MVP-LLMs: Optimizing Intervention Timing and Subsequent Decision Support for Mechanical Ventilation Parameter Control Using Large Language Models

Teqi Hao[1*], Xiaoyu Tan[2*], Bin Li[1], Xuemin Wang[3], Chao Qu[2], Yinghui Xu[4], and Xihe Qiu[1✉]

[1] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China
qiuxihe1993@gmail.com
[2] INFLY TECH(Shanghai)Co.,Ltd, Shanghai, 200030, China
[3] Department of Emergency and Critical Disease, Songjiang Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 201600, China
[4] AI³ Institute, Fudan University, Shanghai 200082, China

Abstract. Since the COVID-19 outbreak, global health systems have faced unprecedented challenges, with mechanical ventilation playing a critical role in supporting patients in ICUs. However, precise adjustment of ventilation parameters remains complex, requiring continuous monitoring and personalized interventions by clinicians. This paper introduces a novel formulation of ventilator parameter adjustment as a composite problem involving optimal stopping and subsequent decision optimization, supported by a domain-specific dataset reflecting real-world scenarios. We propose a framework utilizing Large Language Models (LLMs) to enhance interactivity and interpretability, leveraging their extensive clinical knowledge from large text corpora for informed decision-making. The framework addresses two key tasks: developing scheduled prompts for optimal stopping to replicate clinical observation processes and implementing Best Action Imitation Learning for robust ventilator parameter optimization. Experimental results show significant improvements in LLMs' ability to predict optimal stopping points and optimize decision-making, advancing clinical ventilator control. To our knowledge, this is the first application of LLMs to this dual-task paradigm.

Keywords: Mechanical Ventilation · Optimal Stopping · Decision Optimization · Large Language Models.

1 Introduction

Since the onset of the COVID-19 pandemic, global healthcare systems have faced immense challenges, with pneumonia severely damaging patients' respiratory

* These authors contributed equally to this work.
✉ Corresponding author.

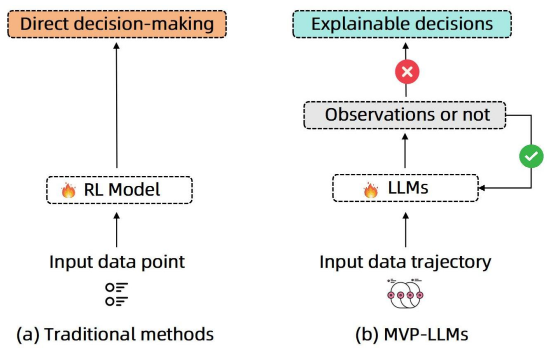


Fig. 1: This paper outlines the key differences between our new task and prior research. Specifically, (a) earlier studies on mechanical ventilation parameter adjustment relied on reinforcement learning and single observational data, while (b) our task simulates the process in a real ICU setting using Large Language Models (LLMs). It requires doctors to evaluate data over time for more precise decisions.

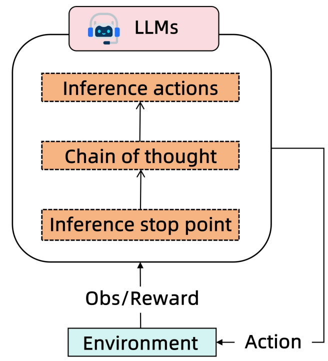


Fig. 2: Our framework uses LLMs for precise mechanical ventilation adjustments, predicting optimal stopping points via scheduled prompts to guide timely observation cessation. We also utilize CoT to enhance reasoning for subsequent actions, boosting decision-making quality.

systems and posing significant threats to life. In ICUs, mechanical ventilation has become a vital intervention for supporting patients with compromised respiration [12,3,8]. However, adjusting ventilation parameters accurately is a highly complex task requiring continuous monitoring and personalized interventions by healthcare professionals. Incorrect settings not only risk failing to support the patient's breathing effectively but may also cause severe complications, potentially resulting in life-threatening outcomes [13,21].

Machine learning, particularly supervised and reinforcement learning, is increasingly applied for efficient mechanical ventilation adjustment [16,22,14,20,5]. However, these methods face limitations: supervised learning requires large datasets and struggles with continuous data, while reinforcement learning is hampered by complex reward definitions and sparse data [16]. Current studies often use discrete data, whereas clinicians observe patients over time (Figure 1), creating a need to determine an optimal observation duration to avoid intervention delays or adjustment errors [9,11]. This challenge highlights the need to imitate clinicians' complex decision-making, a task well-suited for Large Language Models (LLMs) [15,26,23]. With exceptional memory and reasoning, LLMs possess rich knowledge from pre-training on vast text corpora. They can be fine-tuned on limited domain-specific data for effective generalization [17,4], and their reasoning is markedly enhanced by Chain-of-Thought (CoT) techniques that construct adaptive reasoning trajectories [10,24]. **Therefore, imitating clinicians' decision-making logic with LLMs, grounded in prior ventilation knowledge, offers a promising pathway to optimize ventilator parameter adjustments effectively.**

In this paper, we propose MVP-LLMs for **M**echanical **V**entilation **P**arameter control using **LLMs**, as illustrated in Figure 2. This framework, named MVP-LLMs, reframes the complex problem of ventilator parameter adjustment as a sequential decision-making task comprising two core stages: (i) optimal stopping and (ii) subsequent decision. To support this framework, we constructed a specialized dataset containing multiple real-world clinical observation trajectories to train LLMs to learn and execute ventilation optimization strategies.

Rather than treating the LLMs as a simple black-box predictor, we leverage it as a structured decision agent that simulates the thought process of clinical experts. Specifically, in the optimal stopping stage, we design "scheduled prompts" that encode positional information within the time series into natural language, guiding the model to autonomously determine the best intervention timing during continuous observation. In the subsequent decision stage, we combine best action imitation learning[6] with CoT-like reasoning validated by clinical experts. This method not only determines the optimal parameter adjustment but also generates a clear and clinically logical reasoning path from observation to action, thereby significantly enhancing the interpretability of the entire decision-making process. The main contributions of this study are summarized as follows:

- This study pioneers modeling mechanical ventilation parameter adjustment as an optimal stopping and decision-making problem, mirroring ICU professionals' real-world workflow.
- We built a specialized dataset for predicting ventilation parameters using real clinician data, rigorously cleaned and standardized, with clinician verification for quality assurance.
- We proposed an LLM-based framework for ventilation control and optimal stopping, demonstrating that our scheduled prompt method and imitation learning technique significantly enhance policy performance.

2 Methods

2.1 Data Collection and Preprocessing

Our dataset was constructed under close guidance from clinical experts, based on mechanical ventilation data from 165 ICU patients. To ensure medical validity, all feature selection and data extraction procedures were supervised and verified by ICU professionals, aiming to accurately simulate real-world clinical decision-making standards. The dataset includes both monitoring parameters and their corresponding ventilator settings, organized in the form of multiple patient trajectories. For the process of feature dimension selection, we rigorously followed the clinical experience of our cooperative doctors in making mechanical ventilation parameter decisions to ensure the practicality and representativeness of the selected features for performing ventilator control. Specifically, we have listed the monitoring features of patients and set parameters for each patient as shown in Table 1.

Table 1: Meta Information and Data Composition

Category	**Parameter**
Demographic	Age, Gender, Diagnosis
Respiratory	Timestamp, Inhaled Volume, Exhaled Volume, Respiratory Rate, Spontaneous Rate, Minute Ventilation, Peak Airway Pressure, Mean Airway Pressure, Plateau Pressure, Leak Volume, Rapid Breathing Index
Setting	SIMV Frequency, Ventilation Mode, Expiratory Sensitivity, Flow Sensitivity, O2 Concentration, PEEP, Pressure Support

We segment each trajectory into multiple sub-trajectories based on timestamps, with each sub-trajectory covering several time steps of observation and decision-making. This mirrors the approach of ventilation professionals, who observe first to gather sufficient information before making adjustments for accuracy. Guided by this, we use an optimal action learning strategy to sample the most effective sub-trajectories for dataset creation. Ultimately, we have constructed an initial dataset $\mathcal{D}_t$ comprising 3k trajectories. Among them, the initial dataset $\mathcal{D}_t$ is as follows:

$$\mathcal{D}_t = \{(\mathbf{x}_i, \mathbf{y}_i, R_i, \mathbf{x}'_i) \mid R_i > \tau, i \in I\}, \tag{1}$$

where each trajectory consists of i data points that are interconnected through a temporal sequence, where x is defined as the current state of the patient, $\mathbf{y}$ represents the observation decision and parameter decision taken by the doctor at the moment, $\mathbf{x}'$ denotes the new state following the transition and R is the reward calculated from the subsequent observation $\mathbf{x}'$.

2.2 Leveraging LLMs for Optimal Stopping and Sequential Decision-Making

We address the optimal stopping problem using reinforcement learning principles, guided by reward feedback to align preferences. Based on this, we have developed a scheduled prompt, as shown in Figure 3, hypothesizing that the optimal stopping point s_{opt} corresponds to the maximum reward value. We further assume that this reward value follows a Gaussian distribution, decreasing as the distance from s_{opt} increases. The distribution of reward values can be represented as:

$$R(d_i) = R_{\max} \cdot e^{-\frac{(d-\mu)^2}{2\sigma^2}}, s_i = |s_i - s_{\mathrm{opt}}| \tag{2}$$

where d represents the distance of each data point to the optimal stopping point.

```
system: You are an expert proficient in ICU mechanical ventilation therapy, responsible for observing and
making decisions based on the patient's monitoring parameters and set parameters in a real clinical
environment. In a real ICU scenario, a physician needs to observe a patient [....], then make decisions
accordingly. [....] to continue observing or make a decision at each time step.

[Here are the patient's personal information, initial monitoring parameters, and detailed explanations of
each parameter.]

Obs 1: TimeStamp: 2023-04-30 08:01:00, Monitoring Parameters: VTI=320, VTE=295, RATE=38, FSPN=22,
MVSPN=8.5, PPEAK=30, PMEAN=13, PPLAT=26, MVLEAK=4.5, RSBI=160.
Stop 1: <|START|> This is the first observation point at the beginning. <|OPT_BEFORE|> This observation
point is somewhat distant from the optimal stopping point, 4 units away. <|SOMEWHAT_DISTANT|>.
Thought 1: The patient's inspiratory and expiratory volumes have increased slightly, indicating adequate
ventilation. However, the respiratory rate and spontaneous respiratory [....].
Action 1: obs.

[Here are the data from some intermediate observation points, which have been omitted due to space
constraints.]

Obs 8: TimeStamp: 2023-04-30 08:06:00, Monitoring Parameters: VTI=310, VTE=315, RATE=37, FSPN=18,
MVSPN=8.0, PPEAK=31, PMEAN=14, PPLAT=30, MVLEAK=4.8, RSBI=155.
Stop 8: This observation point is the optimal point to cease observation, 0 units away. <|STOP|>.
Thought 8: There is a slight improvement in the respiratory rate and spontaneous respiratory frequency.
However, the plateau pressure is now at 30, which indicates significant issues with lung compliance. Peak
pressure is slightly [....]
Action 8: act, Parameter Adjustment Strategy: SET_SIMVRR=10, SET_VENTMODE=5119, SET_ETS=25,
SET_TRIGGERFLOW=2.0, SET_OXYGEN=50, SET_PEEP=8, SET_PSUPP=10.
```

Fig. 3: We provide an example of trajectory data simulating an ICU doctor's adjustment of ventilation parameters, incorporating multiple observation and decision stages. Using CoT, we enhance the process's interpretability.

We translated reward values into natural language to help LLMs simulate reinforcement learning processes. We identify the 'optimal stopping observation point' s_{opt} within a trajectory and use scheduled prompts $prompt_{\text{stop}}(t)$ to express distances from this point, incorporating special tokens for abstract distance learning. For instance, **'This observation point is the optimal point to cease observation, 0 units away. `<|STOP|>`'** denotes the optimal point, while **'This observation point is nearly at the optimal stopping point, just 1 unit away. `<|NEARLY_OPTIMAL|>`'** indicates proximity. This scheduled prompt process can be viewed as a natural language transformation of the reward function described in Equation (2):

$$prompt_{\text{stop}}(t) = f(R(d_i)). \tag{3}$$

To further enhance the performance of the stopping strategy, we integrate the CoT [24,19] approach for decision analysis. This method links monitoring parameters with data point positions, helping LLMs better understand sequential relationships. This improves reasoning and the ability to determine the optimal stopping moment. CoT enhancement was applied during decision-making to refine the pre-selected optimal actions. Specifically, we provided the model with the current observation state s_t and the corresponding optimal action a_t, using advanced LLMs (e.g., GPT-4 [1]) p_θ to generate a detailed reasoning process c_t:

$$\mathbf{c}_t \sim p_\theta(\cdot|\mathbf{s}_t, \mathbf{a}_t). \tag{4}$$

To ensure medical accuracy, each CoT segment was reviewed by professional medical collaborators. We then combined the refined prompt with the generated

CoT and integrated them into the data trajectory, resulting in the comprehensive dataset $D_{MVP} = (s_t, a_t, r_t, prompt_{stop}(t), c_t)$. This dataset accurately simulates the real-world workflow of ICU doctors in adjusting ventilation parameters.

2.3 Training and Evaluation

Training Stage During the training phase, we reformat the comprehensive dataset D_{MVP} into a multi-turn dialogue mode to simulate the interaction process between doctors and ventilators. In each dialogue round, LLMs perform stop-point reasoning and action decisions based on the patient's data and observations, with historical information recorded and used as prompts in subsequent rounds to support reasoning.

While much of the existing literature focuses on single data point predictions [14,2], our framework goes beyond this by addressing the entire phase of mechanical ventilation regulation. Instead of limiting the analysis to individual data points, our approach accumulates historical observation information h_t across all t time steps, which serves as the foundation for making adjustment decisions. This process is accomplished through supervised fine-tuning and can be mathematically represented as:

$$\min_{\theta} -\mathbb{E}_{D_{MVP}}[\log \prod_{i=1}^{k} p_\theta(e_i|s_t, h_t, e_{<i})], \tag{5}$$

where $e = (a_t, prompt_{\text{stop}}(t), c_t)$ contains all the reasoning processes of the optimal stopping and final decision formatted under $prompt_{\text{stop}}(t)$, and k represents the number of tokens in e. This learning process not only significantly enhances the accuracy of LLMs' decisions on optimal stopping points and final parameter settings but also provides richer interpretability for clinicians to support real-world mechanical ventilation parameter settings.

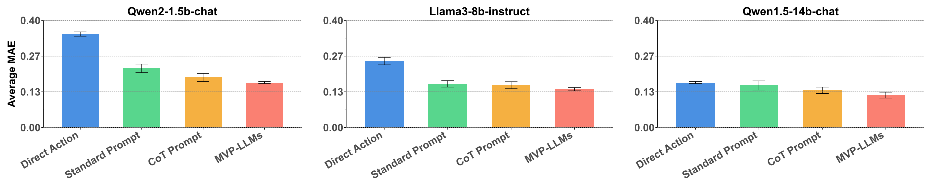
The entire training process was conducted on two Nvidia Tesla A100-80G GPUs [18], with a batch size of 4, a learning rate set at 2e-5, a warmup ratio of 0.03, and a weight decay of 0.05.

Evaluation Stage During evaluation, rewards were assigned to each action based on the dataset's format (Equation (2)), following a Gaussian distribution peaking at the optimal stopping point. We computed actual rewards for LLM-predicted stopping points, using the mean as the evaluation score. For the parameter-setting task, decision accuracy was measured by the standardized mean absolute error (MAE) between predictions and ground truth. These metrics collectively assess LLMs' performance in the observation-decision workflow.

3 Experiment

3.1 Baselines

Our research aims to enhance the interpretability of mechanical ventilation decision-making by simulating the clinical reasoning process. Accordingly, our



(a) The average MAE results of Qwen2-1.5b-chat on decision-making phase tasks.

(b) The average MAE results of Llama3-8b-instruct on decision-making phase tasks.

(c) The average MAE results of Qwen1.5-14b-chat on decision-making phase tasks.

Fig. 4: Presented are the results of mechanical ventilation parameter adjustments under different baselines across three distinct model scales. We comprehensively report the standardized mean MAE for all methods. The experimental data convincingly demonstrate that our proposed MVP-LLMs exhibits optimal performance across all three scales of LLMs.

primary baselines are designed to evaluate different strategies within the LLM-based framework, rather than to make direct comparisons with traditional sequential models that lack inherent interpretability. Based on this objective, we designed the following baselines tailored specifically for evaluating MVP-LLMs: 1) **Direct Action**: Uses only the current state, ignoring history, yielding less effective results. 2) **Standard Prompt**: Includes patient details, diagnostic metrics, and action decisions ("obs" or "stop"), representing conventional LLM decision-making. 3) **CoT Prompt**: Implements CoT between observation-action cycles, generating 10 sequences at temperature 0.7, with medical professionals selecting the best.

Our results show pretrained LLMs effectively encode mechanical ventilation knowledge, analyzing parameter fluctuations systematically. To evaluate the generalization capability of our proposed framework across different model architectures, we tested MVP-LLMs using Qwen2-1.5b-chat [25], Llama3-8b-instruct [7], and Qwen1.5-14b-chat, repeating experiments three times. Reported averages are statistically significant (t-test, $p < 0.05$).

3.2 Multi-step observation is beneficial for LLMs in predicting mechanical ventilation parameters

We assessed multi-step observation's impact on optimizing mechanical ventilation parameter adjustment, following Section 2.3. Figures 4 show that multi-step observation improves prediction accuracy across baselines and MVP-LLMs of varying sizes, underscoring the task's dependence on long-sequence processing. Gathering sufficient information before decision-making aligns with real-world clinical practices, while relying on single data points is less accurate and inconsistent with clinician procedures.

Table 2: The average score performance of three different-sized models in the first-stage optimal stopping task.

Method	Qwen2-1.5b-chat	Llama3-8b-instruct	Qwen1.5-14b-chat
Standard Prompt	42.00 ± 3.84	61.87 ± 4.31	76.37 ± 3.39
CoT Prompt	50.26 ± 5.32	65.87 ± 4.77	82.76 ± 6.32
MVP-LLMs	58.48 ± 3.56	71.13 ± 2.18	87.97 ± 4.17
MVP-LLMs (1-shot)	89.65 ± 2.98	93.48 ± 4.23	96.77 ± 2.88

Empirical results show that integrating CoT reasoning significantly enhances LLMs' reasoning capabilities, demonstrating that pre-trained LLMs hold extensive clinical knowledge on mechanical ventilation. This knowledge can be distilled into models of any size via CoT, with even a 1.5B LLM benefiting substantially.

We observed that the final parameter-setting performance of the proposed LLMs improved significantly with the use of the scheduled prompt $prompt_{stop}$. These prompts aligned LLM responses with observation point positions, ensuring high accuracy by generating responses in semantic order. For instance, if the LLM identified the first point as far from the optimal stopping point (e.g., *"This observation point is very far, 5 units away"*), subsequent judgments decreased incrementally until reaching the optimal point. This positional reasoning improved sequential optimal stopping decisions over long observation contexts.

3.3 LLMs can determine the optimal stopping observation point in mechanical ventilation.

We assessed LLMs in the optimal stopping task across various setups, excluding "Direct Action" due to its lack of an observational step. Table 2 indicates that LLMs can preliminarily determine when to stop observations based on monitoring parameters. CoT reasoning significantly boosted decision-making accuracy, and scheduled prompts improved optimal stopping point predictions. In Section 3.2, we showed that LLMs learn positional associations between observation points via scheduled prompts, influencing subsequent reasoning behaviors.

To validate LLMs' ability to learn positional connections and optimize decisions, we introduced MVP-LLMs (1-shot), providing the true location of the first observation point while maintaining other training aspects as in MVP-LLMs. Results in Table 2 reveal significant improvements: all model sizes excelled in the optimal stopping task. The 1.5B model scored 89.65, and the 14B model achieved 96.77, confirming that scheduled prompts enable LLMs to effectively grasp and utilize sequential positional relationships, showcasing strong performance in the optimal stopping task.

4 Conclusion

This paper proposes using LLMs to optimize mechanical ventilation adjustments in ICUs by mimicking clinicians' decision-making logic. The task is framed as an optimal stopping problem followed by decision-making, enabling MVP-LLMs to achieve superior control through LLMs' contextual understanding and reasoning. A specialized dataset with observation sequences and clinician reasoning annotations was constructed. Empirical results show notable performance gains across LLM sizes using scheduled prompts and CoT-style reasoning. This approach advances ventilator control and underscores AI's potential to improve patient outcomes. To enhance generalization and clinical utility, we seek external collaborations to build a diverse, multi-center dataset.

Acknowledgment. This work is supported by Shanghai Municipal Natural Science Foundation (23ZR1425400, 25ZR1401148) and Shanghai Soft Science Project (25692114700).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Akbulut, F.P., Akkur, E., Akan, A., Yarman, B.S.: A decision support system to determine optimal ventilator settings. BMC Medical Informatics and Decision Making **14**, 1–11 (2014)
3. Chang, R., Elhusseiny, K.M., Yeh, Y.C., Sun, W.Z.: Covid-19 icu and mechanical ventilation patient characteristics and outcomes—a systematic review and meta-analysis. PloS one **16**(2), e0246318 (2021)
4. Chen, Q., Deng, C.: Bioinfo-bench: A simple benchmark framework for llm bioinformatics skills evaluation. bioRxiv pp. 2023–10 (2023)
5. Chen, S., Qiu, X., Tan, X., Fang, Z., Jin, Y.: A model-based hybrid soft actor-critic deep reinforcement learning algorithm for optimal ventilator settings. Information sciences **611**, 47–64 (2022)
6. Chen, X., Zhou, Z., Wang, Z., Wang, C., Wu, Y., Ross, K.: Bail: Best-action imitation learning for batch deep reinforcement learning. Advances in Neural Information Processing Systems **33**, 18353–18363 (2020)
7. Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al.: The llama 3 herd of models. arXiv preprint arXiv:2407.21783 (2024)
8. Esteban, A., Anzueto, A., Alia, I., Gordo, F., Apezteguia, C., Palizas, F., Cide, D., Goldwaser, R., Soto, L., Bugedo, G., et al.: How is mechanical ventilation employed in the intensive care unit? an international utilization review. American journal of respiratory and critical care medicine **161**(5), 1450–1458 (2000)
9. Esteban, A., Ferguson, N.D., Meade, M.O., Frutos-Vivar, F., Apezteguia, C., Brochard, L., Raymondos, K., Nin, N., Hurtado, J., Tomicic, V., et al.: Evolution of mechanical ventilation in response to clinical research. American journal of respiratory and critical care medicine **177**(2), 170–177 (2008)

10. Feng, G., Zhang, B., Gu, Y., Ye, H., He, D., Wang, L.: Towards revealing the mystery behind chain of thought: a theoretical perspective. Advances in Neural Information Processing Systems **36** (2024)
11. Fuller, B.M., Mohr, N.M., Miller, C.N., Deitchman, A.R., Levine, B.J., Castagno, N., Hassebroek, E.C., Dhedhi, A., Scott-Wittenborn, N., Grace, E., et al.: Mechanical ventilation and ards in the ed: a multicenter, observational, prospective, cross-sectional study. Chest **148**(2), 365–374 (2015)
12. Kacmarek, R.M.: The mechanical ventilator: past, present, and future. Respiratory care **56**(8), 1170–1180 (2011)
13. Klompas, M.: Complications of mechanical ventilation—the cdc's new surveillance paradigm. N Engl J Med **368**(16), 1472–1475 (2013)
14. Kondrup, F., Jiralerspong, T., Lau, E., de Lara, N., Shkrob, J., Tran, M.D., Precup, D., Basu, S.: Towards safe mechanical ventilation treatment using deep offline reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 15696–15702 (2023)
15. Koroteev, M.V.: Bert: a review of applications in natural language processing and understanding. arXiv preprint arXiv:2103.11943 (2021)
16. Le, S., Pellegrini, E., Green-Saxena, A., Summers, C., Hoffman, J., Calvert, J., Das, R.: Supervised machine learning for the early prediction of acute respiratory distress syndrome (ards). Journal of Critical Care **60**, 96–102 (2020)
17. Li, Y., Wang, S., Ding, H., Chen, H.: Large language models in finance: A survey. In: Proceedings of the fourth ACM international conference on AI in finance. pp. 374–382 (2023)
18. Lindholm, E., Nickolls, J., Oberman, S., Montrym, J.: Nvidia tesla: A unified graphics and computing architecture. IEEE micro **28**(2), 39–55 (2008)
19. Miao, J., Thongprayoon, C., Suppadungsuk, S., Krisanapan, P., Radhakrishnan, Y., Cheungpasitporn, W.: Chain of thought utilization in large language models and application in nephrology. Medicina **60**(1), 148 (2024)
20. Peine, A., Hallawa, A., Bickenbach, J., Dartmann, G., Fazlic, L.B., Schmeink, A., Ascheid, G., Thiemermann, C., Schuppert, A., Kindle, R., et al.: Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. NPJ digital medicine **4**(1), 32 (2021)
21. Principi, T., Fraser, D.D., Morrison, G.C., Farsi, S.A., Carrelas, J.F., Maurice, E.A., Kornecki, A.: Complications of mechanical ventilation in the pediatric population. Pediatric pulmonology **46**(5), 452–457 (2011)
22. Sayed, M., Riaño, D., Villar, J.: Predicting duration of mechanical ventilation in acute respiratory distress syndrome using supervised machine learning. Journal of clinical medicine **10**(17), 3824 (2021)
23. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al.: Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288 (2023)
24. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems **35**, 24824–24837 (2022)
25. Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., et al.: Qwen2 technical report. arXiv preprint arXiv:2407.10671 (2024)
26. Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., Zhang, Y.: A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. High-Confidence Computing p. 100211 (2024)