# From Complexity to Clarity: Transforming Chest X-ray Reports with Chained Prompting (Student Abstract)

**Sujoy Nath**[1], **Arkaprabha Basu**[2], **Kushal Bose**[3], **Swagatam Das**[3]
[1]Netaji Subhash Engineering College
[2]Institute for Advancing Intelligence, TCG Crest
[3]Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata
arkaprabha17@gmail.com, swagatam.das@isical.ac.in

### Abstract

In the rapidly advancing field of AI-assisted medical diagnosis, the generation of medical reports for Chest X-rays (CXR) has significantly improved with the increased availability of radiographs and their corresponding reports. However, these reports often contain complex medical terminology, making them difficult for patients and non-healthcare professionals to understand. In this study, we introduce a strategy called Chained Prompting for Improved Readability of Medical Reports (CPIR-MR), which translates original medical reports into more comprehensible language. Our primary contribution is the creation of a new extension to the IU X-Ray dataset, providing Simplified Medical Reports (SMRs) generated by CPIR-MR. Additionally, we demonstrate that standard methodologies can effectively produce these simplified reports by proposing a multi-modal text decoder (MTD) that combines BLIP with a classification network to generate simplified medical explanations (SMEs) when fine-tuned on SMRs.

## Introduction

The Medical Report Generation encompassing Chest X-ray images has garnered widespread attention within the research community. The quality of the generated reports typically hinges on the fine-tuning of LLMs and the underlying radiographic features. The prevailing setup predominantly generates reports containing domain-specific medical terminologies creating deterrence for the comprehension of patients or non-healthcare professionals. This fact incentivizes us to design an efficient system that aims to create comprehensible reports, devoid of professional medical terminologies. We introduce CPIR-MR to produce SMRs as an extension of the IU X-Ray dataset (Demner-Fushman et al. 2016) that are not only understandable to any patient but also preserve necessary information. We also observe that standard methods is capable to generate SMEs fine-tuned with SMRs by introducing a model consisting of multi-modal Text Decoder (MTD) with Biomedical Condition Embedding (BCE) prompting which is guided by BLIP embeddings and a classification network as input to MTD.



## Proposed Method

**CPIR-MR** We explore a prompting technique using dynamic, iterative construction, incorporating few-shot (Brown et al. 2020), Chain-of-Thought (CoT) (Wei et al. 2022) and controlled generation. The technique incorporates $n$ exemplars of original medical findings, building prompts sequentially to leverage in-context learning. It employs targeted output structuring for seven-line layperson explanations, explicit constraints for output filtering, and ordinal numbering for coherence. This approach combines prompt chaining with a lower number of API accesses with a larger output token size, constrained generation, and structured formatting to guide the model in producing comprehensive medical explanations. We outline the CPIR-MR as follows.

---

"I have {n} examples of original findings. Add your notions in place of XXXX. Strictly DO NOT SUGGEST MEDICINES and PRACTICES.

The prompt is constructed iteratively as follows:

prompt += "$\{j\}.\{text\}\backslash n$"[1]

prompt= For each finding, generate a more detailed finding for normal human understanding in 7 lines and output those in points

where $n = |\text{texts}|$. Then, for each $j \in [1, n]$:

$$\texttt{prompt} \mathrel{+}= \text{``}\; \overbrace{j}^{\text{index}} \;.\; \overbrace{\texttt{ordinal}_j}^{\text{ordinal number}} \;:\; \underbrace{\texttt{(SMR)}}_{\text{placeholder}} \;\text{''}\;\text{''}$$

---

**Multi-modal Text Decoder** To demonstrate the efficacy of CPIR-MR, we designed a multi-modal text decoder (MTD) that integrates BLIP (Li et al. 2022) embeddings, classification network, and text decoder trained with SMR. The chest X-rays are passed through BLIP encoders, creating embeddings $\mathbf{E}_f$ and $\mathbf{E}_l$, combined into $\mathbf{E}_i$. Concurrently, a classification network generates class-specific probabilities $\mathcal{P} = \{p_1, p_2, ..., p_n\}$. These probabilities are subsequently added with embeddings using the Biomedical Condition Embedding (BCE) prompting technique. With pre-prompting and post-prompting text, this combined information is fed into a QLoRA (Dettmers et al. 2024) model to generate SME.

[1] $j$ is the index of the current iteration and $text$ represents the original findings from IU X-Ray dataset to be added.

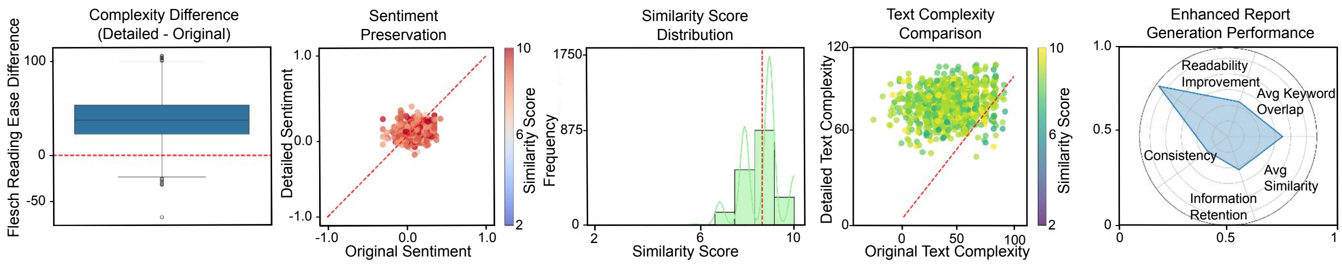


Figure 1: Evaluation of SMR and original findings through (a) complexity distribution, (b) sentiment consistency, (c) similarity scores, (d) text complexity comparison, and (e) performance metrics in different aspects.

## Experiments

The evaluation of CPIR-MR presents a challenge due to the instability of NLP techniques in extracting medically accurate keywords from both original findings and SMR. Therefore, we introduce Chained Prompting for Medical Keyword Extraction (CPMK-E), exploiting the advanced capabilities of *Gemini 1.5 Flash* (Team et al. 2023). CPMK-E employs two CoT mechanisms, incorporating pre-prompting and post-prompting phrases to enhance extraction accuracy. The method consists of $2n$ findings, where $n$ represents the number of original findings and their corresponding SMRs. The prompt progressively analyzes the similarity between the original findings and the SMRs by scoring them on a scale ranging from 1 to 10 with subjective explanation. We noticed that, the corresponding explanation enhanced the performance of the Gemini scoring mechanism. Furthermore, we examine the results in the evaluation stage to show that the original findings and SMRs were semantically similar and could be supported by comparable medical interpretations. Our implementation is available at https://github.com/Thecoder1012/CPIR-MR.

**Quantitative Evaluations** We use FRE (Flesch Reading Ease) score (Farr, Jenkins, and Paterson 1951) and lexicon-based sentiment analysis to determine the complexity difference distribution. Figure 1(a) indicates a higher complexity with possible outliers. This is endorsed by the text complexity comparison (Fig. 1(d)), retaining reduced complexity while ensuring consistency across different levels. Sentiment preservation (Fig. 1(b)) exhibits a strong positive correlation clustered around the diagonal, affirming medical proficiency in maintaining tone while simplifying language. The similarity score distribution (Fig. 1(c)), biased toward higher values (8-10 range), affirms robust content conservation. Unlike other evaluations, Fig. 1(e) reveals exceptional readability improvement and consistency performance, balanced against strong information retention and keyword concordance. This panoptic analysis, followed by CPMK-E, demonstrates the efficacy of SMRs while preserving critical details and sentiment.

**Qualitative Evaluations** Our qualitative analysis illustrates the effectiveness of CPIR-MR in generating SMR that successfully transforms intricate medical findings into simplified language while preserving pertinent information in Fig. 2. It converts technical terms such as *pleural effusion* into a more understandable description `fluid in the lungs`. Structured output format consistently simplifies medical terminologies, replacing *pneumothorax* with `collapsed lung`, while maintaining the original meaning. The context awareness of CPIR-MR is evident in its ability to interpret *cardiomediastinal silhouette* to `The size and shape of the heart and surrounding area`, exemplifying how the technique effectively bridges the gap between professional medical language and simplified explanations. All outputs of CPIR-MR are extracted through API access of *Gemini 1.5 Flash*.

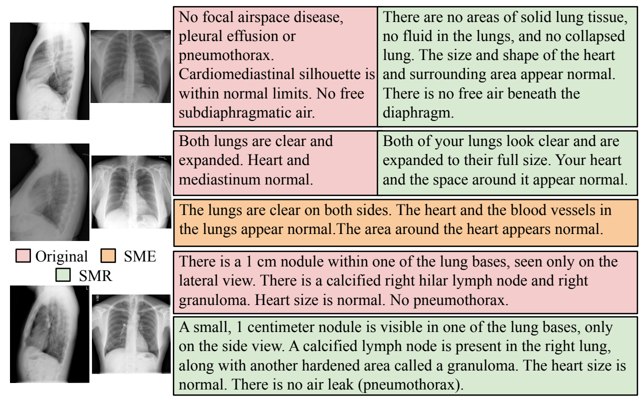


Figure 2: A comparative study between original medical reports, Simplified Medical Reports (SMR) generated by CPIR-MR for chest X-ray images, and Simplified Medical Explanations (SME) from pretrained MTD, demonstrating improved readability while maintaining essential medical information.

## Conclusion & Future Works

We proposed a chained prompting strategy, CPIR-MR, enabling the LLM to generate medical reports that contain comprehensible language for the patients. The reports are free from complicated medical jargon and also preserve critical information. We also demonstrated that existing methods utilizing an extended IU X-ray dataset can generate SMEs by employing appropriate prompt engineering. The approach will further ensure an easy understanding of accurate medical information.

## Acknowledgments

We are grateful to Dr. Sankha Subhra Mullick and the Machine Learning Research Group (MLRG), Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, for their unwavering support and valuable assistance towards the completion of this work.

## References

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language Models are Few-Shot Learners. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33. Curran Associates, Inc.

Demner-Fushman, D.; Kohli, M. D.; Rosenman, M. B.; Shooshan, S. E.; Rodriguez, L.; Antani, S.; Thoma, G. R.; and McDonald, C. J. 2016. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2).

Dettmers, T.; Pagnoni, A.; Holtzman, A.; and Zettlemoyer, L. 2024. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36.

Farr, J. N.; Jenkins, J. J.; and Paterson, D. G. 1951. Simplification of Flesch reading ease formula. *Journal of applied psychology*, 35(5): 333.

Li, J.; Li, D.; Xiong, C.; and Hoi, S. 2022. BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*. PMLR.

Team, G.; Anil, R.; Borgeaud, S.; Alayrac, J.-B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A. M.; Hauth, A.; Millican, K.; et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.