Viewpoint

Governing Embodied AI in Clinical Care: From Digital Recommendation to Recoverability-Gated Physical Action

Jialin Liu[1,2*], MD; Siru Liu[3*], PhD

[1]Information Center, West China Hospital of Sichuan University, Chengdu, Sichuan, China
[2]Department of Otolaryngology-Head and Neck Surgery, West China Hospital of Sichuan University, Chengdu, Sichuan, China
[3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
*all authors contributed equally

Corresponding Author:

Siru Liu, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Ave, Suite 1475
Nashville, TN 37203
United States
Phone: 1 6158755216
Email: siru.liu@vumc.org

Abstract

Clinical AI is expanding from recommending care to participating in physical action through surgical assistance, rehabilitation, patient handling, and AI-connected actuators. These applications may improve precision and support clinical work, but errors can produce immediate, persistent bodily harm. This viewpoint argues that crossing the digital-physical actuation boundary changes the requirements for clinical AI governance. We define embodied clinical AI by its causal coupling to physical action and propose recoverability-gated entrustment: demonstrated competence is necessary, while the recoverability of credible failures constrains permitted action authority. Recoverability comprises detectability, interruptibility, reversibility or clinical compensability, and consequence boundedness. The complete perception-decision-action loop should be the unit of governance and action-grounded validation. Meaningful human control must be assigned where intervention can actually affect outcomes, and risk-proportionate forensic records must support safety learning and fair accountability. Our contribution is to integrate existing AI, medical-device, and robotics safeguards into an action-specific authorization approach, rather than propose a separate regulatory regime. We operationalize this approach through a physical entrustment ladder and PACER-AI: perception reliability; action scope and actuation envelope; clinical context and control locus; evidence through action-grounded validation; and recoverability, responsibility, and forensic sufficiency. An illustrative patient-repositioning decision flow shows how failed recovery criteria require restriction, redesign, or refusal of authorization. Intended for clinicians, institutional governance teams, developers, evaluators, and regulators, this conceptual framework seeks proportionate, interruptible, auditable, and accountable physical action and requires empirical validation.

J Med Internet Res 2026;28:e104088; doi: 10.2196/104088

Keywords: embodied AI; clinical AI; health care robotics; patient safety; AI governance; recoverability; action-grounded validation; accountability; human-machine teaming; forensic sufficiency

Introduction

Clinical AI has largely developed as disembodied intelligence: algorithms classify images, predict clinical deterioration, summarize health records, and recommend actions, while clinicians translate these digital outputs into physical care [1,2]. Two trajectories are converging: clinical AI is becoming more agentic within clinical workflows [3], while learned perception and control are making surgical, rehabilitation, and patient-handling systems more adaptive [4-6]. Their intersection is embodied clinical AI: systems that couple perception, decision-making, and action to alter the physical world, including the patient's body [7].

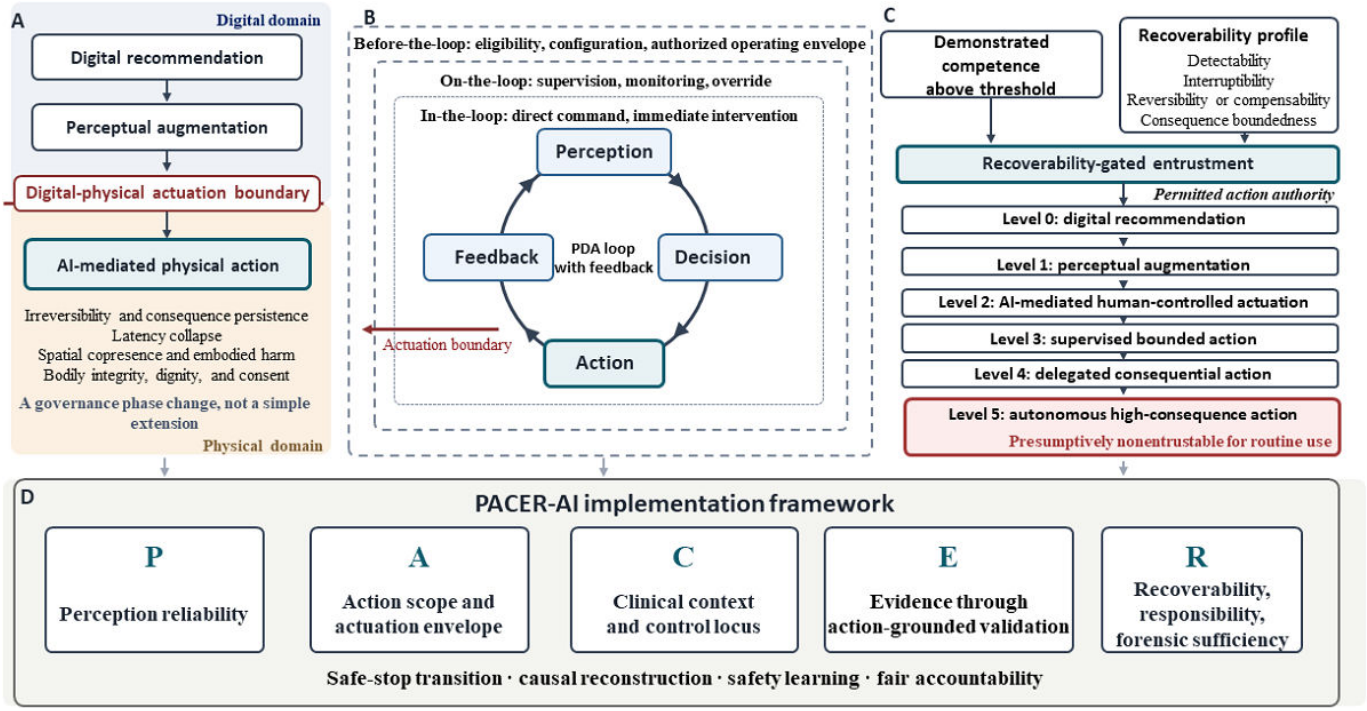
Vision-language-action models and generalist robot policies are expanding the capacity to translate instructions into actions and adapt across tasks and environments [8-10], supported by simulation, demonstration datasets, imitation learning, and reinforcement learning [9,11,12]. However,

reviews of health care applications reveal a heterogeneous and immature evidence base, with limited external, longitudinal, and real-world validation [12-14]. Technical progress therefore does not itself establish readiness for clinical physical action.

Clinical AI governance addresses permitted tasks, supervision, validation, monitoring, and accountability. We use conservation of accountability to denote the normative principle that introducing AI must not create accountability gaps: responsibilities must remain identifiable across clinicians, health care organizations, developers, manufacturers, and other actors involved in the system lifecycle [15-17]. Digital outputs can also cause irreversible harm, but physical actuation changes the timing and feasibility of intervention: human review may come too late once an output becomes force, motion, contact, positioning, or treatment delivery [4,18]. Crossing the digital-physical actuation boundary changes the object of governance: AI outputs enter the causal pathway of physical action. Justifying action authority, therefore, requires assessment of both task performance and the recoverability of credible failures.

Building on accounts of embodied AI as a closed-loop coupling of perception, decision-making, and action [12-14] and studies of robotic autonomy and meaningful human control [4,18,19], this viewpoint proposes an action-specific authorization approach grounded in recoverability. We aim to help clinicians, institutional governance teams, developers, evaluators, and regulators determine which physical actions may be entrusted to AI, for which patients, and under what conditions. We advance five linked claims as follows: (1) embodied clinical AI should be defined by its causal coupling to physical action, not its appearance; (2) crossing the digital-physical actuation boundary constitutes a governance phase change; (3) the complete perception-decision-action (PDA) loop should be the unit of governance and action-grounded validation; (4) competence is necessary for entrustment, but permitted action authority must also be constrained by recoverability; and (5) accountability requires forensic sufficiency and meaningful human control located where intervention is physically feasible (Figure 1).

Figure 1. Integrated governance model for embodied AI in clinical care. (A) The digital-physical actuation boundary marks the transition from informational outputs to AI-mediated physical action. (B) The perception-decision-action loop constitutes the unit of governance, with meaningful human control located in, on, or before the loop. (C) Demonstrated competence and action-specific recoverability jointly determine permitted action authority. (D) PACER-AI operationalizes these principles across five implementation domains. PACER-AI: perception reliability, action scope and actuation envelope, clinical context and control locus, evidence through action-grounded validation, and recoverability, responsibility, and forensic sufficiency; PDA: perception-decision-action.



Embodied AI as Action-Coupled Clinical AI

Defining Action-Coupled Clinical AI

For clinical governance, we define embodied clinical AI as action-coupled clinical AI: an AI-enabled system that integrates situated perception with inference, planning, or control within a physical actuation pathway in a clinical or care environment. Depending on their implementation, such systems may include surgical platforms, rehabilitation robots, smart beds, robotic imaging devices, mobile robots, intelligent prostheses, infusion controllers, or distributed systems that link digital intelligence to sensors and actuators [12,13]. The defining feature is not physical morphology but the causal coupling of AI-enabled processing to

physical actuation, such that system-generated outputs can initiate, modify, constrain, or terminate physical action rather than serve solely as information requiring human-mediated translation into action.

Clarifying Boundary Cases

This action-coupled definition clarifies several boundary cases. A chatbot or virtual avatar explaining rehabilitation exercises remains informational because a person must interpret its output and translate it into action. A conventional teleoperated surgical platform may be robotic without exercising AI-mediated control [4]. Conversely, a digital agent whose output directly adjusts an infusion rate or sets a dose for execution, not merely proposing a rate for clinician review, crosses the actuation boundary, even without a visibly robotic embodiment. Clinician review and authorization govern whether such action is permitted; they do not remove its coupling to actuation. What matters is whether AI output informs human action or enters a control pathway that physically alters the patient, a device, or the care environment, with potential benefit or harm to the patient.

Governing Action Authority Before Full Autonomy

Action coupling is not synonymous with full autonomy. AI may participate in physical action by stabilizing an instrument, adjusting resistance, constraining a trajectory, navigating under supervision, or executing a bounded task [4,18]. Perceptual augmentation alone remains outside the actuation boundary unless its outputs enter the actuation control pathway. Governance therefore begins before full autonomy, because even partial participation can alter trajectory, timing, force, dosage, or contact. The relevant questions are how much action authority the system exercises, where it enters the PDA loop, whether timely interruption and recovery are feasible, and whether responsibilities are clearly assigned.

The Digital-Physical Actuation Boundary

Overview

We use the term digital-physical actuation boundary to denote the threshold at which an AI output ceases to be solely informational and enters the actuation control pathway to initiate, modify, constrain, or terminate physical action. Crossing this boundary should trigger additional requirements for mechanical safety, environmental robustness, defined operating envelopes, meaningful human control, failure recovery, and postincident causal reconstruction [16,18,19]. Four properties of AI-mediated physical action underpin our argument that this transition constitutes a governance phase change.

Irreversibility and Consequence Persistence

Digital outputs can often be revised or rejected before they affect care: an alert can be dismissed, a draft edited, or an order canceled. Physical actions produce effects in the patient's body and care environment, ranging from readily reversible bed-angle adjustments, through persistent injuries sustained during patient transfer, to the effectively irreversible effects of tissue division or medication bolus delivery. The governing variable is therefore not physicality alone but consequence persistence: how long an effect endures, including after the action has stopped. Persistence informs recoverability but does not determine it; recoverability also depends on timely detection and interruption and the capacity to reverse, compensate for, or contain harmful consequences.

Latency Collapse

We use latency collapse to describe situations in which the available intervention window is shorter than the time required for effective human intervention. In some physical human-robot interaction architectures, sensing and torque-command loops operate at 1 kHz, updating system states and control commands every millisecond [20]. This frequency does not itself determine the intervention window. However, when a hazard develops faster than a person can detect it, decide how to respond, and stop the action, nominal human-in-the-loop oversight cannot provide timely protection. Meaningful human control must therefore extend beyond real-time manual intervention to include upstream task selection, system configuration, action limits, and predefined handback conditions, supported by automated safeguards within the control loop [19,20].

Spatial Copresence and Embodied Harm

Embodied systems share physical space with patients, staff, caregivers, and bystanders. Potential harms include collisions, obstruction, excessive or misdirected force, shear injury, dropped objects, compromised sterility, and disruption of care. Such harms may emerge from interactions among software, sensors, actuators, mechanical components, physical layout, clinical workflows, environmental conditions, and human behavior [21,22]. Safety should therefore be treated as an emergent property of the complete sociotechnical system in which physical action occurs, not merely as an attribute of an isolated model or robot [21].

Bodily Integrity, Dignity, and Consent

AI-mediated action on a person's body engages autonomy, bodily integrity, dignity, privacy, and the lived experience of being handled by a machine. These concerns are heightened for patients who are sedated, frail, cognitively impaired, immobilized, or unable to express discomfort or objection in real time [23,24]. Consent to data processing should therefore be distinguished from consent to AI-mediated physical care; authorization for one should not be presumed to authorize the other. Governance must specify who may authorize such action, how patients may refuse or withdraw permission, how authorized clinicians may pause, assume control

of, or terminate the system, and how patient vulnerability should shape the permissible action envelope and required safeguards.

The Perception-Decision-Action Loop as the Unit of Governance

Governing the Complete Action Loop

Embodied AI requires governance of the complete PDA loop. Perception estimates patient, device, environmental, and clinical states. Decision integrates these estimates with clinical goals and constraints, plans actions, estimates uncertainty, and selects, modifies, defers, or abstains from action. Action translates commands into movement, force, contact, positioning, or dose delivery. Feedback returns observed consequences, deviations, and signs of distress to perception and decision, supporting adaptation, interruption, or termination [12,13]. Human control is meaningful only when responsible actors have sufficient information, authority, time, interface support, and practical means to set action limits or intervene where intervention can be effective [19].

Failure Propagation Across the Loop

Failures can originate at any point in the PDA loop and propagate through subsequent stages. Sensors may be occluded, miscalibrated, or drifting; perception may misestimate the patient's state or posture; planning may produce clinically inappropriate actions; and actuators may deviate from commands. Feedback may miss deviations or distress, while interfaces may obscure the system's operating mode or action authority [18]. An erroneous state estimate can thus produce an inappropriate plan that is faithfully executed and remains uncorrected if feedback fails. Validation, monitoring, change control, incident investigation, and accountability should address interactions across the complete PDA loop, not merely isolated components [16,21].

Translating Uncertainty Into Safe Behavior

This loop-based view connects technical uncertainty with clinical governance. The SA-ROC (Safety-Aware Receiver Operating Characteristic) framework illustrates how prespecified reliability targets delineate rule-in and rule-out safe zones for automated diagnostic decisions, alongside a gray zone requiring human review [25]. For embodied systems, we propose extending this selective-automation principle to physical action authority: uncertainty should constrain permitted actions according to the severity and recoverability of credible harm. When perception, assessment of contraindications, prediction of action consequences, or confirmation of intended effects is insufficiently reliable, the system should restrict action or initiate a predefined fallback or handover, accounting for the risks of both continued action and delay or cessation. Abstention and controlled de-escalation are therefore safety functions when clinically appropriate, rather than evidence of inadequate intelligence.

Fragmentation Across Governance Layers

Existing frameworks provide complementary foundations. Trustworthy clinical AI frameworks address safety, fairness, accountability, and lifecycle monitoring [16]. Medical-device risk management addresses hazards, controls, and residual risk [26], while robotics equipment standards address basic safety and essential performance [27,28]. Device guidance addresses clinical evidence and change control [29,30]. These safeguards must be integrated across the components, people, environments, and workflows of the clinical action system [21]. The gap we address is how to connect them around specific physical actions and their consequences.

Our proposed contribution is an action-specific authorization logic linking competence and recovery evidence to permitted action authority. The complete PDA loop defines the unit of action-specific assessment. Intervention feasibility determines the human-control arrangement, extending meaningful-human-control frameworks by linking control locus to the recovery window, while forensic evidence supports investigation and accountability. This approach complements established risk management without independently establishing acceptable risk. Material changes in action scope, autonomy, environment, human control, failure pathways, or recoverability should trigger proportionate reassessment and revalidation, even when the nominal intended purpose remains unchanged.

Recoverability-Gated Entrustment

From Competence to Recoverability

Clinical AI evaluation and governance commonly begin with measures of competence: whether a system is accurate, calibrated, robust, and clinically useful for a specified task [16]. Competence remains necessary, but it cannot, by itself, justify authority to perform physical actions. Even highly competent systems can fail, and physical actuation can convert an infrequent error into immediate bodily or environmental harm. Competence characterizes task performance; recoverability determines whether failure consequences remain governable.

We define recoverability as the action system's capacity to manage unsafe actions and their consequences through four interacting dimensions: (1) detectability of deviation or failure; (2) interruptibility within the effective intervention window; (3) reversibility, clinical compensability, or mitigation of physical effects; and (4) boundedness of harmful consequences within clinically acceptable limits under credible failures. Complete reversibility is not required, but the dimensions must be assessed jointly. The profile depends on the action, the patient, the environment, monitoring, control configuration, and rescue resources [21,26].

Recoverability as a Gate to Entrustment

Under recoverability-gated entrustment, failure to meet any safety-critical recoverability criterion precludes authorization

at the proposed action scope, even when competence requirements are met. A bounded-energy, readily interruptible, and reversible action envelope may support broader authority if clinical benefit and residual risk justify it. Where credible failures unfold rapidly, are difficult to detect or interrupt, or produce irreversible or poorly compensable effects, the authorized envelope should be constrained by these recovery limitations, even when measured task performance is high. Authorization should be reduced or suspended when changes in monitoring, patient condition, staffing, or rescue capacity invalidate the approved recovery arrangements, even without a software change.

Operationalizing Recoverability

We propose representing recoverability as a structured, multidimensional profile with dimension-specific acceptance criteria, rather than a compensatory aggregate score. A highly reversible action may still be unsafe if a hazardous deviation cannot be detected in time, whereas reliable detection provides little protection when harm develops faster than effective intervention. Candidate measures depend on the action and context. Detection and interruption measures include deviation-detection time, intervention latency, and stopping time and distance. Measures of physical consequences include maximum unintended force or energy transferred to the patient and residual harm severity and duration. Recovery measures include physiological recovery achievable through reversal or clinical mitigation, the time required for effective rescue and its probability of success, and control-transfer reliability [20,26-28]. Acceptance criteria established for one action or patient context should not automatically transfer to another. Where effects are irreversible, authorization requires a justified clinical benefit, adequate prevention and containment measures, and acceptable residual risk; financial compensation does not constitute physical recovery.

A Recoverability-Gated Physical Entrustment Ladder

Table 1 translates recoverability-gated entrustment into graded categories of permitted action authority, with the digital-physical actuation boundary between levels 1 and 2. These categories describe risk-proportionate authorization, not technological progress; greater autonomy is not inherently preferable. At level 3, the system executes a clinician-selected action, including parameter adjustment within its authorized envelope; at level 4, it selects a clinical action or clinically meaningful target under a delegated protocol. Both require action-specific recoverability evidence; level 4 additionally requires validation of action-selection safety. Functions meeting level 5 criteria remain level 5 regardless of protocol-based delegation and are presumptively nonentrustable. Exceptional authorization requires all conditions specified in Table 1 to be met. These include meeting minimum recoverability criteria, demonstrating acceptable residual risk, and showing that no safer lower-level arrangement can reasonably achieve the clinical objective.

Table 1. Recoverability-gated physical entrustment ladder across the digital-physical boundary. Examples indicate authorized functions rather than fixed classifications of devices. Mixed systems should be classified by authorized function, with interactions assessed across the complete perception-decision-action (PDA) loop.

Level	Form of AI participation	Illustrative examples	Primary governance requirements
0: Digital recommendation	• AI informs care • A clinician independently decides whether and how to act	• Procedure guidance, rehabilitation education, and risk prediction	• Output validity, calibration, transparency, usability, and appropriate interpretation
1: Perceptual augmentation	• AI senses or interprets states without entering the actuation control pathway	• Instrument tracking, posture detection, surgical phase recognition, and anatomical localization	• Sensor validity, uncertainty handling, subgroup performance, environmental robustness, and failure communication
2: AI-mediated human-controlled actuation	• AI modifies direct human commands • The human retains immediate control and selects the clinical action	• AI-assisted teleoperation, tremor filtering, haptic guidance, virtual fixtures	• Command fidelity, interface integrity, validated latency, force and speed limits, and effective stopping
3: Supervised bounded action	• AI executes a clinician-selected action within authorized bounds • A supervisor can intervene within the validated recovery window	• Clinician-selected gait assistance, camera positioning, bed adjustment, or repositioning • Destination-authorized logistics navigation	• Validated action and recovery envelope, runtime monitoring, automated safeguards, effective override, reliable control transfer, and near-miss surveillance
4: Delegated consequential action	• AI selects a clinical action or clinically meaningful target under a delegated protocol	• Protocol-based selection of rehabilitation targets, imaging actions, repositioning maneuvers, or infusion adjustments	• Level 3 execution-safety and recoverability requirements plus validation of action-selection safety, including credible selection errors; patient eligibility, escalation and abstention

Level	Form of AI participation	Illustrative examples	Primary governance requirements
	• Human control operates primarily before and on the loop		rules, forensic sufficiency, and defined accountability
5: Autonomous high-consequence action with limited recoverability	• AI performs high-consequence interventions without immediate, case-specific authorization • Credible errors may escape timely control	• Autonomous invasive interventions or high-risk closed-loop therapy under conditions of limited recoverability	• Presumptively nonentrustable • Exceptional authorization requires independently replicated evidence and independent safety review, minimum recoverability criteria, acceptable residual risk, stringent restrictions, regulatory oversight, and justification that no safer lower-level arrangement can reasonably achieve the objective

Action-Grounded Validation

From Output Validation to Action-Grounded Validation

Model-level and output-level validation assess whether a prediction or recommendation is accurate, reliable, and clinically useful. We define action-grounded validation as evaluation of the complete PDA loop for safety, effectiveness, interruptibility, recoverability, equity, and clinical value within the intended context of use, including appropriate abstention and fallback behavior. It builds on model- and output-level validation by assessing the integrated performance of models, sensors, actuators, interfaces, human-control arrangements, workflows, and rescue mechanisms. The strength and scope of evidence should be commensurate with the entrusted action authority and authorized action scope, the characteristics and vulnerability of the intended patient population, and the severity, time course, and recoverability of credible failure consequences [16,26,29].

Specifying the Intended Action

The intended action must be specified at the level at which clinical benefit and harm arise. A device category such as a rehabilitation robot is too broad to define the validation target: gait support, resistance adjustment, balance perturbation, and sit-to-stand assistance differ in clinical objectives, hazards, operating conditions, and recovery requirements. Similarly, camera control, instrument stabilization, and tissue manipulation require distinct evidence, while scene recognition and trajectory warnings should be evaluated according to their contribution to the action pathway. Evidence established for one action should not be presumed to apply to another merely because both are performed by the same platform. Evidence transfer requires an explicit bridging analysis addressing similarities and material differences in action scope, actuation envelope, patient population, operating environment, human-control arrangements, credible failure modes, and recoverability, with additional validation where warranted [26,29].

Validating the Complete PDA Loop

Loop-level validation should cover normal, degraded, and emergency states under representative clinical variation, including occlusion, clutter, lighting changes, atypical anatomy, workflow interruptions, and network degradation [26-28]. Controlled fault and scenario testing should examine sensor dropout, communication delays, actuator faults, uncertain or out-of-distribution inputs, unexpected patient movement, and emergency-stop activation. Outcomes should include task performance, deviation detection, safe-state transitions, recovery, and safe resumption, together with control-transfer integrity, timeliness, and reliability, where human takeover is part of the design. Simulation and digital twins can expand coverage of rare, hazardous, and boundary conditions but should complement staged physical and clinical evaluation, with progression to real-world use being conditional on predefined safety and performance criteria [31].

Evaluating Human Control and Interaction

Human-system interaction is an integral component of system performance. Evaluation should assess whether clinicians understand the system's operating mode, action authority, and control locus, and recognize uncertainty, degraded states, and mode transitions. Where human takeover is part of the design, testing should establish whether override, emergency-stop, and takeover mechanisms can be used without unsafe delay or loss of control [19,21]. It should also assess patient experience, including comfort, dignity, and meaningful opportunities to refuse or withdraw permission; appropriate caregiver involvement; and human-factors outcomes such as calibrated trust, teamwork, situation awareness, workload, and supervision burden [21,23,24]. Task-level gains may not yield net clinical value if offset by increased monitoring demands, distraction, coordination costs, or unrecognized additional work.

Measuring Failure and Recovery

Safety evaluation should capture near misses, harm-preventing interventions, unexpected stops, workarounds, nuisance alarms, delayed or abandoned tasks, failed control transfers, and failures to reach or maintain a safe state. Absence of recorded injury in a small pilot provides limited assurance when clinicians repeatedly intervene to prevent harm. Planned assistance should be distinguished from unplanned takeover or rescue. Their frequency relative to task exposure

or operating time, timeliness, effectiveness, and burden should be reported as distinct safety outcomes, not subsumed within aggregate task-success rates [21,26,32]. Recovery outcomes, from deviation detection and interruption through control transfer to restoration of a safe state, including time to effective rescue and residual patient harm, should be measured with the same rigor as nominal task performance.

Demonstrating Clinical Value

Technical task success is not equivalent to patient benefit. Evaluation should compare the system with appropriate standard care across patient-centered and functional outcomes, adverse events, clinician workload, workflow and implementation burden, resource use, and, where appropriate, cost-effectiveness. It should also assess the distribution of benefits, burdens, and failures across patient groups and care settings [21,32]. Postdeployment monitoring should assess whether clinical performance, recoverability, and meaningful human control remain adequate as system components and conditions of use change [16,32]. Responsible clinical translation requires evidence that AI-mediated physical action provides net clinical value, not merely that a machine can complete a movement.

Forensic Sufficiency and the Conservation of Accountability

Conserving Accountability Across the Actuation Boundary

As AI redistributes responsibilities across the system lifecycle, governance must specify the obligations, decision authority, control capacity, and answerability of developers, manufacturers, health care institutions, and those who deploy, supervise, maintain, or use the system [16,17]. Under conservation of accountability, shared or overlapping responsibilities must remain identifiable, with obligations aligned to each actor's role, authority, and capacity to manage risk. In embodied care, accountability may be difficult to enforce when the circumstances and causal pathways of a physical event cannot be reconstructed. Responsibility and blame may then fall disproportionately on the clinician closest to the event, such as a bedside nurse or supervising surgeon, even when contributing risks originated in upstream design, authorization, configuration, deployment, updating, or maintenance decisions [33,34].

From Moral Crumple Zones to Forensic Sufficiency

A moral crumple zone arises when an operator bears disproportionate blame for an automated-system failure despite limited capacity to prevent it [33]. We define forensic sufficiency as evidence adequate to reconstruct the sequence and causal pathways of consequential actions, decisions, and control transitions for safety learning and fair accountability. Aviation recording and investigation provide a useful but incomplete analogy [35].

Synchronized, provenance-preserving records should capture relevant perception inputs and sensor status, system versions and configuration, and calibration, maintenance, and updates. They should also document task goals, operating modes, and action authority; planned, commanded, and measured actuation; safety constraints and uncertainty cues; and human authorizations and interventions. Linked to clinical consequences, these records should establish what information was available and whether timely intervention was feasible [36,37]. Records should be risk-proportionate, purpose-limited, data-minimized, and tamper-evident, with secure retention, privacy protections, and access for authorized independent investigation [16,37].

Beyond Explainability

Forensic sufficiency is not equivalent to model explainability. Model explanations, such as saliency maps or textual rationales, may help interpret an output [16], but cannot by themselves reconstruct an embodied system's operational history or establish whether timely intervention was feasible. Reconstruction requires the operational records defined above, together with evidence of institutional authorization and recovery arrangements [27,28,30,36,37]. This evidentiary standard supports causal investigation, safety learning, and fair accountability without itself determining legal liability or moral blame.

Locus of Control Migration and the Safe-Stop Principle

Redistributing Human Control Across the Loop

Meaningful human control does not require authorization of every microaction. It requires informed, authorized people to define permissible actions, ensure that limits are monitored and enforced, and retain effective means to interrupt or terminate unsafe behavior [19]. Control may be exercised in, on, or before the loop, depending on action consequences, the intervention window, and recoverability.

In-the-loop control requires direct human command or action-specific approval with adequate information and time. On-the-loop supervision requires a trained supervisor, informative monitoring, intervention authority, and a tested interface enabling effective intervention within the validated recovery window. Before-the-loop control requires designated personnel to authorize patient eligibility, action bounds, safeguards, and deployment conditions, with enforceable runtime limits and responsibility for reassessment. A human role in a policy does not by itself establish meaningful control.

For hazards relying on human intervention, the conservatively estimated time from hazardous deviation to unacceptable harm must exceed the validated total time required to detect the deviation, recognize the need for intervention, respond, and achieve stopping or stabilization, plus a justified safety margin under representative workload conditions. Faster hazards require validated automated protection and

authorized action limits. If neither protection nor a clinically acceptable fallback can contain the hazard, then human supervision cannot justify deployment [19,20]. Institutions remain responsible for staffing, training, maintenance, and rescue capacity [16,17].

The Safe-Stop Principle

We define the safe-stop principle as restricting action and initiating a validated transition to a task-specific safe state when uncertainty, deviation, or contextual change may compromise competence, authorized operating limits, or recoverability. The transition may involve cessation, control transfer, or graceful degradation [26-28,32]. A patient refusal or withdrawal should prompt a clinically appropriate, validated transition to a safe state, subject to applicable emergency and incapacity arrangements [19,23,24].

A safe stop, including one prompted by refusal or withdrawal, does not necessarily require immediate cessation. Abruptly stopping a patient transfer, interrupting a critical infusion, or withdrawing gait support may itself cause harm. Task-specific procedures should validate appropriate combinations of controlled deceleration, stabilization, maintenance of essential functions, control transfer, rescue, and communication. Testing should cover realistic failures, timely activation, and safe resumption or termination [20,26-28,32]. Protective stopping, abstention, transfer, and degradation should be distinguished from unnecessary stops, failed transitions, and the clinical costs of interruption.

The PACER-AI Framework

From Governance Principles to Implementation

PACER-AI (perception reliability, action scope and actuation envelope, clinical context and control locus, evidence through action-grounded validation, and recoverability, responsibility, and forensic sufficiency) integrates these requirements into 5 implementation domains (Table 2). Each domain should specify the evidence required to support action-specific authorization and identify accountable roles with clearly defined responsibilities. Table 2 maps each domain to the evidence and accountable roles required for action-specific authorization. PACER-AI is intended to support implementation within existing medical-device, robotics, and AI-governance processes, rather than create a separate regulatory category.

Table 2. PACER-AI[a] implementation framework for embodied clinical AI.

Domain	Core question	Required evidence and controls
P: Perception reliability	Can the system reliably sense the patient, environment, and task state and recognize when perception is unreliable?	• Sensor validity and calibration • Multimodal fusion and uncertainty estimation • Testing under occlusion, noise, and sensor dropout • Performance across relevant patient groups and environments • Provenance of clinical-record and sensed inputs • Abstention or escalation under unreliable perception
A: Action scope and actuation envelope	What actions may the system perform, where, for whom, and within what physical limits?	• Task-specific intended actions • Action authority and autonomy level; force, speed, energy, proximity, and contact limits • Prohibited actions; spatial, temporal, and patient-eligibility boundaries • Scope-change controls • Independently enforced runtime envelope
C: Clinical context and control locus	Is the control arrangement appropriate for the clinical workflow, patient vulnerability, and validated intervention time?	• Workflow and human-factors testing • Role clarity, staffing, and training; infection-control and emergency pathways • Consent, refusal, and withdrawal; in-the-loop, on-the-loop, and before-the-loop responsibilities • Validated intervention times and safety margins • Effective override, takeover, handback, safe-stop, and rescue mechanisms
E: Evidence through action-grounded validation	Has the complete PDA[b] loop demonstrated clinical value, safe performance, and recovery under realistic conditions?	• Bench, simulation, and staged clinical testing; prospective and external validation • Near misses, unplanned takeover or rescue, and residual harm • Patient-centered outcomes • Workload, equity, and cost-effectiveness where appropriate • Drift monitoring; revalidation after material changes
R: Recoverability, responsibility, and forensic sufficiency	Are credible failures acceptably recoverable, and can consequential events be reconstructed for safety learning and fair accountability?	• Action-specific criteria for detectability, interruptibility, reversibility or clinical compensability, and consequence boundedness • Validated safe-state transitions, control transfer, and graceful degradation • Downtime and rescue plans • Synchronized, provenance-preserving, tamper-evident records • Maintenance, update, and investigation responsibilities • Disclosure, privacy, retention, and access controls

[a]PACER-AI: perception reliability, action scope and actuation envelope, clinical context and control locus, evidence through action-grounded validation, and recoverability, responsibility, and forensic sufficiency.
[b]PDA: perception-decision-action.

Putting the Framework to Work

Consider an AI-enabled repositioning system for immobile patients. This decision flow is illustrative, not a validated clinical protocol.

Step 1 defines patient eligibility, intended benefit, care alternatives, proposed action authority, and physical limits (force, motion, shear, and support). It distinguishes execution of a clinician-selected maneuver from delegated system selection of a clinical action or clinically meaningful target.

Step 2 establishes competence within the proposed envelope and identifies credible failures, including excessive shear, abnormal resistance, loss of support, falls, and action-selection errors where selection authority is proposed. Inadequate competence precludes authorization at the proposed scope; redesign or restriction requires reassessment.

Step 3 evaluates each credible failure scenario against the 4 recoverability dimensions. Protective responses must operate within the effective intervention window while maintaining essential support. Every safety-critical criterion must pass; success on another dimension cannot offset failure. Unmet criteria require revision of the action envelope, safeguards, or control and rescue arrangements, followed by reassessment. If no acceptable action envelope remains, decline authorization.

Step 4 confirms or reduces the proposed authority according to the evidence and the allocation of action selection and execution control. If execution safeguards are validated but action-selection evidence is insufficient, authorize only level 3, provided all its requirements are met, and require clinician selection. Level 4 additionally requires validated action-selection safety. Functions meeting level 5 criteria remain subject to the restrictions and exceptional authorization conditions in Table 1.

Step 5 documents control arrangements, rescue resources, forensic recording, and reassessment triggers based on the approved recovery arrangements.

Developers and manufacturers should provide recovery evidence and enforceable safeguards. Clinicians should assess patient eligibility and exercise assigned control authority through validated intervention mechanisms without becoming the default moral crumple zones [33,34]. Institutions should authorize local action envelopes, ensure staffing and rescue capacity, and support investigations. Regulators should consider action-specific evidence within existing device and AI lifecycle processes [26-30].

Clinical Translation Across Care Domains

Surgery and Invasive Procedures

Many surgical and invasive actions are clinically valuable but have limited recoverability: tissue injury, bleeding, and thermal damage may be difficult to reverse, and delayed rescue may worsen outcomes. Near-term entrustment should therefore favor perceptual support and bounded action assistance, such as scene understanding, instrument tracking, camera control, and enforceable safety-zone constraints, over broader autonomous intervention [4,32]. Capabilities that materially expand action authority or alter risk require action-grounded validation, feasible human-control arrangements, and risk-proportionate forensic records supporting causal investigation [26,27,30,32].

Rehabilitation and Patient Handling

Rehabilitation and patient-handling systems offer opportunities for repetitive, individualized assistance, but action recoverability varies substantially. A resistance adjustment may be readily reversed, yet its consequences depend on patient vulnerability; a balance perturbation or transfer may cause a fall before a therapist can intervene [12,28]. Validation should extend beyond movement accuracy and repetition counts to pain, fatigue, functional independence, participation, adherence, adverse events, and sustained benefit [38]. Patient-specific entrustment is essential because weakness, spasticity, osteoporosis, cognitive impairment, and fluctuating attention can alter how rapidly harm develops, its severity, and the feasibility of recovery.

Nursing Support and Bedside Care

Bedside systems may assist with repositioning, mobility support, monitoring, supply delivery, and environmental adjustment [12,39,40]. Bedside nursing nevertheless depends on relational care and context-sensitive clinical judgment amid frequent interruptions. Systems should seek a nursing assessment when predefined limits are reached or contextual judgment is required. Patient discomfort or refusal should prompt assessment and a clinically appropriate safe transition, rather than the continued pursuit of task completion. Evaluation should quantify the nursing work added by supervision, coordination, maintenance, and troubleshooting alongside any time saved [21,39,40].

Hospital Logistics, Telepresence, and Home Care

These applications may appear lower risk when they do not directly deliver treatment, yet their failures can disrupt care, compromise privacy, or delay rescue. Logistics robots may block emergency routes, delay time-critical specimens

or supplies, or compromise infection-control pathways. Telepresence systems may enable intrusive, persistent, or unauthorized observation and, where remote actuation is available, unsafe remote control. Home systems may fail in unstructured environments with limited supervision and delayed access to rescue [41-43]. These settings warrant conservative action envelopes, reliable remote escalation pathways, accessible local stop mechanisms, caregiver-inclusive design, and validation across varied layouts, connectivity conditions, user capabilities, caregiver availability, and resource levels.

Human Factors, Equity, and Privacy

Designing Human-Machine Coordination

Embodied AI reorganizes human and machine activity across clinical spaces, workflows, and institutional roles. It can alter movement patterns and spatial routines, shape patient and clinician behavior, and redistribute attention, authority, coordination, and work [21,39,40]. We use automation choreography to describe this sociotechnical coordination. Effective choreography may reduce burden and improve timing; poor choreography may create hidden work, delays, role ambiguity, miscalibrated trust, and unsafe workarounds, or assign supervisory responsibility to staff without the capacity to exercise it.

Human-factors evaluation should assess workload, interruptions, situation awareness, communication, role clarity, trust calibration, training, control transfer, and emergency response [19,21,32]. Training for personnel who supervise action-coupled systems should cover the scope and limits of authorized actions, credible failure modes, system uncertainty and operating-mode transitions, override and safe-stop procedures, control transfer, and review of action records. Personnel should also understand when effective control depends on upstream decisions about system configuration, deployment rules, or institutional policy.

Equity Across Bodies, Settings, and Workflows

Environmental robustness is both a technical and an equity requirement. Systems validated in spacious, well-lit, and well-resourced settings may perform less reliably or be less feasible in crowded wards, older facilities, rural clinics, or homes with unstable connectivity and limited rescue capacity [16,21,43]. Perception and physical-control performance may also vary with body size, skin tone when optical sensing is used, mobility patterns, contractures, tremor, atypical gait, assistive-device use, and, where relevant, speech, language, or cognitive status [44,45].

Equity evaluation should therefore extend beyond subgroup differences in model accuracy to examine eligibility, compatibility with diverse bodies and environments, distribution of benefits and operational burdens, and timely access to human assistance, rescue, complaint processes, and incident investigation. Similar measured performance across demographic groups does not establish equity if infrastructure, staffing, supervision, or recovery requirements systematically exclude particular populations or care settings [21,46].

Privacy and Proportionate Observability

Continuous or pervasive sensing can create privacy risks for patients, staff, family members, caregivers, and bystanders. Cameras, microphones, depth maps, force signals, room layouts, physiological streams, location traces, and action records may reveal intimate behavior, disability, vulnerability, workplace practices, and social relationships [42,47]. Governance should specify what data are collected or derived, what processing occurs locally, and what information is transmitted or retained. It should also define who may access or reuse the data and how notice, applicable consent requirements, and meaningful opportunities for refusal are managed, particularly when nonpatients may be recorded or meaningful consent is difficult to obtain.

Forensic sufficiency does not justify indiscriminate surveillance or unrestricted retention. Recording should be proportionate to entrusted action authority and potential consequences, limited to evidence necessary for safety monitoring and investigation, and protected through controls for confidentiality, access, integrity, and provenance. Explicit retention and deletion rules are required. We use proportionate observability to describe evidence sufficient to investigate consequential events while minimizing unnecessary surveillance. Depending on the risk and investigative need, this may be supported by local processing, data segmentation, role-based access, secure rolling buffers, or preservation of relevant preevent and postevent data [37,47].

Limitations and Future Directions

Empirical Validation of the Core Concepts

This viewpoint is conceptual. Recoverability-gated entrustment, forensic sufficiency, and PACER-AI have not been validated as measurement or decision instruments. Research should establish whether recoverability profiles can be measured reliably, predict adverse outcomes, and improve authorization decisions. The contribution of forensic records to causal reconstruction, safety learning, and fair accountability also requires evaluation. Prospective comparative studies should evaluate the framework's effects on safety outcomes and implementation burden across actions, populations, and care settings [21,32].

Operational Taxonomies and Recovery Thresholds

The boundaries of action-coupled AI, adjacent authority levels, and recovery thresholds require operational specification and empirical refinement. Taxonomies should classify systems by entrusted actions rather than hardware appearance, considering patient vulnerability, action authority, consequence severity and persistence, intervention feasibility, and available recovery resources. Action-specific thresholds

should account for the comparative risks of initiating, continuing, modifying, stopping, delaying, or withholding action, including harm caused by interruption.

Implementation Across Diverse Care Settings

Applicability across health systems remains untested. Participatory evaluation should include resource-constrained settings from the outset and assess whether local workflows, infrastructure, staffing, training, maintenance, connectivity, rescue, and investigation capacity can sustain the proposed safeguards [16,43,46]. Implementation studies should evaluate integration into existing governance processes, effects on human control, patient refusal and withdrawal opportunities, privacy, equity, and costs. Comparative research should distinguish the framework's contribution to authorization and oversight from the clinical value of particular systems relative to standard care or simpler alternatives. Whether recovery and recording requirements can be met in resource-constrained settings without transferring poorly recoverable risks remains unresolved.

Conclusions

Embodied AI extends clinical AI from recommending care to participating in physical action, where errors can cause immediate and persistent harm. Physical action authority should therefore depend on demonstrated competence and the recoverability of credible failures, supported by evidence sufficient for safety learning and fair accountability. We propose governing and validating the complete PDA loop, constraining authority through recoverability-gated entrustment, and assigning meaningful human control where intervention can be effective. The entrustment ladder and PACER-AI integrate these requirements with existing safeguards to support proportionate, auditable, equitable, and accountable physical action. This conceptual approach requires empirical validation; authority to act on a patient must be justified rather than assumed.

Generative AI Use

Following initial drafting, ChatGPT 5.5 (OpenAI) was used to assist with language editing. All AI-assisted revisions were subsequently reviewed, modified where necessary, and approved by the authors. The authors assume full responsibility for the final manuscript.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Authors' Contributions

JL and SL jointly conceptualized and designed the viewpoint, conducted the literature analysis, and prepared the initial draft. All authors critically reviewed and revised the work for important intellectual content and approved the final version.

Conflicts of Interest

None declared.

References

1. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health. May 2024;6(5):e367-e373. [doi: 10.1016/S2589-7500(24)00047-5] [Medline: 38670745]
2. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res. Jun 28, 2023;25:e48568. [doi: 10.2196/48568] [Medline: 37379067]
3. Hinostroza Fuentes VG, Karim HA, Tan MJT, AlDahoul N. AI with agency: a vision for adaptive, efficient, and ethical healthcare. Front Digit Health. 2025;7:1600216. [doi: 10.3389/fdgth.2025.1600216] [Medline: 40400543]
4. Lee A, Baker TS, Bederson JB, Rapoport BI. Levels of autonomy in FDA-cleared surgical robots: a systematic review. NPJ Digit Med. Apr 26, 2024;7(1):103. [doi: 10.1038/s41746-024-01102-y] [Medline: 38671232]
5. Slade P, Kochenderfer MJ, Delp SL, Collins SH. Personalizing exoskeleton assistance while walking in the real world. Nature. Oct 2022;610(7931):277-282. [doi: 10.1038/s41586-022-05191-1] [Medline: 36224415]
6. Erickson Z, Clever HM, Gangaram V, et al. Characterizing multidimensional capacitive servoing for physical human–robot interaction. IEEE Trans Robot. 2023;39(1):357-372. [doi: 10.1109/TRO.2022.3190217]
7. Bartolozzi C, Indiveri G, Donati E. Embodied neuromorphic intelligence. Nat Commun. Feb 23, 2022;13(1):1024. [doi: 10.1038/s41467-022-28487-2] [Medline: 35197450]
8. Zitkovich B, Yu T, Xu S, et al. RT-2: vision-language-action models transfer web knowledge to robotic control. In: Tan J, Toussaint M, Darvish K, editors. The 7th Conference on Robot Learning (CoRL 2023). Proceedings of Machine

Learning Research (PMLR); 2023:2165-2183. URL: https://proceedings.mlr.press/v229/zitkovich23a/zitkovich23a.pdf [Accessed 2026-11-13]
9. O'Neill A, Rehman A, Maddukuri A, et al. Open X-Embodiment: robotic learning datasets and RT-x models. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). Institute of Electrical and Electronics Engineers (IEEE); 2024:6892-6903. [doi: 10.1109/ICRA57147.2024.10611477]
10. Kim MJ, Pertsch K, Karamcheti S, et al. OpenVLA: an open-source vision-language-action model. In: Agrawal P, Kroemer O, Burgard W, editors. Proceedings of the 8th Conference on Robot Learning. Proceedings of Machine Learning Research (PMLR); 2025:2679-2713. URL: https://proceedings.mlr.press/v270/kim25c.html [Accessed 2026-11-13]
11. Xiao X, Liu J, Wang Z, et al. Robot learning in the era of foundation models: a survey. Neurocomputing. Jul 2025;638:129963. [doi: 10.1016/j.neucom.2025.129963]
12. Mir BA, Nishwa DE, Lee SW. Embodied artificial intelligence in healthcare: a systematic review of robotic perception, decision-making, and clinical impact. Healthcare (Basel). Feb 25, 2026;14(5):572. [doi: 10.3390/healthcare14050572] [Medline: 41827523]
13. Liu Y, Cao X, Chen T, et al. From screens to scenes: a survey of embodied AI in healthcare. Inf Fusion. Jul 2025;119:103033. [doi: 10.1016/j.inffus.2025.103033]
14. Shu W, Zhou Y, Sun E, Deng L, Ye X. Embodied intelligence applications in health care populations: scoping review. J Med Internet Res. Jun 12, 2026;28:e83871. [doi: 10.2196/83871] [Medline: 42284471]
15. Labkoff S, Oladimeji B, Kannry J, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. J Am Med Inform Assoc. Nov 1, 2024;31(11):2730-2739. [doi: 10.1093/jamia/ocae209] [Medline: 39325508]
16. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. Feb 5, 2025;388:e081554. [doi: 10.1136/bmj-2024-081554] [Medline: 39909534]
17. Koski E, Das A, Hsueh PYS, et al. Towards responsible artificial intelligence in healthcare-getting real about real-world data and evidence. J Am Med Inform Assoc. Nov 1, 2025;32(11):1746-1755. [doi: 10.1093/jamia/ocaf133] [Medline: 40999782]
18. Cho SM, Zou X, Fleig L, Unberath M. Mini-review on human-centered assurance in robot-assisted orthopedics and neurosurgery. Front Robot AI. 2026;13:1755883. [doi: 10.3389/frobt.2026.1755883] [Medline: 41809634]
19. Cavalcante Siebert L, Lupetti ML, Aizenberg E, et al. Meaningful human control: actionable properties for AI system development. AI Ethics. Feb 2023;3(1):241-255. [doi: 10.1007/s43681-022-00167-3]
20. Küçüktabak EB, Pons JL, Lynch KM, Zarrin RS. A safety-aware framework for physical human-robot-human interaction with hierarchical constraints. IEEE Access. 2024;12:124931-124942. [doi: 10.1109/ACCESS.2024.3442093]
21. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cogn Eng Decis Mak. Dec 2022;16(4):194-206. [doi: 10.1177/15553434221097357] [Medline: 36704421]
22. SMBPB S, Valori M, Legnani G, Fassi I. Assessing Safety in Physical Human–Robot Interaction in Industrial Settings: A Systematic Review of Contact Modelling and Impact Measuring Methods. Robotics. ;14(3):27. [doi: 10.3390/robotics14030027]
23. Felber NA, Pageau F, McLean A, Wangmo T. The concept of social dignity as a yardstick to delimit ethical use of robotic assistance in the care of older persons. Med Health Care Philos. Mar 2022;25(1):99-110. [doi: 10.1007/s11019-021-10054-z] [Medline: 34822097]
24. Palmier C, Rigaud AS, Ogawa T, et al. Identification of ethical issues and practice recommendations regarding the use of robotic coaching solutions for older adults: narrative review. J Med Internet Res. Jun 18, 2024;26:e48126. [doi: 10.2196/48126] [Medline: 38888953]
25. Kim YT, Kim H, Bahl M, et al. Defining operational safety in clinical artificial intelligence systems. NPJ Digit Med. Feb 20, 2026;9(1):281. [doi: 10.1038/s41746-026-02450-7] [Medline: 41720960]
26. ISO 14971:2019 medical devices—application of risk management to medical devices. International Organization for Standardization (ISO); 2019. URL: https://www.iso.org/standard/72704.html [Accessed 2026-06-15]
27. IEC 80601-2-77:2019+AMD1:2023 CSV medical electrical equipment—part 2-77: particular requirements for the basic safety and essential performance of robotically assisted surgical equipment. International Electrotechnical Commission (IEC); 2023. URL: https://preview.iec.ch/pub/pdf/preview/info_iec80601-2-77%7Bed1.1%7Db.pdf [Accessed 2026-05-16]
28. IEC 80601-2-78:2019+AMD1:2024 CSV medical electrical equipment—part 2-78: particular requirements for basic safety and essential performance of medical robots for rehabilitation, assessment, compensation or alleviation. International Electrotechnical Commission (IEC); 2024. URL: https://preview.iec.ch/pub/pdf/preview/info_iec80601-2-78%7Bed1.1%7Db.pdf [Accessed 2026-06-20]

29. Software as a Medical Device Working Group. Software as a medical device (SaMD): clinical evaluation. International Medical Device Regulators Forum (IMDRF); 2017. URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf [Accessed 2026-06-15]
30. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions: guidance for industry and food and drug administration staff. US Food and Drug Administration (FDA); 2025. URL: https://www.fda.gov/media/166704/download [Accessed 2026-06-25]
31. Lin J, Sun T, Song R, et al. Simulation-based testing of autonomous robotic systems for surgical applications. Actuators. 2025;14(9):439. [doi: 10.3390/act14090439]
32. Marcus HJ, Ramirez PT, Khan DZ, et al. The IDEAL framework for surgical robotics: development, comparative evaluation and long-term monitoring. Nat Med. Jan 2024;30(1):61-75. [doi: 10.1038/s41591-023-02732-7] [Medline: 38242979]
33. Elish MC. Moral crumple zones: cautionary tales in human-robot interaction. Engaging Sci Technol Soc. 2019;5:40-60. [doi: 10.17351/ests2019.260]
34. Toro-Tobon D, Ponte OP, Montori VM, Brito JP. Clinician in the loop: a flawed solution for AI oversight. BMJ. May 5, 2026;393:e089213. [doi: 10.1136/bmj-2025-089213] [Medline: 42086290]
35. Annex 13 to the convention on International Civil Aviation: aircraft accident and incident investigation. International Civil Aviation Organization (ICAO); 2024. URL: https://www.bazl.admin.ch/dam/en/sd-web/F31sDf6UEpy9/icao_annex_13_aircraftaccidentandincidentinvestigation.pdf [Accessed 2026-06-15]
36. Møller KE, Sørensen JL, Topperzer MK, et al. Implementation of an innovative technology called the OR Black Box: a feasibility study. Surg Innov. Feb 2023;30(1):64-72. [doi: 10.1177/15533506221106258] [Medline: 36112770]
37. Medical Device Cybersecurity Working Group. Principles and practices for medical device cybersecurity. International Medical Device Regulators Forum (IMDRF); 2020. URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-200318-pp-mdc-n60.pdf [Accessed 2026-06-15]
38. Kiyono K, Tanabe S, Hirano S, et al. Effectiveness of robotic devices for medical rehabilitation: an umbrella review. J Clin Med. Nov 4, 2024;13(21):6616. [doi: 10.3390/jcm13216616] [Medline: 39518755]
39. Georgadarellis GL, Benjamin E, Roberts SC, Vital CJ, Sup Iv FC. Evaluating nursing work systems and identifying barriers for robotic technology integration: observational study. J Med Internet Res. Jun 1, 2026;28:e89409. [doi: 10.2196/89409] [Medline: 42224675]
40. Lim YW, Tan SW, Tan CYB, et al. An assessment of an inpatient robotic nurse assistant: a mixed-method study. J Med Syst. Oct 22, 2024;48(1):99. [doi: 10.1007/s10916-024-02117-4] [Medline: 39436455]
41. Zachariae A, Plahl F, Tang Y, Mamaev I, Hein B, Wurll C. Human-robot interactions in autonomous hospital transports. Rob Auton Syst. Sep 2024;179:104755. [doi: 10.1016/j.robot.2024.104755]
42. Niemelä M, van Aerschot L, Tammela A, Aaltonen I, Lammi H. Towards ethical guidelines of using telepresence robots in residential care. Int J Soc Robot. Jun 2021;13(3):431-439. [doi: 10.1007/s12369-019-00529-8]
43. Abou Allaban A, Wang M, Padır T. A systematic review of robotics research in support of in-home care for older adults. Information. 2020;11(2):75. [doi: 10.3390/info11020075]
44. Gao S, Chen J, Xia Y, et al. Wearable technologies for assisted mobility in the real world. Nat Commun. 2025;16(1):10988. [doi: 10.1038/s41467-025-67126-4]
45. Setchfield K, Gorman A, Simpson AHRW, Somekh MG, Wright AJ. Effect of skin color on optical properties and the implications for medical optical technologies: a review. J Biomed Opt. Jan 2024;29(1):010901. [doi: 10.1117/1.JBO.29.1.010901] [Medline: 38269083]
46. Stanley EAM, Tsang RY, Gillett H, et al. Connecting algorithmic fairness and fair outcomes in a sociotechnical simulation case study of AI-assisted healthcare. Nat Commun. Dec 19, 2025;17(1):788. [doi: 10.1038/s41467-025-67470-5] [Medline: 41419484]
47. Grabler R, Koeszegi ST. Privacy beyond data: assessment and mitigation of privacy risks in robotic technology for elderly care. J Hum Robot Interact. Mar 31, 2025;14(1):1-23. [doi: 10.1145/3689216]

Abbreviations

PACER-AI: perception reliability, action scope and actuation envelope, clinical context and control locus, evidence through action-grounded validation, and recoverability, responsibility, and forensic sufficiency
PDA: perception-decision-action
SA-ROC: Safety-Aware Receiver Operating Characteristic

Edited by Ivan Steenstra; peer-reviewed by Goodness Nzeigwe, Robert Marshall; submitted 17.Jul.2026; initial comments to author 08.Oct.2026; final revised version received 16.Oct.2026; accepted 30.Oct.2026; published 26.Nov.2026

Please cite as:
Liu J, Liu S
Governing Embodied AI in Clinical Care: From Digital Recommendation to Recoverability-Gated Physical Action
J Med Internet Res 2026;28:e104088
URL: https://www.jmir.org/2026/1/e104088
doi: 10.2196/104088

© Jialin Liu, Siru Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.