

Enhancing Large Language Models for Clinical Decision Support by Incorporating Clinical Practice Guidelines

David Oniani^{1,*}, Xizhi Wu^{1,*}, Shyam Visweswaran^{2,3,4}, Sumit Kapoor⁵, Shravan Kooragayalu⁶,
Katelyn Polanska¹, and Yanshan Wang^{1,2,3,4,†}

¹*Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA, USA*

²*Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA*

³*Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, USA*

⁴*Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, USA*

⁵*Department of Critical Care Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, USA*

⁶*Department of Pulmonary Medicine, University of Pittsburgh Medical Center Western Maryland, Cumberland, MD, USA*

Abstract

Background

Large Language Models (LLMs), enhanced with Clinical Practice Guidelines (CPGs), can significantly improve Clinical Decision Support (CDS). However, methods for incorporating CPGs into LLMs are not well studied.

Methods

We develop three distinct methods for incorporating CPGs into LLMs: Binary Decision Tree (BDT), Program-Aided Graph Construction (PAGC), and Chain-of-Thought-Few-Shot Prompting (CoT-FSP). To evaluate the effectiveness of the proposed methods, we create a set of synthetic patient descriptions and conduct both automatic and human evaluation of the responses generated by four LLMs: GPT-4, GPT-3.5 Turbo, LLaMA, and PaLM 2. Zero-Shot Prompting (ZSP) was used as the baseline method. We focus on CDS for COVID-19 outpatient treatment as the case study.

Results

All four LLMs exhibit improved performance when enhanced with CPGs compared to the baseline ZSP. BDT outperformed both CoT-FSP and PAGC in automatic evaluation. All of the proposed methods demonstrated high performance in human evaluation.

Conclusion

LLMs enhanced with CPGs demonstrate superior performance, as compared to plain LLMs with ZSP, in providing accurate recommendations for COVID-19 outpatient treatment, which also highlights the potential for broader applications beyond the case study.

*Equal contribution

†Corresponding author: yanshan.wang@pitt.edu

Introduction

Large Language Models (LLMs)¹ have opened new opportunities in many fields, with applications ranging from creative writing [1] to chemical research [2]. LLMs have also been studied in the context of healthcare, where they have proven successful at numerous tasks, including clinical reasoning [3], question-answering [4], and medical evidence summarization [5], among other applications. Moreover, recent research produced powerful healthcare LLMs, such as GatorTron [6] and Med-PaLM [7].

The success and popularity of LLMs are at least partially due to their In-Context Learning (ICL) capability [8]. ICL is a learning method that does not require model parameter updates but “learns” from the given context. It can be thought of as a higher-order function that takes several few-shot examples (i.e., the context) to produce the predictor function [9]. The emergence of ICL-capable LLMs has provided an alternative to a typical pre-train, fine-tune, and predict pipeline, with prompting replacing fine-tuning. Prompting is a novel paradigm where, with the use of textual prompts, downstream tasks are modeled as those typically solved during pre-training [10]. It has given rise to Prompt Engineering (PE) [11], a field that aims to design prompts that enhance ICL and improve LLM reasoning. One of several widely used PE methods [12, 13, 14] is Chain-of-Thought (CoT) [15] prompting, which attempts to improve LLM reasoning by providing intermediate reasoning steps as part of the prompt.

Clinical Decision Support (CDS) supports healthcare decision-making by enabling the timely delivery of evidence-based guidelines at the point of care [16, 17]. Optimizing CDS has the potential to significantly improve clinical workflows, and it has been one of the major topics of discussion in healthcare [18]. The importance of CDS, paired with its recent focus on patient-centered solutions [19, 20], demands novel methods for improving the quality of healthcare decision-making, ultimately facilitating better health outcomes. At the same time, the literature on LLM-driven algorithmic methods that incorporate CPGs is limited. LLMs offer many benefits for CDS, such as timely decision support, ease of use due to their interactive nature, and the ability to capture both Clinical Practice Guidelines (CPGs) and patient-specific information.

In this paper, we propose three methods for improving LLMs for CDS by incorporating CPGs: Binary Decision Tree (BDT), Program-Aided Graph Construction (PAGC), and Chain-of-Thought-Few-Shot Prompting (CoT-FSP). Specifically, we enhance LLMs by incorporating CPGs and evaluate the performance of the proposed approaches on synthetic patient descriptions. The evaluation was done by two physicians from the University of Pittsburgh Medical Center (UPMC). We used Zero-Shot Prompting (ZSP) as the baseline. The experimental results show that the proposed methods provide significant improvements over the baseline, which speaks to the effectiveness of the methods.

Related Work

The field of using LLMs for CDS is rapidly evolving. Here we summarized related work up to the point when this paper was drafted. A recent study considered ChatGPT [21] LLM for improving CDS [22], reaching a conclusion that LLMs can be important complementary parts for optimizing CDS. In another study, researchers examined LLMs for improving CDS in personalized oncology [23], reaching a similar conclusion that LLMs did not reach the quality and credibility of human experts, but generated helpful responses that could complement established procedures. There have also been other studies evaluating LLMs for CDS in specific fields, such as cardiology [24] and orthopedics [25]. It should be noted that none of the studies augmented LLMs with CPGs, nor did they consider advanced prompting methods

¹Note that in this paper, the term “LLM” refers to a *generative* Large Language Model (LLM).

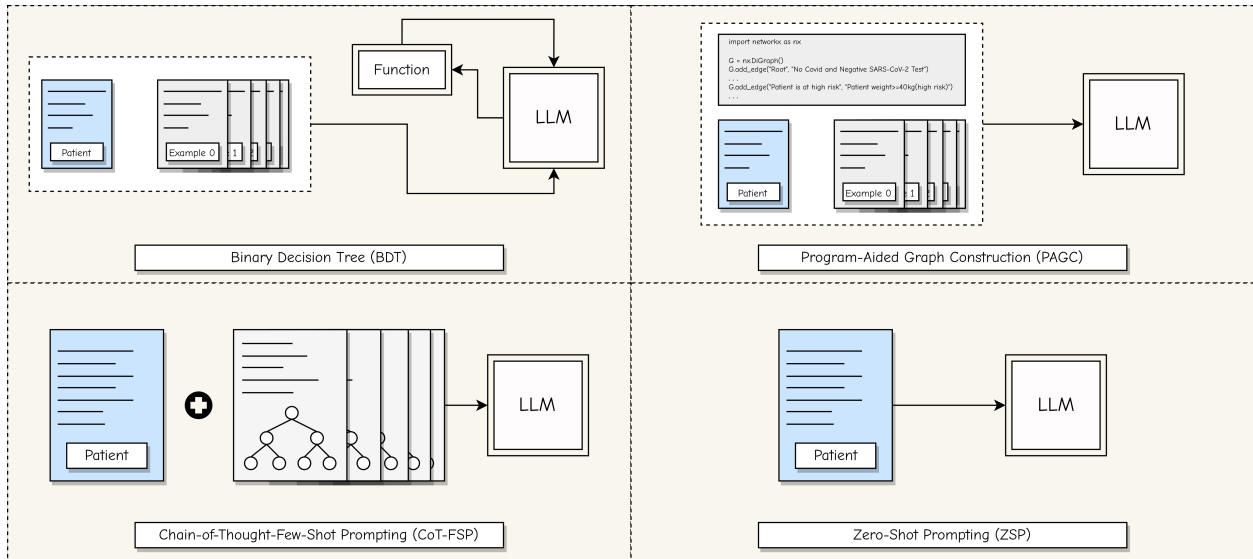


Figure 1: The figure shows the three proposed methods and the baseline Zero-Shot Prompting (ZSP) method. In the case of the Binary Decision Tree (BDT) method, we use a recursive function to call the LLM with prompts. For Program-Aided Graph Construction (PAGC), a program is a part of the prompt passed to the LLM. Chain-of-Thought-Few-Shot Prompting (CoT-FSP) uses several few-shot examples for guiding the LLM. Finally, ZSP only takes the patient description to produce the result. *Note that for BDT, PAGC, and CoT-FSP, the prompt typically contains a task description, patient description, and several few-shot examples besides additions specific to a method (e.g., a program in the case of PAGC).*

such as the ones proposed in our study. To the best of our knowledge, our study is the first one to develop and evaluate such methods.

Methods

We propose three methods enhanced by CPGs as well as a baseline Zero-Shot Prompting (ZSP) method. Figure 1 illustrates the methods.

COVID-19 Treatment Guidelines

We focused on supporting outpatient treatment decisions for Coronavirus Disease 2019 (COVID-19). Because COVID-19 was a novel disease at the start of the pandemic, clinical practice guidelines evolved rapidly in response to new discoveries and treatments. Keeping up with the frequent guideline updates was difficult for healthcare providers. In such a situation, a CDS tool like LLM can be valuable in providing reliable support for decision making in clinical care. For the COVID-19 outpatient treatment guidelines, we used the Centers for Disease Control and Prevention (CDC) and Infectious Diseases Society of America (IDSA) COVID-19 Outpatient Treatment Guidelines [26]. The CPGs provide a comprehensive step-by-step approach that outlines outpatient treatment options. We slightly revised the guidelines to reflect insights and recommendations from our physicians. Body weight, in particular, was used as a key determinant for COVID-19 treatment since many medication dosages are weight-based. We considered the patient's underlying medical conditions, COVID-19 related symptoms, and the duration of symptoms to determine if they were in the high-risk versus low-risk category. Based on this information, the guidelines offer three outpatient treatment options for high-risk patients: *Paxlovid*, *Remdesivir*, or *Molnupiravir*, and one treatment suggestion for low-risk patients: *Supportive Care*. The revised COVID-19 Outpatient Treatment

Guidelines were used to build LLM-enhanced CDS systems for COVID-19 outpatient treatment decision support and can be found in the [Appendix](#).

Binary Decision Tree (BDT)

Binary Decision Tree (BDT) is a recursive algorithm guided by CPGs. We model CPGs as a binary tree and use an LLM to navigate the tree. There are two calls to the LLM. In the first call, an input to an LLM contains a description of the task, several few-shot examples, and a question from CPGs (i.e., node value). The generated output is stored. In the second call, we use the response from the first call to ask a yes-no question of whether the answer was affirmative or negative. Depending on the generated response, we recurse to the left or right subtree till reaching one of the leaf nodes. Algorithm 1 is the pseudocode for the method.

Algorithm 1 Binary Decision Tree (BDT) algorithm

Require: P_1 : task description, patient description, and several few-shot examples

Require: P_2 : task description and several few-shot examples for YES/NO question

Require: M : pre-initialized Large Language Model (LLM)

```

1: procedure BDT(node, responses=())
2:   if not (node.left or node.right) then                                ▷ Base case
3:     return (node.value, responses)
4:   end if
5:    $O_1 \leftarrow M(P_1 + \text{node.question})$                                 ▷ First LLM call
6:   responses.add( $O_1$ )
7:    $O_2 \leftarrow M(\text{"Response YES or NO?"} + P_2 + \text{node.question} + O_1)$     ▷ Second LLM call
8:   if  $O_2 = \text{"YES"}$  then                                             ▷ Recursive step: recurse to left or right subtree
9:     (result, responses) = BDT(node.left, responses)
10:  else if  $O_2 = \text{"NO"}$  then
11:    (result, responses) = BDT(node.right, responses)
12:  else
13:    throw Exception
14:  end if
15:  return (result, responses)
16: end procedure

```

Chain-of-Thought-Few-Shot Prompting (CoT-FSP)

In this method, we use a prompt to describe the CPGs. Specifically, we construct a prompt containing a description of the task, five few-shot examples, and an if-else description of CPGs.

The if-else description is an algorithmic seven-step CoT prompt. At every step, CoT-FSP decides between selecting a specific treatment and progressing to the next step. The selection of a specific treatment ends the chain. In the first step, the algorithm checks whether the patient has COVID-19, and if they do, it moves to the second step. Otherwise, it provides the treatment suggestion "Vaccination and booster is recommended". In the second step, the algorithm checks whether the patient needs hospitalization or increased oxygen. In the third step, the algorithm looks for risk factors for severe COVID-19. If the hypothetical patient has risk factors, it goes to the fourth, fifth, and seventh steps to find an appropriate treatment such as Paxlovid, Remdesivir, or Molnupiravir. If none of the treatment options are available for the patient, it will suggest the patient get "monitoring and supportive care" and end the process.

The five few-shot examples are five hypothetical patient descriptions and five answers on providing treatment suggestions after going through the step-by-step if-else process. We carefully selected more complex patient cases to improve ICL and handle challenging scenarios. Algorithm 2 shows the pseudocode for the method.

Algorithm 2 Chain-of-Thought-Few-Shot Prompting (CoT-FSP) algorithm

Require: P_1 : patient description, task description, and several CPG-enhanced few-shot examples

Require: P_2 : text description of the tree via if-else statements

Require: M : pre-initialized Large Language Model (LLM)

1: **procedure** CoT-FSP

2: $P \leftarrow P_1 + P_2$

3: $O \leftarrow M(P)$

4: **return** O

5: **end procedure**

Program-Aided Graph Construction (PAGC)

Inspired by Program-aided Language Models (PAL) [27], the PAGC approach defines a `networkx`² graph directly in the prompt and uses five few-shot examples to select candidates.

The `networkx` graph represents the revised COVID-19 CPGs, with eight leaf nodes representing eight COVID-19 outpatient treatment suggestions and each internal node representing the patient's possible medical condition or characteristics. The method is accompanied by a candidate selection algorithm that selects candidate nodes in the `networkx` graph that match the patient description. After selecting all the candidate nodes, a path to one of the leaf nodes is created.

The five few-shot examples are designed to be challenging, improving the LLM's ability to handle complex cases. Algorithm 3 illustrates the pseudocode for the method.

Algorithm 3 Program-Aided Graph Construction (PAGC) algorithm

Require: P : task description and patient description

Require: C : code description of the algorithm

Require: M : pre-initialized Large Language Model (LLM)

1: **procedure** PAGC

2: $O \leftarrow LLM(P + C)$

3: **return** O

4: **end procedure**

Zero-Shot Prompting (ZSP)

Zero-Shot Prompting (ZSP) method constructs a prompt that contains a task description and patient description, followed by a query. In contrast with BDT, CoT-FSP, and PAGC methods, ZSP does not utilize CPGs or few-shot examples. Instead, ZSP produces the output directly from a patient description. Algorithm 4 describes this method.

²<https://networkx.org/>

Algorithm 4 Zero-Shot Prompting (ZSP) algorithm

Require: P : task description and patient description

Require: M : pre-initialized Large Language Model (LLM)

```
1: procedure ZSP
2:    $O \leftarrow LLM(P)$ 
3:   return  $O$ 
4: end procedure
```

Synthetic Patient Dataset

For rigorous evaluation of the proposed methods, we created a synthetic patient dataset based on CPGs. We modeled CPGs as a binary tree with eight leaf nodes. There were thirteen different paths to reach one of the leaves. Considering three different difficulty categories (*easy*, *medium*, and *hard*), we created $3 \times 13 = 39$ synthetic patients to examine all possible paths. All the synthetic patients, including their *difficulty*, are listed in the [Appendix](#).

In the *easy* category, patient descriptions adhered closely to the phrases used in the COVID-19 CPGs. Such an approach allowed for assessing methods in making correct decisions at each checkpoint and identifying the correct treatment path following the guidelines.

The *medium* category contained patient descriptions that used synonyms and phrases semantically equivalent to those in the CPGs. For example, instead of describing a patient as one “who had a positive COVID-19 test”, the description mentioned “a positive Polymerase Chain Reaction (PCR) test”. The goal was to evaluate the ability to comprehend the semantic meaning in the patient description, correctly map it to the relevant checkpoint, and make the correct decision.

The *hard* category included patient descriptions that differed considerably from the phrases used in the CPGs. We also intentionally included unrelated and subtle information. Hence, the methods would not only need to extract symptoms from the description but also filter out irrelevant information.

Evaluation

The evaluation has two stages: best method selection and human annotation. We considered four LLMs, GPT-4 [28], GPT-3.5 Turbo [29], LLaMA [30], and PaLM 2 [31], for a more holistic evaluation.

In the method selection stage, we automatically evaluate the approaches and select those with an F-score greater than 0.5. Systematically evaluating few-shot classification performance is challenging as predictions on small datasets can be unstable [32]. To ensure the reproducibility and robustness of results, we performed four runs with random seeds (9631, 4603, 6367, and 4057) and computed the mean for each metric³.

The human evaluation has two rounds: a round for ensuring agreement among two UPMC physicians (authors S.K. and S.K.) and the final LLM response annotation round⁴. We used random seed 8747 for generating the responses. In the first round, the annotators rated ten questions for evaluation. We used Gwet’s AC1 for computing Inter-Annotator Agreement (IAA) [33]. For interpreting IAA scores, we used Landis and Koch scale [34].

³Note that PaLM 2 does not provide a way to set the seed, and we just re-ran the model several times. Besides, GPT-based models do not always produce the same output when called with the same random seed, which is a known issue: <https://community.openai.com/t/seed-param-and-reproducible-output-do-not-work/487245>.

⁴Note that model names (e.g., GPT-4, PaLM 2, etc.) were hidden from annotators to ensure fair and unbiased evaluation.

Question	Score	Interpretation
Presence of Incorrect Medical Content	0.87	Almost Perfect
Omission of Content	0.60	Moderate
Presence of Possible Harmful Content	0.77	Substantial

* Gwet’s AC1 scores for three different questions asked to the physicians. Landis and Koch interpretations for the scores are also included. “Omission of Content” is the only category with *Moderate* agreement. It should also be noted that even in this case, the score signifies borderline substantial agreement (i.e., would have been substantial if the obtained score was 0.61).

For physician annotations, we have three different evaluation categories: *presence of incorrect medical content*, *ommission of content*, and *presence of possible harmful content* [7]. We designed these categories to comprehensively capture and represent the errors and mistakes we found in our testing process. The human evaluation scale ranges from 0 to 2 for all three question categories and is described in detail in the [Appendix](#).

The *presence of incorrect medical content* category aims to discern the accuracy of the generated answers concerning medical content. The objective is to determine whether the responses contain information conflicting with established medical guidance. The *omission of content* evaluates the completeness of the generated answers and aims to identify instances where essential information is either inadequately addressed or omitted entirely. Understanding the model’s ability to provide comprehensive responses is crucial for its practical utility in clinical decision-making. The *presence of possible harmful content* is an essential consideration in our evaluation framework as the identification of potential harm that LLMs could pose to users relying on the generated answers. This category is responsible for ensuring the safety and ethical use of the methods in real-world applications, particularly in sensitive domains such as healthcare.

Table 1 shows the results for the first round of annotations. Since the score for “omission of content” was *Moderate* per Landis and Koch scale, we held a meeting where the physicians discussed discrepancies, reaching the perfect agreement on the annotations for this specific category. We note that 0.61 would have already been a substantial agreement, but we got 0.60, which was 0.01 less.

Results

LLM	BDT	CoT-FSP	PAGC	ZSP
GPT-4	1.00 ¹	0.97 ²	0.83 ⁴	0.47
GPT-3.5 Turbo	0.85 ³	0.69 ⁶	0.38	0.26
LLaMA-13b	0.37	0.31	0.42	0.31
PaLM 2	0.71 ⁵	0.58 ⁷	0.41	0.01

* F-score metrics averaged over four different runs with four randomly generated four-digit prime random seeds: 9631, 4603, 6367, 4057. The selection of methods based on F-score was done using this table. Only those with F-score > 0.5 were selected. Selected methods have a rank number as superscript.

We first performed the automatic evaluation and then picked the seven most performant methods for the human evaluation stage, namely *GPT-4 BDT*, *GPT-4 CoT-FSP*, *GPT-3.5 Turbo BDT*, *GPT-4 PAGC*,

PaLM 2 BDT, GPT-3.5 Turbo CoT-FSP, and PaLM 2 CoT-FSP.

Pairing our methods with LLaMA-13b did not yield promising results. As a result, the model was excluded entirely from the human evaluation stage. In its defense, however, LLaMA-13b is significantly smaller than the other models, and a larger LLaMA model could perform better. ZSP performed poorly across all models, and aw also excluded it due to this reason. Table 2 shows and ranks the selected methods.

In the human evaluation stage, two UPMC physicians (authors S.K. and S.K.) annotated 49 responses generated using the proposed methods. We generated seven responses using every method and computed the mean score for each evaluation category. Table 3 shows the human evaluation results.

Table 3: Human evaluation scores.*							
Category	GPT-4 BDT	GPT-4 CoT- FSP	GPT- 3.5 Turbo BDT	GPT-4 PAGC	PaLM 2 BDT	GPT- 3.5 Turbo CoT- FSP	PaLM 2 CoT- FSP
Presence of Incorrect Medical Content	1.14	2.00	1.71	2.00	2.00	1.14	1.14
Omission of Content	0.86	2.00	2.00	2.00	2.00	1.43	2.00
Presence of Possible Harmful Content	1.29	2.00	1.86	2.00	2.00	1.14	1.14
Overall Average	1.10	2.00	1.86	2.00	2.00	1.24	1.43

* Human evaluation results for the methods selected via automatic evaluation. We computed the average score for each evaluation category. GPT-4 CoT-FSP, GPT-4 PAGC, and PaLM 2 BDT obtained the perfect scores, with the highest quality of responses (rating of 2) for all queries. GPT-3.5 Turbo BDT was next, with an overall average of 1.86. PaLM 2 CoT-FSP, GPT-3.5 Turbo CoT-FSP, and GPT-4 BDT showed the worst performance, with average scores of 1.43, 1.24, and 1.10, respectively.

GPT-4 CoT-FSP, GPT-4 PAGC, and PaLM 2 BDT obtained the perfect scores, with the highest quality of responses (rating of 2) for all queries. GPT-3.5 Turbo BDT was next, with an overall average of 1.86. PaLM 2 CoT-FSP, GPT-3.5 Turbo CoT-FSP, and GPT-4 BDT showed the worst performance, with average scores of 1.43, 1.24, and 1.10, respectively.

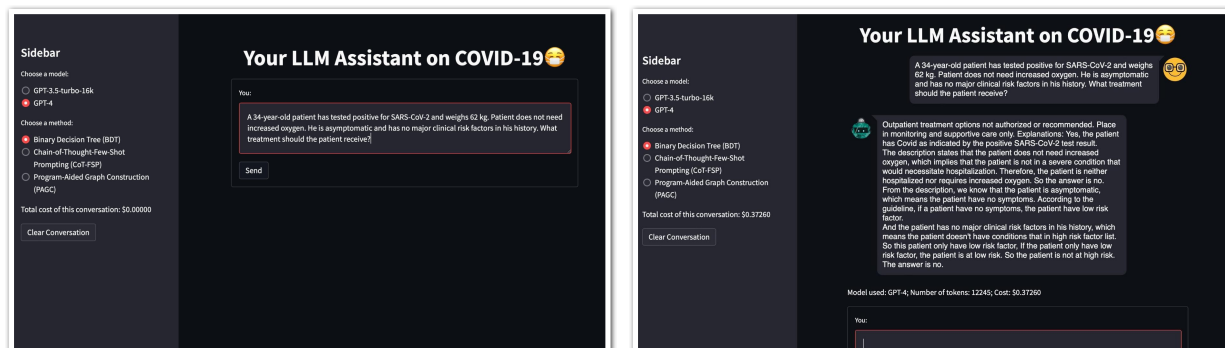
We should note that in all cases, the average score was above 1.00, which speaks to the promising performance of the approaches. Interestingly, the best method in the automatic evaluation, GPT-4 BDT, did not perform as well in human evaluation. At the same time, among models that utilized BDT, PaLM 2 and GPT-3.5 Turbo demonstrated strong performance across all queries.

Discussion

Aiming to improve CDS, we proposed three new methods for incorporating CPGs into LLMs. We took COVID-19 CDS as the case study and created a set of synthetic patients. We first selected methods with an F-score higher than 0.5 and then performed a rigorous human evaluation of the approaches. The proposed methods outperformed the baseline ZSP and have shown high performance across the tasks. Our work also opens future directions of research.

The promising performance of the proposed methods allows for incorporating them as part of software systems, such as chatbots, that streamline CDS. As part of this work, we have also built a prototype, shown in Figure 2. The system takes the prompt from a medical professional as input and provides treatment recommendations, enhancing the clinical pipeline and facilitating accurate healthcare delivery.

A case study exploring the systems utilizing the proposed methods in real-world clinical settings is another interesting future direction of research.



(a) Prototype chatbot interface *before* generating the response. (b) Prototype chatbot interface *after* generating the response.

Figure 2: The figure shows the user interface of the chatbot system that implements the proposed methods. Figure 2b shows the interface with the prompt but *before* generating the response. Figure 2a shows the interface *after* after generating the response. We developed the system as part of the research effort to demonstrate real-world implementation of the methods and collect user feedback.

Future research can also focus on improving the proposed methods. An example of such improvement could be tree pruning or node skipping in the case of BDT. In other words, some responses may contain enough information to answer several questions, allowing for shortcuts across the tree. The proposition of completely different LLM-based approaches for enhancing CDS can also be another avenue of exploration.

In addition, we would like to emphasize the need for transparent and privacy-aware development of LLMs, which is especially important in the case of healthcare [35], where even a small-scale accidental leak of Protected Health Information (PHI) can be dangerous. Open-source models offer transparency, allowing researchers to audit the codebase if necessary. On the other hand, proprietary LLMs typically offer the Application Programming Interface (API) model, where the codebase is not publicly available, and there is no clarity on where the user or user request data is kept (and for how long). Among the four models used in our study, only LLaMA⁵ and PaLM⁶ were open-source.

Finally, we also note that experiments with LLMs can be costly. In our case, the associated costs for conducting experiments with GPT-4 and GPT-3.5 Turbo were approximately \$500 and \$150, respectively. Thus, we paid \$650 for two models alone, with GPT-4 being roughly 3.3 times more expensive than GPT-3.5 Turbo. And this is with only half of the models requiring a payment. The other two LLMs, LLaMA and PaLM 2, were free-to-use and open-source, with no expenses besides utility costs for running a server. For LLaMA, we used our lab server with Quadro RTX 8000 GPUs. As for PaLM 2, we utilized the provided API⁷.

Acknowledgements

None.

⁵<https://github.com/facebookresearch/llama>

⁶<https://github.com/conceptofmind/PaLM>

⁷<https://ai.google.dev/models/palm>

Contributions

D.O. led the study, designed the experiments, helped conduct the experiments, analyzed the results, and wrote, reviewed, and revised the paper. X.W. designed and conducted the experiments, analyzed the results, and wrote, reviewed, and revised the paper. S.V. conceptualized the study and revised the paper. S.K. and S.K. annotated the results and reviewed the paper. K.P. analyzed the results and reviewed the paper. Y.W. conceptualized the study and reviewed and revised the paper.

References

- [1] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. “Wordcraft: Story Writing With Large Language Models.” In: *27th International Conference on Intelligent User Interfaces*. IUI '22. Helsinki, Finland: Association for Computing Machinery, 2022, pp. 841–852. ISBN: 9781450391443. DOI: [10.1145/3490099.3511105](https://doi.org/10.1145/3490099.3511105). URL: <https://doi.org/10.1145/3490099.3511105>.
- [2] Daniil A. Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. “Autonomous chemical research with large language models.” In: *Nature* 624.7992 (Dec. 2023), pp. 570–578. ISSN: 1476-4687. DOI: [10.1038/s41586-023-06792-0](https://doi.org/10.1038/s41586-023-06792-0). URL: <https://doi.org/10.1038/s41586-023-06792-0>.
- [3] Eric Strong, Alicia DiGiammarino, Yingjie Weng, et al. “Chatbot vs Medical Student Performance on Free-Response Clinical Reasoning Examinations.” In: *JAMA Internal Medicine* 183.9 (Sept. 2023), pp. 1028–1030. ISSN: 2168-6106. DOI: [10.1001/jamainternmed.2023.2909](https://doi.org/10.1001/jamainternmed.2023.2909). eprint: https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/2806980/jamainternal_strong_2023_ld_230023_1693517126.00366.pdf. URL: <https://doi.org/10.1001/jamainternmed.2023.2909>.
- [4] Lambert T. Li, Margaret A. Sinkler, Jeremy M. Adelstein, James E. Voos, and Jacob G. Calcei. “ChatGPT Responses to Common Questions about Anterior Cruciate Ligament Reconstruction Are Frequently Satisfactory.” In: *Arthroscopy: The Journal of Arthroscopic & Related Surgery* (2024). ISSN: 0749-8063. DOI: <https://doi.org/10.1016/j.arthro.2023.12.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0749806323010137>.
- [5] Liyan Tang, Zhaoyi Sun, Betina Ildnay, et al. “Evaluating large language models on medical evidence summarization.” In: *npj Digital Medicine* 6.1 (Aug. 2023), p. 158. ISSN: 2398-6352. DOI: [10.1038/s41746-023-00896-7](https://doi.org/10.1038/s41746-023-00896-7). URL: <https://doi.org/10.1038/s41746-023-00896-7>.
- [6] Xi Yang, Aokun Chen, Nima PourNejatian, et al. “A large language model for electronic health records.” In: *npj Digital Medicine* 5.1 (Dec. 2022), p. 194. ISSN: 2398-6352. DOI: [10.1038/s41746-022-00742-2](https://doi.org/10.1038/s41746-022-00742-2). URL: <https://doi.org/10.1038/s41746-022-00742-2>.
- [7] Karan Singhal, Shekoofeh Azizi, Tao Tu, et al. “Large language models encode clinical knowledge.” In: *Nature* 620.7972 (Aug. 2023), pp. 172–180. ISSN: 1476-4687. DOI: [10.1038/s41586-023-06291-2](https://doi.org/10.1038/s41586-023-06291-2). URL: <https://doi.org/10.1038/s41586-023-06291-2>.
- [8] Tom Brown, Benjamin Mann, Nick Ryder, et al. “Language Models are Few-Shot Learners.” In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf.
- [9] Shivam Garg, Dimitris Tsipras, Percy Liang, and Gregory Valiant. “What Can Transformers Learn In-Context? A Case Study of Simple Function Classes.” In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho. 2022. URL: <https://openreview.net/forum?id=f1NZJ2e0et>.

- [10] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. “Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing.” In: *ACM Comput. Surv.* 55.9 (Jan. 2023). ISSN: 0360-0300. DOI: [10.1145/3560815](https://doi.org/10.1145/3560815). URL: <https://doi.org/10.1145/3560815>.
- [11] Laria Reynolds and Kyle McDonell. “Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm.” In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI EA '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380959. DOI: [10.1145/3411763.3451760](https://doi.org/10.1145/3411763.3451760). URL: <https://doi.org/10.1145/3411763.3451760>.
- [12] Long Ouyang, Jeffrey Wu, Xu Jiang, et al. “Training language models to follow instructions with human feedback.” In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho. 2022. URL: <https://openreview.net/forum?id=TG8KACxEON>.
- [13] Xuezhi Wang, Jason Wei, Dale Schuurmans, et al. “Self-Consistency Improves Chain of Thought Reasoning in Language Models.” In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=1PL1NIMMrw>.
- [14] Shunyu Yao, Dian Yu, Jeffrey Zhao, et al. *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*. 2023. arXiv: [2305.10601](https://arxiv.org/abs/2305.10601) [cs.CL].
- [15] Jason Wei, Xuezhi Wang, Dale Schuurmans, et al. “Chain of Thought Prompting Elicits Reasoning in Large Language Models.” In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho. 2022. URL: https://openreview.net/forum?id=_VjQlMeSB_J.
- [16] Prashila Dullabh, Krysta Heaney-Huls, David F Lobach, et al. “The technical landscape for patient-centered CDS: progress, gaps, and challenges.” In: *Journal of the American Medical Informatics Association* 29.6 (Mar. 2022), pp. 1101–1105. ISSN: 1527-974X. DOI: [10.1093/jamia/ocac029](https://doi.org/10.1093/jamia/ocac029). eprint: <https://academic.oup.com/jamia/article-pdf/29/6/1101/43674306/ocac029.pdf>. URL: <https://doi.org/10.1093/jamia/ocac029>.
- [17] Reed T. Sutton, David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. “An overview of clinical decision support systems: benefits, risks, and strategies for success.” In: *npj Digital Medicine* 3.1 (Feb. 2020), p. 17. ISSN: 2398-6352. DOI: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y). URL: <https://doi.org/10.1038/s41746-020-0221-y>.
- [18] Jerome A. Osheroff, Jonathan M. Teich, Blackford Middleton, Elaine B. Steen, Adam Wright, and Don E. Detmer. “A Roadmap for National Action on Clinical Decision Support.” In: *Journal of the American Medical Informatics Association* 14.2 (Mar. 2007), pp. 141–145. ISSN: 1067-5027. DOI: [10.1197/jamia.M2334](https://doi.org/10.1197/jamia.M2334). eprint: <https://academic.oup.com/jamia/article-pdf/14/2/141/2563507/14-2-141.pdf>. URL: <https://doi.org/10.1197/jamia.M2334>.
- [19] Prashila Dullabh, Shana F Sandberg, Krysta Heaney-Huls, et al. “Challenges and opportunities for advancing patient-centered clinical decision support: findings from a horizon scan.” In: *Journal of the American Medical Informatics Association* 29.7 (May 2022), pp. 1233–1243. ISSN: 1527-974X. DOI: [10.1093/jamia/ocac059](https://doi.org/10.1093/jamia/ocac059). eprint: <https://academic.oup.com/jamia/article-pdf/29/7/1233/44062158/ocac059.pdf>. URL: <https://doi.org/10.1093/jamia/ocac059>.
- [20] Dean F Sittig, Aziz Boxwala, Adam Wright, et al. “A lifecycle framework illustrates eight stages necessary for realizing the benefits of patient-centered clinical decision support.” In: *Journal of the American Medical Informatics Association* 30.9 (July 2023), pp. 1583–1589. ISSN: 1527-974X. DOI: [10.1093/jamia/ocad122](https://doi.org/10.1093/jamia/ocad122). eprint: <https://academic.oup.com/jamia/article-pdf/30/9/1583/51141471/ocad122.pdf>. URL: <https://doi.org/10.1093/jamia/ocad122>.
- [21] OpenAI. “Introducing ChatGPT.” In: (). URL: <https://openai.com/blog/chatgpt> (visited on 01/03/2024).

- [22] Siru Liu, Aileen P Wright, Barron L Patterson, et al. "Using AI-generated suggestions from ChatGPT to optimize clinical decision support." In: *Journal of the American Medical Informatics Association* 30.7 (Apr. 2023), pp. 1237–1245. ISSN: 1527-974X. DOI: [10.1093/jamia/ocad072](https://doi.org/10.1093/jamia/ocad072). eprint: <https://academic.oup.com/jamia/article-pdf/30/7/1237/50634330/ocad072.pdf>. URL: <https://doi.org/10.1093/jamia/ocad072>.
- [23] Manuela Benary, Xing David Wang, Max Schmidt, et al. "Leveraging Large Language Models for Decision Support in Personalized Oncology." In: *JAMA Network Open* 6.11 (Nov. 2023), e2343689–e2343689. ISSN: 2574-3805. DOI: [10.1001/jamanetworkopen.2023.43689](https://doi.org/10.1001/jamanetworkopen.2023.43689). eprint: https://jamanetwork.com/journals/jamanetworkopen/articlepdf/2812097/benary_2023_oi_231267_1699629981.22404.pdf. URL: <https://doi.org/10.1001/jamanetworkopen.2023.43689>.
- [24] Paul C. Lee, Samin K. Sharma, Shreya Motaganahalli, and Andy Huang. "Evaluating the Clinical Decision-Making Ability of Large Language Models Using MKSAP-19 Cardiology Questions." In: *JACC: Advances* 2.9 (2023), p. 100658. DOI: [10.1016/j.jacadv.2023.100658](https://doi.org/10.1016/j.jacadv.2023.100658). eprint: <https://www.jacc.org/doi/pdf/10.1016/j.jacadv.2023.100658>. URL: <https://www.jacc.org/doi/abs/10.1016/j.jacadv.2023.100658>.
- [25] Srijan Chatterjee, Manojit Bhattacharya, Soumen Pal, Sang-Soo Lee, and Chiranjib Chakraborty. "ChatGPT and large language models in orthopedics: from education and surgery to research." In: *Journal of Experimental Orthopaedics* 10.1 (Dec. 2023), p. 128. ISSN: 2197-1153. DOI: [10.1186/s40634-023-00700-1](https://doi.org/10.1186/s40634-023-00700-1). URL: <https://doi.org/10.1186/s40634-023-00700-1>.
- [26] Infectious Diseases Society of America. *COVID-19 Outpatient Treatment Guidelines Roadmap*. https://www.idsociety.org/covid-19-real-time-learning-network/therapeutics-and-interventions/covid-19-outpatient-treatment-guidelines-roadmap/#/+0/publishedDate_na_dt/asc/. Last Updated: February 2, 2023. Accessed: 2023-12-28.
- [27] Luyu Gao, Aman Madaan, Shuyan Zhou, et al. "PAL: Program-aided Language Models." In: *Proceedings of the 40th International Conference on Machine Learning*. Ed. by Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett. Vol. 202. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 10764–10799. URL: <https://proceedings.mlr.press/v202/gao23f.html>.
- [28] Josh Achiam, Steven Adler, Sandhini Agarwal, et al. *GPT-4 Technical Report*. 2023. arXiv: [2303.08774](https://arxiv.org/abs/2303.08774) [cs.CL].
- [29] OpenAI. *GPT-3.5*. <https://platform.openai.com/docs/models/gpt-3-5>. Accessed: 2023-12-28.
- [30] Hugo Touvron, Thibaut Lavril, Gautier Izacard, et al. *LLaMA: Open and Efficient Foundation Language Models*. 2023. arXiv: [2302.13971](https://arxiv.org/abs/2302.13971) [cs.CL].
- [31] Rohan Anil, Andrew M. Dai, Orhan Firat, et al. *PaLM 2 Technical Report*. 2023. arXiv: [2305.10403](https://arxiv.org/abs/2305.10403) [cs.CL].
- [32] Tianyu Gao, Adam Fisch, and Danqi Chen. "Making Pre-trained Language Models Better Few-shot Learners." In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Ed. by Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli. Online: Association for Computational Linguistics, Aug. 2021, pp. 3816–3830. DOI: [10.18653/v1/2021.acl-long.295](https://doi.org/10.18653/v1/2021.acl-long.295). URL: <https://aclanthology.org/2021.acl-long.295>.
- [33] Nahathai Wongpakaran, Tinakon Wongpakaran, Danny Wedding, and Kilem L. Gwet. "A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples." In: *BMC Medical Research Methodology* 13.1 (Apr. 2013), p. 61. ISSN: 1471-2288. DOI: [10.1186/1471-2288-13-61](https://doi.org/10.1186/1471-2288-13-61). URL: <https://doi.org/10.1186/1471-2288-13-61>.

- [34] J. Richard Landis and Gary G. Koch. "The Measurement of Observer Agreement for Categorical Data." In: *Biometrics* 33.1 (Mar. 1977). Full publication date: Mar., 1977, pp. 159–174. DOI: [10.2307/2529310](https://doi.org/10.2307/2529310). URL: <https://doi.org/10.2307/2529310>.
- [35] David Oniani, Jordan Hilsman, Yifan Peng, et al. "Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare." In: *npj Digital Medicine* 6.1 (Dec. 2023), p. 225. ISSN: 2398-6352. DOI: [10.1038/s41746-023-00965-x](https://doi.org/10.1038/s41746-023-00965-x). URL: <https://doi.org/10.1038/s41746-023-00965-x>.

Appendix

Human Evaluation Guidelines for COVID-19 Treatment Recommendation

We designed three human evaluation categories with categorical ratings:

- Presence of Incorrect Medical Content
 1. Yes, great clinical significance (there is incorrect medical content and it will make great clinical significance on the treatment recommendation, such as an incorrect treatment suggestions based on patient's conditions, like recommending patient who does not wish to go to hospital for Remdesivir infusion treatment)
 2. Yes, little clinical significance (there is incorrect medical content and it will make little clinical significance on the treatment recommendation)
 3. No (i.e., no incorrect medical content)
- Omission of Content
 1. Yes, great clinical significance (there is omission of the input patient's condition or characteristics and it will make great clinical significance on the treatment recommendation, such as changing the result of recommendation)
 2. Yes, little clinical significance (there is omission of the patient's condition or characteristics and it will make little clinical significance on the treatment recommendation, for example, even though omission of content occurred, it does not change the ultimate result)
 3. No (i.e., no omission of content)
- Presence of Possible Harmful Content
 1. Death or severe harm
 2. Moderate or mild harm
 3. No harm (i.e., no harmful content)

Paper Graphic

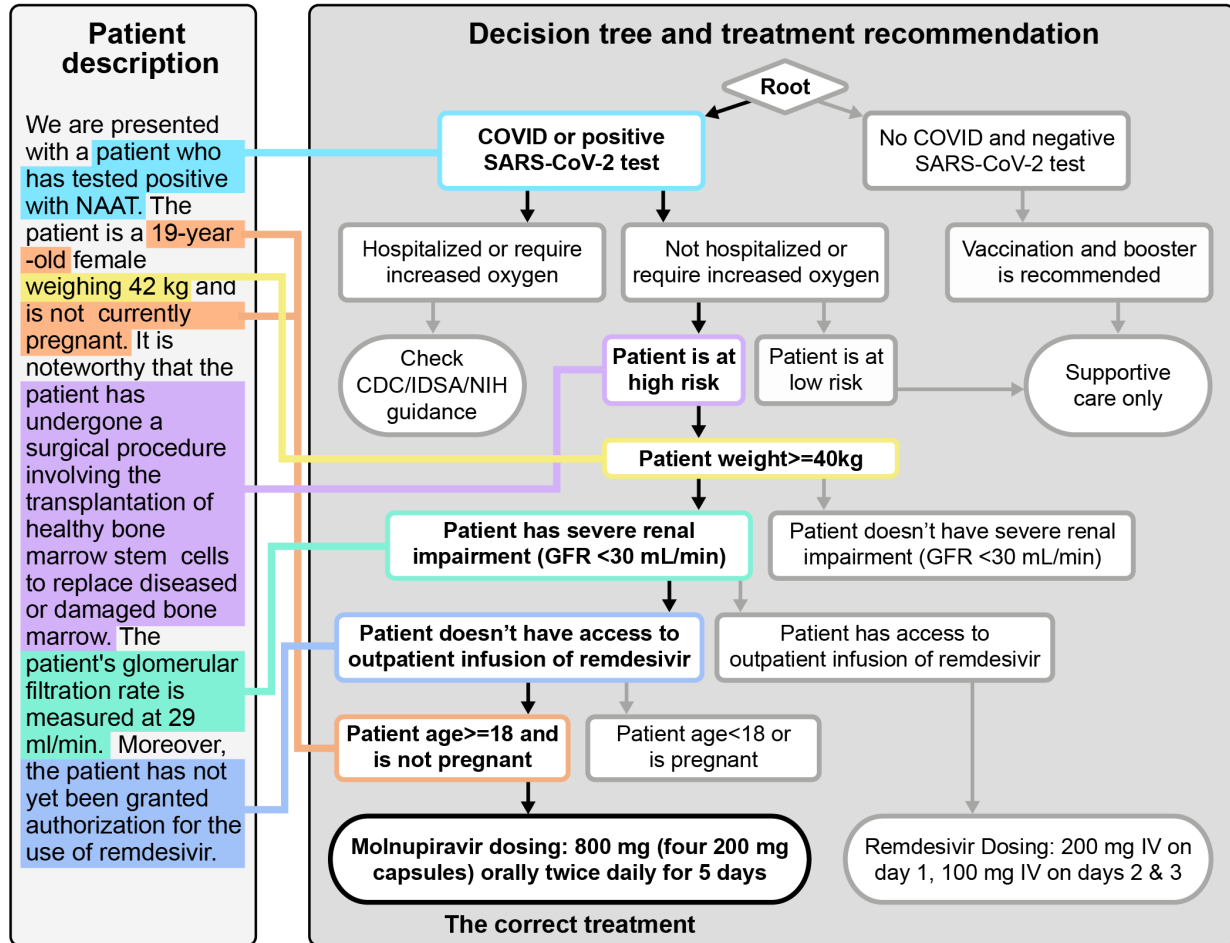


Figure 3: A graphic for the paper.

Clinical Practice Guidelines

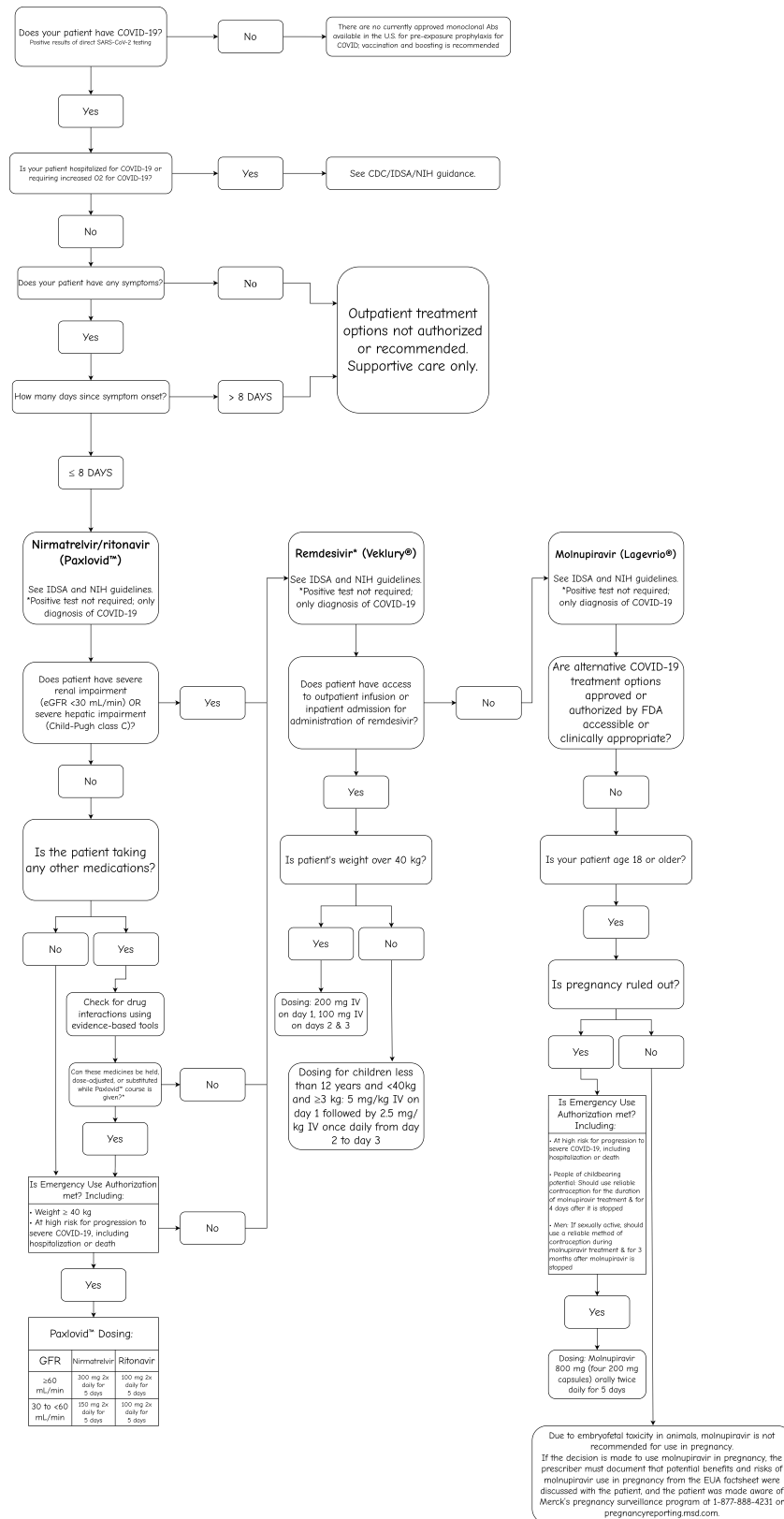


Figure 4: Clinical Practice Guidelines (CPGs) for COVID-19. We used modified the Centers for Disease Control and Prevention (CDC) and Infectious Diseases Society of America (IDSA) COVID-19 Outpatient Treatment Guidelines.

Synthetic Patients

Table 4: All 39 synthetic patients used in the study.

Table 1: All 39 synthetic patients used in the study.*		
Synthetic Patient	Correct Answer	Difficulty
Patient is positive for covid-19, is hypoxic and needs supplemental oxygen. Patient has a history of renal transplant. His eGFR is 61 ml/min. He takes medications that interact with Paxlovid. He has no contraindications for remdesivir and can receive it at nearby infusion center.	Check CDC/IDSA/NIH Guidance	hard
An 18-year-old patient has tested positive for Covid-19. He weighs 60 kg and/ asymptomatic. The patient does not need hospitalization. He has eGFR of 72 ml/min. He also takes medications that interact with Paxlovid. He cannot take remdesivir as he has no access to infusion center and does not want to be admitted to hospital.	Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only	hard
An 18-year-old boy/male, weighing 60 kg tested for Covid-19 and results are negative. He has a history of genetic blood disorder and chronic kidney disease with eGFR of 32 ml/min. Patient takes medication that induces CYP3A4 enzyme and interacts with paxlovid. He cannot take remdesivir as he does not have nearby infusion center and does not want to be admitted to the hospital.	Vaccination and booster is recommended	hard
A 28-year-old biological female tested positive for Covid-19. She weighs 65 kg. She is not pregnant currently. Patient is not hospitalized and do not require supplemental oxygen. Patients received a deceased donor kidney transplant 3 years back, which make her a high risk patient. The patient's eGFR=61 ml/min. And the patient do not have any liver issues. And the patient is taking other medications that can be coordinated/given with Paxlovid. And the patient can also get/ has access to outpatient infusion or inpatient admission for remdesivir.	Paxlovid Dosing: Nirnatrelvir 300 mg 2x daily for 5 days and Ritonavir 100 mg 2x daily for 5 days	hard

Continued on next page

Table 4: (Continued)

<p>A patient, born on May 11, 1989, underwent a diagnostic test (NAAT) for Covid-19 and tested positive. The patient's weight stands at 40 kg and has a peculiar genetic blood disorder. A crucial indicator of kidney function, known as the glomerular filtration rate (GFR), is abnormally low, measuring at 32 ml/min, signifying impaired kidney function. And the patient's liver is functioning normally. Patient do not require increased oxygen. Presently, the patient refrains from the use of any additional medications, can get remdesivir. Given these intricate circumstances, determine the appropriate course of treatment for this patient. What treatment should the patient receive?</p>	<p>Paxlovid Dosing: Nirma- relvir 150 mg 2x daily for 5 days and Ritonavir 100 mg 2x daily for 5 days</p>	<p>hard</p>
<p>We are currently presented with a female patient who has recently undergone a Covid-19 antigen test, yielding a positive result. The patient does not need hospitalization. The patient is 18 years old and weighs 58 kg, not pregnant. Notably, the patient has received a liver transplant from a deceased donor. Now the patient's liver is very healthy. In terms of kidney function, the patient's glomerular filtration rate stands at 61 ml/min. Additionally, the patient is taking multiple medications, some of which may interact adversely with Paxlovid. However, the patient has physician's permission for remdesivir.</p>	<p>Remdesivir Dosing: 200 mg IV on day 1, 100 mg IV on days 2 & 3</p>	<p>hard</p>
<p>We are currently presented with a female patient who has recently undergone a Covid-19 antigen test, yielding a positive result. The patient does not need increased oxygen. The patient is 18 years old and weighs 58 kg. Notably, the patient received a liver transplant from a deceased donor. In terms of kidney function, the patient's glomerular filtration rate stands at 21 ml/min. The patient does not have liver issues. Additionally, the patient is taking multiple medications, some of which may interact adversely with Paxlovid. However, the patient has been granted permission for the administration of remdesivir and have access to it. Given these intricate circumstances, What treatment should the patient receive?</p>	<p>Remdesivir Dosing: 200 mg IV on day 1, 100 mg IV on days 2 & 3</p>	<p>hard</p>

Continued on next page

Table 4: (Continued)

<p>A patient has recently undergone NAAT test for Covid-19 and received a positive result. The patient is a female, not pregnant, and is only 12 years old, weighing 40 kg. The patient does not need increased oxygen. It is worth mentioning that the patient has a genetic blood disorder. Her glomerular filtration rate stands at 32 ml/min, indicating impaired kidney function. And the patient's liver is in a healthy state. Moreover, the patient is currently taking other medications that conflict with the usage of Paxlovid. Additionally, the patient does not have access for the administration of remdesivir. Given these complex circumstances, What treatment should the patient receive?</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>hard</p>
<p>We are confronted with a patient who has yielded a positive result in a Covid-19 antigen test, signaling a true acute infection. But the patient does not need hospitalization. The patient is a female, not currently pregnant, and is in her 19th year, with a weight of 40 kg. Pertinently, the patient carries a hereditary blood disorder. The glomerular filtration rate is alarmingly low, measuring at 36 ml/min, highlighting impaired kidney function. Patient has always pay attention to her liver health, and her liver thanks her for that. Patient is taking drugs conflict with Paxlovid. Furthermore, the patient does not have access for the infusion center.</p>	<p>Molnupiravir dosing: 800 mg (four 200 mg capsules) orally twice daily for 5 days</p>	<p>hard</p>
<p>We are faced with a patient who has obtained a true positive result with Covid-19 NAAT testing. The patient is not hospitalized. The patient is a female, currently not pregnant, and is 17 years old, with a weight of 40 kg. Notably, the patient is afflicted with a genetic blood disorder. A concerning aspect is the patient's glomerular filtration rate, which stands at 32 ml/min, indicating compromised kidney function. Patient have liver failures. Furthermore, the patient has not yet been granted authorization for the administration of remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>hard</p>

Continued on next page

Table 4: (Continued)

<p>We are presented with a patient who has tested positive with Covid-19 NAAT. The patient does not require increased oxygen or hospitalization. The patient is a 19-year-old female weighing 42 kg and is not currently pregnant. It is noteworthy that the patient has undergone a surgical procedure involving the transplantation of healthy bone marrow stem cells to replace diseased or damaged bone marrow. The patient's glomerular filtration rate is measured at 29 ml/min, indicating compromised kidney function. Moreover, the patient does not have access to the use of remdesivir.</p>	<p>Molnupiravir dosing: 800 mg (four 200 mg capsules) orally twice daily for 5 days</p>	<p>hard</p>
<p>We encountered a patient who has tested positive with Covid-19 NAAT. The patient is a 17-year-old female weighing 32 kg and is not currently pregnant. Patient does not need hospitalization. Importantly, the patient has been diagnosed with a debilitating lung disease characterized by airway inflammation and damage, leading to breathing difficulties. Additionally, the patient's glomerular filtration rate (GFR) is measured at 32 ml/min, indicating impaired kidney function. In the patient's liver examinations, there are some figures above or below the standard range. It is noteworthy that the patient has been granted permission for the administration of remdesivir and have access to it.</p>	<p>Remdesivir Dosing: 5 mg/kg IV on day 1 followed by 2.5 mg/kg IV once daily from day 2 to day 3</p>	<p>hard</p>
<p>We are faced with a patient who has tested positive with Covid-19 NAAT. Patient does not require increased oxygen or hospitalization. The patient is an 11-year-old female weighing 32 kg and is not currently pregnant. Notably, the patient has been diagnosed with a challenging lung disease (bronchiectasis) characterized by airway inflammation and damage, resulting in breathing problems. Patient's liver is very healthy. Moreover, the patient has not yet been granted authorization for the use of remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>hard</p>
<p>Patient has tested positive for covid-19 and requires supplemental oxygen.</p>	<p>Check CDC/IDSA/NIH Guidance</p>	<p>medium</p>

Continued on next page

Table 4: (Continued)

<p>An 18-year-old patient tested positive for Covid-19, weighs 60 kg. He is asymptomatic.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>medium</p>
<p>An 18-year-old tested negative for Covid-19. He weighs 60 kg.</p>	<p>Vaccination and booster is recommended</p>	<p>medium</p>
<p>We have a patient who has a positive Covid-19 antigen test, female, 28 years old and weighs 65 kg. Patient does not need increased oxygen. Patient received a deceased donor kidney transplant 3 years back and has GFR=61 ml/min. She takes other medications that can be coordinated / given with Paxlovid. And the patient has been approved for remdesivir.</p>	<p>Paxlovid Dosing: Nirmatrelvir 300 mg 2x daily for 5 days and Ritonavir 100 mg 2x daily for 5 days</p>	<p>medium</p>
<p>We have a 31-year-old thin patient testing positive for Covid-19 NAAT. She weighs 40 kg. She does not need increased oxygen. Patient has a history of genetic blood disorder. The patient has chronic kidney disease with eGFR=32 ml/min. She does not take other medications. And the patient has insurance's permission for remdesivir infusion so has access to it.</p>	<p>Paxlovid Dosing: Nirmatrelvir 150 mg 2x daily for 5 days and Ritonavir 100 mg 2x daily for 5 days</p>	<p>medium</p>
<p>We have an 18-year-old female patient who has a positive antigen test, weighing 58 kg. Patient does not need hospitalization. Patient received a liver transplant from her brother 2 years back and it was very successful. The patient has a GFR=61 ml/min. And the patient takes other medications that conflicts with Paxlovid. She has her insurance's permission for remdesivir for outpatient infusion.</p>	<p>Remdesivir Dosing: 200 mg IV on day 1, 100 mg IV on days 2 & 3</p>	<p>medium</p>
<p>We have a female patient who has a true positive result for Covid-19 antigen test, not pregnant, 17 years old and weighs 40 kg. She does not need hospitalization. Patient has a genetic blood disorder. Her GFR=29 ml/min. She took medications conflict with Paxlovid. And the patient has been granted permission for the infusion of remdesivir.</p>	<p>Remdesivir Dosing: 200 mg IV on day 1, 100 mg IV on days 2 & 3</p>	<p>medium</p>

Continued on next page

Table 4: (Continued)

<p>We have a female patient testing positive for Covid-19 NAAT test. Patient does not need increased oxygen. She is 12 years old, not pregnant and weighs 41 kg. Patient has a genetic blood disorder and GFR=32 ml/min and does not have any liver problem. She takes medications that conflict with Paxlovid. She does not have permission for remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>medium</p>
<p>We have a patient who has a positive Covid-19 antigen test, female, not pregnant, 18 years old and weighs 58 kg. Patient does not need hospitalization. Patient received a donated liver from her uncle and it was a successful transplant surgery. Her liver is functioning well. The patient has a GFR=61 ml/min. And the patient takes other medications that conflict and interact with Paxlovid. Patient does not have permission for remdesivir infusion.</p>	<p>Molnupiravir dosing: 800 mg (four 200 mg capsules) orally twice daily for 5 days</p>	<p>medium</p>
<p>We have a patient who has a true result for Covid-19 NAAT, female, not pregnant, 17 years old and weighs 40 kg. She does not need hospitalization. Patient has a genetic blood disorder. The patient has a GFR=29 ml/min. And she does not yet have permission for remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>medium</p>
<p>We have a patient with a positive Covid-19 NAAT, female, 19 years old and weighs 42 kg. Patient does not need hospitalization. Patient is not pregnant. Patient had a surgery involving replacing a person's diseased or damaged bone marrow with healthy bone marrow stem cells. The patient has a GFR=29 ml/min. And the patient does not yet have permission/authorization for remdesivir. What treatment should the patient receive?</p>	<p>Molnupiravir dosing: 800 mg (four 200 mg capsules) orally twice daily for 5 days</p>	<p>medium</p>
<p>We have a 17-year-old female patient with a positive Covid-19 NAAT. Patient does not need hospitalization. She weighs 32 kg. Patient is not pregnant. Patients has a history of lung disease that makes it difficult to breathe due to inflammation and damage to the airways. The patient has a GFR=29 ml/min. And the patient received permission for remdesivir and have access to it. What treatment should the patient receive?</p>	<p>Remdesivir Dosing: 5 mg/kg IV on day 1 followed by 2.5 mg/ kg IV once daily from day 2 to day 3</p>	<p>medium</p>

Continued on next page

Table 4: (Continued)

<p>We have a patient with a positive NAAT Covid-19 test, female, 11 years old and weighs 32 kg. Patient does not need hospitalizations. Patient is not pregnant. Patient has history of chronic lung disease (bronchiectasis) that makes it difficult to breathe due to inflammation and damage to the airways. Patient's GFR is 4mL/min and does not take any drugs that cause conflict with Paxlovid. And the patient has not received authorization for remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>medium</p>
<p>The patient has covid-19, oxygen saturation (SpO2) at room air is low and needs supplemental oxygen.</p>	<p>Check CDC/IDSA/NIH Guidance</p>	<p>easy</p>
<p>A 34-year-old patient has tested positive for SARS-CoV-2 and weighs 62 kg. Patient does not need increased oxygen. He is asymptomatic and has no major clinical risk factors in his history.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>easy</p>
<p>A 16-year-old boy tested negative for Covid-19. His weight is 44 kg.</p>	<p>Vaccination and booster is recommended</p>	<p>easy</p>
<p>A 28-year-old female has a past medical history of kidney transplant and takes immunosuppression drugs. She weights 65 kg and has tested positive for Covid-19. She does not need hospitalization. And she does not have any chronic kidney disease with a GFR of 94 mL/min. Her immunosuppressive medications do not interact with paxlovid and she can hold few other home medicines while taking paxlovid. She can also take remdesivir at nearest infusion center.</p>	<p>Paxlovid Dosing: Nirmatrelvir 300 mg 2x daily for 5 days and Ritonavir 100 mg 2x daily for 5 days</p>	<p>easy</p>
<p>A 31-year-old female, weighing 40 kg, tested positive for covid-19. The patient does not need hospitalization. She has a history of genetic blood disorder and cardiovascular disease. She has a GFR of 32 mL/min and does not have any hepatic impairment. But does not take any medications. She has access to outpatient infusion or inpatient admission for remdesivir.</p>	<p>Paxlovid Dosing: Nirmatrelvir 150 mg 2x daily for 5 days and Ritonavir 100 mg 2x daily for 5 days</p>	<p>easy</p>

Continued on next page

Table 4: (Continued)

<p>We have a patient who is covid-19 positive, female, 18 years old and weighs 58 kg. Patient is taking immunosuppressive drugs. Patient is not hospitalized. The patient has a GFR=61 ml/min. Patient does not have hepatic impairment. The patient is taking other medications that cannot be held, dose adjusted or substituted while Paxlovid is given. The patient has access to outpatient infusion or inpatient admission for administration of remdesivir.</p>	<p>Remdesivir Dosing: 200 mg IV on day 1, 100 mg IV on days 2 & 3</p>	<p>easy</p>
<p>We have an 18-year-old non pregnant female patient who has a positive covid-19 test and weighs 58 kg. Patient is not hospitalized. The patient is a bone marrow transplant recipient at the age of 8 years. The patient has a GFR=31 ml/min. But the patient has hepatic impairment. She takes other medications that cannot be dose-adjusted with Paxlovid. Patient can get remdesivir.</p>	<p>Remdesivir Dosing: 200 mg IV on day 1, 100 mg IV on days 2 & 3</p>	<p>easy</p>
<p>We have a 12-year-old non pregnant female weighing 40 kg testing positive for Covid-19. The patient does not need increased oxygen. She has chronic kidney disease and is at high risk for disease progression due to Covid-19. Patient has a genetic blood disorder and GFR=32 ml/min. Patient does not have any liver issues. She takes medications that conflict with Paxlovid. And the patient does not have access to remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>easy</p>
<p>We have a patient who has a positive result for covid-19, female, not pregnant, 18 years old and weighs 40 kg. Patient does not require increased oxygen. The patient is at high risk for Covid-19 disease progression. She has GFR of 36 mL/min, does not have hepatic impairment. The patient is taking other drug conflict with Paxlovid. And the patient does not has authorization for outpatient infusion for remdesivir.</p>	<p>Molnupiravir dosing: 800 mg (four 200 mg capsules) orally twice daily for 5 days</p>	<p>easy</p>
<p>A 17-year-old non pregnant female weighs 40 kg and tested positive for Covid-19. Patient does not need hospitalization. The patient is at high risk for clinical deterioration. The patient has severe renal impairment. And she does not yet have authorization for administration of remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>easy</p>

Continued on next page

Table 4: (Continued)

<p>We have a patient with a positive covid-19 test, female, 19 years old and weighs 42 kg. Patient is not hospitalized. Patient is not pregnant. Patients had a recent surgery which puts her at high risk for clinical decompensation from Covid-19. The patient has a GFR=29 ml/min. And she does not yet have access for administration for remdesivir.</p>	<p>Molnupiravir dosing: 800 mg (four 200 mg capsules) orally twice daily for 5 days</p>	<p>easy</p>
<p>We have a patient with a positive covid-19 test, female, 17 years old and weighs 32 kg. Patient does not need increased oxygen. Patient is not pregnant. Patient has a history of chronic lung disease which puts him at high risk. The patient also has renal impairment. And the patient has access for administration for remdesivir.</p>	<p>Remdesivir Dosing: 5 mg/kg IV on day 1 followed by 2.5 mg/ kg IV once daily from day 2 to day 3</p>	<p>easy</p>
<p>We have a patient with a positive covid-19 test, female, 11 years old and weighs 32 kg. Patient is not hospitalized. Patient is not pregnant. Patient is at high risk for decompensation from Covid-19. And the patient has not received authorization for remdesivir.</p>	<p>Outpatient treatment options not authorized or recommended. Place in monitoring and supportive care only</p>	<p>easy</p>