ORIGINAL RESEARCH **Open access**

From Protocols to Preferences: Why Reinforcement Learning from Human Feedback Must Replace Fixed Weaning Protocols for Prolonged Mechanical Ventilation

Luis Herrera[1*], Daniela Rojas[1], Andres Castro[2]

Abstract

Prolonged mechanical ventilation (PMV), affecting 5–15% of ICU patients, is associated with high mortality (30–50%), long-term disability, and substantial healthcare costs exceeding $100,000 per admission. These patients often require extended respiratory support beyond 14–21 days and consume significant ICU resources. Current weaning strategies rely on fixed spontaneous breathing trial (SBT) criteria (e.g., RSBI thresholds, oxygenation, respiratory rate), which fail to account for the heterogeneous and evolving physiology of PMV patients. This reduces weaning to discrete events rather than a continuous adaptive process. We propose reinforcement learning from human feedback (RLHF) as a superior framework for weaning, enabling AI systems to learn sequential decision-making policies from clinician preferences across patient trajectories. Traditional protocols ignore temporal dependencies such as prior SBT outcomes, sedation exposure, and respiratory muscle trends. While standard reinforcement learning supports sequential optimization, it depends on difficult-to-define reward functions. RLHF overcomes this by learning reward signals directly from clinician comparisons, aligning model behavior with real-world clinical judgment. Research should shift toward RLHF-based dynamic weaning policies rather than static prediction models. Clinical stakeholders should support data collection and prospective evaluation of RLHF-guided weaning versus standard protocols. RLHF offers a necessary advancement for personalized PMV weaning, addressing limitations of rigid protocols and improving alignment with clinical decision-making.

Keywords Clinical decision support, Reinforcement learning, Sequential decision-making, Mechanical ventilation, Prolonged weaning, Human feedback

*Correspondence:
Luis Herrera
luis.herrera@gmail.com
[1] Department of Intelligent Healthcare Analytics, Pontifical Catholic University of Chile, Santiago, Chile
[2] Department of AI Clinical Systems, University of Concepcion, Concepcion, Chile

Introduction

Mechanical ventilation is among the most common interventions in intensive care medicine, with approximately 30-50% of ICU patients receiving invasive respiratory support during their admission [1, 2]. For those who survive the acute phase of critical illness, the process of weaning—gradually transferring respiratory work from the machine back to the patient—consumes nearly 40% of total ventilation time and represents a period of persistent vulnerability [3, 4]. Failure to wean within 14 days defines prolonged mechanical ventilation (PMV), a condition associated with dramatically worsened outcomes including 30-50% mortality, mandatory tracheostomy in many cases, and extended stays in long-term acute care facilities [2, 3].

PMV patients represent a distinct clinical population with unique pathophysiology that distinguishes them from those who wean successfully within the first week [3, 4]. Prolonged ventilation induces diaphragmatic atrophy, oxidative stress injury to respiratory muscles, and neuromuscular weakness that perpetuates ventilator dependence [4, 5]. These patients frequently suffer from multiple comorbid conditions including chronic obstructive pulmonary disease, congestive heart failure, obesity hypoventilation syndrome, and neuromuscular disorders that interact unpredictably during weaning attempts [5, 6]. The complexity of PMV demands personalized, adaptive strategies that current protocols cannot provide [7].

Contemporary weaning protocols rely on spontaneous breathing trials (SBTs) conducted for 30-120 minutes using T-piece, continuous positive airway pressure (CPAP), or low-level pressure support ventilation [5, 6]. Success criteria include RSBI below 105 breaths per minute per liter, SpO2 above 90% on FiO2 below 0.4, respiratory rate below 35, and absence of hemodynamic instability or marked anxiety [6, 8]. While these criteria perform adequately for simple patients who wean within days, they fail systematically for PMV patients whose physiology violates the assumptions underlying fixed thresholds [8, 9]. We contend that protocol-based weaning represents a one-size-fits-all failure that must be replaced by adaptive approaches [10].

This position paper advances a specific thesis: reinforcement learning from human feedback (RLHF) should replace fixed SBT protocols for weaning PMV patients [1, 11]. I argue that weaning is fundamentally a sequential decision-making problem where each action (adjusting pressure support, initiating an SBT, terminating a trial, deciding to extubate) constrains and enables subsequent options [8, 9]. RL captures this sequential structure naturally, while human feedback resolves the otherwise intractable problem of specifying what constitutes a good weaning trajectory [1].

Ventilator Weaning Landscape

Spontaneous breathing trials

Spontaneous breathing trials evaluate whether a patient can sustain adequate gas exchange and respiratory mechanics without significant ventilator support [5, 6]. The three most common SBT methods are T-piece ventilation (no pressure support), CPAP at 5 cm H2O, and low-level pressure support at 5-8 cm H2O, with evidence suggesting pressure support SBTs reduce weaning failure compared to T-piece in some populations [6, 8]. Standardized success criteria require RSBI below 105, SpO2 above 90% on FiO2 ≤0.4, respiratory rate below 35, heart rate ≤140 or not increased more than 20% from baseline, systolic blood pressure between 90-180 mmHg, and absence of diaphoresis, marked anxiety, or worsening dyspnea [5, 9]. Failure during an SBT manifests as tachypnea, desaturation, tachycardia, hypertension or hypotension, arrhythmias, or visible signs of increased work of breathing including accessory muscle use and paradoxical abdominal motion [6, 8].

Clinicians must decide not only whether an SBT succeeds or fails but also how to respond to failure [8, 9]. A patient who fails an SBT after 20 minutes with mild tachypnea but stable oxygenation might benefit from continued pressure support at a higher level, whereas a patient who fails rapidly with desaturation and hemodynamic instability requires return to full support and investigation of the underlying cause [9, 12]. Current protocols typically mandate a fixed rest period of 24 hours following SBT failure before the next attempt, ignoring the possibility that some patients recover respiratory muscle strength more quickly while others deteriorate further during prolonged rest [5, 8]. We contend that this rigid structure reflects administrative convenience rather than physiological optimization [13].

Prolonged mechanical ventilation

Prolonged mechanical ventilation is most commonly defined as requiring invasive ventilation for 14 or more consecutive days, though some definitions use 21 days to distinguish extreme PMV [2, 3]. The incidence of PMV ranges from 5-15% of all mechanically ventilated ICU patients, but these patients consume a vastly disproportionate share of ICU resources including up to 50% of total ventilation days in some units [3, 4]. Risk factors for developing PMV include advanced age, pre-existing chronic lung disease (particularly COPD with hypercapnia), neuromuscular disorders, obesity, sepsis on admission, and acute respiratory distress syndrome requiring prolonged deep sedation [4, 5].

Outcomes for PMV patients are substantially worse than for those who wean earlier, with in-hospital mortality ranging

from 30-50% depending on the population and definition used [2, 3]. Among survivors, tracheostomy is performed in 50-80% of PMV cases, and many patients require transfer to long-term acute care hospitals or skilled nursing facilities for ongoing ventilator management [3, 4]. Quality of life after PMV is frequently poor, with persistent functional disability, recurrent respiratory infections, and high rates of readmission [4, 5]. The economic impact is severe: estimates place the mean cost per PMV admission above $100,000, with the cumulative national expenditure exceeding $20 billion annually in the United States alone [2, 3]. These stark outcomes underscore the urgent need for better weaning strategies [7].

Limitations of Current Approaches

One-size-fits-all protocols

Fixed SBT criteria assume that the thresholds distinguishing readiness from unreadiness are invariant across patients, ignoring the substantial heterogeneity in physiology, reserve, and recovery trajectories among PMV patients [4, 5]. A patient with severe COPD and chronic hypercapnia may tolerate an RSBI of 120 with minimal distress, whereas a previously healthy patient with ICU-acquired weakness may fail at RSBI of 90 due to different mechanisms of respiratory failure [5, 8]. By applying identical thresholds to all patients, protocols systematically delay weaning in those with chronic adaptations while prematurely challenging those with acute deconditioning [6, 8].

The absence of adaptation mechanisms in current protocols means that weaning decisions do not improve with experience [8, 9]. An ICU that successfully weans 80% of its simple patients but struggles with PMV continues to use identical SBT criteria for both populations, learning nothing from repeated failures [9, 12]. We contend that this represents a fundamental design flaw: protocols cannot update their parameters based on observed outcomes because they contain no learning mechanism [12, 14]. Each patient effectively starts from scratch, with no transfer of information about which SBT timing and duration strategies succeeded for similar previous patients [14, 15].

Furthermore, fixed protocols cannot accommodate the dynamic changes that occur during PMV [4, 8]. A patient's optimal SBT duration may increase from 20 minutes to 60 minutes over two weeks as respiratory muscles strengthen, then decrease again following an intercurrent infection [5, 9]. Protocols that mandate fixed 30-minute SBTs regardless of this trajectory will either under-challenge the patient (prolonging ventilation unnecessarily) or over-challenge them (precipitating failure and requiring another rest day) [6, 12]. This rigidity is not a bug but a feature of protocol-based weaning—and it is precisely why protocols fail for complex, dynamic patients [8, 14].

Ignoring sequential decision-making

Weaning from mechanical ventilation is inherently sequential: the decision to attempt extubation depends not only on current SBT performance but also on the trajectory of RSBI over multiple trials, the cumulative duration of pressure support reduction attempted in preceding days, and the patient's recovery pattern from previous failures [8, 9]. Clinicians implicitly model these dependencies when they decide to attempt a fourth SBT after three marginal failures with improving trends, or to postpone extubation for a patient whose RSBI meets criteria but who has deteriorated after each previous attempt [9, 12]. Current protocols treat each SBT as an independent event, evaluating success criteria without reference to the patient's response trajectory [5, 8].

This failure to model sequential structure leads to predictable errors [8, 14]. A patient who marginally passes three consecutive SBTs but exhibits increasing tachypnea each time is likely failing, yet protocols would classify each trial as "passed" and eventually recommend extubation—which will likely fail [9, 12]. Conversely, a patient who marginally fails two SBTs but shows clear improvement in respiratory rate and oxygenation from trial to trial may be on a successful trajectory, yet protocols mandate rest days after each failure, potentially interrupting a positive trend [8, 14]. We contend that these errors arise not from clinician incompetence but from the inherent limitations of non-sequential decision rules [9, 15].

Figure 1 illustrates the fundamental structural shift from rigid, protocol-based weaning toward an RLHF-driven sequential decision framework that integrates patient state trajectories and clinician preferences.

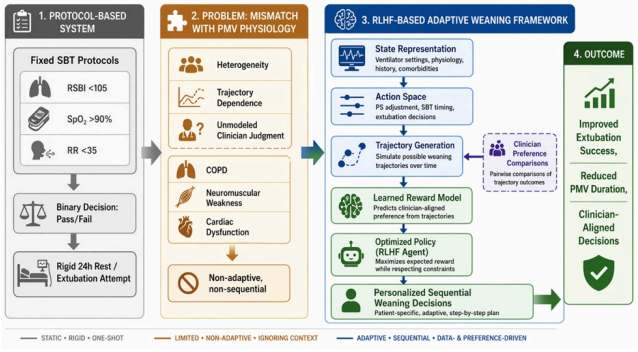

Figure 1. Hierarchical Transformation of Ventilator Weaning: From Static Protocols to RLHF-Driven Sequential Decision Systems

Sequential decision-making also requires reasoning about when to stop weaning and return to full support [9, 15]. A patient who develops fever and increased secretions during weaning may require treatment of a new pneumonia before further attempts, and attempting to continue weaning risks clinical deterioration [8, 12]. Protocols that specify only success/failure criteria without incorporating state variables indicating infection, hemodynamic stability, or neurological status cannot make this distinction [14, 15]. The result is either premature termination of weaning for patients with transient issues or persistence in weaning for patients who require medical stabilization first [9, 16].

Underutilization of clinician expertise

Experienced critical care clinicians detect subtle signs of weaning readiness and failure that are not captured by any standardized scoring system [9, 12]. The quality of a patient's cough, the pattern of accessory muscle recruitment during tidal breathing, the trajectory of mental status over the preceding shift, and the subjective assessment of work of breathing all inform clinical judgment but appear in no protocol [12, 14]. These signs are not mysterious intuitions but legitimate clinical data that resist easy quantification yet predict outcomes as well or better than RSBI in some studies [9, 15]. Protocols that ignore this information are not merely incomplete but systematically biased against the holistic assessment that defines expert practice [17].

Current weaning protocols were designed to reduce unwarranted variation and speed weaning for simple patients, not to capture expert reasoning about complex cases [12, 16]. For PMV patients, who present precisely the kind of complexity that requires expert judgment, protocols offer little guidance beyond "perform SBTs and monitor these five variables" [14, 15]. The implicit message is that weaning PMV patients does not require special expertise—a claim that contradicts both clinical experience and outcome data showing wide variation in PMV weaning success across ICUs [8, 13]. We contend that this de-skilling of weaning through protocolization has harmed PMV patients by substituting statistical norms for clinical reasoning [18].

Moreover, protocols do not learn from clinician expertise over time [12, 19]. An ICU where experienced clinicians consistently adjust SBT duration based on patient phenotype cannot encode this knowledge into the protocol without formal research studies and guideline updates [14, 15]. The protocol treats the first PMV patient and the hundredth identically, whereas clinicians improve with experience [9, 16]. RLHF offers a solution to this problem by learning directly from clinician preferences, effectively distilling expertise into a decision policy that improves as more preference data accumulate [1, 11].

Reinforcement Learning with Human Feedback

Standard RL for weaning

Standard RL formulates weaning as an MDP where state includes ventilator settings, physiological variables (RSBI, SpO2, respiratory rate, heart rate, blood pressure), SBT metrics, and clinical context [14, 15]. Actions include pressure support adjustments (increase/decrease by 2-5 cm H2O), SBT initiation/termination, extubation, and return to full support [11, 15]. The reward function typically combines successful extubation, avoidance of reintubation, and minimized ventilation days [14, 15]. The agent learns a policy maximizing cumulative reward, capturing sequential dependencies protocols miss [15, 16]. However, standard RL faces sparse and delayed rewards—extubation success occurs only at episode end, providing little guidance for intermediate decisions [14, 19]. Reward shaping requires hand-engineering sub-behaviors, returning to the problem of fixed criteria [16, 19].

Limitations of reward specification

Reward engineering for weaning is notoriously difficult because patient welfare is multidimensional, context-dependent, and contested among stakeholders [15, 16]. Trade-offs between extubation success, ventilation

duration, reintubation risk, and infection risk admit no unique correct answer, and preferences vary across clinicians, patients, and families [11, 16]. Sparse rewards mean an agent learning from scratch needs thousands of episodes to learn that gradual reduction works better than abrupt discontinuation, but each episode is a real patient [14, 19]. Offline RL reduces but does not eliminate this problem [14, 15]. Moreover, patient-centered outcomes like comfort, sleep quality, delirium avoidance, and communication ability rarely appear in protocols but require quantifying the incommensurable [11, 16].

Static machine learning models have attempted to predict weaning success using various approaches, including ventilator mode shifting prediction [20], biosignal-based digital biomarkers [21], continuous ventilator parameters during SBTs [22], and explainable machine learning for extubation prediction [23]. However, these remain single-point predictions rather than sequential decision-making frameworks. Recent work has also explored reduction in PEEP levels [24], time-series ventilator parameters [25], and convolutional neural networks [26], but these share the same fundamental limitation: they predict an outcome rather than learning a dynamic policy.

RLHF solution

RLHF learns the reward function from clinician pairwise comparisons of weaning trajectories rather than requiring advance specification [1, 11]. A clinician indicates which of two trajectories they prefer, providing training data for a neural network that predicts reward for any trajectory [1, 11]. The RL agent then optimizes its policy to maximize predicted reward, learning to behave in ways clinicians judge superior. Pairwise comparisons are far easier for humans than absolute reward values—a clinician can readily judge whether trajectory A is better than trajectory B without assigning numeric weights [1, 11]. For weaning, RLHF incorporates subtle multidimensional aspects of quality that protocols ignore, captures variation in preferences across clinicians and contexts, and sidesteps the need for explicit reward specification [1, 16].

Separate reward models can be developed for different preference groups [11, 19]. Third, the framework naturally supports continuous improvement: as more preference data accumulate, the reward model becomes more accurate, and the resulting policies improve [1, 11]. We contend that RLHF transforms weaning from a protocol-driven procedure into a learning system that becomes smarter with every patient [5].

MDP Formulation for Weaning

State space

The state vector must capture four categories: ventilator settings (pressure support, PEEP, FiO2, mode), continuous physiological variables (RSBI, SpO2, respiratory rate, heart rate, blood pressure, temperature), categorical clinical variables (sedation status, delirium, command-following, cough strength), and weaning history (days intubated, SBT attempts, last SBT duration and reason for termination) [14, 16]. Temporal dynamics are critical, so rolling windows of key variables like last three RSBI measurements with timestamps enable the agent to distinguish improving from deteriorating trajectories, mirroring how clinicians reason [14, 19]. Patient heterogeneity is accommodated via static features including COPD, heart failure, neuromuscular disease, obesity, reason for intubation, sepsis presence, and cumulative sedation dose [16, 19].

Action space

The discrete action set for pressure support weaning includes: decrease PS by 2 or 4 cm H2O, maintain current level, increase by 2 cm H2O, initiate SBT at current or reduced level (5-8 cm H2O), terminate SBT with return to previous or increased support, extubate, or return to full support for 24 hours [11, 15]. The agent can act at any 4-hour interval with "no change" as an explicit action, enabling variable timing while preventing forced actions [15, 19]. Hard safety constraints override the learned policy: never recommend extubation with RSBI >130, vasopressor-dependent hemodynamic instability, or SpO2 <88% on maximal FiO2 [14, 16]. This "safety cage" filters recommendations through clinically validated exclusion criteria, preserving RL benefits while providing guaranteed safety boundaries [14, 15].

Integrating Clinician Feedback

Preference elicitation

Preference elicitation for RLHF requires presenting clinicians with pairs of weaning trajectories and asking

which they judge superior, a task that maps naturally to existing clinical review processes [1, 11]. Each trajectory includes the sequence of states visited and actions taken over a complete weaning episode, from the decision to begin active weaning through extubation and 72 hours post-extubation to capture reintubation outcomes [1, 16]. Clinicians reviewing these pairs consider multiple dimensions simultaneously: time to extubation, patient comfort during SBTs, avoidance of reintubation, preservation of respiratory muscle function, and the pattern of pressure support reduction [11, 19]. We contend that this holistic judgment, rendered in seconds by an experienced clinician, contains more information about weaning quality than any hand-engineered reward function could encode [27].

The volume of preference data required for effective RLHF is substantial but achievable through systematic collection during routine clinical operations [1, 11]. Christiano *et al*. demonstrated that reward models trained on thousands of pairwise comparisons can effectively approximate human preferences across complex sequential tasks [1]. For weaning, a single large academic ICU might generate 100-200 PMV weaning episodes annually, each providing multiple potential pairwise comparisons [14, 16]. Supplementing this with retrospective comparisons of historical trajectories and prospective comparisons during silent deployment phases could accumulate the necessary dataset within 12-24 months [11, 15]. We argue that this investment is modest relative to the potential benefits of improved weaning outcomes [28].

Preference elicitation must address known biases in human comparisons, including recency bias (overweighting late events) and salience bias (overweighting dramatic events like reintubation over gradual processes like pressure support reduction) [11, 19]. We propose using structured presentation formats that highlight all phases of the weaning episode equally, along with explicit training for clinicians on how to compare trajectories holistically [1, 16]. Additionally, collecting preferences from multiple clinicians for the same trajectory pairs enables estimation of inter-rater agreement and identification of systematic biases [11, 19]. The reward model can then learn a central tendency that reflects shared clinical judgment while accommodating reasonable variation [1, 11].

Reward learning from preferences

The reward learning component of RLHF uses pairwise comparison data to train a neural network that assigns scalar rewards to weaning trajectories [1, 11]. The network takes as input a trajectory representation (a sequence of state-action pairs) and outputs a predicted reward value, with the training objective ensuring that trajectories preferred by clinicians receive higher predicted rewards than dispreferred ones [1, 19]. This formulation does not require the network to output clinically interpretable reward values—only that the ordering of trajectories by predicted reward matches clinician preferences [11, 19]. The learned reward function can then be used to evaluate and optimize weaning policies without further clinician input [1, 16].

Architecturally, the reward model must handle variable-length trajectories and capture temporal dependencies that determine weaning quality [1, 16]. We propose using a recurrent neural network or transformer architecture that processes the sequence of states and actions, with attention mechanisms allowing the model to focus on critical events such as SBT failures, extubation decisions, and reintubation episodes [11, 19]. The final layer produces a scalar reward, and the network is trained using a pairwise ranking loss that penalizes inversions of the clinician's preference order [1, 11]. Regularization techniques prevent overfitting to idiosyncratic preferences in the training data [19].

Once trained, the reward model enables policy optimization without requiring further clinician input [1, 11]. Standard RL algorithms (e.g., proximal policy optimization) can optimize the weaning policy to maximize expected reward as predicted by the model, effectively learning to behave in ways that clinicians previously judged superior [1, 19]. Importantly, the policy is optimized offline using historical data or simulation, not through online experimentation with patients [11, 19]. We contend that this two-stage approach—first learning reward from preferences, then optimizing policy against learned reward—preserves the safety of offline learning while achieving clinician-aligned objectives that hand-engineered rewards cannot capture [5, 6].

Table 1 provides a conceptual mapping between clinical weaning practice and its formal representation within an RLHF-based computational framework.

Table 1. Conceptual Decomposition of the RLHF Weaning System: Mapping Clinical Reality to Computational Architecture

System Component	Clinical Analog	Computational Representation	Functi
Patient State	Bedside assessment (vitals, ventilator settings, mental status)	High-dimensional state vector (continuous + categorical + temporal features)	Encod and h patient
Action Space	Clinician interventions (PS adjustment, SBT initiation, extubation)	Discrete action set with safety constraints	De pern deci
Trajectory	Patient weaning course over days	Sequence of state-action pairs	Ca ten evo
Clinician Preference	Expert judgment comparing patient courses	Pairwise comparison dataset	Sup rewar tra
Reward Model	Implicit valuation of outcomes and processes	Neural network approximating preference ordering	Assig rew traje
Policy	Clinician decision-making strategy	Learned mapping from states to actions	Ger recomm
Safety Constraints	Clinical contraindications (e.g., instability)	Hard-coded rule filters	Preven ac
Learning Loop	Experience accumulation in ICU practice	Iterative retraining with new data	Con impro

Counterarguments Addressed

"RLHF is too complex for ICU deployment"

Critics argue RLHF requires specialized expertise unavailable in most ICUs [14, 16]. This confuses development complexity with use complexity: intensive computation occurs offline, while the deployed agent shows simple recommendations clinicians can accept, modify, or reject [1, 11]. EHRs already collect needed state variables, and inference requires minimal computational resources [11, 14]. The barriers are organizational and cultural, not technical [15, 16]. Current PMV weaning is already extraordinarily complex, requiring coordination across multiple teams [9, 16]. RLHF reduces cognitive load rather than increasing it [11, 14].

"Clinicians won't trust AI recommendations"

Critics worry clinicians will ignore opaque systems [12, 16]. RLHF builds trust because it is explicitly trained to align with clinician preferences—trust develops through experience across many cases rather than requiring interpretability [1, 11]. The preference learning framework lets clinicians override errors, collecting implicit feedback to adapt to local practice through periodic retraining [1, 19]. Many clinicians already deviate from protocols they distrust, implementing private weaning policies [9, 15]. Clinicians will trust an RLHF system because it learned from people like them, not from guideline committees [1, 11].

"SBT protocols work well enough"

For simple patients weaning within 7 days, protocols work adequately [5, 9]. But PMV patients (5-15% of ventilated patients) see protocol failure in 30-50% of cases [2, 3]. Each unnecessary ventilation day risks pneumonia, barotrauma, and delirium while consuming ICU resources [4, 9]. Even reducing PMV failure from 40% to 30% would prevent hundreds of deaths annually [2, 3]. Current protocols are not the ceiling—some centers achieve 70-80% success while others achieve 40-50% [8, 11]. RLHF captures and disseminates best practices across centers [1, 11]. Accepting current performance as "good enough" condemns PMV patients to preventable harm [8, 12].

Mplementation Pathway

Data collection phase

The first phase retrospectively extracts state variables, actions, and outcomes from EHRs for all PMV weaning episodes over 2-3 years, constructing complete trajectories

from weaning initiation to 72 hours post-extubation [14, 16]. These enable offline policy training without patient exposure [1, 19]. Concurrently, 5-10 clinicians provide preferences on 1,000-2,000 trajectory pairs varying by early vs. late extubation, gradual vs. abrupt pressure support reduction, and multiple short vs. fewer long SBTs [1, 11]. Chart review validates extracted representations against clinical reality, adjusting for discrepancies like undocumented coughs terminating SBTs [14, 15].

Prospective silent mode

For 3-6 months or 50-100 episodes, the agent generates recommendations recorded but not shown to clinicians [11, 14]. The team compares agent recommendations to clinician actions, investigating disagreements due to missing state information or genuine philosophical differences addressed by additional preference data [14, 15]. Clinician deviations from agent recommendations provide abundant but noisy implicit preference data that, combined with explicit comparisons, improve sample efficiency and alignment [11, 14].

Active deployment

The interface shows recommendations with rationale and allows accept/modify/override, maintaining clinician-in-the-loop control [11, 14]. Override reasons are recorded for future learning [11, 16]. Continuous monitoring tracks extubation failure, reintubation, and 7-day mortality with predefined stopping rules and statistical process control to detect deviations rapidly [14, 15]. Low clinician acceptance triggers investigation and retraining [1, 16]. A continuous learning loop retrains monthly or on-demand using accumulated preference data, with validation on holdout before redeployment [1, 11].

Conclusion

Protocol-based weaning fails PMV patients (5-15% of ventilated ICU patients, 30-50% mortality) by ignoring sequential dependencies and clinical judgment. RLHF captures sequential structure by formulating weaning as a Markov decision process and learns reward functions from pairwise clinician comparisons instead of hand-engineered criteria. The technology and data exist. What is missing is investment in preference collection and prospective trials. PMV patients cannot wait. While static machine learning models for weaning prediction have demonstrated value, they address the wrong question—predicting success rather than prescribing actions. RLHF represents the necessary evolution toward dynamic, sequential, preference-aligned decision support.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 17 Oct 2023 Revised: 21 Dec 2023 Accepted: 22 Jan 2024
Published online: 20 July 2024

References

Christiano PF, Leike J, Brown T, Martic M, Legg S, Amodei D. Deep reinforcement learning from human preferences. Adv Neural Inf Process Syst. 2017;30.

Stiennon N, Ouyang L, Wu J, Ziegler D, Lowe R, Voss C, et al. Learning to summarize with human feedback. Adv Neural Inf Process Syst. 2020;33:3008-21.

Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. Adv Neural Inf Process Syst. 2022;35:27730-44.

Sendak M, Elish MC, Gao M, Futoma J, Ratliff W, Nichols M, et al. "The human body is a black box": supporting clinical decision-making with deep learning. In: Proc 2020 Conf Fairness Accountability Transparency. 2020. p. 99-109.

Yu C, Liu J, Zhao H. Inverse reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. BMC Med Inform Decis Mak. 2019;19(Suppl 2):57.

Yu C, Ren G, Dong Y. Supervised-actor-critic reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. BMC Med Inform Decis Mak. 2020;20(Suppl 3):124.

Huang HY, Huang CY, Li LF. Prolonged mechanical ventilation: outcomes and management. J Clin Med. 2022;11(9):2451.

den Hengst F, Otten M, Elbers P, van Harmelen F, François-Lavet V, Hoogendoorn M. Guideline-informed reinforcement learning for mechanical ventilation in critical care. Artif Intell Med. 2024;147:102742.

Roggeveen LF, Hassouni AE, de Grooth HJ, Girbes AR, Hoogendoorn M, Elbers PW, et al. Reinforcement learning for intensive care medicine: actionable clinical insights from novel approaches to reward shaping and off-policy model evaluation. Intensive Care Med Exp. 2024;12(1):32.

Burns KEA, Rochwerg B, Seely AJ. Ventilator weaning and extubation. Crit Care Clin. 2024;40(2):391-408.

Lin MY, Li CC, Lin PH, Wang JL, Chan MC, Wu CL, et al. Explainable machine learning to predict successful weaning among patients requiring prolonged mechanical ventilation: a retrospective cohort study in central Taiwan. Front Med (Lausanne). 2021;8:663739.

Marshall DC, Komorowski M. Is artificial intelligence ready to solve mechanical ventilation? Computer says blow. Br J Anaesth. 2022;128(2):231-3.

Jhou HJ, Chen PH, Ou-Yang LJ, Lin C, Tang SE, Lee CH. Methods of weaning from mechanical ventilation in adult: a network meta-analysis. Front Med (Lausanne). 2021;8:752984.

Misseri G, Piattoli M, Cuttone G, Gregoretti C, Bignami EG. Artificial intelligence for mechanical ventilation: a transformative shift in critical care. Ther Adv Pulm Crit Care Med. 2024;19:29768675241298918.

Balagopalan A, Baldini I, Celi LA, Gichoya J, McCoy LG, Naumann T, et al. Machine learning for healthcare that matters: reorienting from technical novelty to equitable impact. PLOS Digit Health. 2024;3(4):e0000474.

Sblendorio E, Dentamaro V, Cascio AL, Germini F, Piredda M, Cicolini G. Integrating human expertise and automated methods for a dynamic and multi-parametric evaluation of large language models' feasibility in clinical decision-making. Int J Med Inform. 2024;188:105501.

Lee CS, Chen NH, Chuang LP, Chang CH, Li LF, Lin SW, et al. Hypercapnic ventilatory response in the weaning of patients with prolonged mechanical ventilation. Can Respir J. 2017;2017(1):7381424.

Ghiani A, Paderewska J, Sainis A, Crispin A, Walcher S, Neurohr C. Variables predicting weaning outcome in prolonged mechanically ventilated tracheotomized patients: a retrospective study. J Intensive Care. 2020;8(1):19.

Liao KM, Ko SC, Liu CF, Cheng KC, Chen CM, Sung MI, et al. Development of an interactive AI system for the optimal timing prediction of successful weaning from mechanical ventilation for patients in respiratory care centers. Diagnostics (Basel). 2022;12(4):975.

Cheng KH, Tan MC, Chang YJ, Lin CW, Lin YH, Chang TM, et al. The feasibility of a machine learning approach in predicting successful ventilator mode shifting for adult patients in the medical intensive care unit. Medicina (Kaunas). 2022;58(3):360.

Park JE, Kim TY, Jung YJ, Han C, Park CM, Park JH, et al. Biosignal-based digital biomarkers for prediction of ventilator weaning success. Int J Environ Res Public Health. 2021;18(17):9229.

Park JE, Kim DY, Park JW, Jung YJ, Lee KS, Park JH, et al. Development of a machine learning model for predicting weaning outcomes based solely on continuous ventilator parameters during spontaneous breathing trials. Bioengineering (Basel). 2023;10(10):1163.

Pai KC, Su SA, Chan MC, Wu CL, Chao WC. Explainable machine learning approach to predict extubation in critically ill ventilated patients: a retrospective study in central Taiwan. BMC Anesthesiol. 2022;22(1):351.

Sheikhalishahi S, Kaspar M, Zaghdoudi S, Sander J, Simon P, Geisler BP, et al. Predicting successful weaning from mechanical ventilation by reduction in positive end-expiratory pressure level using machine learning. PLOS Digit Health. 2024;3(3):e0000478.

Huang KY, Hsu YL, Chen HC, Horng MH, Chung CL, Lin CH, et al. Developing a machine-learning model for real-time prediction of successful extubation in mechanically ventilated patients using time-series ventilator-derived parameters. Front Med (Lausanne). 2023;10:1167445.

Jia Y, Kaul C, Lawton T, Murray-Smith R, Habli I. Prediction of weaning from mechanical ventilation using convolutional neural networks. Artif Intell Med. 2021;117:102087.

Torrini F, Gendreau S, Morel J, Carteaux G, Thille AW, Antonelli M, et al. Prediction of extubation outcome in critically ill patients: a systematic review and meta-analysis. Crit Care. 2021;25(1):391.

Leonov Y, Kisil I, Perlov A, Stoichev V, Ginzburg Y, Nazarenko A, et al. Predictors of successful weaning in patients requiring extremely prolonged mechanical ventilation. Adv Respir Med. 2020;88(6):477-84.