ReMedi: Reasoner for Medical Clinical Prediction

Yushi Cao[1], Yiming Chen[1]*, Hongchao Jiang[1], Hung-yi Lee[2], Robby T. Tan[1,3]
[1]ASUS Intelligent Cloud Services (AICS), Singapore
[2]NTU Artificial Intelligence Center of Research Excellence (NTU AI-CoRE), Taiwan
[3]National University of Singapore, Singapore
yushi_cao@asus.com Hongchao_Jiang@asus.com MattYM_Chen@asus.com
hungyilee@ntu.edu.tw robby.tan@nus.edu.sg

Abstract

Predicting future clinical outcomes from electronic health records (EHR) remains challenging due to the complexity and heterogeneity of patient data. LLMs have shown strong potential for such predictive tasks, yet existing approaches mainly focus on enhancing medical knowledge through distillation or RAG while relying on the model's internal ability to interpret contextual information. In this work, we present **ReMedi** (Reasoner for Medical Clinical Prediction), a framework for improving clinical outcome prediction from EHR. **ReMedi** generates rationale–answer pairs using a challenging sample re-generation mechanism for complex clinical questions, which leverages ground-truth answers as hints to enhance reasoning for further supervised fine-tuning and preference tuning. **ReMedi** integrates ground-truth outcome guidance into the preference data construction loop, regenerating rationale-answer variants. By tuning on these rationale-answer pairs, the model improves its predictive performance on clinical prediction tasks. Experiments on multiple EHR prediction tasks demonstrate substantial gains of up to 19.9% over state-of-the-art baselines in terms of F1 score, underscoring **ReMedi**'s effectiveness in real-world clinical prediction.

1 Introduction

Predicting future outcomes such as readmission, length of stay, and mortality is crucial for improving patient care and hospital resource management. Electronic Health Records (EHRs) (Johnson et al., 2016, 2023; Wornow et al., 2023) provide rich longitudinal data but are difficult to model due to their structured and heterogeneous nature. While Large Language Models (LLMs) show promise in processing medical data (Steinberg et al., 2021; Dwivedi et al., 2024; Rasmy et al., 2021), they still struggle to interpret EHR inputs effectively.[1]

Despite recent progress in clinical prediction using LLMs (Xu et al., 2024; Jiang et al., 2024), most approaches underutilize the models' reasoning abilities, which have proven effective in complex problem-solving domains (Guo et al., 2025a; Wang et al., 2025; Wu et al., 2025). For instance, KARE (Jiang et al., 2025b) incorporates reasoning through knowledge distillation and structured medical graphs, but its reliance on proprietary teacher models and predefined ontologies hinders scalability and real-world deployment.

In this work, we propose **ReMedi** (Reasoner for Medical Clinical Prediction), a simple and effective framework that improves clinical outcome prediction through self-generated supervised fine-tuning (SFT) and Direct Preference Optimization (DPO) (Rafailov et al., 2023). **ReMedi** leverages a challenging sample re-generation process to utilize complex and difficult clinical prediction questions for training data construction. We further devise **iReMedi**, an iterative variant of ReMedi, which follows an iterative refinement process that progressively enhances the predictive accuracy across multiple training rounds.

We evaluate the proposed **ReMedi** on three clinical prediction tasks, where it significantly outperforms existing methods. By incorporating iterative training, **iReMedi** further improves overall performance. Notably, we observe that directly applying existing general-domain self-improvement approaches (e.g., STaR (Zelikman et al., 2022)) to clinical prediction tasks yields only marginal gains due to training inefficiency. Finally, we conduct ablation studies to examine the contribution of each sub-component.

Our main contributions are as follows: (1) We propose **ReMedi**, an efficient and effective framework for clinical prediction tasks. **ReMedi** introduces outcome-guided preference construction, integrating ground-truth outcomes into the generation of rationale–answer pairs and forming an im-

*Corresponding author.
[1]See APPX. A for a detailed discussion of related works.

Findings of the Association for Computational Linguistics: ACL 2026, pages 20232–20241

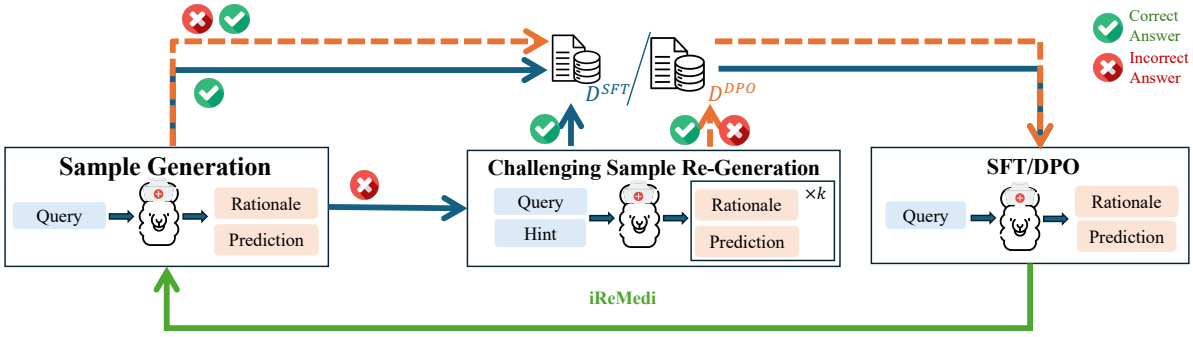

Figure 1: Overview of **ReMedi**, which operates iteratively across three stages: (1) Sample Generation, (2) Challenging Sample Re-Generation, and (3) Model Training. The dotted orange line represents the data processing pipeline for DPO, while the solid blue line denotes the pipeline for SFT.

provement process optimized via SFT and DPO. (2) **ReMedi** incorporates challenging sample regeneration to focus learning on difficult clinical prediction tasks. (3) Experiments on three standard EHR benchmarks show that ReMedi outperforms existing methods by up to 19.9% in terms of F1 score, and extensive ablation analyses validate its advantages in terms of predictive performance.

2 ReMedi

Overview The overview of **ReMedi** is illustrated in Figure 1. **ReMedi** enhances clinical prediction through a three-stage framework: (1) Sample Generation, (2) Challenging Sample Re-Generation, and (3) Model Training. We further introduce an iterative variant, **iReMedi**, which repeatedly executes these three stages to progressively refine the model's reasoning and prediction ability.

2.1 Sample Generation

The **ReMedi** cycle begins with the sample generation stage. Given an initial dataset $D = \{(q_i, a_i)\}_{i=1}^{N}$, where q_i denotes a query, a_i its ground-truth answer, and N the total number of samples, the generator model M is prompted to produce rationale–answer pairs: $\{(\hat{r}_i, \hat{a}_i) \sim M(q_i)\}_{i=1}^{N}$. Since ground-truth answers are available, we can evaluate and filter the generated pairs. Intuitively, high-quality rationales tend to produce correct answers. We retain those pairs where the generated answer matches the ground truth to construct a supervised fine-tuning dataset: $D^{\text{SFT}} = \{(q_i, \hat{r}_i, \hat{a}_i) \mid \hat{a}_i = a_i, q_i \in D\}$. To improve data efficiency, we also collect rationale–answer pairs with incorrect predictions and pair them with correct ones to form the DPO dataset: $D^{\text{DPO}} = \{(q_i, \hat{r}_i, \hat{a}_i, \hat{r}'_i, \hat{a}'_i) \mid \hat{a}_i = a_i, \hat{a}'_i \neq a_i, q_i \in D\}$.

To ensure a rapid adaptation to clinical prediction tasks, we leverage a warm-start (WS) phase for the initial model generator M. Specifically, during this phase, we directly apply both label rationalization and repeated sampling to all queries, and then gather correct rationale-answer pairs to ensure the quality of the generated data.

2.2 Challenging Sample Re-Generation

Queries that the model answers incorrectly typically represent challenging cases, and leveraging such samples has been shown to significantly improve model performance (Muennighoff et al., 2025). To effectively utilize these difficult cases, we employ two strategies: *label rationalization* and *repeated sampling*. Label rationalization differs from standard answer generation by providing the model with additional grounding. During this process, the model receives both the original query and the ground-truth label as hints, helping it generate more coherent rationales, particularly in the early stages when domain knowledge is still limited.

To further maximize the value of these challenging queries, we generate k responses per query and select one correct rationale–answer pair from the multiple samples. However, this approach risks the model explicitly referencing the hint during reasoning. To mitigate this, we filter out any rationale that explicitly mentions the given hint. Samples that consistently fail to produce the correct answer, even when hints are provided, are discarded. The remaining rationale–answer pairs are incorporated into D^{SFT} for supervised fine-tuning, while D^{DPO} is augmented with the newly generated, more challenging samples.

2.3 Model Training

After generating the synthetic data, we fine-tune the model in two stages: SFT and DPO. First, the base model M is fine-tuned on the supervised dataset D_{SFT} to obtain the SFT model. Next, the SFT model is used to collect preference data, forming the dataset D_{DPO}, which is then employed to further optimize the model through DPO training built upon the SFT model. The model optimization follows standard SFT and DPO objectives (Rafailov et al., 2023), where SFT minimizes the cross-entropy loss over correct rationale–answer pairs, and DPO optimizes the log-preference ratio between preferred and dispreferred samples.

2.4 Iterative ReMedi (iReMedi)

Building on the success of iterative training for improving LLM reasoning, we propose **iReMedi**, an iterative extension of **ReMedi**. In each iteration, the model undergoes the complete ReMedi optimization pipeline to progressively enhance both reasoning quality and predictive accuracy. The updated model M^* then serves as the generator for the next round of sample generation. To mitigate overfitting and maintain generalization, the training process is reinitialized from the original base model M in each ReMedi round.

3 Experiments

3.1 Experimental Setup

Implementation Details We leverage HuatuoGPT-o1 as the base model to utilize its enormous medical knowledge learned through its pre-training and post-training stages, reducing the need for complex medical knowledge retrieval (Jiang et al., 2025b, 2024). We fine-tune our model using the TRL (von Werra et al., 2020), Tranformers (Wolf et al., 2020)), DeepSpeed (Rasley et al., 2020), and Flash-Attention2 (Dao, 2024) frameworks. Following (Jiang et al., 2025b), for all the fine-tuning models we use set the initial learning rate to 5e-6 using AdamW (Loshchilov and Hutter, 2017) optimizer with a batch size of 16. We follow (Jiang et al., 2025b) for data processing. We use the Clinical Classifications Software to map the medical codes of conditions and procedures to natural language and use the Anatomical Therapeutic Chemical system for medication. The train, test, and validation sets are split in a 0.8/0.1/0.1 ratio.

Dataset and Tasks We process the publicly available MIMIC-IV EHR dataset (Johnson et al., 2023) into three standard clinical prediction tasks following (Jiang et al., 2025b, 2024; Johnson et al., 2023): (1) *Mortality Prediction*: predicting whether a patient will die during the next hospital visit; (2) *Readmission Prediction*: predicting whether the patient will be readmitted within 15 days after discharge; (3) *Length of Stay*: predicting the duration of hospitalization for the current visit. We follow (Jiang et al., 2025b) for data processing. More specifically, we retain 10,000 samples for the readmission task (with 5,000 samples labeled as readmission, and 5,000 as no-readmission). For mortality prediction, 2,701 patients with a mortality outcome and 7,299 patients with a survival outcome are retained. For Length of Stay, we set four classes: less than one day, one to seven days, one to two weeks, and more than two weeks. We obtain a total number of 10,000 samples (with 2,500 samples for each class).

Annotation Instruction Given the generated rationale-answer pairs, we check if the content in the rationale matches the prediction results. For example, if the rationale thinks the patient will be re-admitted, then the prediction should be 1 (readmission), otherwise, it is treated as a misalignment.

Baselines We compare against two categories of methods: (1) *Prompting-based LLMs*: Zero-shot and Few-shot prompting (with reasoning trace) with Gemini-2.5-Flash (Team et al., 2023), HuatuoGPT-o1-7B (Chen et al., 2024a), MedReason (Wu et al., 2025), and MedGemma-27B (Sellergren et al., 2025); Direct prediction (without reasoning trace) with Gemini-2.5-Flash and HuatuoGPT-o1-72B. (2) *Fine-tuned LLMs*: KARE (Jiang et al., 2025b) and a HuatuoGPT-o1-7B variant trained directly on ground-truth data.

3.2 Main Results

Table 1 summarizes the main results of **ReMedi** compared with all baselines. We observe that **ReMedi** and **iReMedi** consistently achieve the best performance across all tasks. Notably, all prompting-based methods exhibit near-random guessing behavior, reflecting their inability to produce high-quality data for clinical prediction tasks unfamiliar to general-purpose LLMs. This highlights the difficulty of clinical prediction tasks.

ReMedi outperforms both fine-tuning baselines, KARE and the SFT variant of HuatuoGPT-o1-7B

Type	Method	Mortality Prediction				Readmission Prediction				Length of Stay	
		Acc.	F1	TPR.	TNR.	Acc.	F1	TPR.	TNR.	Acc.	F1
Prompting	Direct Prediction										
	Gemini-2.5-Flash	49.3	33.7	99.6	0.80	62.8	62.6	94.1	44.0	31.7	23.6
	HuotuoGPT-o1-72B	61.5	60.6	62.5	60.0	61.0	55.9	77.8	34.8	39.8	25.9
	Zero-shot (CoT)										
	Gemini-2.5-Flash	68.0	67.9	83.5	58.9	50.3	36.1	99.2	3.12	41.5	36.6
	HuotuoGPT-o1-7B	48.2	44.6	97.6	18.3	49.9	36.8	97.1	4.35	26.3	16.4
	MedReason	37.4	28.0	100.0	1.00	50.1	35.6	98.3	2.68	27.3	25.1
	MedGemma	63.8	63.8	89.6	50.7	50.1	33.6	99.2	0.26	23.6	18.7
	Few-shot (CoT)										
	Gemini-2.5-Flash	71.3	71.0	82.7	64.5	49.8	35.1	99.2	2.21	43.9	38.7
	HuotuoGPT-o1-7B	75.2	73.9	78.6	73.5	52.2	41.8	96.8	9.73	31.4	24.6
	MedReason	45.6	41.8	96.6	15.9	49.9	36.6	95.1	4.18	29.7	24.6
	MedGemma	51.4	44.6	55.6	50.7	51.5	39.2	98.2	6.71	29.4	20.2
Fine-tuning	SFT	88.9	88.3	82.8	92.9	69.2	66.4	91.4	43.5	39.9	36.6
	KARE	95.9	95.5	95.1	97.5	81.2	81.3	83.7	78.3	40.4	35.9
	ReMedi (Ours)	97.7(+1.8)	97.6 (+2.1)	94.3(-0.8)	100.0 (+2.5)	90.5(+9.3)	90.4(+9.1)	80.6(-3.1)	100.0 (+21.7)	55.6(+15.2)	55.5(+19.6)
	iReMedi (Ours)	97.8 (+1.9)	97.6 (+2.1)	94.1(-1.0)	100.0 (+2.5)	91.5 (+10.3)	91.4 (+10.3)	83.8(+0.1)	100.0 (+21.7)	56.1 (+15.7)	55.8 (+19.9)

Table 1: Comparisons between ReMedi and baselines. Following (Jiang et al., 2025b, 2024), we report the Accuracy, Macro F1, True Positive Rate (TPR), and True Negative Rate (TNR). The best performance is highlighted. The values in brackets are the absolute performance gain between our method and KARE.

(that is, fine-tuning on ground-truth data once only). For mortality and readmission prediction, **ReMedi** achieves absolute accuracy improvements of 1.8% and 9.1%, respectively, compared to KARE. For the length-of-stay task, it yields larger gains of 15.2% in accuracy and 19.6% in F1 score. Additionally, **iReMedi** further outperforms ReMedi and other baselines, which demonstrates the effectiveness of the three-stage learning framework.

3.3 Discussion

Effect of Individual Components We conduct an ablation study to evaluate the contribution of each component to **ReMedi**'s performance. Additionally, we also consider the iterative self-taught learning method (STaR (Zelikman et al., 2022)). Results of each setting are summarized in Table 2. Incorporating DPO training leads to consistent performance gains for both **ReMedi** and **iReMedi**. Compared with STaR, the proposed **iReMedi** achieves markedly better results, primarily attributed to the challenging sample regeneration component, which effectively utilize the complex and challenging samples.

Method	Acc.	F1	TPR.	TNR.
ReMedi	90.5	90.4	80.6	100.0
ReMedi w/o DPO	84.4	84.4	85.3	83.6
iReMedi	91.5	91.4	83.8	100.0
iReMedi w/o DPO	86.8	86.8	83.7	89.9
STaR	59.1	53.2	96.1	23.4

Table 2: Effect of different sub-components. The best performance is highlighted.

Method	No Readmission		Readmission		Avg.	
	Human	Gemini	Human	Gemini	Human	Gemini
KARE	25.0	14.0	95.0	90.0	60.0	52.0
ReMedi	85.0	80.6	100.0	100.0	92.5	90.0

Table 3: Alignment between thinking and prediction.

Alignment Between Thinking and Prediction We manually evaluated responses from both KARE and the proposed **ReMedi** to examine the consistency between reasoning processes and final predictions. Specifically, 40 instances were randomly sampled from each of the readmission and no-readmission classes. Additionally, we utilize Gemini-2.5-Flash to evaluate the extent of misalignment across all test samples, and the results are consistent with the manual evaluation. The results, summarized in Table 3, show that for the no-readmission class, both KARE and **ReMedi** exhibit strong consistency between reasoning and prediction. However, in the readmission class, both methods display varying degrees of misalignment. Notably, KARE suffers from more severe misalignment, likely due to its multi-task learning design (Jiang et al., 2025b), where reasoning generation and label prediction are trained jointly. In contrast, **ReMedi**, with its end-to-end learning scheme, produces rationale–answer pairs that are better aligned and more coherent, leading to reduced misalignment. Nonetheless, occasional misalignment may still arise during data generation and filtering.

4 Case Study

To investigate why certain cases are particularly challenging, we conducted a case study involving three patients who were admitted to the hospital due to appendicitis. The differences are that: Patient A has no other chronic diseases (no readmission); Patient B has asthma (no readmission); Patient C has cancer (readmission). The difficulty lies in how to differentiate the nuanced clinical judgment about different conditions. The HuatuoGPT-o1 can predict Patient A and C correctly, but incorrectly predicts Patient B as readmission. It means that existing LLMs, such as HuatuoGPT-o1, do not differentiate the subtle distinctions in risk assessment and overestimate the risks (e.g., typical appendectomy recovery and stable chronic conditions). This also aligns with the results in Table 1, where the TPR (True Positive Rate) is high and the TNR (True Negative Rate) is low. By contrast, after fine-tuning, ReMedi demonstrates the capacity to perform more structured clinical reasoning. It evaluates risk by integrating both the severity of individual symptoms and their typical clinical trajectories. This suggests that ReMedi has learned to make finer-grained risk assessments by aligning observed symptoms and treatments with expected outcomes, thereby distinguishing between high-risk and low-risk cases with greater nuance.

Huatuo Response to Patient B: "....With the chronic condition, the potential for future hospital visits is definitely there. Plus, the appendectomy itself might have introduced new health concerns or complications... So, considering all this, the patient's appendectomy and recovery stage, along with the asthma, I'm leaning towards the prediction that this patient might be readmitted within 15 days."

ReMedi Response to Patient B: "... the most prominent factor here is the recent appendectomy. It requires close monitoring, and any complications could lead to readmission. The patient's current condition and medication regimen suggest they are managing well. Asthma could be a potential risk factor, but there's no clear indication of an imminent exacerbation as no specific procedures or medication are taken for asthma.... Therefore, the likelihood of readmission is not high."

5 Conclusion

We present **ReMedi**, a framework for improving clinical outcome prediction from EHR. **ReMedi** integrates supervised fine-tuning and DPO, leveraging the challenging sample re-generation mechanism, where complex clinical prediction questions are effectively utilized. This design enhances both predictive accuracy and task adaptability across diverse clinical settings. Experiments on multiple EHR benchmarks show that **ReMedi** achieves substantial gains over state-of-the-art methods. For future work, we intend to explore extending **ReMedi** to more challenging, open-question clinical tasks as well as multimodal medical prediction tasks.

Limitations

Despite the promising results achieved by **ReMedi**, several limitations warrant consideration. First, **ReMedi** exhibits minor misalignment between the rationale and the prediction. How to further mitigate such misalignment towards a coherent rationale and prediction needs to be investigated. Second, the assessment focuses on well-established clinical prediction tasks, such as mortality and hospital readmission, which have definitive outcome labels. The capability of the model (and the self-consistency strategy) to handle more complex, open-ended clinical reasoning tasks remains an open question. Third, this work primarily leverages relatively small-scale language models (e.g., HuatuoGPT-o1-7B); the performance and scalability of the approach when applied to significantly larger models (e.g., those exceeding 70 billion parameters) have yet to be systematically evaluated. Lastly, due to the limited availability of medical experts, the human evaluation is restricted to assessing the alignment between the model's reasoning and its final prediction. We acknowledge that rigorously verifying the clinical correctness and validity of the reasoning process with domain experts is an important direction for future work.

Ethical Considerations

An important ethical consideration arises from the use of sensitive electronic health record (EHR) data obtained from the MIMIC-IV dataset. To ensure responsible data handling, we have completed the required Data Use Agreement[2] and strictly adhere to its terms and conditions. The Gemini-2.5-Flash is accessed via Vertex, which is approved by the MIMIC4 dataset provider[3]. Furthermore, all model

[2]https://physionet.org/about/licenses/physionet-credentialed-health-data-license-150/

[3]https://physionet.org/news/post/gpt-responsible-use

training was conducted locally using secured, in-house computational resources, thereby minimizing potential risks associated with data privacy and external data transmission.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Kuluhan Binici, Abhinav Ramesh Kashyap, Viktor Schlegel, Andy T. Liu, Vijay Prakash Dwivedi, Thanh-Tung Nguyen, Xiaoxue Gao, Nancy F. Chen, and Stefan Winkler. 2025. Medsage: Enhancing robustness of medical dialogue summarization to asr errors with llm-generated synthetic dialogues. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(22):23496–23504.

Collin Burns, Pavel Izmailov, Jan Hendrik Kirchner, Bowen Baker, Leo Gao, Leopold Aschenbrenner, Yining Chen, Adrien Ecoffet, Manas Joglekar, Jan Leike, and 1 others. 2024. Weak-to-strong generalization: eliciting strong capabilities with weak supervision. In *Proceedings of the 41st International Conference on Machine Learning*, pages 4971–5012.

Junying Chen, Zhenyang Cai, Ke Ji, Xidong Wang, Wanlong Liu, Rongsheng Wang, Jianye Hou, and Benyou Wang. 2024a. Huatuogpt-o1, towards medical complex reasoning with llms. *arXiv preprint arXiv:2412.18925*.

Yiming Chen, Xianghu Yue, Xiaoxue Gao, Chen Zhang, Luis Fernando D'Haro, Robby T. Tan, and Haizhou Li. 2024b. Beyond single-audio: Advancing multi-audio processing in audio large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 10917–10930, Miami, Florida, USA. Association for Computational Linguistics.

Zixiang Chen, Yihe Deng, Huizhuo Yuan, Kaixuan Ji, and Quanquan Gu. 2024c. Self-play fine-tuning converts weak language models to strong language models. In *International Conference on Machine Learning*, pages 6621–6642. PMLR.

Pengyu Cheng, Yong Dai, Tianhao Hu, Han Xu, Zhisong Zhang, Lei Han, Nan Du, and Xiaolong Li. 2024. Self-playing adversarial language game enhances llm reasoning. *Advances in Neural Information Processing Systems*, 37:126515–126543.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. 2017. Gram: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 787–795.

Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in neural information processing systems*, 29.

Edward Choi, Zhen Xu, Yujia Li, Michael Dusenberry, Gerardo Flores, Emily Xue, and Andrew Dai. 2020. Learning the graphical structure of electronic health records with graph convolutional transformer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 606–613.

Tri Dao. 2024. FlashAttention-2: Faster attention with better parallelism and work partitioning. In *International Conference on Learning Representations (ICLR)*.

Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum, and Tong Zhang. 2023. Raft: Reward ranked finetuning for generative foundation model alignment. *arXiv preprint arXiv:2304.06767*.

Vijay Prakash Dwivedi, Viktor Schlegel, Andy T Liu, Thanh-Tung Nguyen, Abhinav Ramesh Kashyap, Jeng Wei, Wei-Hsian Yin, Stefan Winkler, and Robby T Tan. 2024. Representation learning of structured data for medical foundation models. *arXiv preprint arXiv:2410.13351*.

Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M Glass, and Jimeng Sun. 2020. Stagenet: Stage-aware neural networks for health risk prediction. In *Proceedings of the web conference 2020*, pages 530–540.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025a. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025b. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Lin Lawrence Guo, Ethan Steinberg, Scott Lanyon Fleming, Jose Posada, Joshua Lemmon, Stephen R Pfohl, Nigam Shah, Jason Fries, and Lillian Sung. 2023. Ehr foundation models improve robustness in the presence of temporal distribution shift. *Scientific Reports*, 13(1):3767.

Jiaxin Huang, Shixiang Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2023. Large

language models can self-improve. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1051–1068.

Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. 2024. Large language models cannot self-correct reasoning yet. In *The Twelfth International Conference on Learning Representations*.

Daniel P Jeong, Saurabh Garg, Zachary Chase Lipton, and Michael Oberst. 2024. Medical adaptation of large language and vision-language models: Are we making progress? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12143–12170, Miami, Florida, USA. Association for Computational Linguistics.

Hongchao Jiang, Yiming Chen, Yushi Cao, Hung-yi Lee, and Robby T Tan. 2025a. Codejudgebench: Benchmarking llm-as-a-judge for coding tasks. *arXiv preprint arXiv:2507.10535*.

Pengcheng Jiang, Cao Xiao, Adam Cross, and Jimeng Sun. 2024. Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. In *The Twelfth International Conference on Learning Representations*.

Pengcheng Jiang, Cao Xiao, Minhao Jiang, Parminder Bhatia, Taha Kass-Hout, Jimeng Sun, and Jiawei Han. 2025b. Reasoning-enhanced healthcare predictions with knowledge graph community retrieval. In *The Thirteenth International Conference on Learning Representations*.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, and 1 others. 2023. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

Geunwoo Kim, Pierre Baldi, and Stephen McAleer. 2023. Language models can solve computer tasks. *Advances in Neural Information Processing Systems*, 36:39648–39677.

Zhiming Li, Yushi Cao, Xiufeng Xu, Junzhe Jiang, Xu Liu, Yon Shin Teo, Shang-Wei Lin, and Yang Liu. 2024. Llms for relational reasoning: How far are we? In *Proceedings of the 1st international workshop on large language models for code*, pages 119–126.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Ning Miao, Yee Whye Teh, and Tom Rainforth. 2023. Selfcheck: Using llms to zero-shot check their own step-by-step reasoning. *arXiv preprint arXiv:2308.00436*.

Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori Hashimoto. 2025. s1: Simple test-time scaling. *arXiv preprint arXiv:2501.19393*.

Shuai Niu, Jing Ma, Liang Bai, Zhihua Wang, Li Guo, and Xian Yang. 2024. Ehr-knowgen: Knowledge-enhanced multimodal learning for disease diagnosis generation. *Information Fusion*, 102:102069.

Chao Pang, Xinzhuo Jiang, Krishna S Kalluri, Matthew Spotnitz, RuiJun Chen, Adler Perotte, and Karthik Natarajan. 2021. Cehr-bert: Incorporating temporal information from structured ehr data to improve prediction tasks. In *Machine Learning for Health*, pages 239–260. PMLR.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

Vyas Raina, Adian Liusie, and Mark Gales. 2024. Is llm-as-a-judge robust? investigating universal adversarial attacks on zero-shot llm assessment. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7499–7517.

Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. 2020. Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 3505–3506.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. 2021. Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, 4(1):86.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cían Hughes, Charles Lau, and 1 others. 2025. Medgemma technical report. *arXiv preprint arXiv:2507.05201*.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36:8634–8652.

Avi Singh, John D Co-Reyes, Rishabh Agarwal, Ankesh Anand, Piyush Patil, Xavier Garcia, Peter J Liu, James Harrison, Jaehoon Lee, Kelvin Xu, and 1 others. 2023. Beyond human data: Scaling self-training for problem-solving with language models. *Transactions on Machine Learning Research*.

Ethan Steinberg, Ken Jung, Jason A Fries, Conor K Corbin, Stephen R Pfohl, and Nigam H Shah. 2021. Language models are an effective representation learning technique for electronic health record data. *Journal of biomedical informatics*, 113:103637.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, and 1 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Dennis Thomas Ulmer, Elman Mansimov, Kaixiang Lin, Justin Sun, Xibin Gao, and Yi Zhang. 2024. Bootstrapping llm-based task-oriented dialogue agents via self-talk. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 9500–9522. Association for Computational Linguistics.

Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Gallouédec. 2020. Trl: Transformer reinforcement learning. `https://github.com/huggingface/trl`.

Bingning Wang, Haizhou Zhao, Huozhi Zhou, Liang Song, Mingyu Xu, Wei Cheng, Xiangrong Zeng, Yupeng Zhang, Yuqi Huo, Zecheng Wang, and 1 others. 2025. Baichuan-m1: Pushing the medical capability of large language models. *arXiv preprint arXiv:2502.12671*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, and 3 others. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Michael Wornow, Rahul Thapa, Ethan Steinberg, Jason Fries, and Nigam Shah. 2023. Ehrshot: An ehr benchmark for few-shot evaluation of foundation models. *Advances in Neural Information Processing Systems*, 36:67125–67137.

Juncheng Wu, Wenlong Deng, Xingxuan Li, Sheng Liu, Taomian Mi, Yifan Peng, Ziyang Xu, Yi Liu, Hyunjin Cho, Chang-In Choi, and 1 others. 2025. Medreason: Eliciting factual medical reasoning steps in llms via knowledge graphs. *arXiv preprint arXiv:2504.00993*.

Xiancheng Xie, Yun Xiong, Philip S Yu, and Yangyong Zhu. 2019. Ehr coding with multi-scale feature attention and structured knowledge graph propagation. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 649–658.

Ran Xu, Wenqi Shi, Yue Yu, Yuchen Zhuang, Bowen Jin, May Dongmei Wang, Joyce Ho, and Carl Yang. 2024. RAM-EHR: Retrieval augmentation meets clinical predictions on electronic health records. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 754–765, Bangkok, Thailand. Association for Computational Linguistics.

Muchao Ye, Suhan Cui, Yaqing Wang, Junyu Luo, Cao Xiao, and Fenglong Ma. 2021. Medretriever: Target-driven interpretable health risk prediction via retrieving unstructured medical text. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 2414–2423.

Zheng Yuan, Hongyi Yuan, Chengpeng Li, Guanting Dong, Keming Lu, Chuanqi Tan, Chang Zhou, and Jingren Zhou. 2023. Scaling relationship on learning mathematical reasoning with large language models. *Preprint*, arXiv:2308.01825.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. 2022. Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488.

Jesse Zhang, Jiahui Zhang, Karl Pertsch, Ziyi Liu, Xiang Ren, Minsuk Chang, Shao-Hua Sun, and Joseph J Lim. 2023. Bootstrap your own skills: Learning to solve new tasks with large language model guidance. *arXiv preprint arXiv:2310.10021*.

Yinghao Zhu, Changyu Ren, Zixiang Wang, Xiaochen Zheng, Shiyun Xie, Junlan Feng, Xi Zhu, Zhoujun Li, Liantao Ma, and Chengwei Pan. 2024. Emerge: Enhancing multimodal electronic health records predictive modeling with retrieval-augmented generation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 3549–3559.

A Related Works

Traditional Clinical Prediction Models Electronic health record (EHR) data contains a vast amount of detailed patient information (normally stored as medical codes), making it a valuable source in the medical and clinical domains (Johnson et al., 2023, 2016; Wornow et al., 2023). With the development of deep learning, researchers started to learn and capture the complex latent patterns of the EHR data (Steinberg et al., 2021; Dwivedi et al., 2024; Pang et al., 2021). Methods like CLMBER (Steinberg et al., 2021; Guo et al., 2023) utilize sophisticated network structures to learn the latent representations of the EHR data for various downstream clinical prediction tasks (Choi et al., 2016, 2017; Gao et al., 2020). Another line of methods focuses on constructing graphs to link all the information centered on the patient, aiming

to improve prediction accuracy (Jiang et al., 2024; Choi et al., 2020; Xie et al., 2019). However, such traditional models are often inflexible and require specific construction on the training and input data, which is inadequate for the dynamically changing healthcare domain.

LLM for Clinical Predictions LLMs (Achiam et al., 2023; Grattafiori et al., 2024; Guo et al., 2025a) have demonstrated strong capabilities in various domains (Jeong et al., 2024; Chen et al., 2024b; Binici et al., 2025; Jiang et al., 2025a; Li et al., 2024). Motivated by this, researchers are actively harnessing LLMs for precise clinical prediction tasks. Most approaches utilize retrieval-augmented generation (RAG) to obtain useful external knowledge from public medical databases and construct prompts together with patient information (Xu et al., 2024; Zhu et al., 2024; Ye et al., 2021; Niu et al., 2024). To create more specific and high-quality prompts, Jiang et al. (Jiang et al., 2025b) construct knowledge graphs from external medical knowledge tailored by patient information to create more fine-grained and precise content. Such methods mainly focus on retrieving related information and constructing augmented input for prompting or reasoning distillation from larger models. Therefore, the model's intrinsic reasoning capabilities are often overlooked. In contrast, our approach focuses on fully harnessing recent medical reasoning models' internal reasoning abilities with little to no access to larger models.

Reasoning Enhancement With the evolution of large language models (LLMs), a variety of approaches have been proposed to enhance their reasoning capabilities. One prominent direction is Self-improvement, which is closely related to self-taught learning. This approach aims to enhance the reasoning capabilities of large language models (LLMs) through the use of synthetic data, thereby reducing dependence on ground-truth labels (Huang et al., 2023; Chen et al., 2024c; Singh et al., 2023; Burns et al., 2024). This paradigm primarily relies on the ability of powerful, off-the-shelf LLMs to evaluate or select candidate answers—often through techniques such as majority voting—to improve output quality (Huang et al., 2023). However, the reliability of these judgments is not guaranteed and may inadvertently reinforce incorrect or biased behavior if the underlying model is flawed (Raina et al., 2024; Cheng et al., 2024). In our work, we provide ground truth labels as hints so that the reasoning behavior and final prediction are aligned and guided by the ground truth, mitigating the potential intrinsic behavior/bias of the LLMs.

Self-correction has also shown promise in improving the quality of LLMs' outputs. Many previous works (Shinn et al., 2023; Huang et al., 2024; Kim et al., 2023; Miao et al., 2023) employ multiple rounds of feedback and correction to correct their own outputs. However, as noted in (Huang et al., 2024), current LLMs continue to face challenges in identifying and correcting errors without external feedback. Moreover, these models may convert correct responses into incorrect ones during the self-correction process. Additionally, self-correction typically incurs longer inference times due to the need for multiple rounds of internal evaluation. In contrast, our approach prompts the model to generate both the reasoning trace and the final answer in a single pass, eliminating the need for iterative feedback and correction mechanisms.

Bootstrapping is a widely adopted strategy for enhancing the reasoning capabilities of large language models (LLMs). It typically involves generating new training samples using the model itself, which are then used to improve performance through supervised fine-tuning (SFT) or reinforcement learning (RL) (Yuan et al., 2023; Dong et al., 2023; Zelikman et al., 2022; Guo et al., 2025b; Ulmer et al., 2024; Zhang et al., 2023). When ground-truth labels are available, rejection sampling is often employed to filter out high-quality data for post-training (Yuan et al., 2023; Dong et al., 2023; Zelikman et al., 2022; Guo et al., 2025b). In scenarios where ground-truth labels are unavailable, many approaches depend on existing models to either generate data or guide its refinement (Ulmer et al., 2024; Zhang et al., 2023). In our work, we improve the efficiency of the bootstrapping process by including a warm-up stage. This is especially useful when the initial models are not well-adapted to the task (e.g., clinical prediction using EHR data).

Comparison between ReMedi and STaR The distinctions between ReMedi and STaR can be summarized into three key aspects:

- **Pipeline design:** ReMedi's main method is not iterative. It performs a single SFT stage followed by DPO. STaR, in contrast, depends on iterative self-training loops as its primary mechanism. Our iterative variant (iReMedi)

was added, yielding marginal improvements (90.5->91.5), suggesting that the main performance gains arise from the design of ReMedi itself.

- **Data construction and supervision:** In STaR, the model first attempts to generate a rationale-answer pair without hints; for the examples it still gets wrong, a hint is then provided, and the model regenerates a rationale-answer pair. Only those pairs that ultimately yield the correct answer are retained, while unsuccessful attempts are discarded. ReMedi differs in two key ways: (i) for each challenging query, we generate candidate rationale-prediction pairs and select one usable pair to maximize the value of each challenging query; and (ii) during our DPO stage we leverage both correct and incorrect rationale-prediction pairs rather than training only on the successful ones, thus providing a richer/informative supervision signal.

- **Clinical-domain adaptation:** We acknowledge the effectiveness of self-improvement frameworks demonstrated in prior work. However, directly applying STaR-like strategies to clinical prediction tasks is non-trivial. As shown in Table 2, a direct application of the STaR framework yields only 59.1% accuracy—far below the performance achieved by ReMedi. This highlights that the success of clinical reasoning tasks requires more than simply transferring existing self-training methods. This motivates our key design choices, which are tailored to novel tasks like clinical prediction tasks and are essential for strong performance.

B Prompts

Prompt with ground truth hint

```
Given the following task description, patient EHR
context, please provide a step-by-step reasoning
process that leads to the prediction outcome based on
the patient's context.
After the reasoning process, provide the prediction
strictly follow this format:
# Prediction # Your prediction
=====================================================
# Task #
Readmission Prediction Task:
Objective: Predict if the patient will be readmitted
to the hospital within 15 days of discharge.
Labels:  1 = readmission within 15 days, 0 = no
readmission within 15 days
Note:  Analyze the information comprehensively to
determine the likelihood of readmission. The goal is to
accurately distinguish between patients who are likely
to be readmitted and those who are not.
=====================================================
# Patient EHR Context #
[[Patient EHR]]
=====================================================
# Ground Truth #
# Prediction # [[Ground Truth Hint]]
=====================================================
Please provide a step-by-step reasoning process that
leads to the correct prediction based on the patient's
context and the ground truth.
The reasoning should be comprehensive, medically sound,
and clearly explain how the patient's information
leads to the predicted outcome. Your reasoning process
must align with the ground truth provided. You cannot
mention the ground truth in your reasoning process.

Then, provide your final prediction label in this
format:
# Prediction # Your answer

**Important Notes:**
- You must follow the ground truth to generate the
reasoning process!!
- Pretend that you do not know about the ground truth and
do not mention the ground truth label in the reasoning
process!!
```

Prompt with ground truth hint

```
Given the following task description, patient EHR
context, please provide a step-by-step reasoning
process that leads to the prediction outcome based on
the patient's context.
After the reasoning process, provide the prediction
strictly follow this format:
# Prediction # Your prediction
=====================================================
# Task #
Readmission Prediction Task:
Objective: Predict if the patient will be readmitted
to the hospital within 15 days of discharge.
Labels:  1 = readmission within 15 days, 0 = no
readmission within 15 days
Note:  Analyze the information comprehensively to
determine the likelihood of readmission. The goal is to
accurately distinguish between patients who are likely
to be readmitted and those who are not.
=====================================================
# Patient EHR Context #
[[Patient EHR]]
=====================================================
Please provide a step-by-step reasoning process that
leads to the correct prediction based on the patient's
context.
The reasoning should be comprehensive, medically sound,
and clearly explain how the patient's information
leads to the predicted outcome.

Then, provide your final prediction label in this
format:
# Prediction # Your answer
```