the MIMIC-IV database and MIMIC-IV-Note dataset on PhysioNet. [5] It contains three categories of Question Answer pairs: image-related, table-related, and image-table-related. The answers are SQL queried

from the database, thus providing only key words as ground truth. Most useful part of this dataset is the image-related QA pairs, which could be useful for further data generation methods. For instance, given the SQL-queried answer, we could utilize other LLMs to generate comprehensive descriptions of the answer, thus providing long-form responses for the model to learn.

Another dataset that is useful for facilitating model reasoning is the MedQA.[4] MedQA is a multiple-choice type of QA dataset for solving medical problems with real-world scenario type questions. The questions typically require a deep understanding of the medical textbooks. The most useful part of the sample pairs is the evidence section. Aside from the correct answer choice, it also contains an explanation section giving the reasoning behind the answer. Although the dataset size is only 60k samples, the quality of the evidence is believed to be very beneficial.

3 Related Work

Since the release of reasoning models such as GPT-o1 and DeepSeek-R1 in late 2024 and early 2025, various efforts has been put into integrating these models into the medical domain. Lai et. al. introduced Med-R1, which conducts reinforcement learning (RL) to improve generalizable medical reasoning in vision-language models.[6] Pan et. al. introduced MedVLM-R1, which extends Med-R1 to incentivize deeper reasoning in VLM through explicit reward shaping and advanced policy regularization.[9] Their method focuses on image to text reasoning and factual grounding. Other works such as RaRL combines RL with LoRA to enhance medical VLM generalization under limited data and compute. [10] In more recent months, several works has been conducted on incorporating RAG with RL as an attempt to increase report generation quality. Patho-AgenticRAG designs a multi-modal RAG framework for pathology VLMs guided by RL. [13] Med-R^3 introduces progressive RL for retrieval-augmented reasoning in LLMs. [8] To better facilitate RAG, others have developed RAG frameworks for the medical domain, for instance RAGAS and MedRAG. [3] [12]These frameworks have been adopted as reward model or can be used as information retrieval querying by the VLM.

4 Method

Due to the reasoning requirements of the VLM, the DeepSeek-R1-Distill-Llama-8B is best suited to be the base model. [2] This model will also be used as a baseline to compare. Overall, there are three steps to the method. First, the most simple method to give model exposure to medical data is via SFT. Thus, I plan on conducting SFT using one of the datasets described above and evaluate its results. Second, to address the hallucination issues and give the model access to relevant medical domain information, I will incorporate self-ask information retrieval into the model to allow the model to search for any disease or symptoms

it's not familiar with. I envision this as an API call to existing RAG frameworks such as MedRAG or Ragas. The third and last implementation I would like to conduct is Reinforcement Learning (RL) via optimization methods such as PPO or GRPO. This will further encourage the model to think longer and take more time to do research on the symptoms. The goal is to improve the diagnosis accuracy by doing more Test Time Compute.

5 Evaluation Methods

The model could be evaluated in various methods. Both Med-R_3 and Patho-AgenticRAG conducted comprehensive benchmarking on several datasets. The datasets contain real world QA situations as well as medical school exam type questions. Metrics such as BLEU, which measures the token-level precision can be utilized to measure the relevant keywords present in the generated response and the ground truth. The evaluation will also include summary tables with comparison between the various methods attempted in this work as well as the baseline model, DeepSeek-R1-Distill-Llama-8B. In addition, training results throughout each epoch or iteration will be graphed to show the increase in accuracy over time. In addition, exemplar of the trained model during inference will be shown to display the thinking process and how the model reason about the patient information and radiology results.

6 Hypotheses

There are several hypotheses drawn from this work conducted. First, due to a more powerful RAG framework, it is able to improve the base model's knowledge pool and improve diagnosis accuracy without any Reinforcement Learning. Second, by incorporating a more powerful external RAG framework, it is able to facilitate more effective Reinforcement Learning.

References

[1] Seongsu Bae, Daeun Kyung, Jaehee Ryu, Eunbyeol Cho, Gyubok Lee, Sunjun Kweon, Jungwoo Oh, Lei Ji, Eric Chang, Tackeun Kim, and Edward Choi. Ehrxqa: A multi-modal question answering dataset for electronic health records with chest x-ray images (version 1.0.0). PhysioNet, 2024. RRID:SCR_007345.

[2] DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli

Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jiawei Wang, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Shengfeng Ye, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wanjia Zhao, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanhong Xu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.

[3] ExplodingGradients. Ragas: Supercharge your llm application evaluations. `https://github.com/explodinggradients/ragas`, 2024.

[4] Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *arXiv preprint arXiv:2009.13081*, 2020.

[5] Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Aaron Gayles, Ahmad Shammout, Steven Horng, Tom J. Pollard, Siqi Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo Anthony Celi, and Roger G. Mark. Mimic-iv, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1, 2023.

[6] Yuxiang Lai, Jike Zhong, Ming Li, Shitian Zhao, and Xiaofeng Yang. Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. *arXiv preprint arXiv:2503.13939*, 2025.

[7] Yuan Li, Xiaodan Liang, Zhiting Hu, and Eric P Xing. Hybrid retrieval-generation reinforced agent for medical image report generation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[8] Keer Lu, Zheng Liang, Youquan Li, Jiejun Tan, Da Pan, Shusen Zhang, Guosheng Dong, Zhonghai Wu, Huang Leng, Bin Cui, and Wentao Zhang. Med-r^3: Enhancing medical retrieval-augmented reasoning of llms via progressive reinforcement learning, 2025.

[9] Jiazhen Pan, Che Liu, Junde Wu, Fenglin Liu, Jiayuan Zhu, Hongwei Bran Li, Chen Chen, Cheng Ouyang, and Daniel Rueckert. Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning, 2025.

[10] Tan-Hanh Pham and Chris Ngo. Rarl: Improving medical vlm reasoning and generalization with reinforcement learning and lora under data and hardware constraints, 2025.

[11] Ziyue Wang, Junde Wu, Linghan Cai, Chang Han Low, Xihong Yang, Qiaxuan Li, and Yueming Jin. Medagent-pro: Towards evidence-based multi-modal medical diagnosis via reasoning agentic workflow, 2025.

[12] Guangzhi Xiong, Qiao Jin, Zhiyong Lu, and Aidong Zhang. Benchmarking retrieval-augmented generation for medicine. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6233–6251, Bangkok, Thailand, 2024. Association for Computational Linguistics.

[13] Wenchuan Zhang, Jingru Guo, Hengzhe Zhang, Penghao Zhang, Jie Chen, Shuwan Zhang, Zhang Zhang, Yuhao Yi, and Hong Bu. Patho-agenticrag: Towards multimodal agentic retrieval-augmented generation for pathology vlms via reinforcement learning, 2025.

[14] Xiaoman Zhang, Chaoyi Wu, Ziheng Zhao, Weixiong Lin, Ya Zhang, Yanfeng Wang, and Weidi Xie. Pmc-vqa: Visual instruction tuning for medical visual question answering. *arXiv preprint arXiv:2305.10415*, 2023.