ORIGINAL RESEARCH **Open access**

Deep Reinforcement Learning with Inverse Reinforcement Learning for Learning Optimal Personalized Rehabilitation Exercise Prescriptions from Physical Therapist Demonstrations

Lucas Fernandez[1*], Diego Martinez[1], Pablo Ruiz[2]

Abstract

Personalized rehabilitation exercise prescriptions are essential for recovery after neurological injury, orthopedic surgery, and chronic decline. While physical therapists have valuable expertise, translating it into scalable computational systems is challenging. Standard deep reinforcement learning relies on manually defined reward functions, but in rehabilitation, clinically significant goals like movement quality, fatigue, pain, safety, motivation, and adherence are difficult to quantify. This paper introduces a framework combining inverse reinforcement learning (IRL) and deep reinforcement learning (DRL) to learn personalized rehabilitation prescriptions from therapist demonstrations. IRL would derive expert-aligned rewards, and DRL would use these to create adaptive exercise plans. The framework encompasses therapist demonstration collection, movement trajectory representation, reward inference, policy learning, safety constraints, and clinical oversight. Demonstrations would include exercise selection, progression decisions, and therapist responses to patient fatigue, pain, or adherence issues. IRL could capture implicit clinical priorities, while DRL would adjust prescriptions based on patient conditions such as fatigue, progress, and engagement. The framework aims to create scalable, personalized rehabilitation prescriptions, offering a conceptual model for future rehabilitation robotics, exergaming, and home-based digital rehabilitation systems.

Keywords Deep reinforcement learning, Inverse reinforcement learning, Personalized rehabilitation, Physical therapist demonstrations, Exergaming, Reward learning

*Correspondence:
Lucas Fernandez
lucas.fernandez@gmail.com
[1] Department of Intelligent Healthcare Analytics, University of Madrid, Madrid, Spain
[2] Department of AI Clinical Systems, University of Barcelona, Barcelona, Spain

Introduction

Rehabilitation exercise prescription is central to recovery after stroke, orthopedic surgery, neurological injury, and chronic mobility impairment because patients often require repeated practice, progressive challenge, and movement-quality feedback. Home-based guidance systems, exergaming platforms, and telerehabilitation robots have been developed to extend therapy beyond the clinic, but these systems still require careful personalization to patient ability, safety, and adherence patterns [1-4]. Physical therapists remain essential because they integrate clinical observation, patient goals, compensatory movement patterns, and recovery trajectories into exercise decisions. However, the scarcity of expert supervision and the growing demand for remote rehabilitation motivate computational frameworks that could scale personalized exercise prescription while preserving therapist-aligned clinical reasoning [5-7].

Deep reinforcement learning offers a principled approach for sequential decision-making because an agent can learn policies that adapt actions to changing patient states over time. In healthcare, reinforcement learning has been applied to personalized treatment recommendation and just-in-time adaptive intervention design, showing how sequential policies could support individualized care rather than static protocols [8, 9]. In rehabilitation, deep reinforcement learning has also been explored for locomotion simulation, exoskeleton control, assistive ankle-foot orthosis design, and adaptive robotic assistance, suggesting relevance for exercise progression and motor recovery support [10-16]. Yet a major limitation is that reward functions must usually be specified manually, while clinical goals such as safe movement quality, motivation, fatigue control, pain avoidance, and adherence are difficult to encode as explicit scalar objectives [17, 18].

A structural comparison of learning paradigms highlights how the proposed IRL-driven deep reinforcement learning framework uniquely integrates expert knowledge with adaptive decision-making **(Table 1).**

Table 1. Structural Comparison of Learning Paradigms for Rehabilitation Exercise Prescription: From Rule-Based Systems to IRL-Driven Deep RL

Dimension	Rule-Based Systems	Supervised Learning	Reinforcement Learning (Manual Reward)	IRL + Deep RL Framework (Proposed)
Knowledge Source	Explicit clinical rules	Labeled datasets	Hand-designed reward	Therapist demonstrations
Personalization	Limited	Moderate (depends on data)	High (state-dependent)	High (state-dependent + expert-aligned)
Adaptation Over Time	Static	Limited	Dynamic	Dynamic and context-sensitive
Reward Specification	None (rules only)	Implicit in labels	Manually engineered	Inferred from expert behavior
Handling of Tacit Knowledge	Poor	Partial	Poor	Strong (via IRL)
Interpretability	High	Moderate	Low–Moderate	Moderate–High (if reward decomposed)
Safety Integration	Rule-based constraints	Data-dependent	Requires explicit constraints	Integrated with constraints + demonstrations
Robustness to Patient Variability	Low	Moderate	Moderate	High
Data Requirements	Low	High labeled data	Moderate interaction data	High-quality demonstration data
Clinical Alignment	High initially	Variable	Often weak	Strong (expert-aligned)
Scalability	Limited	High	High	High (with supervision)
Key Limitation	Inflexibility	Label dependency	Reward design difficulty	Demonstration quality + reward ambiguity

Inverse reinforcement learning addresses this reward-design challenge by inferring the latent reward structure underlying expert demonstrations rather than requiring all clinical priorities to be hand-coded. Surveys and methodological studies describe IRL as a family of approaches for recovering reward functions from observed behavior, including settings where expert actions are stochastic, contextual, or high-dimensional [19-21]. Physical therapists naturally generate demonstrations through exercise selection, manual guidance, corrective feedback, progression decisions, and adaptation to patient response. These demonstrations could allow an IRL module to infer what therapists implicitly value, such as prioritizing form before intensity, rest before overload, or adherence before aggressive progression [22, 23].

This manuscript proposes a conceptual framework that combines inverse reinforcement learning with deep reinforcement learning to learn personalized rehabilitation exercise prescriptions from physical therapist demonstrations. The framework would first collect therapist-guided demonstrations, infer expert-aligned rewards through IRL, and then train a deep RL policy that recommends patient-specific exercises, intensity, duration, rest, and progression decisions [17, 22, 24]. The article is conceptual and does not report experiments, dataset sizes, reward values, or performance outcomes. The remaining sections describe the rehabilitation background, high-level architecture, demonstration collection strategy, reward-learning module, and how the inferred reward could support later patient-specific policy learning [8, 18, 25].

Background

Rehabilitation exercise prescription

Rehabilitation exercise prescription typically combines strength training, range-of-motion practice, balance tasks, coordination activities, gait training, and task-specific motor practice. Clinical progression depends on patient-specific factors such as baseline function, movement compensation, pain response, fatigue tolerance, motivation, and adherence history. Digital rehabilitation and exergaming studies illustrate how personalized motor-cognitive training, home guidance, and adaptive interfaces can support repeated practice while tracking exercise execution outside conventional therapy sessions [1-3]. A reinforcement learning framework for rehabilitation would need to represent these exercise categories and progression principles as sequential decisions rather than isolated recommendations [17, 18].

Deep reinforcement learning for healthcare

Deep reinforcement learning is relevant to healthcare because many clinical decisions are sequential, patient-specific, and dependent on delayed responses. Personalized treatment recommendation and adaptive digital intervention research show that RL can model individualized decision policies, but they also expose the challenge of designing reward signals that reflect complex clinical goals [8, 9]. In rehabilitation, deep RL has been used conceptually or computationally for simulated locomotion control, exoskeleton assistance, ankle-foot orthosis policy design, and robot-assisted exercise adaptation [10-16]. These studies motivate the present framework, while also showing why reward engineering must be clinically grounded before RL-generated exercise prescriptions can be trusted.

Inverse reinforcement learning

Inverse reinforcement learning seeks to infer a reward function from expert behavior, allowing a learning system to recover the preferences that make observed demonstrations appear purposeful. Contemporary IRL research emphasizes challenges such as reward ambiguity, stochastic demonstrations, contextual decision processes, and the relationship between imitation learning and reward inference [19-21]. Maximum entropy IRL is useful when expert behavior may vary across contexts, whereas adversarial and deep IRL approaches are attractive when state and action spaces include high-dimensional trajectories or sensor-derived representations. In rehabilitation, these properties matter because therapist decisions are adaptive, context-sensitive, and shaped by patient movement quality, fatigue, pain, and adherence constraints [22, 23].

Learning from demonstration in robotics

Learning from demonstration has been widely used in robotics to transfer expert behavior into policies or reward models, especially when direct manual programming is impractical. Example-guided deep reinforcement learning and apprenticeship-style methods show how demonstrated trajectories can shape skill acquisition in complex movement domains [24, 26]. Robotic manipulation studies using IRL illustrate a pathway from expert demonstrations to inferred reward functions that guide subsequent policy learning [24, 27]. In healthcare, therapist demonstrations could be collected through motion capture, wearable sensing, teleoperation, video observation, or therapist-supervised robotic rehabilitation sessions, allowing clinical expertise to be represented as demonstrations rather than verbal rules [22, 23, 28].

Personalization in rehabilitation

Personalization is essential because patients differ in age, baseline function, diagnosis, comorbidities, motivation, home environment, pain sensitivity, fatigue response, and recovery trajectory. Personalized exergames and home rehabilitation systems have therefore emphasized tailoring task difficulty, feedback, and interaction design to individual patient capabilities rather than relying on uniform protocols [1-7, 29, 30]. Reinforcement learning is attractive in this setting because the agent would learn a policy conditioned on patient state, making prescription decisions responsive to changing recovery patterns [17, 18]. However, personalization must remain clinically constrained because the safest action for one patient may be inappropriate for another with different surgical precautions, neurological impairments, or adherence barriers [2, 4].

Framework Overview

High-level architecture

The proposed architecture begins with physical therapist demonstrations of rehabilitation sessions, including exercise selection, movement correction, progression decisions, rest periods, and responses to patient-reported symptoms. An inverse reinforcement learning module would use these demonstrations to infer a reward function that reflects the therapist's implicit clinical priorities, building on the principle that rewards can be recovered from observed expert behavior [19, 20, 22]. A deep reinforcement learning policy would then use the inferred reward to recommend personalized exercise prescriptions based on each patient's current state, including capability, fatigue, pain, adherence history, and recovery stage [8, 17, 18]. After the patient completes prescribed exercises through a clinical, home-based, robotic, or exergaming platform, newly observed state information would update the next prescription cycle [3, 4, 7].

The proposed framework integrates therapist demonstrations, inverse reinforcement learning–based reward inference, and deep reinforcement learning–based policy learning into a unified hierarchical architecture for personalized rehabilitation prescription **(Figure 1).**

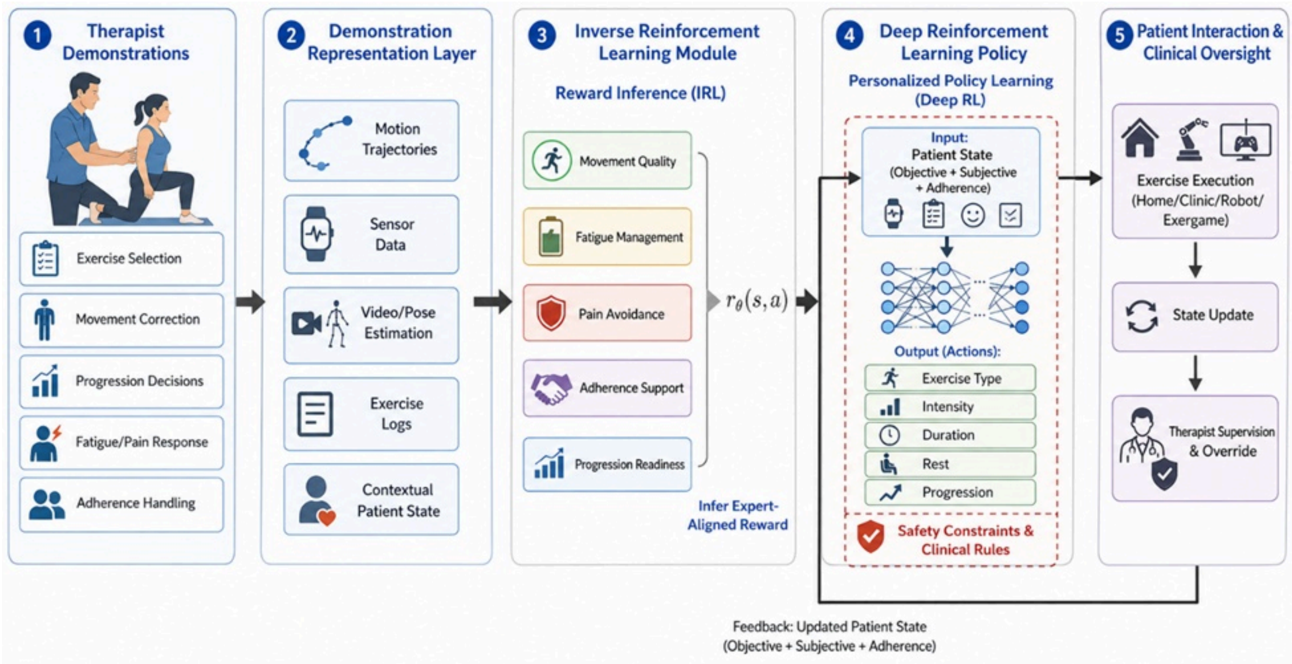

Figure 1. Hierarchical Architecture of Inverse Reinforcement Learning–Driven Deep Reinforcement Learning for Personalized Rehabilitation Exercise Prescription

Core assumptions

The framework assumes that therapist demonstrations can be captured with sufficient fidelity to represent clinically meaningful decisions and patient responses. Demonstration data could include motion trajectories, wearable sensor streams, exercise logs, video-based pose estimates, therapist feedback, and contextual information about pain, fatigue, and readiness [3, 25, 31]. It also assumes that policy learning can occur in a simulator, constrained digital environment, retrospective decision model, or closely supervised clinical platform before any autonomous recommendation is used in practice [10, 15, 17]. Because rehabilitation actions have safety implications, the framework treats therapist oversight and constrained exploration as design requirements rather than optional extensions [14, 18].

Design principles

The first design principle is expert alignment, meaning that the reward function should reflect therapist demonstrations rather than arbitrary numerical proxies. The second is patient-specific adaptation, where the policy would condition prescriptions on individual functional status, response history, and engagement patterns instead of applying a fixed protocol [9, 17, 18]. The third is safety-constrained optimization, in which the action space would be limited by clinical rules, therapist-defined contraindications, and observed patient tolerance [12-14]. The fourth is interpretability, because the recovered reward should be decomposable into clinically understandable components such as movement quality, adherence support, fatigue management, pain avoidance, and progression readiness [19, 21].

Expert Demonstration Collection

Demonstration modalities

Therapist demonstrations could be collected through multiple modalities, including motion capture of therapist-guided exercises, inertial sensors worn by patients, robotic device logs, video-based pose estimation, and structured exercise records. Deep learning systems for rehabilitation assessment show that movement quality can be represented from sensor or vision data, supporting the idea that demonstrations can include both kinematic execution and qualitative performance features [25, 31]. Home-based guidance and telerehabilitation platforms further suggest that patient movement, feedback, and exercise completion can be observed beyond the clinic when appropriate sensing and user interfaces are available [3, 4]. These modalities would allow the IRL module to observe not only what exercise was selected, but also how the therapist responded to patient performance during the session [22, 23].

Demonstration content

The content of a therapist demonstration would include exercise choice, exercise order, progression decisions, dosage adjustments, rest timing, movement-quality correction, and adaptation to fatigue or pain. For example, a therapist might reduce intensity when compensatory movement appears, repeat a lower-level task when coordination deteriorates, or progress difficulty when form remains stable across practice. These decisions are analogous to expert demonstrations used in IRL and learning-from-demonstration systems, where observed trajectories reveal preferences that may not be explicitly documented [24, 26, 27]. In rehabilitation, the demonstration should therefore encode both observable movement trajectories and clinical decisions about when to challenge, maintain, regress, or stop an exercise [17, 18, 28].

Inverse Reinforcement Learning

Reward function representation

The reward function could be represented as a linear combination of clinically interpretable features or as a neural reward model over high-dimensional patient and movement states. Interpretable features might include exercise completion, movement smoothness, range-of-motion quality, pain avoidance, fatigue management, adherence support, balance safety, and progression readiness. A neural reward model could capture more complex interactions among movement trajectories, patient state, and therapist actions, especially when demonstrations include video, sensor streams, and longitudinal adherence histories [19, 21, 25]. The conceptual goal is not to assign fixed reward values, but to allow the framework to infer what therapists appear to prioritize when selecting and adapting rehabilitation exercises [22, 23].

The inferred reward structure can be analytically decomposed into clinically meaningful components that shape policy behavior across competing rehabilitation objectives **(Table 2).**

Table 2. Analytical Decomposition of Therapist-Inferred Reward Components and Their Functional Roles in Rehabilitation Policy Learning

Reward Component	Clinical Meaning	Observable Proxies	Policy-Level Influence	Potential Trade-offs
Movement Quality	Prioritization of correct biomechanics and form	Kinematic smoothness, joint alignment, trajectory error	Encourages form-preserving exercise selection and regression when compensation occurs	May reduce intensity progression speed
Fatigue Management	Avoidance of excessive physical exhaustion	Heart rate trends, repetition decline, rest frequency	Introduces rest periods and moderates workload	May limit training volume
Pain Avoidance	Prevention of symptom exacerbation	Patient-reported pain, movement guarding	Reduces intensity or switches exercise modality	Risk of under-challenging patient
Adherence Support	Sustaining long-term engagement	Session completion, dropout patterns	Adjusts difficulty and pacing to maintain participation	May compromise optimal physical challenge

Progression Readiness	Timing of increased difficulty	Stable performance metrics, low variability	Gradual increase in intensity and complexity	Risk of delayed advancement
Safety Compliance	Respect for clinical constraints	Contraindications, surgical restrictions	Masks unsafe actions and restricts policy space	Limits exploration and learning diversity
Motivation Optimization	Enhancing patient willingness to continue	Self-reported motivation, engagement signals	Tailors feedback and task difficulty	Hard to quantify reliably
Recovery Efficiency	Balancing speed and safety of recovery	Multi-session improvement trends	Aligns long-term policy toward functional gains	May conflict with short-term comfort

IRL algorithm selection

Maximum entropy IRL would be appropriate when therapist behavior is expected to vary across patients or sessions while still reflecting coherent clinical preferences. Adversarial IRL and imitation-learning-related approaches would be relevant when patient states and movement trajectories are high-dimensional, as in sensor-based exergaming, robotic therapy, or video-derived movement analysis [20, 25, 26]. Contextual IRL could also support the framework because therapist choices depend on patient-specific context, including diagnosis, impairment level, fatigue, pain, adherence history, and recovery stage [17, 21]. Algorithm selection should therefore depend on the demonstration modality, state dimensionality, desired interpretability, and the degree to which therapist behavior is expected to be stochastic rather than deterministic [19, 24].

Recovered reward interpretation

The recovered reward should be interpreted as an expert-aligned clinical preference model rather than a direct substitute for therapist judgment. It could reveal that therapists implicitly prioritize movement form before repetition count, rest after deteriorating control, gradual intensity escalation after stable performance, or adherence-preserving task difficulty when motivation is fragile [5, 18, 22]. Such interpretation would be especially valuable in exergaming and home rehabilitation platforms, where automated systems must balance challenge, safety, usability, and sustained engagement [7, 29, 30, 32]. By making latent reward components explicit, the IRL module could provide a bridge between tacit therapist expertise and the later deep RL policy that would generate personalized prescriptions [8, 17, 19].

Deep RL for Personalization

Policy architecture

The deep reinforcement learning policy would map patient states to rehabilitation prescription actions, including exercise type, intensity, duration, rest timing, feedback emphasis, and progression level. A value-based architecture such as a deep Q-network could be used when prescriptions are represented as discrete exercise choices, while policy-gradient or actor-critic architectures could support continuous adjustment of intensity or assistance [8, 17]. The state would include patient capabilities, pain, fatigue, adherence history, recovery stage, and sensor-derived movement quality, while the action would represent the next recommended rehabilitation exercise prescription [10, 12, 18]. In robotic and exergaming environments, similar policy architectures could be adapted to prescribe assistance, task difficulty, or motor challenge while remaining grounded in therapist-inferred rewards [11, 13, 16].

Reward from IRL

The deep RL agent would use the reward function recovered by inverse reinforcement learning rather than relying on manually engineered clinical scores. This reward could encode therapist-aligned priorities such as safe movement quality, appropriate challenge, fatigue management, adherence support, and cautious progression after setbacks [19, 22, 23]. Because reward design is a major difficulty in healthcare RL, using demonstrations to infer reward structure could make the policy more clinically meaningful than a system optimized only for exercise completion or intensity escalation [8, 9, 20]. The agent would therefore learn prescriptions that are shaped by observed therapist behavior while still adapting to each patient's evolving state [17, 18].

Personalization mechanism

Personalization would occur by conditioning the policy on patient-specific state features and by adapting the policy as new patient responses are observed. A patient who progresses slowly, reports fatigue, or shows compensatory movement would receive different prescriptions than a patient who maintains movement quality and tolerates increased challenge [18, 25, 31]. In exergaming and home-based rehabilitation, this mechanism could adjust task difficulty, feedback, pacing, and engagement strategies without assuming that all patients benefit from the same trajectory [1, 5, 29]. Fine-tuning could be constrained by therapist-defined safety rules so that adaptation remains individualized but does not become clinically unsafe [2, 13, 14].

Patient State Representation

Objective measures

Objective patient state measures would include range of motion, strength, gait quality, balance performance, heart rate response, task completion, and sensor-derived movement quality. Wearable sensors, camera-based pose estimation, robotic device logs, and home guidance systems could support continuous or session-based observation of these features [3, 25, 31]. In neurological and orthopedic rehabilitation, such measures would help the agent distinguish between successful challenge, excessive difficulty, compensatory movement, and recovery-related improvement [10, 15, 16]. The state representation should therefore combine biomechanical capability, exercise execution quality, and recent response history rather than treating prescription as a static classification problem [17, 18].

Subjective measures

Subjective state measures would include pain, fatigue, perceived exertion, motivation, confidence, fear of movement, and patient-reported readiness. These variables matter because a prescription that is biomechanically feasible may still be inappropriate if the patient reports high pain, low motivation, or excessive fatigue. Personalized exergaming and digital rehabilitation studies emphasize that engagement, usability, and perceived challenge influence whether patients continue training outside supervised clinical settings [1, 5, 6, 29]. In the proposed framework, subjective measures would complement sensor-derived movement data so that the policy could adapt to both physical performance and patient experience [7, 9].

Adherence tracking

Adherence tracking would represent whether the patient completes prescribed sessions, follows exercise order, maintains duration, responds to reminders, and sustains participation over time. A deep RL policy could use adherence history as part of the patient state, allowing the agent to recommend exercises that preserve engagement rather than escalating difficulty in a way that undermines participation [5, 9]. Home-based systems, telerehabilitation platforms, and exergaming interfaces provide natural channels for capturing adherence-related signals, including session completion, missed sessions, early stopping, and patient feedback [3, 4, 30]. This feature is important because an expert-aligned rehabilitation policy should optimize for feasible long-term participation as well as immediate movement performance [2, 7].

Safe Exploration and Adaptation

Constrained policy

A constrained policy would prevent the agent from recommending unsafe exercises, excessive intensity, inappropriate resistance, or contraindicated movements. Action masking could be defined by therapist-specified rules, diagnosis-specific precautions, surgical restrictions, pain thresholds, fatigue indicators, and patient readiness criteria [12-14]. In rehabilitation robotics and exoskeleton control, safety constraints are essential because assistance and challenge must remain compatible with the patient's physical capacity and clinical condition [10, 11, 16]. The proposed framework would therefore treat the learned policy as operating inside a clinically bounded action space rather than as an unconstrained optimizer [17, 18].

Progressive adaptation

Progressive adaptation would begin from conservative prescriptions and expand challenge only when the patient demonstrates readiness through stable movement quality, tolerable fatigue, manageable pain, and adequate adherence. The agent would explore within therapist-defined safety bounds, selecting modest variations in exercise difficulty, duration, rest, or feedback rather than abrupt changes that could increase risk [13, 14]. Adaptive robotic assistance and personalized reaching rehabilitation show how reinforcement learning concepts can support individualized challenge while still requiring clinical safeguards [13, 18]. In home-based or exergaming contexts, this gradual progression would be especially important because therapist supervision may be intermittent rather than continuous [3, 7, 29].

Clinical Integration

Exercise delivery platform

The exercise delivery platform could be a mobile application, exergaming system, rehabilitation robot, wearable-supported home program, or hybrid clinic-home interface. The RL policy would recommend the daily or session-level prescription, while sensors and patient reports would update the state after the patient performs the exercise [3, 4, 30]. Exergaming and extended-reality rehabilitation designs show how interactive platforms can combine feedback, motivation, and task practice, making them suitable channels for adaptive exercise prescription [7, 29, 32]. The platform should present recommendations in clinically interpretable terms so that patients and therapists understand the intended exercise goal, safety boundary, and progression rationale [5, 6].

Therapist supervision

Therapist supervision would remain central because the framework is intended to support, not replace, clinical judgment. Therapists could review recommendations, override unsafe or inappropriate prescriptions, label unusual patient responses, and update constraints when medical status changes [2, 4, 28]. This supervision would also create new demonstrations and corrective examples, allowing the reward model and policy to remain aligned with evolving clinical practice [22, 23]. Exception handling would be required for injury, severe pain, non-adherence, unexpected neurological change, or any situation where automated prescription should defer to direct clinical evaluation [17, 18].

Evaluation Strategy

Reward recovery metrics

Reward recovery evaluation would examine whether the inferred reward reflects therapist-stated preferences and produces prescription patterns that clinicians judge to be reasonable. Conceptual metrics could include feature agreement between recovered reward components and therapist explanations, qualitative review of policy behavior after removing selected features, and clinician assessment of whether the reward emphasizes movement quality, safety, adherence, and progression readiness [19, 21, 22]. Because multiple reward functions can explain similar demonstrations, evaluation should focus on interpretability, clinical plausibility, and robustness rather than assuming a single uniquely correct reward [20]. The purpose would be to determine whether the reward model captures therapist-aligned priorities sufficiently to guide safe policy learning [23, 27].

Policy performance

Policy evaluation would begin in simulated patient models or supervised digital environments where different recovery trajectories, fatigue responses, pain patterns, and adherence behaviors can be represented conceptually. The learned policy could be compared against fixed protocols, therapist-only decision rules, or hand-coded reward policies to assess whether it generates clinically coherent adaptation without claiming experimental outcomes [8, 17]. Rehabilitation simulation, musculoskeletal modeling, and robotic control studies motivate the use of simulated or constrained environments before deployment in patient-facing systems [10, 15, 16]. Such evaluation would focus on policy logic, safety adherence, and responsiveness to patient state rather than reporting unsupported performance numbers [12, 14].

Clinical pilot metrics

A future clinical pilot would require prospective evaluation under therapist supervision, with attention to feasibility, safety, usability, adherence, patient satisfaction, therapist workload, and functional recovery indicators. Relevant functional measures might include timed mobility tasks, gait-related assessments, range-of-motion tracking, movement-quality review, and patient-reported experience, but the present conceptual article does not report outcomes or numerical findings [1, 2, 6]. Home guidance, telerehabilitation, and exergaming studies suggest that feasibility and engagement are important early evaluation targets before broader clinical adoption [3, 4, 7]. Any pilot should be designed to verify whether therapist-aligned AI recommendations can be integrated into real rehabilitation workflows without compromising patient safety [28, 32].

Limitations

Technical limitations

The framework assumes that therapist demonstrations are sufficiently high quality and representative, yet therapists may vary in style, experience, specialty, and tolerance for risk. Inverse reinforcement learning also faces reward unidentifiability, because more than one reward function may explain the same observed behavior [19, 21]. Deep RL introduces additional challenges, including distribution shift, simulation-to-reality gaps, sparse or noisy patient observations, and difficulty validating policies across heterogeneous rehabilitation populations [8, 15, 17]. These limitations mean that the proposed system should be viewed as a clinical decision-support architecture rather than an autonomous replacement for expert therapists [18, 22].

Clinical limitations

Patient safety is the central clinical limitation because exercise prescriptions can cause harm if they ignore surgical precautions, pain escalation, neurological deterioration, cardiovascular risk, or unexpected complications. Automated recommendations cannot anticipate every rare event, and home-based systems may miss subtle clinical cues that a therapist would detect in person [2-4]. Regulatory review, prospective validation, explainability, privacy protection, and clear responsibility boundaries would be necessary before such a framework could influence care [7, 28]. The system would also need to avoid widening access disparities by ensuring that sensing, exergaming, and mobile delivery platforms remain usable for patients with different impairments, resources, and digital literacy levels [1, 29, 30].

Conclusion

This manuscript has proposed a conceptual framework combining inverse reinforcement learning from physical therapist demonstrations with deep reinforcement learning for personalized rehabilitation exercise prescription. The central idea is that therapists would demonstrate expert decision-making, the framework could infer the implicit reward structure behind those decisions, and a deep reinforcement learning policy would use that reward to generate adaptive prescriptions.

The key advantage of the framework is that it could capture tacit clinical expertise without requiring every rehabilitation objective to be manually engineered. It could also support patient-specific adaptation by conditioning exercise decisions on changing movement quality, fatigue, pain, adherence, readiness, and recovery stage.

Important limitations remain. Demonstration quality, reward ambiguity, safety constraints, simulation-to-reality transfer, prospective validation, and regulatory oversight would all shape whether such a framework could be responsibly integrated into rehabilitation care.

Future work should implement this framework in rehabilitation robotics, exergaming systems, and home-based digital therapy platforms under therapist supervision. Prospective clinical trials should then evaluate feasibility, safety, interpretability, workflow integration, and patient-centered value before any autonomous deployment is considered.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 Jan 2026 Revised: 28 Jan 2026 Accepted: 19 Mar 2026
Published online: 20 July 2026

References

Huber SK, Held JP, de Bruin ED, Knols RH. Personalized motor-cognitive exergame training in chronic stroke patients—a feasibility study. Front Aging Neurosci. 2021;13:730801.
https://doi.org/10.3389/fnagi.2021.730801.

Nguyen AV, Ong YL, Luo CX, Thuraisingam T, Rubino M, Levin MF, et al. Virtual reality exergaming as adjunctive therapy in a sub-acute stroke rehabilitation setting: facilitators and barriers. Disabil Rehabil Assist Technol. 2019;14(8):1-12.
https://doi.org/10.1080/17483107.2019.1618403.

Lu HY, Wang X, Hu C, Lau CC, Tong RK. Home-based guidance training system with interactive visual feedback using Kinect on stroke survivors with moderate to severe motor impairment. J Neuroeng Rehabil. 2024;21(1):189.
https://doi.org/10.1186/s12984-024-01475-5.

Aguirre-Ollinger G, Chua KS, Ong PL, Kuah CW, Plunkett TK, Ng CY, et al. Telerehabilitation using a 2-D planar arm rehabilitation robot for hemiparetic stroke: a feasibility study of clinic-to-home exergaming therapy. J Neuroeng Rehabil. 2024;21(1):207.
https://doi.org/10.1186/s12984-024-01502-5.

Huber SK, Manser P, de Bruin ED. PEMOCS: theory derivation of a concept for PErsonalized MOtor-Cognitive exergame training in chronic Stroke—a methodological paper with an application example. Front Sports Act Living. 2024;6:1397949.
https://doi.org/10.3389/fspor.2024.1397949.

Huber SK, Knols RH, Held JP, Betschart M, de Bruin ED. PEMOCS: Evaluating the effects of a concept-guided, PErsonalised, MOtor-Cognitive exergame training on cognitive functions and gait in chronic Stroke—study protocol for a randomised controlled trial. Trials. 2024;25(1):451.
https://doi.org/10.1186/s13063-024-08289-2.

Lorenz EA, Bråten Støen A, Lie Fridheim M, Alsos OA. Design recommendations for XR-based motor rehabilitation exergames at home. Front Virtual Real. 2024;5:1340072.
https://doi.org/10.3389/frvir.2024.1340072.

Liu M, Shen X, Pan W. Deep reinforcement learning for personalized treatment recommendation. Stat Med. 2022;41(20):4034-56.
https://doi.org/10.1002/sim.9473.

Gönül S, Namlı T, Coşar A, Toroslu İH. A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions. Artif Intell Med. 2021;115:102062.
https://doi.org/10.1016/j.artmed.2021.102062.

Luo S, Androwis G, Adamovich S, Nunez E, Su H, Zhou X, et al. Robust walking control of a lower limb rehabilitation exoskeleton coupled with a musculoskeletal model via deep reinforcement learning. J Neuroeng Rehabil. 2023;20(1):34.
https://doi.org/10.1186/s12984-023-01157-8.

Luo S, Androwis G, Adamovich S, Su H, Nunez E, Zhou X. Reinforcement learning and control of a lower extremity exoskeleton for squat assistance. Front Robot AI. 2021;8:702845.
https://doi.org/10.3389/frobt.2021.702845.

Han JI, Lee JH, Choi HS, Kim JH, Choi J. Policy design for an ankle-foot orthosis using simulated physical human-robot interaction via deep reinforcement learning. IEEE Trans Neural Syst Rehabil Eng. 2022;30:2186-97.
https://doi.org/10.1109/TNSRE.2022.3195657.

Li A, Minto R, Dölling M, Boschetti G, Zanotto D. Personalized adaptive assistance with reinforcement learning control enhances engagement, performance, and retention in robot-assisted arm-reaching exercises. IEEE Trans Neural Syst Rehabil Eng. 2025;34:532-42.
https://doi.org/10.1109/TNSRE.2025.3528893.

Khan NA, Jamwal PK, Hussain F, Ghayesh MH, Hussain S. Reinforcement learning-driven path generation for ankle rehabilitation robot using musculoskeletal-informed energy optimization. IEEE Trans Neural Syst Rehabil Eng. 2025;33:1-10.

https://doi.org/10.1109/TNSRE.2025.3567891.

Song S, Kidziński Ł, Peng XB, Ong C, Hicks J, Levine S, et al. Deep reinforcement learning for modeling human locomotion control in neuromechanical simulation. J Neuroeng Rehabil. 2021;18(1):126.
https://doi.org/10.1186/s12984-021-00912-2.

De Vree L, Carloni R. Deep reinforcement learning for physics-based musculoskeletal simulations of healthy subjects and transfemoral prostheses' users during normal walking. IEEE Trans Neural Syst Rehabil Eng. 2021;29:607-18.
https://doi.org/10.1109/TNSRE.2021.3063301.

Ye D, Luo H, Winstein C, Schweighofer N. Towards AI-based precision rehabilitation via contextual model-based reinforcement learning. J Neuroeng Rehabil. 2025;22(1):263.
https://doi.org/10.1186/s12984-025-01622-8.

Pelosi AD, Roth N, Yehoshua T, Itah D, Braun Benyamin O, Dahan A. Personalized rehabilitation approach for reaching movement using reinforcement learning. Sci Rep. 2024;14(1):17675.
https://doi.org/10.1038/s41598-024-68479-z.

Arora S, Doshi P. A survey of inverse reinforcement learning: Challenges, methods and progress. Artif Intell. 2021;297:103500.
https://doi.org/10.1016/j.artint.2021.103500.

Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst. 2021;33(12):6999-7019.
https://doi.org/10.1109/TNNLS.2021.3084827.

Belogolovsky S, Korsunsky P, Mannor S, Tessler C, Zahavy T. Inverse reinforcement learning in contextual MDPs. Mach Learn. 2021;110(9):2295-334.
https://doi.org/10.1007/s10994-021-06027-8.

Bharadwaj VR, Miller BA, Novak VD, Jiang C. Learning skill training schedules from domain experts in a rehabilitation gym using inverse reinforcement learning. In: 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob). IEEE; 2024. p. 1815-21.
https://doi.org/10.1109/BIOROB60548.2024.10601609.

Luciani B, Pedrocchi A, Braghin F, Gandolla M. Learn from therapists' demonstrations approaches for robotic rehabilitation exercises. In: International Conference on NeuroRehabilitation. Cham: Springer; 2024. p. 583-7.
https://doi.org/10.1007/978-3-031-77589-8_92.

Naranjo-Campos FJ, Victores JG, Balaguer C. Expert-trajectory-based features for apprenticeship learning via inverse reinforcement learning for robotic manipulation. Appl Sci. 2024;14(23):11131.
https://doi.org/10.3390/app142311131.

Liao Y, Vakanski A, Xian M. A deep learning framework for assessing physical rehabilitation exercises. IEEE Trans Neural Syst Rehabil Eng. 2020;28(2):468-77.
https://doi.org/10.1109/TNSRE.2020.2966249.

Peng XB, Abbeel P, Levine S, Van de Panne M. DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Trans Graph. 2018;37(4):1-14.
https://doi.org/10.1145/3197517.3201311.

Ozalp R, Ucar A, Guzelis C. Advancements in deep reinforcement learning and inverse reinforcement learning for robotic manipulation: toward trustworthy, interpretable, and explainable artificial intelligence. IEEE Access. 2024;12:51840-58.
https://doi.org/10.1109/ACCESS.2024.3384457.

Martínez-Pascual D, Catalán JM, Lledó LD, Blanco-Ivorra A, García-Aracil N. A deep learning model for assistive decision-making during robot-aided rehabilitation therapies based on therapists' demonstrations. J Neuroeng Rehabil. 2025;22(1):18. https://doi.org/10.1186/s12984-025-01501-2.

Amiri Z, Sekhavat YA, Goljaryan S. StepAR: A personalized exergame for people with multiple sclerosis based on video-mapping. Entertain Comput. 2022;42:100487. https://doi.org/10.1016/j.entcom.2022.100487.

Vallejo D, Gómez-Portes C, Albusac J, Glez-Morcillo C, Castro-Schez JJ. Personalized exergames language: a novel approach to the automatic generation of personalized exergames for stroke patients. Appl Sci. 2020;10(20):7378. https://doi.org/10.3390/app10207378.

Mennella C, Maniscalco U, De Pietro G, Esposito M. A deep learning system to monitor and assess rehabilitation exercises in home-based remote and unsupervised conditions. Comput Biol Med. 2023;166:107485. https://doi.org/10.1016/j.compbiomed.2023.107485.

van Doorn PF, Nyman E Jr, Wishaupt K, van der Krogt MM, Roerdink M. Deriving motor states and mobility metrics from gamified augmented reality rehabilitation exercises in people with Parkinson's disease. Sensors. 2025;25(23):7172. https://doi.org/10.3390/s25237172.