Synthetic Electronic Health Record Generation of Rare Disease With Reinforcement Learning

Xiaofeng Lin[*1] **Jun Han**[2] **Melika Emami**[2] **Brian L. Hill**[2] **Robert Tillman**[2]

[1] University of California, Los Angeles [2]Optum AI

bernardo1998@g.ucla.edu {jun_han, melika.emami, brian.l.hill, rob.tillman}@optum.com

Abstract

The generation of synthetic electronic health record (EHR) data using large foundation models (FMs) holds immense potential for mitigating data scarcity in healthcare, particularly in addressing the critical challenge of modeling rare diseases. However, the inherent imbalance in EHR data, where rare diseases are underrepresented, limits the ability of FMs to accurately generate these crucial data samples. This quality gap affects the usability of synthetic data in downstream applications, such as predictive modeling for rare diseases. To tackle this challenge, we propose Reinforcement Learning with Target Feedback (RLTF), a reinforcement learning-based framework designed to fine-tune FMs specifically for generating high-quality synthetic EHR data. By leveraging Direct Preference Optimization (DPO), RLTF optimizes the generative model to favor sequences that closely replicate real-world patterns of rare disease groups, ensuring their accurate representation. Experimental results demonstrate that RLTF significantly outperforms base model and other state-of-the-art methods in generating rare diagnostic codes and improves the utility of synthetic data for downstream tasks, such as rare disease prediction.

Introduction

The generation of synthetic electronic health record (EHR) data has emerged as a critical solution for addressing data scarcity and privacy concerns in healthcare. In many medical applications, accessing large, diverse datasets is challenging due to the sensitive nature of patient information. Synthetic data provides an avenue to bypass these limitations, enabling innovation while preserving privacy (Baowaly et al. 2019; Murtaza et al. 2023). Despite advancements in deep generative models, a significant challenge remains: the accurate generation of synthetic data for rare diseases. Rare conditions are often underrepresented in training datasets, leading existing models to struggle in faithfully capturing their intricate patterns (Al-Dhamari, Abu Attieh, and Prasser 2024; Chen et al. 2024). This quality gap in synthetic data generation poses serious limitations for downstream machine learning tasks, particularly those requiring rare disease modeling (He et al. 2024a; Peña-Guerrero, Nguewa, and García-Sosa 2021).

To bridge this gap, we propose a novel **Reinforcement Learning (RL) framework with Target Feedback (RLTF)**, designed specifically to enhance the generation of rare disease data in synthetic structured datasets. RLTF represents a flexible pipeline that can be integrated with various generative approaches to fine-tune models towards the accurate synthesis of rare disease patterns. For demonstration purposes, we implemented RLTF using a transformer-based foundation model (Hill et al. 2023), leveraging Direct Preference Optimization (DPO) (Rafailov et al. 2024) to guide the model. By explicitly prioritizing sequences containing rare ICD codes, RLTF ensures that synthetic data reflects the real-world occurrence and distribution of rare diseases more faithfully. This reinforcement learning framework enables fine-tuning of generative models to address the inherent limitations of baseline approaches in modeling rare events.

The key contribution of RLTF lies in its ability to substantially improve the representation of rare ICD codes in synthetic EHR data, effectively addressing the limitations of baseline generative models. By explicitly prioritizing sequences containing rare diseases, RLTF enhances the fidelity of rare disease modeling, resulting in improved performance in downstream tasks of rare disease prediction. This framework provides a valuable solution for rare data synthesis and augmentation, advancing machine learning applications that rely on high-quality synthetic healthcare data.

Related Work

Deep Generative Models for Electronic Health Record(EHR) Generation Recent advancements in synthetic EHR data generation have focused on generative modeling, particularly GAN-based methods. Early work like medGAN (Choi et al. 2017; Baowaly et al. 2019) and EHRGAN (Che et al. 2017) applied GANs to generate realistic EHRs. Diffusion models, such as MedDiff (He et al. 2023), leverage denoising diffusion probabilistic models (DDPM) with conditional sampling to ensure label consistency, while EHRDiff (Yuan, Zhou, and Yu 2023) enhances this by addressing diverse data types. Recently, diffusion models have been proposed to synthesize mixed-type time-series EHRs (Tian et al. 2024; Suh et al. 2024; He et al. 2024b). Transformer models are also gaining traction for EHR generation, with HALO (Theodorou, Xiao, and Sun 2023) learning patient visit sequences, and

[*]Work was done during an internship at Opum AI.

TabFormer (Padhi et al. 2021) and CHIRon (Hill et al. 2023) tokenizing medical codes and using autoregressive modeling for conditional generation based on patient history.

Reinforcement Learning in Generative Modeling Reinforcement Learning (RL) has been applied to synthetic data generation, focusing on aligning generated data with real data properties (Bauer et al. 2024). Notable examples include LSTM agents for dialogue generation (Li et al. 2016) and ORGAN, which combines RL with adversarial training for molecule and music generation (Guimaraes et al. 2017). In language models, methods like InstructGPT (Ouyang et al. 2022) and Factually Augmented RLHF (Sun et al. 2023) optimize outputs using human or AI feedback, while Direct Preference Optimization (DPO) (Rafailov et al. 2023) further simplifies preference learning by using labeled datasets. While these techniques are well-established in language and multi-modal tasks, their application to structured tabular data generation remains underexplored. Our work is the first to extend RLHF and DPO principles to EHR data, prioritizing rare disease prediction, a crucial aspect of synthetic healthcare data, and specifically prioritize machine learning utility.

Methodology

In this work we build upon the CHIRon model, a decoder-only generative foundation model for structured sequential medical data where each code is represented as a token. CHIRon is trained on a large healthcare institution's de-identified clinical and claims dataset and utilizes diagnosis, procedure, and medication codes along with lab results and other metadata including patient demographics. Additional embeddings, such as visit, age, and place of service, were also incorporated to provide more context for each medical code, as detailed in prior work (Hill et al. 2023). The protocol and supporting materials representing this work were prospectively submitted to the UnitedHealth Group Office of Human Research Affairs for IRB review and were approved.

Problem Statement Let $\mathcal{D} = \{S\}$ denote a dataset of structured event sequences, where each sequence $S = \{E_1, E_2, \ldots, E_L\}$ consists of events E_i drawn from a predefined vocabulary. The goal of generative models is to learn a synthetic joint distribution $\tilde{P}(S)$ that approximates the true distribution $P(S)$, enabling the generation of synthetic sequences.

For downstream classification tasks, the focus shifts to accurately modeling the conditional probability $P(E_{\text{target}}|S_{\text{context}})$, where S_{context} excludes the target event E_{target}. Precise estimation of the full sequence distribution $P(S)$ is not essential for achieving high utility in classification tasks (Xu, Sun, and Cheng 2023). According to statistical learning theory (Ng and Jordan 2001; Vapnik 2013), focusing on estimating $P(E_{\text{target}}|S_{\text{context}})$ is generally more effective than estimating $P(S)$, which includes unnecessary sequence details.

Therefore, we aim to use the alignment of sequence-target relationships in synthetic examples as a reward signal to fine-tune a generator initially trained on the autoregressive objective. Our objective is to refine the synthetic conditional distribution $\tilde{P}(E_{\text{target}}|S_{\text{context}})$ to closely match the real conditional distribution $P(E_{\text{target}}|S_{\text{context}})$, while minimizing disruptions to the overall sequence distribution $\tilde{P}(S)$ learned during pre-training.

Preference Labeling Dataset Construction: For each target code, we extract N random sequences with and without the target from a large healthcare institution's dataset, ensuring balanced representation to avoid bias in fine-tuning, creating a training set S. Given a sequence of length L from S, we randomly select a split point j to divide it into a prompt sequence $P = \{E_1, E_2, \ldots, E_j\}$ and a continuation sequence $C = \{E_{j+1}, \ldots, E_L\}$. If the target code E_{rare} is present, j is chosen from 0 to $r-1$, where $E_r = E_{\text{rare}}$, ensuring no target information leaks into P; otherwise, j is selected from 0 to L. We used $N = 10000$ for our main results as it enables both fast and effective training.

We construct the preference dataset with favored and less favored pairs by directly manipulating the presence of real rare codes to emphasize the correct versus incorrect target relationships. Specifically, for sequences containing E_{rare}, we remove it to generate a less favored (rejected) sequence R; for sequences S without E_{rare}, we randomly insert the rare code to create R. These perturbed sequences present incorrect relationships between event sequences and the occurrence of the rare code, which the model should learn to avoid.

In preliminary experiments, we also considered training an additional tree-based reward model to assess whether the presence of the target code in generated sequences is realistic. However, this approach introduced extra training complexity and potential prediction noise without yielding improvements in generation quality. Therefore, we opted for the simpler method of directly constructing preference pairs based on the presence or absence of E_{rare}.

Direct Preference Optimization (DPO): In rare data scenarios, training a reliable reward model to guide the generator is challenging. To overcome this, we apply direct preference optimization (DPO) (Rafailov et al. 2024), which directly uses preference-labeled data without requiring a reward model. Here, the patient's historical sequence P, the preferred sequence C, and the rejected sequence R correspond to the prompt, chosen, and less favored sequences, respectively. The DPO framework optimizes the following objective:

$$
\begin{aligned}
L_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = &- \mathbb{E}_{(P,C,R)\sim D}\left[\log \sigma\left(\beta \log \frac{\pi_\theta(C \mid P)}{\pi_{\text{ref}}(C \mid P)}\right.\right. \\
&\left.\left. -\beta \log \frac{\pi_\theta(R \mid P)}{\pi_{\text{ref}}(R \mid P)}\right)\right] \\
&- \lambda \cdot \max\left(0, \log \frac{\pi_{\text{ref}}(C \mid P)}{\pi_\theta(C \mid P)}\right),
\end{aligned}
$$

where D is the distribution of preference-labeled data, and C and R represent the preferred and rejected sequences, respectively. Following DPO-Positive (Pal et al. 2024), we introduce a regularization term to ensure the model not only reduces the likelihood of rejected sequences R but also maintains the likelihood of preferred sequences C. We use hyperparameter $\lambda = 0.1$ to control the regularization strength.

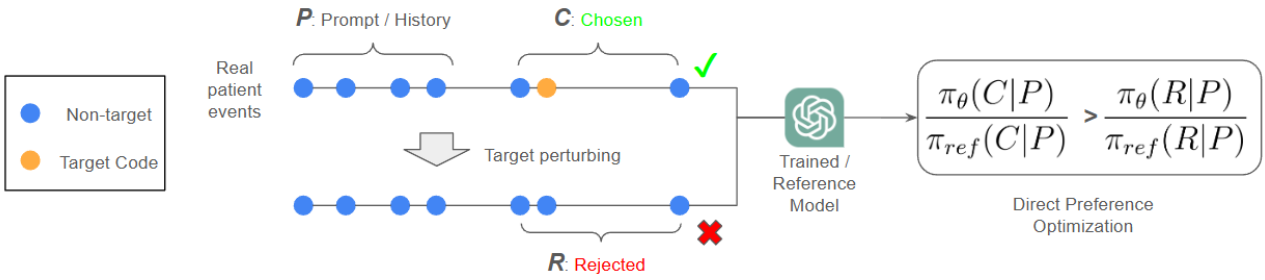

Figure 1: Overview of the proposed reinforcement fine-tuning pipeline.

Experiment

Target Cohort: Our analysis includes a total of 15 disease cohorts, covering both rare and more common chronic conditions to provide a balanced evaluation of our methodology. In this paper, we highlight seven rare conditions with the least frequency in the dataset: Sickle Cell Disease (0.036%), Shock (0.341%), Pleurisy, Pneumothorax, and Pulmonary Collapse (0.516%), Acute Myocardial Infarction (0.524%), Acute Cerebrovascular Disease (0.648%), Complications of Surgical Procedures or Medical Care (0.64%), and Septicemia Except in Labor (0.871%). These rare conditions are of particular interest due to their clinical significance and the challenges they pose for existing methods. Additionally, our analysis encompasses more common chronic conditions, such as Diabetes Mellitus Without Complication (11.13%) and Disorders of Lipid Metabolism (4.857%), to ensure robust performance evaluation across a spectrum of disease prevalence. The ICD codes and their summary statistics for all 15 cohorts are detailed in Table 4. Our goal is to demonstrate that the proposed methodology can effectively address both rare and common conditions, overcoming limitations of prior methods that struggled with rare disease modeling.

Fine-tuning and Evaluation: We fine-tune the pre-trained CHIRon model (Hill et al. 2023) using the preference-labeled data, reinforcing its ability to generate rare codes based on past medical history. We train a different model checkpoint for each target code. For evaluation, we select 100,000 examples from the a large healthcare institution's test set, truncate the last T medical codes in each sequence, and prompt the model to generate the remaining sequence. We assess the model's performance in two ways: (1) the ability to correctly generate rare disease codes, and (2) utility of synthetic data in disease classification, where an Catboost classifier is trained to predict presence of disease code from count of non-target medical codes. For the utility evaluation, we split the test examples into a modeling set and a validation set (70%/30% split). The models are prompted to complete the sequences in the modeling set, which are then used to train a Catboost classifier. The classifier's performance is evaluated on the unseen validation set, providing an assessment of how well the synthetic data supports disease classification tasks.

Comparative Analysis: We compare our method against the baseline CHIRon model, Long Short Term Memory network-based generator, and HALO (Theodorou, Xiao, and Sun 2023), a transformer-based model that represents codes from a single visit as binary vectors. HALO and LSTM are only trained on the modeling set as described above. Supervised finetuning are performed separately for target codes like RL training. Note that as the RLTF framework is not necessarily specific to any model formulation, the primary goal of this analysis is demonstrating its effectiveness in boosting rare code generation over the base model, rather than simply outperforming existing baselines.

Fidelity Table 1 presents the generation performance for target codes alongside the ROUGE-1 scores. The DPO method demonstrates a significant improvement in recall, particularly for rare codes such as Septicemia and Sickle Cell, resulting in a marked increase in the F1 score. This enhancement in recall was achieved with minimal degradation in fidelity for other codes, as evidenced by the ROUGE-1 score remaining nearly unchanged compared to the baseline CHIRon method (0.372 vs. 0.369).

In contrast, both the HALO and LSTM baselines exhibited poor performance in conditional generation tasks. These methods struggled to generate accurate sequences for rare codes, leading to significantly lower F1 scores. For instance, both HALO and LSTM performed poorly on Sickle Cell (F1 scores of 0.02 and 0.00, respectively). Furthermore, HALO and LSTM exhibited poor performance in conditional generation quality for other codes, as evidenced by their drastically lower ROUGE-1 scores (0.283 and 0.147, respectively).

These results emphasize the importance of RL-based approaches, such as DPO, to effectively address the challenges of rare code generation while preserving fidelity to the original code distribution. The improvements facilitated by DPO-positive regularization highlight its potential for balancing rare and common code generation. The fidelity scores for non-rare disease are reported in table 5 with the same trend displayed.

Machine Learning Utility Figure 2 presents the AUROC scores for disease classification models trained on synthetic data generated by different methods. The RLTF method consistently demonstrates significant improvements, particularly in rare disease codes such as Sickle Cell (SCD), resulting in competitive AUROC scores. For SCD, RLTF achieves an AUROC of 52.35%, outperforming HALO (44.76%), LSTM (47.24%), and CHIRon (43.91%), showcasing its ability to better handle rare diseases.

For common diseases like Septicemia (Sep.) and Diabetes Mellitus (DM), RLTF also demonstrates strong performance.

Table 1: Rouge-1, Precision, Recall, and F1 Scores for Different Methods and Target Codes on conditional generation of target code. The highest score for each disease is bolded.

Method	ROUGE-1	Target Codes						
		SCD	Surg.	Sep.	AMI	CVD	Pleu.	Shock
HALO	0.283	1.00/0.01/0.02	0.21/0.02/0.04	0.69/0.11/0.19	0.74/0.06/0.11	0.73/0.02/0.04	0.65/0.04/0.08	0.82/0.04/0.08
LSTM	0.147	0.00/0.00/0.00	0.91/0.02/0.03	0.90/0.06/0.12	0.81/0.07/0.14	0.86/0.03/0.05	0.71/0.03/0.06	0.90/0.03/0.05
CHIRon	0.369	1.00/0.03/0.05	0.13/0.07/0.09	0.64/0.23/0.34	0.72/0.08/0.14	0.44/0.04/0.07	0.52/0.05/0.09	0.62/0.12/0.21
RLTF	**0.372**	0.50/0.25/**0.33**	0.09/0.24/**0.13**	0.62/0.35/**0.45**	0.44/0.40/**0.42**	0.41/0.24/**0.30**	0.06/0.18/**0.09**	0.52/0.34/**0.41**

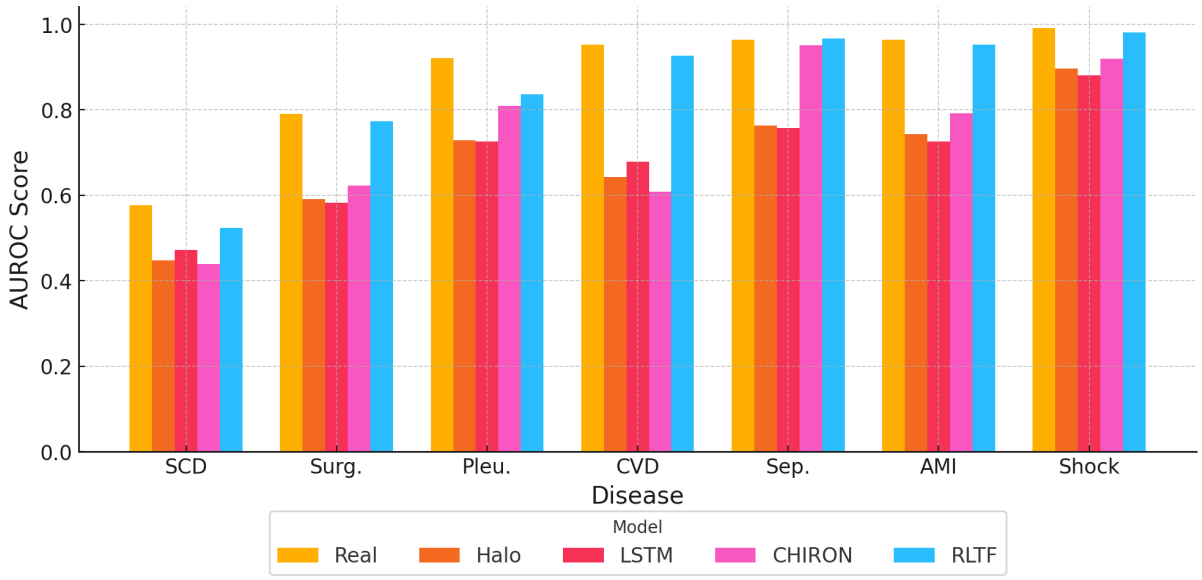

Figure 2: Performance of Methods on AUROC for Different Codes.

RLTF achieves an AUROC of 96.69% for Sep., nearly matching the real data AUROC of 96.38%, and outperforms HALO (76.33%), LSTM (75.74%), and CHIRon (95.10%). Similarly, for Diabetes Mellitus (not explicitly shown in the scores above but referenced in prior analysis), RLTF maintains high fidelity and classification performance, aligning closely with the real data levels.

In other target diseases, RLTF shows its robustness. For Septicemia, RLTF achieves the highest AUROC score of 96.69%, outperforming all other methods. For rare diseases like Acute Myocardial Infarction (AMI) and Chronic Venous Disease (CVD), RLTF achieves 95.27% and 92.62%, respectively, again surpassing the baseline methods and coming close to real data levels (96.44% and 95.22%). For Pleural Effusion (Pleu.) and Shock, RLTF demonstrates competitive scores of 83.63% and 98.08%, respectively, further highlighting its superior ability to handle both rare and common diseases. These trends are consistent across various datasets, with RLTF outperforming HALO, LSTM, and CHIRon in AUROC and AUPRC, particularly for challenging cases like Sickle Cell and Shock.

Overall, RLTF achieves the highest AUROC scores across most target diseases, showcasing its robustness and superior ability to handle challenging cases, as reported in Table 6. Similar effects have been shown in Table 7 under data augmentation cases, where the synthetic data is combined with real data in training classifier.

Table 2: Generation F1 and Machine Learning AUROC Scores for Different Number of RL Examples (N) and Proportion of Positive Examples (P) on Sickle Cell.

Training	N	P	F1	AUROC	ROUGE-1
RL	10000	0.5	**0.41**	0.61	0.372
RL	1000	0.5	0.01	0.57	0.318
RL	100	0.5	0.39	**0.73**	**0.396**
RL	10000	1	0.15	0.54	0.362
SFT	10000	0.5	0.05	0.58	0.327

Ablation Study: The ablation study results, as shown in Table 2, indicate that the default setting of $N = 10000$ and $P = 0.5$ yields the best or equivalent performance across most metrics. Specifically, Generation F1 and ROUGE-1 scores either remain consistent or are improved under this configuration. Other settings, including variations in the number of RL examples (N) or the proportion of positive examples (P), fail to achieve better results. Increasing the number of RL examples is beneficial for improving rare code generation, as evidenced by higher Generation F1 scores. Maintaining a balanced proportion of positive examples ($P = 0.5$) also proves critical for optimizing both generation fidelity and utility in downstream tasks, as seen in the Machine Learn-

ing AUROC and AUPRC scores. The last row reflects the results of supervised fine-tuning (SFT), where $N = 10000$ and $P = 0.5$. The SFT method lags behind the RL method in terms of Generation F1, downstream utility and ROUGE-1 scores, indicating that supervised fine-tuning is less effective at capturing the nuances of rare code generation. This underscores the importance of reinforcement learning (RL) in optimizing both the generation quality and downstream task performance, which simple re-balancing of training data fails.

Data Memorization: To assess potential privacy risks associated with fine-tuning on specific examples from real data, we analyzed the ROUGE-1 scores for conditional generation on both the training and testing data. A substantial increase in the ROUGE-1 score for the training set relative to the testing set would suggest an increased likelihood of the model memorizing the training data. Table 3 presents the results for the CHIRon and RLTF methods. While the ROUGE-1 scores are slightly higher on the training data compared to the testing data (e.g., 0.381 vs. 0.303 for CHIRon, and 0.384 vs. 0.311 for RLTF), the differences are not significant enough to indicate concerning levels of memorization. These results suggest that the model maintains generalization and does not overly memorize training data, mitigating potential privacy concerns.

Table 3: Average rouge-1 score on conditional generation of RLTF training and testing sequences.

Method	Average	
	Train	Test
CHIRon	0.381	0.303
RLTF	**0.384**	**0.311**

Discussion

In this paper, we introduce a novel reinforcement learning (RL) approach for improving the generation of rare medical codes, particularly focusing on rare but clinically significant conditions. Our method leverages RL to adjust the generation process, targeting recall improvements for these rare codes while maintaining high fidelity to the overall data distribution. Through extensive experiments, we demonstrate that the RL-based approach (RLTF) significantly enhances both recall and F1 scores for rare codes, outperforming baseline methods like Halo, LSTM and CHIRon. Additionally, RLTF shows minimal loss in data fidelity, as evidenced by unchanged ROUGE scores, and leads to better utility scores for downstream machine learning tasks, as indicated by higher AUROC values. This novel method not only proves effective for handling rare code generation in healthcare but also presents potential for broader applications in other domains, such as ensuring fairness and promoting diversity in data generation.

References

Al-Dhamari, I.; Abu Attieh, H.; and Prasser, F. 2024. Synthetic datasets for open software development in rare disease research. *Orphanet Journal of Rare Diseases*.

Baowaly, M. K.; Lin, C.-C.; Liu, C.-L.; and Chen, K.-T. 2019. Synthesizing electronic health records using improved generative adversarial networks. *Journal of the American Medical Informatics Association*.

Bauer, A.; Trapp, S.; Stenger, M.; Leppich, R.; Kounev, S.; Leznik, M.; Chard, K.; and Foster, I. 2024. Comprehensive exploration of synthetic data generation: A survey. *arXiv preprint arXiv:2401.02524*.

Che, Z.; Cheng, Y.; Zhai, S.; Sun, Z.; and Liu, Y. 2017. Boosting deep learning risk prediction with generative adversarial networks for electronic health records. In *2017 IEEE International Conference on Data Mining (ICDM)*, 787–792. IEEE.

Chen, Z.; Han, J.; Li, Y.; Kou, Y.; Halperin, E.; Tillman, R. E.; and Gu, Q. 2024. Guided Discrete Diffusion for Electronic Health Record Generation. *arXiv preprint arXiv:2404.12314*.

Choi, E.; Biswal, S.; Malin, B.; Duke, J.; Stewart, W. F.; and Sun, J. 2017. Generating multi-label discrete patient records using generative adversarial networks. In *Machine learning for healthcare conference*, 286–305. PMLR.

Guimaraes, G. L.; Sanchez-Lengeling, B.; Outeiral, C.; Farias, P. L. C.; and Aspuru-Guzik, A. 2017. Objective-reinforced generative adversarial networks (organ) for sequence generation models. *arXiv preprint arXiv:1705.10843*.

He, D.; Wang, R.; Xu, Z.; Wang, J.; Song, P.; Wang, H.; and Su, J. 2024a. The use of artificial intelligence in the treatment of rare diseases: A scoping review. *Intractable & Rare Diseases Research*, 13(1): 12–22.

He, H.; Xi, Y.; Chen, Y.; Malin, B.; and Ho, J. 2024b. A Flexible Generative Model for Heterogeneous Tabular EHR with Missing Modality.

He, H.; Zhao, S.; Xi, Y.; and Ho, J. C. 2023. MedDiff: Generating electronic health records using accelerated denoising diffusion model. *arXiv preprint arXiv:2302.04355*.

Hill, B. L.; Emami, M.; Nori, V. S.; Cordova-Palomera, A.; Tillman, R. E.; and Halperin, E. 2023. CHIRon: A Generative Foundation Model for Structured Sequential Medical Data.

Li, J.; Monroe, W.; Ritter, A.; Galley, M.; Gao, J.; and Jurafsky, D. 2016. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.

Murtaza, H.; Ahmed, M.; Khan, N. F.; Murtaza, G.; Zafar, S.; and Bano, A. 2023. Synthetic data generation: State of the art in health care domain. *Computer Science Review*, 48: 100546.

Ng, A.; and Jordan, M. 2001. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 14.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P. F.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in*

Neural Information Processing Systems, volume 35, 27730–27744. Curran Associates, Inc.

Padhi, I.; Schiff, Y.; Melnyk, I.; Rigotti, M.; Mroueh, Y.; Dognin, P.; Ross, J.; Nair, R.; and Altman, E. 2021. Tabular transformers for modeling multivariate time series. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3565–3569. IEEE.

Pal, A.; Karkhanis, D.; Dooley, S.; Roberts, M.; Naidu, S.; and White, C. 2024. Smaug: Fixing failure modes of preference optimisation with dpo-positive. *arXiv preprint arXiv:2402.13228*.

Peña-Guerrero, J.; Nguewa, P. A.; and García-Sosa, A. T. 2021. Machine learning, artificial intelligence, and data science breaking into drug design and neglected diseases. *Wiley Interdisciplinary Reviews: Computational Molecular Science*.

Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; and Finn, C. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 53728–53741. Curran Associates, Inc.

Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; and Finn, C. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Suh, N.; Yang, Y.; Hsieh, D.-Y.; Luan, Q.; Xu, S.; Zhu, S.; and Cheng, G. 2024. TimeAutoDiff: Combining Autoencoder and Diffusion model for time series tabular data synthesizing. *arXiv preprint arXiv:2406.16028*.

Sun, Z.; Shen, S.; Cao, S.; Liu, H.; Li, C.; Shen, Y.; Gan, C.; Gui, L.-Y.; Wang, Y.-X.; Yang, Y.; et al. 2023. Aligning large multimodal models with factually augmented rlhf. *arXiv preprint arXiv:2309.14525*.

Theodorou, B.; Xiao, C.; and Sun, J. 2023. Synthesize high-dimensional longitudinal electronic health records via hierarchical autoregressive language model. *Nature communications*, 14(1): 5305.

Tian, M.; Chen, B.; Guo, A.; Jiang, S.; and Zhang, A. R. 2024. Reliable generation of privacy-preserving synthetic electronic health record time series via diffusion models. *Journal of the American Medical Informatics Association*.

Vapnik, V. 2013. *The nature of statistical learning theory*. Springer science & business media.

Xu, S.; Sun, W. W.; and Cheng, G. 2023. Utility theory of synthetic data generation. *arXiv preprint arXiv:2305.10015*.

Yuan, H.; Zhou, S.; and Yu, S. 2023. Ehrdiff: Exploring realistic ehr synthesis with diffusion models. *arXiv preprint arXiv:2303.05656*.

Cohort Detail

Table 4 shows the disease considered in our finetuning, their ICD code as well as frequencies in the de-identified dataset used from CHIRon model training.

Table 4: ICD-10 code and frequency in training data for disease cohort considered.

Abbreviation	Disease	ICD-10 Codes (DIAG_ICD10)	Frequency
Sep.	Septicemia Except In Labor	A4101, A4150, A4151, A4189, A419, R7881, R6520	0.871%
Surg.	Complications Of Surgical Procedures Or Medical Care	D6481, T82818A, I9581, J95811, K9423, K9189, K912, R5082, T889XXS, I9789, I973, T888XXA, T8130XA, T8131XA, T814XXA, T8189XA, Z283	0.64%
SCD	Sickle Cell Disease	D573	0.036%
DM	Diabetes Mellitus Without Complication	E109, E139, E119, R7301, R7302, R7309, R81, Z9641	11.13%
Lipid.	Disorders Of Lipid Metabolism	E780, E781, E782, E785	4.857%
Fluid.	Fluid And Electrolyte Disorders	E870, E871, E872, E873, E874, E869, E860, E861, E8779, E875, E876, E878, T783XXA	21.5%
AMI	Acute Myocardial Infarction	I2109, I2119, I2129, I214, I213	0.524%
Cond.	Conduction Disorders	I442, I4430, I440, I441, I4469, I447, I4510, I454, I452, I4581, I4589, I459, Z959, Z950, Z95810, Z95818, Z45018, Z4502	1.767%
Dys.	Cardiac Dysrhythmias	I471, I472, I479, I4891, I4892, I491, I4949, R001, I499, R000, R002	6.643%
CVD	Acute Cerebrovascular Disease	I609, I619, I6200, I629, I63239, I6330, I669, I6340, I6350, I6789	0.648%
Pneu.	Pneumonia Except That Caused By Tuberculosis Or Sexually Transmitted Disease	J17, J129, J181, J158, J156, J159, J168, J180, J189	1.611%
Pleu.	Pleurisy; Pneumothorax; Pulmonary Collapse	J869, R091, J942, J948, J918, J939, J9819	0.516%
Resp.	Respiratory Failure; Insufficiency; Arrest Adult	J9600, J9620, J9690, J80, J9610, Z9911, Z9981	1.167%
GI Hem.	Gastrointestinal Hemorrhage	K228, K254, K625, K920, K921, K922	1.517%
Renal.	Acute And Unspecified Renal Failure	N170, N178, N179, N19	1.319%
Shock	Shock	R579, R570, R6521, R578	0.341%

Implementation Details

CHIRon + RLTF: For RLTF, we finetune a separate checkpoint for each target code using a batch size of 8, training for 6 epochs with an initial learning rate of 1e-5. The training is conducted on a computing node equipped with a single Nvidia V100 GPU (16GB GPU memory) and 128GB of RAM. The fine-tuning process takes approximately 5 minutes per checkpoint. For **CHIRon + RLTF**, we modify the `DPOTrainer` from Hugging Face's TRL library, adding a positive regularization term to the DPO loss function.

LSTM: We implement an LSTM-based model for sequence generation, using an embedding size of 128, a hidden size of 256, and 3 layers for the LSTM. The model is trained with a batch size of 128 for 50 epochs, using an initial learning rate of 1e-3 and Adam optimizer. The training is performed on a computing node equipped with a single Nvidia V100 GPU (16GB GPU memory). For generation, we use a beam search strategy and stop sequence generation when an end-of-sequence (EOS) token is

predicted or when the maximum sequence length of 512 tokens is reached. The same tokenizer and vocabulary from CHIRon is used.

HALO: HALO utilizes stacked transformer decoder model with multi granularity to generate longitudinal medical code sequence at both the visit level and code level based on the history of all previous visits. We follow the official implementation of HALO (Theodorou, Xiao, and Sun 2023) to train the model on the same training data as used in our method. The label codes in HALO are the target codes of interest in this paper. The vocabulary size is 7954, which is the same as our method. The maximal number of visits we consider is also 48. The other model and optimization parameters are set as the default values used in the https://github.com/btheodorou99/halo_inpatient.

Table 5: Precision, Recall and F1 Scores for Different Methods and Target Codes on Generation of Target Code.

Disease	Halo	LSTM	CHIRON	DPO
Sep.	0.69/0.11/0.19	0.90/0.06/0.12	0.64/0.23/0.34	0.62/0.35/0.45
SCD	1.00/0.01/0.02	0.00/0.00/0.00	1.00/0.03/0.05	0.50/0.25/0.33
Surg.	0.21/0.02/0.04	0.91/0.02/0.03	0.13/0.07/0.09	0.09/0.24/0.13
DM	0.64/0.15/0.24	0.75/0.14/0.23	0.78/0.24/0.37	0.66/0.57/0.61
Lipid.	0.55/0.09/0.15	0.63/0.07/0.13	0.63/0.20/0.30	0.57/0.36/0.44
Fluid.	0.67/0.03/0.06	0.77/0.03/0.05	0.46/0.04/0.08	0.11/0.19/0.14
AMI	0.74/0.06/0.11	0.81/0.07/0.14	0.72/0.08/0.14	0.44/0.40/0.42
Cond.	0.71/0.05/0.09	0.82/0.04/0.08	0.65/0.12/0.20	0.22/0.29/0.25
Dys.	0.69/0.04/0.08	0.74/0.07/0.13	0.77/0.06/0.11	0.48/0.36/0.41
CVD	0.73/0.02/0.04	0.86/0.03/0.05	0.44/0.04/0.07	0.41/0.24/0.30
Pneu.	0.80/0.07/0.13	0.88/0.06/0.11	0.37/0.21/0.27	0.09/0.26/0.14
Pleu.	0.65/0.04/0.08	0.71/0.03/0.06	0.52/0.05/0.09	0.06/0.18/0.09
Resp.	0.81/0.05/0.09	0.93/0.07/0.12	0.68/0.27/0.38	0.63/0.37/0.47
GI Hem.	0.72/0.02/0.04	0.77/0.02/0.05	0.67/0.03/0.07	0.18/0.22/0.20
Renal.	0.78/0.06/0.11	0.87/0.07/0.13	0.58/0.20/0.30	0.58/0.29/0.38
Shock	0.82/0.04/0.08	0.90/0.03/0.05	0.62/0.12/0.21	0.52/0.34/0.41

Table 6: AUROC Scores for Different Methods and Target Codes on Downstream Utility Evaluation.

Model Disease	Real	Halo	LSTM	CHIRON	DPO
Sep.	0.963883	0.763294	0.757402	0.950971	0.966863
SCD	0.576595	0.447568	0.472366	0.439055	0.523485
Surg.	0.789956	0.591065	0.583039	0.622276	0.772995
DM	0.954576	0.912782	0.920986	0.936084	0.939164
Lipid.	0.870144	0.786727	0.766577	0.828785	0.834174
Fluid.	0.849671	0.680248	0.689911	0.690835	0.729934
AMI	0.964426	0.743526	0.725831	0.791430	0.952727
Cond.	0.928802	0.828109	0.837644	0.852052	0.896331
Dys.	0.870332	0.734622	0.771347	0.729618	0.854042
CVD	0.952163	0.641987	0.678373	0.607578	0.926226
Pneu.	0.907858	0.772562	0.752410	0.878474	0.835806
Pleu.	0.920464	0.728271	0.726257	0.809138	0.836274
Resp.	0.971665	0.810249	0.796211	0.959314	0.971213
GI Hem.	0.817911	0.569127	0.605476	0.528093	0.770189
Renal.	0.948784	0.874256	0.864835	0.928585	0.938689
Shock	0.991505	0.896535	0.880568	0.919362	0.980769

Table 7: AUROC Scores for Different Methods and Target Codes with Data Augmentation.

Model Disease	Real	Halo	LSTM	CHIRON	DPO
Sep.	0.963883	0.957284	0.963633	0.964703	0.968165
SCD	0.576595	0.552546	0.577321	0.578732	0.593915
Surg.	0.789956	0.758249	0.769419	0.779733	0.805474
DM	0.954576	0.950263	0.951639	0.951164	0.951946
Lipid.	0.870144	0.861792	0.866152	0.865382	0.866047
Fluid.	0.849671	0.816476	0.828953	0.845331	0.825845
AMI	0.964426	0.951325	0.953470	0.954856	0.969395
Cond.	0.928802	0.920657	0.922438	0.923409	0.923387
Dys.	0.870332	0.857298	0.860837	0.861953	0.869598
CVD	0.952163	0.926174	0.930452	0.942206	0.944967
Pneu.	0.907858	0.882461	0.897954	0.904490	0.878901
Pleu.	0.920464	0.912675	0.920021	0.918651	0.932653
Resp.	0.971665	0.964218	0.968671	0.972935	0.976146
GI Hem.	0.817911	0.790863	0.792729	0.806536	0.820504
Renal.	0.948784	0.942732	0.946204	0.945539	0.948128
Shock	0.991505	0.987896	0.991323	0.992183	0.991389