

Article

Healthcare AI Governance as a Closed-Loop: A Simulation Based Analysis of Human-Centered Experience Engineering

Minseong Kim and Joongho Chang *

Program in AI Convergence Engineering, aSSIST University, Seoul 03767, Republic of Korea;
mskim924@stud.assist.ac.kr

* Correspondence: jhchang@assist.ac.kr

Abstract

As artificial intelligence becomes increasingly embedded in healthcare operations, governance can no longer be treated solely as a static set of principles or regulatory requirements. This study proposes a Human-Centered Experience Engineering (HCEE)-based healthcare AI governance architecture designed as a closed-loop operational control system for stabilizing and adaptively managing AI-enabled healthcare systems. Rather than treating patient and employee experience as downstream outcomes, the framework repositions them as governance input signals, operationalized as the Patient Experience Index (PXI) and Employee Experience Index (EXI). The architecture integrates Policy, Governance, Control, and Experience through recurrent feedback, trigger-control rules, and adaptive learning mechanisms that translate experiential signals into ongoing operational adjustment. To examine how this architecture affects system behavior, agent-based simulation compares four scenarios: S1 (no governance), S2 (sense-only), S3 (full HCEE closed loop), and S4 (full HCEE under exogenous stress). Results show a consistent pattern of $S1 < S2 < S3$ in both PXI and EXI, indicating that sensing alone yields limited improvement, whereas feedback-coupled sensing, control, and learning produce stronger stabilization and performance gains. In S4, the same architecture demonstrates recovery, re-stabilization, and adaptive reinforcement under shock. These findings suggest that effective healthcare AI governance depends on whether feedback loops are architected to function operationally. These results are based on synthetic experience indices and are interpreted as conceptual, mechanism-level evidence.

Keywords: healthcare AI governance; Human-Centered Experience Engineering (HCEE); closed-loop governance; operational control system; patient experience index (PXI); employee experience index (EXI); adaptive governance; agent-based simulation

Academic Editor: Polinapilinho
Katina

Received: 1 May 2026

Revised: 3 June 2026

Accepted: 11 June 2026

Published: 3 July 2026

Copyright: © 2026 by the author.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As the adoption of artificial intelligence (AI) in healthcare rapidly expands, AI is no longer merely an auxiliary tool; rather, it is being rapidly integrated across multiple operational domains within healthcare organizations, including clinical support, operational management, decision support, and service coordination [1]. Accordingly, the question of the standards and procedures under which healthcare AI can be stably operated has become increasingly important. Beyond simply allowing or restricting the adoption and use of AI, there is now a need for a healthcare AI governance structure capable

of continuously coordinating and recalibrating AI operations in real-world deployment settings [2].

To date, discussions on healthcare AI governance have made substantial progress in articulating sophisticated normative principles such as accountability, transparency, fairness, safety, and regulatory compliance [3–5]. This accumulated body of work is meaningful in that it has provided a foundational reference point for what healthcare organizations should consider when adopting and deploying AI. However, beyond the articulation of principles themselves, there remains insufficient explanatory depth regarding how such principles are actually implemented within the operational context of healthcare practice through processes of sensing, judgment, and adjustment [6]. As a result, compared with normative coherence, operational specificity has remained relatively weak, and it is precisely at this juncture that the gap between principles and practice becomes apparent [7,8].

This gap becomes even more evident when the structural characteristics of healthcare settings are considered. In care environments where variability in patient conditions, the exercise of professional judgment, interdependence across organizational units, and repeated human–system interactions are intricately layered, it may prove difficult to adequately capture the accumulated tensions and frictions in AI operation, subtle imbalances, and signs indicating the need for intervention if healthcare AI governance is understood only as a list of predefined principles or a set of ex post inspection procedures. Healthcare AI governance, therefore, needs to be reconfigured from static norms into a dynamic structure capable of detecting changes in the course of operation and adjusting accordingly [9,10].

This study seeks to reposition the status of human experience as one of the starting points for such reconfiguration. In the existing literature, patient experience and employee experience have generally been treated as ex post outcome indicators used to evaluate satisfaction, acceptability, convenience, and service quality after AI adoption [11]. Such an approach is certainly useful for examining implementation effects. However, considering that in real-world healthcare settings the signs of problems may emerge earlier, not as declines in performance metrics but as anxiety and distrust perceived by patients and as workflow friction and a reduced sense of control experienced by staff, experience needs to be understood not simply as an evaluative endpoint, but as an early signal revealing failures of alignment within the system and the need for adjustment [12,13]. In other words, the core task of healthcare AI governance lies less in adding new norms than in clarifying how experience-based information generated during operations can be sensed, interpreted, and converted into input for iterative adjustment.

Against this background, this paper proposes Human-Centered Experience Engineering (HCEE) as a design framework for healthcare AI governance. HCEE is not intended merely to emphasize conventional user-centered or human-centered technology design. Rather, it is positioned as an operational framework for linking human-centered experiential variation to the operational logic of governance. Within this framework, patient experience and employee experience are operationalized as the Patient Experience Index (PXI) and the Employee Experience Index (EXI), respectively, and these two indicators are intended to function not simply as reportable performance values but as input signals that diagnose the current state of the system and help determine the need for intervention and the intensity of adjustment. In particular, it is also important to address patient experience and employee experience together as outcome values derived from an interdependent relationship so as to avoid errors that may arise from managing only one dimension of experience in isolation. This is because as the burden on healthcare staff and coordination fatigue intensify, patient experience may also become unstable;

conversely, as friction and distrust accumulate on the patient side, the response burden on staff may likewise increase [14].

On this basis, this paper conceptualizes healthcare AI governance as a closed-loop structure in which the four layers of Policy, Governance, Control, and Experience are mutually interconnected. Experience is the layer that reveals the state of the field, Control is the layer that adjusts operational levers, and Governance is the layer that interprets sensed signals and determines the priority and level of intervention. Policy, at a higher level, provides the value criteria and direction. In this structure, the key issue does not lie only in how precisely experience is measured. More fundamentally, the central question is whether measured experiential information can actually be incorporated into pathways of judgment and control so as to alter the system's subsequent state. From this structural perspective, healthcare AI governance can be interpreted not as a static set of rules, but as an adaptive system in which sensing, judgment, adjustment, and learning circulate continuously.

This line of inquiry extends directly into the empirical design. The focus is not on making a simple comparison between the presence and absence of governance. Rather, it is on distinguishing different system states according to the extent to which experience-based feedback is institutionalized within the system. Accordingly, four scenarios (S1–S4) are constructed. S1 is set as a baseline state in which the experience-based adjustment structure does not substantially operate; S2 is a state in which the sensing and measurement of experience are introduced only partially; and S3 is a state in which the circuits of sensing, judgment, control, and learning are more tightly connected. By additionally introducing S4, in which exogenous stress is imposed, the study also examines the extent to which a fully closed-loop structure may exhibit recovery and readjustment under disruptive conditions. Therefore, this comparison is not a simple parallel arrangement of conditions, but rather forms a structural axis of comparison that examines how deeply experience-based information is coupled with the governance system's capacity for control and adjustment.

Accordingly, this paper first examines how differences in the level of activation of an experience-based governance structure produce changes in the experience-centered outcomes and temporal trajectories of healthcare AI systems, and second, analyzes the extent to which HCEE-based closed-loop governance can demonstrate the possibility of recovery and re-stabilization after temporary decline under exogenous stress conditions. Ultimately, the study seeks to determine whether healthcare AI governance can be advanced beyond the level of normative declaration or ex post evaluation and reframed as a problem of system architecture that receives experience signals and activates adjustment rules within the course of actual operation. That said, this discussion is proposed within the scope of design and simulation, and it does not immediately warrant direct generalization to healthcare organizations as a whole.

To this end, Section 2 reviews the theoretical foundations of healthcare AI governance from the perspectives of human experience, feedback and adaptation, and leverage points. Section 3 presents the research model and scenario logic for structurally comparing experience-based governance. Section 4 elaborates this into a four-layer architecture and a trigger–control mechanism and operationalizes it in an experimentally observable form. Section 5 compares terminal outcomes and temporal trajectories across scenarios, with particular attention to patterns of resilience and readjustment under exogenous stress conditions. Finally, Section 6 concludes the paper by summarizing the theoretical and practical implications of the findings and discussing the limitations of the study.

2. Theoretical Background

2.1. From Human-Centered AI to Experience-Based Governance in Healthcare

The discourse on human-centered AI has provided an important turning point for research and practice in healthcare AI. This body of work has encouraged scholars and practitioners to view AI not merely as a computational tool, but as a system that must be considered in relation to usability, explainability, human oversight, and compatibility with clinical context [15–17]. This shift is especially important in healthcare. Healthcare AI does not operate as an isolated technical object; rather, it is deeply embedded in care processes, professional judgment, interactions with patients, and the operational order of healthcare organizations [7,10,18]. Accordingly, the value of healthcare AI cannot be sufficiently explained by accuracy or efficiency alone. It also needs to be examined in terms of how acceptably and sustainably it functions within real-world clinical and operational environments [16,17,19,20].

However, recent discussions suggest that human-centeredness frequently remains at the level of design principles or the language of acceptability. While such discussion is clearly necessary, it remains insufficient to explain how healthcare organizations should manage and adjust AI in everyday operations [7,10,21,22]. In practice, problems often emerge not initially through technical performance limitations, but through conflicts with workflow, ambiguity in role responsibilities, imbalances of trust, and the additional burdens imposed on frontline personnel [23–26]. In other words, the problem of healthcare AI often arises less from the technology itself than from the way it interfaces with the people, procedures, expectations, and routines of healthcare organizations. In this respect, human-centeredness needs to move beyond a declarative principle and be extended into a question of operational order [7,10,18,19].

At this point, patient experience and employee experience again become significant. In the existing literature, they have largely been treated as ex post evaluation items such as satisfaction, acceptability, convenience, and service quality [27–30]. This approach is useful for assessing implementation effects, but it remains somewhat limited in explaining the complex operational realities of healthcare organizations. In practice, on the patient side, changes such as anxiety, confusion, insufficient explanation, and declining trust may appear first [28,31–33], while on the employee side, work friction, interpretive burden, coordination fatigue, and a weakened sense of control may emerge as earlier signals [29,30,34–37]. These changes are not merely subjective reactions; they may instead function as early signals indicating how smoothly AI is operating within the organization [27,29,35,38].

Accordingly, this study does not treat experience simply as an implementation outcome, but as an important clue for understanding healthcare AI governance. This does not mean that existing criteria such as safety, accountability, fairness, and regulatory compliance should be replaced [16,21,39,40]. Rather, it suggests that if one seeks to detect, at an earlier point, when such criteria begin to destabilize in practice, it is necessary to attend carefully to the operational conditions as experienced by patients and staff [16,39–41]. Patient-side experience and employee-side experience are not separate dimensions, but are deeply intertwined, and the more tension accumulates on one side, the more instability on the other side may also intensify. In this sense, experience becomes an important observation point for understanding how well healthcare AI remains aligned within the organization [18,27,29,30,38].

This line of reasoning does not reject the discourse on human-centered AI; rather, it seeks to advance the discussion to the next stage. Instead of merely explaining why human-centered AI matters, it asks how healthcare organizations can position human experience within the grounds of governance judgment and the logic of adjustment. This

question naturally leads to the next discussion: namely, that healthcare AI governance should be understood not as a static system of oversight, but as a socio-technical order that reads and responds to changing operational conditions [16,19,20,42].

2.2. Feedback, Adaptation, and Closed-Loop Governance in Socio-Technical Healthcare Systems

Healthcare AI operates in environments characterized by high uncertainty and interdependence. Clinical workflows are not static, patient conditions and workloads continually change, and the degree of trust in and reliance on technology may also fluctuate over time. Under such conditions, it is insufficient to understand governance merely as one-time approval, ex post inspection, or compliance verification [16,19,20,40]. Healthcare AI governance needs to be understood as a process that observes what kinds of changes emerge in actual operation after implementation, interprets whether those changes remain within acceptable bounds, and enables adjustment when necessary. Recent discussions of healthcare AI governance have also moved beyond the presentation of principles toward issues of monitoring, oversight, post-deployment management, and organizational response [16,20,22,40].

A socio-technical systems perspective makes this point even clearer. Healthcare organizations do not function through technology alone; rather, people, rules, workflows, responsibility structures, and communication practices are interwoven in their operation [7,10,18,25]. Accordingly, AI governance should not be conceived as a separate device that monitors technology from outside, but as a process of coordinating the operational order within the organization. What matters is not simply whether rules exist, but whether the organization can interpret certain signs as meaningful under changing conditions, review them at an appropriate level, and respond appropriately [7,10,23,24]. In this context, feedback does not merely mean the return of data. It is closer to a recurring flow of signals through which the organization can determine whether the relationship between technology and the organization remains operationally viable [29,34–36].

The issue of adaptation also becomes important here. Because the operational environment of healthcare organizations is not static, it is difficult to address all situations with a single standard or procedure [16,17,19,20]. Workload may increase, the skill and trust levels of frontline staff may vary, and patient expectations may also change. Under such conditions, the role of governance is not to eliminate all variation. Rather, it is to distinguish temporary disturbance from structural imbalance and to enable organizational adjustment before small frictions accumulate into larger problems [35,36,39,42]. For this reason, healthcare AI governance is better understood not as a fixed oversight system, but as an adaptive order that manages the balance between sensitivity and stability [40].

This perspective also makes it possible to interpret healthcare AI governance in more rigorous systems terms. In other words, the success or failure of governance depends less on the mere existence of norms than on how continuously the organization can read and respond to changing conditions [16,19,20]. If intervention occurs only after a problem becomes visible as an official failure, governance may exist formally but still respond substantively too late. By contrast, if recurring signals from the field can be connected to interpretation, review, and subsequent adjustment, governance can function not as a static approval device but as a living operational order [29,33,35,36]. In this sense, healthcare AI governance should be understood not as a matter of surveillance, but as a matter of organizational responsiveness and adjustment capacity [7,23,24].

This study adopts this perspective interpreting healthcare AI governance from the standpoint of a socio-technical feedback order. Of course, not every change immediately justifies intervention. What matters is having a framework of judgment capable of distinguishing temporary fluctuation from persistent deterioration, and localized friction

from systemic imbalance [35,39,42]. Ultimately, the core of healthcare AI governance lies not in declaring more principles but in the structural means by which the organization can respond to changing operational realities. This discussion provides the basis for positioning HCEE in the next section not as a simple human-centered discourse, but as a more design-oriented framework [16,20,42].

2.3. Positioning This Study: HCEE as an Architecture-Level Design Framework

Recent healthcare AI literature has addressed issues such as accountability, oversight, implementation barriers, and institutional readiness with considerable breadth. This accumulated body of work is clearly valuable. However, while existing discussions are relatively clear about what healthcare AI governance ought to aim for, they are often less specific about how such aims can actually be aligned and sustained within organizational operation [10,21–24]. In particular, issues such as operational-level adjustment, interconnection within organizations, and the structure of recurring review remain insufficiently systematized. As AI moves beyond experimental introduction into routine use, this gap becomes even more apparent [16,19,20,40].

It is at this juncture that this study positions HCEE. Here, HCEE is not a rhetorical concept intended merely to reiterate the importance of human-centeredness. Nor is it presented merely as a framework for experience evaluation. Rather, this study proposes HCEE as a framework for thinking about healthcare AI governance from a more design-oriented perspective [15,42]. The key point is to treat the experiences of patients and employees not as peripheral reference materials, but as important clues through which healthcare organizations can understand the operational state of AI. In doing so, the human-centered discussion can be extended beyond the principles of technology design to the question of how organizations should manage and adjust AI [17,27–30,38].

This positioning makes the scope of the present study clearer. The study does not reduce healthcare AI governance to a list of ethical principles, nor does it narrow it to a general discussion of user experience. Instead, from a socio-technical systems perspective, it asks how organizations can read and maintain the relationship between human experience and operational reality [15,21,39,42]. In this context, the meaning of HCEE lies not in adding a specific technical function, but in enabling a governance conception through which healthcare organizations can respond to changing conditions. That is, it connects the question of human-centeredness to the question of actual operational order [16,19,20,40].

This perspective also explains why the present study requires the language of design and comparison. If healthcare AI governance is understood not merely at the level of normative desirability but as a structurally designable order, then it becomes necessary to examine what differences emerge from distinct governance conditions. Accordingly, the next chapter translates the preceding literature review into a more explicit research design, and the subsequent chapters examine step by step how that design leads to a particular governance configuration and interpretation of results. In this respect, Chapter 2 serves not as an independent literature overview, but as a theoretical bridge that enables the analyses and explanations that follow [19,20,42].

In sum, HCEE is presented in this study as a conceptual framework that reconnects healthcare AI governance to the issues of human experience, organizational responsiveness, and socio-technical adjustment. It does not reject the discourse on human-centered AI; rather, it attempts to extend that discourse one step further to a level that engages with the actual operation of healthcare organizations. In this precise sense, HCEE functions as the central axis that ties together the subsequent research design, governance configuration, and interpretation of results [15,27,42].

It is therefore useful to state explicitly how HCEE differs from adjacent frameworks. Responsible-AI and AI-ethics frameworks (e.g., WHO guidance, FUTURE-AI) primarily specify normative principles—safety, accountability, fairness, transparency—and the conditions under which AI should be approved or restricted, but they remain largely silent on the run-time control logic by which such principles are maintained during operation. Socio-technical systems perspectives correctly emphasize the interdependence of people, rules, and technology, yet they are typically descriptive rather than operationally specified. HCEE does not replace these contributions; it operationalizes them. Specifically, HCEE (i) repositions patient and employee experience from ex-post outcome indicators to real-time governance input signals (PXI, EXI), (ii) couples these signals to explicit trigger-control rules and adaptive learning within a four-layer closed loop, and (iii) renders the resulting governance behavior observable and testable through simulation. In short, where prior frameworks answer what governance should aim for, HCEE specifies how those aims can be sensed, interpreted, and adjusted as an operating architecture. Table 1 summarizes how HCEE differs from these adjacent governance frameworks.

Table 1. Differentiation of HCEE from adjacent governance frameworks.

Dimension	Responsible-AI/Ethics Frameworks	Socio-Technical Systems View	HCEE (This Study)
Primary object	Normative principles and approval criteria	Interdependence of people, rules, technology	Operational control logic linking experience to adjustment
Status of experience	Ex-post acceptability/satisfaction outcome	Contextual factor	Real-time governance input signal (PXI, EXI)
Mechanism	Guidelines, audits, compliance checks	Conceptual/analytical description	Trigger-control rules + adaptive learning in a closed loop
Temporal mode	Mostly static/periodic review	Variable	Continuous sense–interpret–adjust–learn cycle
Testability	Principle articulation	Framework-level	Simulation-observable behavior and ordering

3. Methods

3.1. Scenario Design and Governance Logic

The purpose of this chapter is to translate the closed-loop governance perspective of HCEE elaborated in the previous chapter into a verifiable research design. To this end, the study first establishes the axes of scenario comparison and the governance logic, then operationalizes the core variables and trigger-control rules, and subsequently specifies the execution procedure and reproducibility conditions. Finally, by clarifying what this design is intended to validate and how the post-processing procedure is organized, this chapter examines the experimental framework within which the architecture described in the next chapter is evaluated.

This study treats healthcare AI governance not as a mere list of norms or a collection of ex post inspections, but rather as a control system capable of receiving experiential signals and adjusting itself accordingly [43]. From this perspective, the core issue in simulation design is not the parallel arrangement of different institutional environments. Rather, by contrasting the degree to which experience-based governance is integrated within the system, the design seeks to examine how different levels of integration among

sensing, adjustment, learning, and shock response reshape the operational trajectory [44]. This repositioning is illustrated in Figure 1.

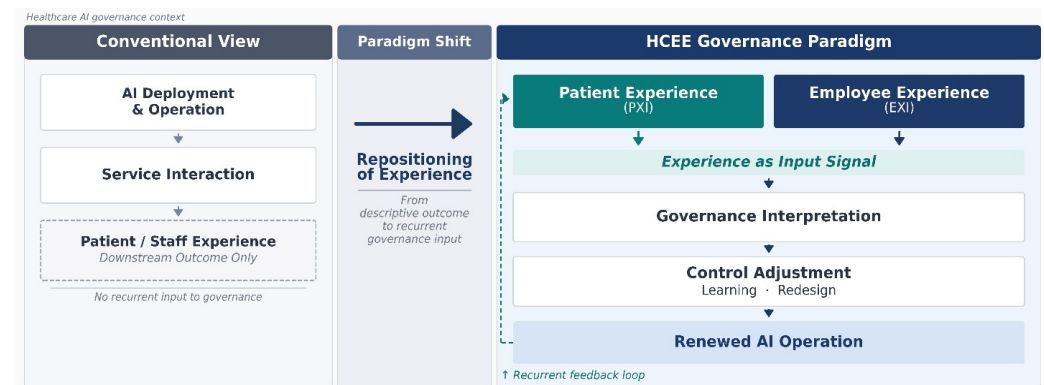


Figure 1. Repositioning Experience from Downstream Outcome to Governance Input in the HCEE Paradigm.

HCEE seeks to move patient and employee experience beyond the category of ex post satisfaction and to reposition them as input signals that governance can read and respond to [45]. In this framework, PXI and EXI are treated not as secondary indicators for reporting results, but as core state variables that reveal operational imbalances and interaction tensions at an early stage. This repositioning constitutes the design logic of shifting experience from a downstream outcome to a governance input, and the scenario comparison that follows is intended precisely to examine what system-level differences this shift actually produces.

To render this logic experimentally testable, four scenarios were constructed [46]. The first scenario, S1, is the baseline condition in which recurrent governance input is absent. Under this condition, even though PXI and EXI are generated, they are neither systematically captured nor linked to subsequent adjustment. AI is simply allowed to operate without intervention. The second scenario, S2, is a partial HCEE condition in which only the sensing and visualization of experience signals are activated. In this design, experiential information is observed, but it is not sufficiently connected to control and learning, leaving the feedback loop only partially open. The third scenario, S3, is the full HCEE condition in which sensing, control, and learning are integrated [47]. In this scenario, experiential signals are not merely recorded but are used as inputs for operational adjustment and rule updating. S3 therefore serves as the key comparison point for determining whether HCEE constitutes not merely a value-oriented declaration, but a designable and observable closed-loop governance structure.

The fourth scenario, S4, retains the full-governance structure of S3 while adding an exogenous shock as a stress-test condition. This scenario does not represent another maturity stage. Rather, it is positioned as an extension condition intended to examine whether a closed loop can establish recovery and readjustment even after disruption.

The above design can be summarized along two comparative axes. One is the governance–maturity axis running from S1 to S3, and the other is the stress–recovery axis running from S3 to S4 [8,48]. The former is intended to examine how adjustment capacity changes as experience-based signals become more deeply embedded in the system. The latter is intended to examine what adaptive path a fully closed-loop structure can establish when confronted with exogenous disturbance. By separating the analytical role of each scenario in this way, the interpretation of subsequent results can be developed not as a simple ranking of performance, but in terms of structural differences and mechanism effects. The full scenario definitions are summarized in Table 2.

Table 2. Experimental Scenario Definition: Governance Module Activation Level and Expected System Dynamics.

Scenario	Governance Module Activation				What Changes in the System	Expected Dynamic Signature
	Sense	Control	Learn	Stress		
S1 No Governance (Baseline)	OFF	OFF	OFF	OFF	Unregulated AI deployment. PXI/EXI signals are generated but neither captured nor routed to governance. No intervention mechanism operates.	Drift + error accumulation: persistent oscillation in PXI/EXI; no corrective feedback; system diverges from safe operating band over time.
S2 Partial HCEE (Sense-only)	ON	OFF	OFF	OFF	PXI/EXI monitored and reported; no automated control action. Signals are visible but the feedback loop remains open—correction depends on manual response.	Delayed response + residual instability: partial improvement in mean PXI/EXI; oscillation persists due to open-loop architecture; no stabilization of variance achieved.
S3 Full HCEE (Closed-loop)	ON	ON	ON	OFF	Full closed-loop activated: PXI/EXI signals trigger automated control adjustments (AAL/HOI/TR/IF); trigger-control rules are updated through the learning module. System self-regulates without external intervention.	Recurrent feedback-based adjustment: PXI/EXI signals are continuously incorporated into control and learning, shaping a more coordinated operational trajectory.
S4 Stress Test (Full + Shock)	ON	ON	ON	ON	All modules active; exogenous shocks injected (demand surge, misdiagnosis event, staff shortage, regulatory change). System responds under full closed-loop governance.	Post-stress adaptive readjustment: following an initial disruption, PXI/EXI signals are re-incorporated into control and learning, shaping a recovery and re-stabilization trajectory under exogenous stress.

Note. Sense = PXI/EXI monitoring; Control = trigger-based operational adjustment; Learn = adaptive rule updating; Stress = exogenous shock injection for resilience testing. ON/OFF indicates module activation status in the simulation. S1 is the no-governance baseline, S2 is partial HCEE (sense-only), S3 is full HCEE, and S4 is the stress-test condition applied to full HCEE. PXI = Patient Experience Index; EXI = Employee Experience Index. This table reports scenario design logic, not empirical results.

In sum, the scenario design in Section 3.1 establishes what is to be evaluated and along which comparative axes. The next section operationalizes the core variables, including PXI and EXI, as well as the minimum trigger–control rules required to translate experiential signals into adjustment so that this comparison can be implemented in practice.

3.2. Variables, Measures, and Trigger–Control Rules

The purpose of this section is to convert the scenario comparison into an actually executable experimental condition [49]. To this end, the key outcome variables were set as PXI and EXI. PXI was used as an integrated indicator of system-level changes in patient experience, and EXI as an integrated indicator of system-level changes in employee experience. Rather than fragmentary evaluations of individual events, both indicators were constructed as weighted indices reflecting the accumulated state of repeated interactions, thereby capturing not only deviations at a particular point in time but also trends over time and the potential for adjustment [50].

Two clarifications govern the status of these variables. First, PXI and EXI are synthetic index variables produced inside the simulation; they are not drawn from real patient records or staff surveys. Their role is to make experience-coupled governance behavior observable under controlled conditions, not to estimate field-level magnitudes. Second, the trigger thresholds ($PXI < 70$, $EXI < 65$, $|PXI - EXI| > 15$) are design parameters of the model rather than externally validated clinical cut-points. They are positioned just below the shared initial state of the simulation ($PXI = 69.88$, $EXI = 65.10$ at tick 0) so that a trigger fires when an

experience signal drifts below its starting band. The lower EXI threshold reflects a deliberate asymmetry: because staff are exposed to the AI system more frequently and more deeply than patients, a slightly lower employee-side threshold reduces false-positive triggering from routine workload fluctuation while still capturing genuine deterioration. The gap threshold ($|PXI - EXI| > 15$) likewise marks cross-domain decoupling that exceeds normal between-instrument variation. Because all three thresholds are model-internal, the relevant validation question is not whether 70 and 65 are the correct real-world values, but whether the comparative ordering across scenarios is preserved when these and related parameters are varied—an issue addressed directly in the sensitivity and robustness analysis (Section 5.3). For field deployment, all thresholds would require site-specific calibration.

An important feature of this design is that PXI and EXI are not treated as mere performance indicators. They are used not as endpoints for reporting results, but as input signals for detecting system state and determining the need for subsequent intervention. Accordingly, the main unit of analysis does not remain limited to mean levels alone. The absolute level of each indicator, the gap between PXI and EXI, the persistence of decline, stabilization status, and the activation state of each scenario are all recorded. These are used to assess whether experience-based governance produces not merely a higher level, but a change in adjustability itself.

Trigger-control rules are the minimum unit that translates these state signals into executable intervention conditions [51]. In the main text, only three core triggers are summarized. TC-01 is activated when PXI falls below a preset threshold and induces immediate control aimed at alleviating patient-side burden. TC-02 is activated when EXI falls below its threshold and calls for adjustment to buffer excessive workload and gaps in oversight. TC-03 operates when $|PXI - EXI|$ exceeds the allowable range, thereby triggering cross-domain readjustment so that optimization in one domain does not undermine the overall balance.

This study fixed the core trigger conditions at $PXI < 70$, $EXI < 65$, and $|PXI - EXI| > 15$. However, rule activation did not depend on a simple one-point judgment [2]. Because excessive signal sensitivity may misread short-term fluctuations as structural risks, a pre-specified detection window and smoothing rule were jointly applied to the judgment process [4,52]. In addition, by combining OR-type rules that capture individual deterioration with composite conditions that account for persistence or cross-domain imbalance, the study sought to treat experience-based governance not as a logic of immediate reaction, but as a problem of balancing sensitivity and stability.

Baseline setting was likewise not treated as a mechanical fixation of a single value. Initial conditions and operational values were arranged by distinguishing among empirical, stylized, calibrated, and stress-induced settings. This was intended to clarify the provenance and purpose of each parameter configuration and to reduce the risk of overinterpreting patterns that emerge accidentally from a specific parameter value as structural characteristics. In other words, the purpose of this section is not to reproduce the field in full, but to present in a reproducible manner which indicators and under what conditions interventions begin.

Once the variables and rules are organized in this way, the next task is to present the execution environment and output structure transparently so that the same design can be run again by others. The next section therefore specifies the pipeline and reproducibility criteria extending from NetLogo implementation and BehaviorSpace repeated execution to CSV export and Python-based post-processing.

3.3. Simulation Execution and Reproducibility

This section specifies the procedure through which the scenarios and operational rules defined above were actually executed. The simulation was implemented in the

NetLogo 7.0.3 environment, and repeated runs and parameter exploration were conducted through BehaviorSpace. The basic pipeline consisted of agent-based implementation of the HCEE-based governance structure, scenario-specific module activation settings, repeated execution in BehaviorSpace, CSV export, and Python-based post-processing [53,54]. The separation of the execution stage from the interpretation stage was intended to allow the computational process and the result-organization process to be re-examined jointly [55].

The basic time unit of this study was the week, and each run was fixed at 364 ticks. This setting corresponds to a 52-week observation period and was intended to allow initial adaptation, mid-term adjustment, and late-stage stabilization to be observed within a single run. All major experiments shared this same time range, thereby ensuring comparability across scenarios.

To ensure reproducibility, the version, execution unit, experiment naming, and data structure were fixed consistently [56]. First, the execution environment was maintained at NetLogo 7.0.3, and the same model file and initial settings were used throughout. In addition, the validation experiments were divided into four blocks: Robustness_SeedCheck, Sensitivity_Alpha, Sensitivity_Agents, and Sensitivity_InteractProb. Robustness_SeedCheck applied sim-seed values from 1 to 30 and was executed a total of 90 times across S1–S3. Sensitivity_Alpha varied the learning rate α across 0.16, 0.18, 0.20, 0.22, and 0.24 and was executed 15 times. Sensitivity_Agents adjusted the population scale across levels 1–5 and was executed 15 times. Sensitivity_InteractProb varied combinations of patient-prob and worker-prob and was executed 60 times. The total number of validation runs across the four experiments was therefore 180.

The CSV output structure was also kept as consistent as possible. The terminal summary file was configured to include run number, scenario identifier, recording point, system-PXI, system-EXI, and auxiliary variables indicating the activation state of each scenario. The weekly tracking file was configured to accumulate changes in PXI and EXI over time by scenario. This schema was designed to link end-point comparisons and path-level comparisons consistently while serving different analytical purposes. Based on this execution and output structure, the study further examined whether the observed differences among scenarios were dependent on a particular seed or a single parameter configuration. Table 3 summarizes the analytical purpose, varied parameters, experimental levels, and total runs of the BehaviorSpace-based validation experiments conducted for this purpose. Table 4 details the BehaviorSpace validation design and experimental settings for these four blocks.

Table 3. Variables, Measures, and Core Trigger Conditions.

Construct/Variable	Operational Definition	Measurement/Data Source	Trigger Condition or Evaluation Rule	Immediate Control Direction	Analytical Use
PXI (Patient Experience Index)	A composite index representing the system-level state of patient-side experience. Used as a governance input signal reflecting cumulative outcomes of repeated interactions, rather than a single satisfaction score.	Weekly and longitudinal PXI outputs within the simulation. Longitudinal snapshots (364 ticks) and weekly tracking are used in parallel.	TC-01: PXI < 70 (weekly moving average)	AAL decrease, TR activation, IF increase	Cross-scenario comparison of patient-side experience outcomes, time-path analysis, trigger activation assessment
EXI (Employee Experience Index)	A composite index representing the system-level state of employee-side	Weekly and longitudinal EXI outputs within the simulation.	TC-02: EXI < 65 (weekly moving average)	HOI increase, IF decrease, conditional TR	Cross-scenario comparison of employee-side experience

	experience. Used as a governance input signal reflecting operational experience changes including workload, sense of control, and interaction tension.	Longitudinal snapshots and weekly tracking are used in parallel.		adjustment as needed	outcomes, detection of workload-related deterioration
PXI-EXI Gap	A cross-domain indicator representing the degree of imbalance between patient experience and employee experience.	Calculated based on the difference between smoothed PXI and EXI values.	TC-03: $ PXI - EXI > 15$ (weekly moving average)	Cross-rebalancing toward the lower-performing domain; AAL↓ when PXI is inferior, HOI↑ when EXI is inferior, TR/IF adjustment as needed	Cross-domain balance evaluation rather than single-metric optimization
Terminal Snapshot Outcomes	Final state at the end of each simulation run.	Final PXI/final EXI at tick 364 (=Week 52).	N/A	N/A	Cross-scenario comparison of terminal outcomes, calculation of improvement rate vs. baseline
Weekly Trajectory Measures	Dynamic data showing adaptive pathways and adjustment patterns over time.	PXI/EXI time series accumulated in the weekly tracking file.	Used to verify threshold crossing, recovery timing, and re-stabilization.	Confirmation of time-path of control action following trigger firing	Interpretation of path dependency, recovery speed, and re-stabilization patterns
Stabilization Indicator	An operational indicator assessing whether PXI/EXI fluctuation is maintained within an acceptable range.	Assessed based on weekly trend, oscillation amplitude, and variance trend.	Determined by whether the system continuously remains within the target band after returning to it.	N/A (evaluation metric)	Review of stabilization effects for RQ2, S1–S3 comparison
Equity/Vulnerable Subgroup Signal	Signal related to vulnerable subgroups or experienced disparities.	Vulnerable subgroup PXI gap, complaint/anomaly signals.	Used as a supplementary signal for governance review when rapid gap widening or threshold breach occurs.	Connected as input for enhanced sampling or higher-level review	Detection of community-facing inequality beyond simple average outcomes
Operational/Event Variables	Operational change variables such as system errors, response delays, and complaint events.	System event logs, anomaly flags, complaint counts.	Refer to separate rulebook when safety-critical events such as AI error clusters occur.	Emergency AAL step-down, HOI reinforcement, etc., as needed	Supplementary interpretation of stress response and operational instability
Signal Smoothing/Detection Window	Signal processing rules to reduce trigger over-sensitivity.	Exponential smoothing ($\alpha = 0.3$), 7-day rolling average.	Applied to the base detection window for TC-01 through TC-03.	False-positive trigger mitigation	Ensuring signal stabilization and reproducible trigger evaluation
Baseline Setting Types	A classification distinguishing the source and purpose of parameter configurations.	Empirical/stylized/calibrated/stress settings.	N/A	N/A	Explicit documentation of parameter provenance and securing reproducibility

Note. AAL = AI Autonomy Level; HOI = Human Oversight Intensity; TR = Trust Recalibration; IF = Information Flow; PXI = Patient Experience Index; EXI = Employee Experience Index; TC = Trigger-Control rule. All trigger thresholds are simulation-derived; operational deployment requires site-specific calibration.

Table 4. BehaviorSpace Validation Design and Experimental Settings.

Experiment	Analytical Purpose	Varied Parameter(s)	Levels Scenarios	Total Runs
Robustness_SeedCheck	Probabilistic robustness	sim-seed (1–30)	30 × 3 scenarios	90
Sensitivity_Alpha	Learning-rate sensitivity	$\alpha \in \{0.16, 0.18, 0.20, 0.22, 0.24\}$	5 × 3 scenarios	15
Sensitivity_Agents	Population-scale sensitivity	n -scale = 1–5	5 × 3 scenarios	15
Sensitivity_InteractProb	Interaction sensitivity	patient-prob × worker-prob	20 × 3 scenarios	60
Total				180

Note. All experiments share the same NetLogo 7.0.3 model and a fixed observation window of 364 ticks (52 weeks). Robustness_SeedCheck covers S1, S2, and S3; Sensitivity experiments cover S2 and S3 only. α = learning rate; n -scale = agent population multiplier.

Implementation stabilization was also treated as part of reproducibility. During validation, two technical problems were identified in an earlier version. One was that BehaviorSpace parameters were overwritten by clear-all, and the other was that a mismatch between the stopping condition and the time limit caused some terminal reporters to return zero. After correcting these issues, the major validation experiments were rerun using the revised version. This revision history shows that the model is not a completed replica of reality, but a research tool that is progressively stabilized. At the same time, it underscores that implementation transparency is as important as result interpretation.

Convergence assessment was also included as a separate verification item. The study did not accept the validity of execution merely because terminal values were produced. Additional checks examined whether all runs reached 364 ticks, whether terminal reporters returned no abnormal zero values, and whether the major trajectories by scenario maintained a consistent directional pattern across repeated runs. This should be understood not as a formal mathematical proof of convergence, but as a practical criterion for distinguishing whether the results are artifacts of implementation error or arise from differences in the designed governance structure. The confirmed experimental metadata and initial-state evidence are summarized in Table 5.

Table 5. Confirmed Experimental Metadata and Initial-State Evidence.

Item	Confirmed Value	Interpretive Meaning
Main terminal experiment	1000 runs per scenario	Used for terminal snapshot comparisons in Table 5
Weekly tracking experiment	50 runs per scenario × ticks 0–364	Used for trajectory summaries and shock–recovery landmarks in Tables 6 and 7
Simulation horizon	52 weeks (1 year)	Analytical horizon for the annual simulation cycle
Terminal snapshot	Tick 364	Reference point for terminal PXI/EXI values
Common initial PXI	69.88 at tick 0 across S1–S4	Shared baseline state for weekly trajectory comparison
Common initial EXI	65.10 at tick 0 across S1–S4	Shared baseline state for weekly trajectory comparison

Having established the execution conditions in this manner, the next task is to explain what the execution was designed to validate and how the CSV outputs were transformed through the analytical procedure into result interpretations.

3.4. Validation Design and Post-Processing

The purpose of this section is to transform the simulation design from a mere set of execution procedures into an explicit validation structure. Rather than claiming absolute predictive accuracy, this study examines whether the HCEE-based closed-loop governance exhibits consistent mechanism effects under a specified set of conditions [57,58]. To this end, the validation objectives were divided into four categories: P1 for mechanism effect, P2 for leverage explanatory power, P3 for resilience under stress, and P4 for structural robustness.

P1 examines whether an experience-based closed-loop structure actually induces changes in PXI and EXI. Accordingly, the main evidence is drawn from the scenario comparison among S1, S2, and S3. P2 does not aim to read outcome differences as mere mean differences, but to examine whether the core levers—AAL, HOI, TR, and IF—can function as an interpretive framework for system adjustment [59]. In this case, the focus is not on independently confirming the micro-level effect of each rule, but on examining whether rule-based intervention logic provides explanatory power for interpreting the results.

P3 examines, through S4, whether the full HCEE structure can establish a path of recovery and re-stabilization after exogenous shock. What is important here is that S4 is not treated as just another general scenario. S4 is a condition for testing resilience when full governance is already operating, and the interpretation in this section should therefore focus less on the magnitude of loss itself than on the existence of a recovery path and the possibility of readjustment.

P4 examines whether the core findings are maintained across variation in seed, learning rate, agent scale, and interaction probability. This is intended to reduce the risk of overinterpreting patterns that emerge incidentally under a particular setting as structural characteristics.

To support this, the result files were reorganized in a Python-based post-processing pipeline by experiment type and scenario. The post-processing sequence consisted of CSV cleanup, calculation of summary statistics, cross-scenario comparison, reconstruction of weekly trajectories, robustness checking, visualization, and log interpretation. Terminal output and weekly tracking output were stored separately at the saving stage, but were configured to be cross-referenced at the interpretation stage. As a result, a single design made it possible to examine, jointly, end-point differences, path-level differences, post-shock recovery patterns, and stability against parameter change.

This study also accounted for the possibility of pseudoreplication when the number of repeated runs is sufficiently large but the amount of substantive variation information is limited. Accordingly, in the subsequent statistical review process, the number of repetitions itself was not overinterpreted as independent evidence. Instead, attention was paid jointly to directional consistency, maintenance of scenario ordering, and stability under sensitivity change. The weekly tracking data were likewise treated not solely as objects of strict probability estimation, but as path-level evidence showing what dynamic patterns differences in governance structure produce. These validation objectives and their corresponding evidence are summarized in Table 6.

Table 6. Validation Objectives, Corresponding Evidence, and Interpretive Scope.

Validation Purpose	Analytical Focus	Corresponding Evidence or Experiment	What Can Be Inferred
P1 Mechanism Effect	Does experience-driven closed-loop governance	Main scenario comparison across S1 (No-HCEE), S2	Whether differences in governance maturity translate into distinct levels

	actually induce measurable changes in PXI and EXI?	(Partial-HCEE), and S3 (Full-HCEE)	and trajectories of experience indicators
P2 Leverage Explanatory Power	Do core levers—AAL, HOI, TR, and IF—contribute meaningfully to explaining system adjustment outcomes?	Trigger–control structure and rule-based intervention logic operationalized in the Full-HCEE condition (S3)	Whether outcome differences can be interpreted as the result of specific adjustment mechanisms, rather than mere performance comparisons
P3 Resilience Under Stress	Does the system demonstrate the potential for recovery and re-stabilization following an exogenous shock?	S4 Stress-Test scenario and week-by-week trajectory analysis post-disruption	Whether the Full-HCEE architecture can form an adaptive recovery path after system-level disruption
P4 Structural Robustness	Are the core directional findings maintained across variations in seed, learning rate, agent scale, and interaction probability?	Robustness_SeedCheck, Sensitivity_Alpha, Sensitivity_Agents, and Sensitivity_InteractProb experiments	Whether results are robust against specific initial conditions or single-parameter configurations

From this point onward, the study explains how the HCEE-based healthcare AI governance architecture—the object of this design—is implemented through layers, interfaces, and adjustment mechanisms.

4. HCEE-Based Healthcare AI Governance Architecture

4.1. Four-Layer Architecture

The purpose of this chapter is to explain what, in concrete terms, the research design established in Section 3 is intended to implement. Whereas the previous chapter established the scenarios, variables, execution procedures, and validation logic, this chapter presents the structure and operating principles of the HCEE-based healthcare AI governance architecture. Accordingly, the focus here is not to repeat the experimental design but to show how experiential signals are connected to judgment and adjustment through multiple layers.

Before describing the layers, it is useful to make explicit how the AI system itself is represented in the model. The architecture is deliberately AI-agnostic: rather than modeling a specific algorithm, the AI is abstracted as an operational decision agent whose behavior is summarized by its AI Autonomy Level (AAL)—the degree to which it acts without human confirmation. Higher autonomy increases throughput but also the rate at which misalignment can propagate into patient- and staff-facing interactions, which the model registers as changes in PXI and EXI. Governance does not act on the AI's internal computation; it acts on the operating envelope around the AI through the four control levers (AAL, HOI, TR, IF). This abstraction lets the framework apply to heterogeneous healthcare AI—clinical decision support, scheduling, triage, documentation—without depending on any single model's internals, while keeping the object of governance well defined.

The architecture proposed in this study consists of four layers: Policy, Governance, Control, and Experience. The purpose of this distinction is not simply to classify roles within an organization. Rather, it is intended to provide a minimal structural framework for clarifying at which level normative direction, operational judgment, executional adjustment, and field-level experience are formed and how they are interconnected. The overall four-layer structure is shown in Figure 2.

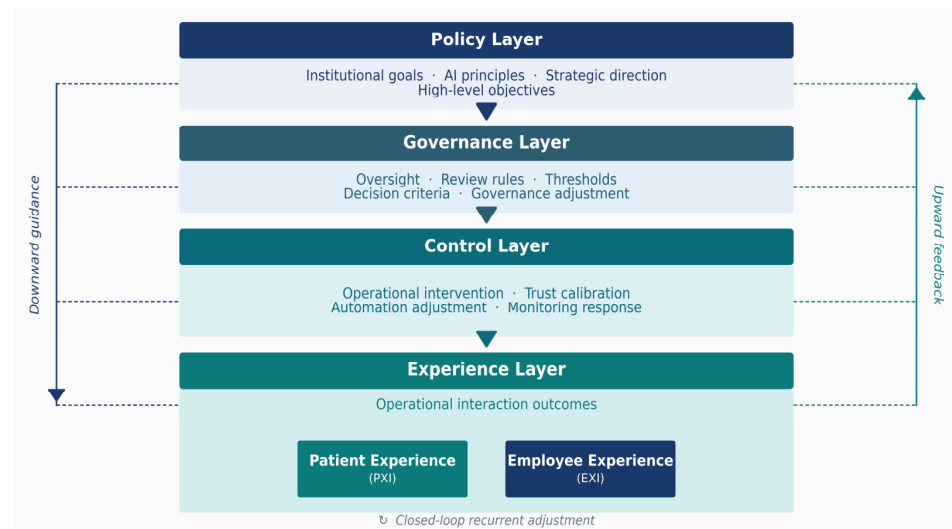


Figure 2. Four-Layer Closed-Loop Governance Structure of the HCEE Architecture.

The Policy Layer presents higher-order criteria such as patient safety, accountability, human oversight, and fairness, and sets the boundaries within which lower-level adjustments may operate [60]. The Governance Layer translates these criteria into operational rules and review procedures [7,10]. The Control Layer is the execution level at which actual interventions are carried out, while the Experience Layer is the field level where patient and staff interactions accumulate and PXI and EXI are formed [20]. In this architecture, the Experience Layer is understood not as the endpoint at which results are collected, but as the sensing point at which upward feedback begins [40].

The explanatory power of the four-layer structure lies less in the number of layers themselves than in the bidirectional connections among them [1]. The upper layers transmit principles, constraints, priorities, and directions for intervention downward, while the lower layers return experiential signals and operational outcomes upward. Accordingly, this structure should be understood not as a one-way oversight system, but as a closed-loop architecture that converts experience into an input within governance itself [61].

4.2. Interlayer Interfaces and Information Flows

If the four-layer structure provides the basic skeleton, the interlayer interfaces specify how that structure actually operates [19]. This study argues that the explanatory power of healthcare AI governance lies less in the mere existence of layers than in the flows of information, rules, responsibility, and adjustment that move across them [62]. Even when the same deterioration in experience occurs, the system's response may differ depending on which signal is transmitted to which level and how that signal is translated into rules. The interlayer interfaces of the four-layer structure are summarized in Table 7.

Table 7. Four-Layer Governance Interface: Inputs, Outputs, Core Responsibility, Upward Feedback, and Downward Guidance.

Layer	Primary Input	Primary Output	Core Responsibility	Upward Feedback	Downward Guidance
Policy Layer	Institutional priorities, strategic goals, escalated governance signals	Governance direction, review criteria, policy priorities	Set normative boundaries and strategic orientation	Receives escalated system-level concerns	Provides high-level principles and policy direction

Governance Layer	Policy direction, PXI/EXI summaries, threshold signals	Oversight rules, escalation decisions, adjustment directives	Interpret system state and determine governance response	Synthesizes experience signals for upper-level review	Translates policy into governance rules and thresholds
Control Layer	Governance directives, trigger conditions, lever settings	Intervention actions, recalibration decisions, implementation status	Execute operational adjustment through controllable levers	Reports implementation effects and residual issues	Alters operational conditions affecting interactions
Experience Layer	AI-enabled service conditions, user interactions, adjusted operating environment	PXI, EXI, experiential signals	Generate experiential outcomes and recurrent signals	Returns lived operational feedback upward	Receives altered service conditions from control decisions

In this study, the interfaces are organized into four types of flow. First, data flow transmits PXI and EXI formed in the Experience Layer to the Governance and Control levels. Second, rule flow connects upper-level principles and lower-level signals in order to define intervention conditions and adjustment scope [16]. Third, responsibility flow clarifies which layer handles which type of issue and under what conditions higher-level review becomes necessary [23,63]. Fourth, adjustment flow ensures that adjustments made by the Control Layer alter the operating conditions of the Experience Layer.

These flows are bound together through the bidirectional circulation of upward feedback and downward guidance. Signals generated in the Experience Layer are transmitted upward, prompting reconsideration of the current operating state and structural risks, while principles and rules transmitted from the upper layers are translated into lower-level operational adjustments [64]. In this process, experience is not merely reported information, but an active signal through which rules and interventions are reconfigured.

A more detailed correspondence between intervention depth and leverage points is best separated into an appendix table. Accordingly, the deeper mapping is presented in Appendix A, Table A1. Leverage Point Mapping Across the Four-Layer HCEE Architecture. The main text focuses on clarifying the existence of the interfaces and their functional significance, while the detailed classification of leverage is provided in the Appendix A for supplementary reference.

4.3. Trigger-Control Mechanism and Core Levers

This section examines the conditions under which the interlayer connections described above are designed to operate. The purpose of the trigger-control mechanism is to ensure that experiential information does remain merely collected or accumulated [65]. In other words, it is important to specify structurally when changes in PXI and EXI should be translated into operational adjustment, and through which levers such adjustment is executed.

Figure 3 illustrates, at the level of operating mechanism, how the HCEE architecture converts experiential signals into governance interventions. In this study, PXI and EXI function not as simple post hoc evaluation results, but as input signals that governance repeatedly reads and interprets [66]. Combined with the PXI–EXI gap, the persistence of warning conditions, and exogenous stress cues, the system becomes capable of sensing

the current operational state in a multidimensional manner [67]. These experiential signals first enter the Trigger Detection stage, where threshold violations, imbalances between patient experience and employee experience, persistence of warning states, and the presence of stress conditions are identified. Importantly, this stage does not remain at the level of simple monitoring. From this point onward, changes in experience may be processed as governance inputs capable of initiating subsequent adjustment.

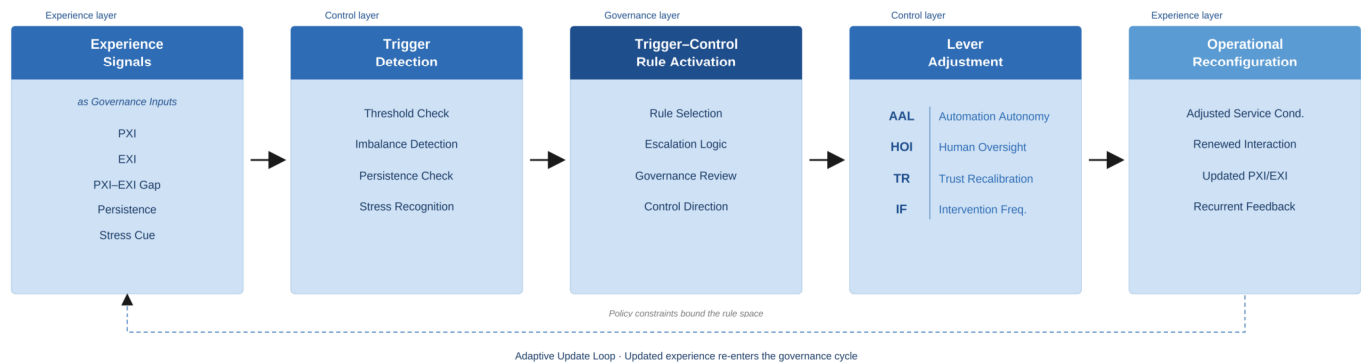


Figure 3. Trigger-Control Logic and Core Levers in the HCEE Architecture.

At the next stage, Trigger-Control Rule Activation, the detected signals are translated into rule selection, escalation judgment, governance review, and the direction of control [15]. In other words, the system does not respond mechanically or immediately to experiential deterioration. Rather, it distinguishes what type of problem is occurring and what level of intervention is required, and then selects an appropriate control pathway.

In this study, the trigger logic can be organized into four categories. Absolute decline refers to the detection of absolute deterioration in PXI or EXI, whereas cross-domain imbalance addresses situations in which the gap between the two indices widens. Persistence provides the criterion for identifying sustained deterioration rather than short-term fluctuation, and stress cue refers to conditions in which exogenous disruption or accumulated instability calls for stronger readjustment [68]. These categories are presented not to repeat the threshold values and detection window introduced in Section 3, but rather at a level that explains why particular signals are interpreted as conditions for intervention. Accordingly, the adjustment stage is executed through four core levers.

Four levers are presented in summary form. AAL (AI Autonomy Level) refers to the degree of autonomy granted to AI within the operating environment and may be lowered when risk, instability, or excessive dependence is detected. HOI (Human Oversight Intensity) refers to the degree to which human review and supervision are reinforced, and it can be increased when experiential or safety-related concerns require closer scrutiny. TR (Trust Recalibration) refers to interventions intended to restore confidence when misunderstanding, anxiety, distrust, or communicative breakdown has accumulated. IF (Information Flow) refers to the intensity and cadence of communication, reporting, or intervention-related information exchange, and may be adjusted in order to reduce coordination gaps and improve interpretive alignment [69].

These levers exist independently, but in practice they operate in combination, and a single signal may invoke multiple directions of intervention. In this sense, the HCEE architecture is better understood not as a simple if-then automation, but as a structure that translates experiential signals into phased and selective adjustments.

The rule-level specification is not fully enumerated in the main text and is instead separated into an appendix. Accordingly, the detailed specification of the full set of conditions and adjustment directions is presented in Appendix A, Table A2. Full Trigger-Control Rulebook for the HCEE Closed-Loop Governance Architecture. By retaining only

the category-level mechanism in the main text, this chapter aims to clarify how triggers and levers bind feedback, leverage, and adaptation into a single operating logic.

Finally, in the Operational Reconfiguration stage, adjusted service conditions and interaction environments are reformed, and the resulting updated PXI and EXI are re-entered as new experiential signals. This iterative return flow is precisely what constitutes the closed-loop character of the HCEE architecture. Accordingly, Figure 3 is not simply a diagram summarizing a list of rules, but a mechanism diagram visualizing an adaptive update loop in which experience-based signals are detected, interpreted, translated into adjustment, and then returned again to experience.

Although the policy level is not presented in Figure 3 as a separate execution box, it operates as a higher-order boundary condition that constrains the rule space. That is, policy does not directly perform control; rather, it functions as an upper-level constraint specifying which trigger-control responses are permitted and which are restricted. In this sense, Figure 3 focuses, within the overall four-layer structure of Policy-Governance-Control-Experience, on the intermediate operating logic through which experiential signals are transformed into actual adjustment mechanisms.

4.4. Scenario-Specific Implementation

The purpose of this section is to organize the implementation state of the HCEE architecture and trigger-control mechanism across scenarios. A critical point is that S1 through S4 are not different systems; they all share the same four-layer HCEE architecture. Differences across scenarios should therefore be understood not as differences in the structure itself, but as differences in which combinations of the sensing, control, learning, and stress modules are activated [8].

As specified in Section 3.1 (Table 2) and summarized in Table 8, S1–S4 are not different systems: they share the same four-layer HCEE architecture and differ only in which modules (sensing, control, learning, stress) are activated. The PXI/EXI levels, stabilization patterns, and shock-recovery patterns reported in the next chapter should therefore be read as outcomes of differing activation states within one architecture, not as comparisons among different models.

Table 8. Scenario-Specific Implementation States of the HCEE Architecture across S1–S4.

Scenario	Sense	Control	Learn	Stress	Active Governance State	Implementation Note
S1 No Governance	OFF	OFF	OFF	OFF	Open-loop baseline	Experience signals are generated but not routed to governance
S2 Partial HCEE	ON	OFF	OFF	OFF	Sensing without closed-loop adjustment	Experience is monitored, but control and learning remain inactive
S3 Full HCEE	ON	ON	ON	OFF	Full closed-loop governance	Sensing, adjustment, and learning are integrated
S4 Stress Test	ON	ON	ON	ON	Full closed-loop under exogenous stress	Full governance retained while stress module is activated

Note. Sense = PXI/EXI monitoring module; Control = automated trigger-control rules (AAL, HOI, TR, IF); Learn = adaptive rule-update mechanism; Stress = exogenous shock injection. ON/OFF reflects module activation within the ABM simulation. All four scenarios share the same four-layer HCEE architecture; differences reflect activation states, not distinct systems. Table 2 (Section 3) presents the comparative logic across scenarios; this table specifies the implementation state of each scenario within the architecture.

5. Results

This chapter examines the performance trajectories generated by the HCEE-based healthcare AI governance architecture operationalized in Section 4 in relation to the research questions established in Section 3. The analysis focuses on four issues. First, it

examines how differences in governance maturity from S1 to S3 shape the terminal outcomes of PXI and EXI. Second, it considers whether these differences appear not only as gaps in final values but also as divergences in temporal trajectories. Third, through S4, it assesses whether the full HCEE architecture demonstrates recovery and re-stabilization under exogenous stress. Fourth, through robustness and sensitivity checks, it evaluates whether the observed patterns are incidental products of a single parameter setting [70].

5.1. Comparative Results Across Governance–Maturity Scenarios

The purpose of this section is to examine whether system-level performance improves more clearly as experience-based governance becomes more deeply institutionalized [10]. The interpretation is centered on the governance–maturity axis from S1 to S3, while S4 is treated separately in the next section as a resilience scenario.

The terminal results show a clear stepwise ordering according to governance maturity. In S1 (No Governance), PXI reached 58.22 and EXI reached 51.90, representing the lowest terminal state. In S2 (Partial HCEE/Sense-only), PXI rose to 65.52 and EXI to 61.08. In S3 (Full HCEE), PXI further increased to 78.29 and EXI to 74.50. These results indicate that the differences across scenarios are not merely score gaps; rather, they are aligned with the degree to which experience signals are captured, interpreted, and connected to corrective action. These terminal outcomes are summarized in Table 9.

Table 9. Terminal PXI/EXI Outcomes across Governance Scenarios (Main Experiment, 1000 Runs per Scenario).

Scenario	Governance Condition	Terminal PXI	Terminal EXI	Change vs. S1	Interpretation
S1	No Governance	58.22	51.90	Reference baseline	Lowest terminal state under open-loop operation
S2	Partial HCEE	65.52	61.08	PXI +7.30; EXI +9.18	Sensing alone improves outcomes, but correction remains limited
S3	Full HCEE	78.29	74.50	PXI +20.07 (+34.5%); EXI +22.60 (+43.6%)	Closed-loop governance yields the strongest non-stress terminal performance
S4	Stress Test	88.02	85.36	PXI +29.80 (+51.2%); EXI +33.46 (+64.5%)	Under stress, the architecture recovers and reaches the highest terminal state

Note. Terminal values are taken from the main BehaviorSpace experiment with 1000 runs per scenario and a terminal snapshot at tick 364, corresponding to a 52-week simulation horizon. S1–S3 constitute the governance–maturity comparison axis, whereas S4 is treated as a stress-response extension scenario rather than a simple next maturity stage.

In particular, the shift from S1 to S3 reflects a structural difference between leaving experience as an ex-post satisfaction indicator and converting it into a governance input signal. Accordingly, the observed performance differences are more appropriately interpreted not as a simple distinction between the presence and absence of governance but as a difference in how deeply experience signals are institutionalized within the system [7].

The magnitude of improvement also supports this interpretation. At the level of the core finding, S3 improved PXI by 34.5% and EXI by 43.6% relative to S1. This indicates that although partial sensing alone can produce a certain degree of improvement, a much larger gain emerges when sensing, control, and learning are combined within a closed-loop architecture [20]. In particular, the larger increase in EXI suggests that HCEE-based governance intervenes not only at the patient interface but also in staff working conditions and the organizational environment for coordination [68]. The corresponding weekly PXI/EXI trajectories across S1–S4 are shown in Figure 4.

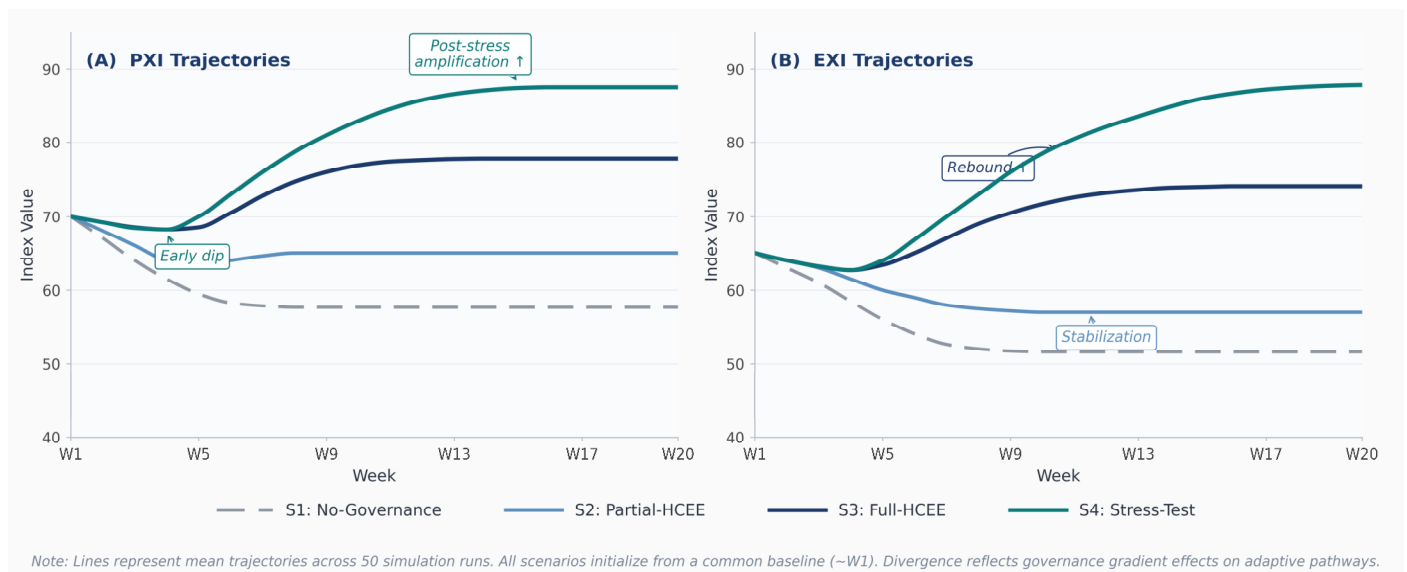


Figure 4. Weekly PXI/EXI Trajectories across S1–S4.

The weekly tracking results indicate that these differences are not confined to the terminal endpoint. Although all four scenarios began from the same initial state, they quickly diverged into markedly different trajectories. S1 became locked into a low-level state after decline. S2 partially buffered the decline and then stabilized at an intermediate level. By contrast, S3 rebounded quickly after a limited initial adjustment and converged at a higher level. In other words, the advantage of full HCEE lies not only in a higher terminal value but also in producing a smaller initial loss, faster recovery, and a higher stabilization level.

Taken together, the comparison across S1–S3 suggests that an architecture that repositions experience as a governance input can improve both system performance and dynamic stabilization. With respect to Research Questions 1 and 2, the results support the interpretation that sensing alone provides only partial improvement, whereas the combination of sensing and control generates a stronger governance effect.

5.2. Shock Response, Recovery, and Re-Stabilization in S4

This section interprets S4 not as the next stage in governance maturity, but as a separate extension scenario designed to examine the resilience of the full HCEE architecture [61]. The central question is whether, under exogenous shock, the same closed-loop structure can absorb losses, recalibrate intervention intensity, and re-establish a stabilization pathway. This stress, recovery, and re-amplification dynamic in S4 is shown in Figure 5.

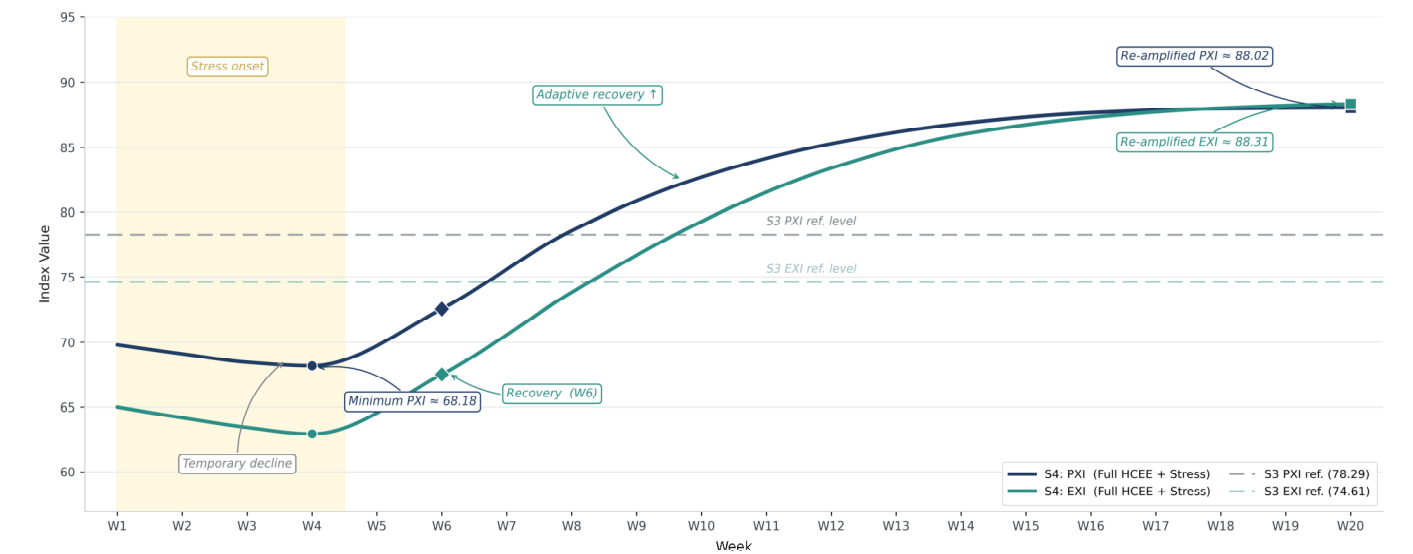


Figure 5. Stress, Recovery, and Re-amplification in S4 under the Full HCEE Architecture.

Within this same architecture (Section 3.1), S4 ultimately reaches a PXI of 88.02 and an EXI of 85.36, the highest terminal state among the four scenarios. The critical point is not that stress itself raises performance; rather, within the present rule structure, the architecture performs adaptive reconfiguration following the shock. The weekly trajectory summary across all scenarios is reported in Table 10.

Table 10. Weekly PXI/EXI Trajectory Summary across S1–S4 (Weekly Tracking Dataset).

Scenario	Start PXI	Minimum PXI (tick)	Final PXI	Start EXI	Minimum EXI (tick)	Final EXI	Trajectory Interpretation
S1	69.88	58.11 (126)	58.22	65.10	51.62 (93)	51.90	Sustained decline followed by convergence to a low operating level
S2	69.88	65.45 (68)	65.52	65.10	60.85 (344)	61.08	Partial buffering of decline, then stabilization at an intermediate level
S3	69.88	68.25 (1)	78.29	65.10	62.65 (1)	74.50	Brief initial adjustment followed by rebound and upward stabilization
S4	69.88	68.18 (1)	88.02	65.10	62.35 (1)	85.36	Initial disruption followed by recovery, reinforcement, and the strongest terminal trajectory

Note. Values are recalculated from the uploaded WeeklyTracking_AllScenarios raw CSV. The weekly tracking dataset contains 50 runs per scenario across ticks 0–364 and is used here to summarize temporal trajectory patterns. All four scenarios begin from the same initial weekly tracking state (PXI = 69.88; EXI = 65.10), so subsequent differences are interpreted as scenario effects rather than differences in starting conditions. This weekly tracking dataset is distinct from the 1000-run terminal dataset used in Table 8.

The weekly trajectory summary clarifies this interpretation. S4 begins from the same initial conditions as the other scenarios and experiences a limited decline at an early stage, with the PXI minimum at 68.18 and the EXI minimum at 62.35. Thereafter, however, the trajectory develops in a direction beyond simple recovery. The recovery-week is recorded as 6, and the system enters the recovery phase within a relatively short time. Subsequently, S4 overtakes S3 and establishes an upward post-shock trajectory.

A closer look at the landmarks of S4 reveals a sequential pattern: an initial decline immediately after the shock; the onset of reinforcement around Week 8; the cross-over above S3 during Weeks 8.3–8.4; and a first post-shock plateau around Weeks 11–14. Further reinforcement phases follow, and adaptive amplification continues into the latter half of the annual trajectory. Accordingly, S4 is more accurately understood not as a scenario that merely endures the shock, but as one in which shock response, recovery, re-stabilization, and adaptive amplification unfold in sequence.

A natural concern is whether S4's terminal advantage over S3 reflects a numerical overshoot or a sensitivity artifact rather than genuine adaptive reinforcement. Three features argue against the overshoot reading. First, the post-shock trajectory is not a single monotonic spike: it shows a bounded early dip (PXI 68.18, EXI 62.35), a first post-shock plateau around weeks 11–14, and discrete, staged reinforcement steps (ticks 57, 113, 169, 225) rather than continuous divergence. Second, no oscillatory ringing or overshoot-and-collapse is observed; the trajectory settles into successive stable bands rather than overshooting a target and correcting. Third, the same qualitative pattern—recovery followed by reinforcement above the non-stress full-governance level—persists across the seed, learning-rate, agent-scale, and interaction-probability variations reported in Section 5.3, which is inconsistent with an artifact tied to one particular parameterization. Mechanistically, the shock activates the full set of control levers simultaneously (a compound response), which would otherwise remain only partially engaged under non-stress conditions; the terminal advantage of S4 therefore reflects the activation of governance capacity that is latent in S3 rather than a numerical artifact. We accordingly interpret S4 as staged adaptive reinforcement under the present rule structure, while explicitly declining the stronger and unsupported claim that exogenous stress is, in general, beneficial.

These results distinguish resilience from resistance. Resistance focuses on minimizing the magnitude of decline, whereas S4 transitions to a higher terminal state after a temporary degradation as adjustment rules are re-activated. This suggests that when experience-based signals, trigger and control rules, and staged reinforcement mechanisms are combined, closed-loop governance can function as an adaptive re-stabilization structure that goes beyond a simple reactive system. At the same time, this study does not extend the result into a general proposition that “stress is beneficial”; the interpretation is confined to the claim that, within the present simulation rules and conditions, the architecture preserved its recovery capacity.

5.3. Robustness and Sensitivity Validation

This section examines whether the main results of this study are contingent artifacts dependent on particular initial seeds or a narrow parameterization. To this end, additional validation was conducted by varying the seed, the learning rate, the population scale, and the interaction probability. The analytic focus was placed not on the identity of individual values, but on whether the ordering stability across scenarios was preserved. In other words, the purpose of this section is not to reproduce identical point estimates across all runs, but to examine whether the direction of the relative effects of HCEE-based closed-loop governance is repeatedly preserved. Such validation serves to defend the claim that the stepwise comparative results in Section 5.1 and the stress-response interpretation in Section 5.2 are not specific to a single run. The validation of scenario ordering under parameter variations is shown in Figure 6.

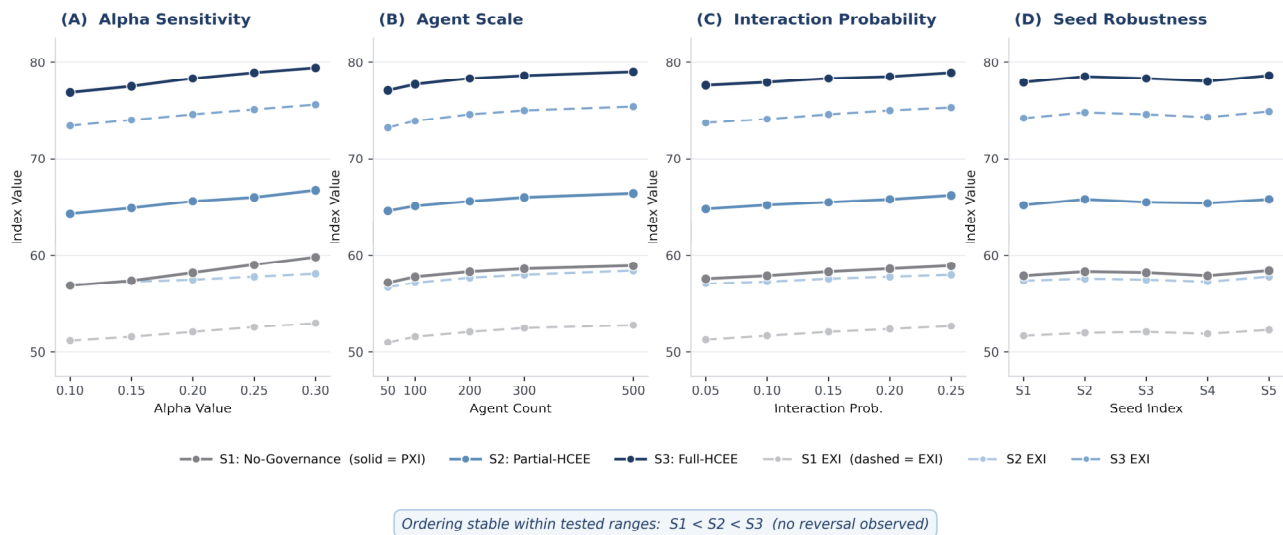


Figure 6. Validation of Scenario Ordering under Parameter Variations.

The robustness and sensitivity results broadly supported the stability of the main pattern. Across the tested ranges, the relative ordering among scenarios was largely preserved, especially the core pattern of $S1 < S2 < S3$. This suggests that the principal findings are not attributable to a specific random seed or a narrowly tuned parameter setting, but are more plausibly associated with the structural differences in governance configuration [69].

More specifically, the seed-robustness check indicated that even when the initial randomization changed, the overall direction of the scenario differences was maintained. Although absolute values fluctuated somewhat across runs, the baseline scenario remained the lowest, the partial HCEE condition occupied an intermediate position, and the full HCEE condition consistently produced the most favorable terminal pattern under non-stress settings [15]. This implies that the observed differences are less likely to be incidental artifacts of one particular initialization.

A similar tendency appeared in the sensitivity tests for learning rate, agent population, and interaction probability. Parameter changes affected the slope and magnitude of trajectories to some extent, but they did not fundamentally reverse the comparative ordering across governance conditions [40]. This point is important because it indicates that the explanatory force of the HCEE architecture does not depend on a single optimized setting. Rather, within the tested ranges, the architecture continued to produce the same directional effect: the more fully experience-based governance was connected to sensing, control, and learning, the more favorable the resulting PXI/EXI trajectory became.

It is also worth explaining why these parameters move the trajectories in the directions observed. The learning rate (α) controls how quickly trigger-control rules are updated from accumulated experience signals: higher α speeds adjustment and steepens the early rebound but does not alter which scenario stabilizes higher, because the closed loop in S3/S4 still receives feedback that the open loop in S1 lacks. Agent population scale affects the averaging and variance of PXI/EXI: larger populations smooth idiosyncratic fluctuation and tighten the bands, while smaller populations widen them, yet the relative spacing of $S1 < S2 < S3$ is retained. Interaction probability sets the density of patient–staff interactions and therefore the rate at which experience signals are generated and fed back; higher density accelerates both deterioration in S1 and correction in S3, again preserving order. In each case the parameter changes the speed or amplitude of the dynamics, not the structural advantage conferred by closing the feedback loop—which is precisely why the ordering is robust.

At the same time, this section does not claim unconditional robustness under every conceivable environment. The validation range remained bounded by the simulation design adopted in this study, and the tested parameter space was necessarily selective rather than exhaustive [49]. Therefore, the present results should be interpreted not as proof that any version of HCEE will always dominate under all conditions, but as evidence that the proposed architecture maintains directional stability across meaningful variations in initial conditions and key parameters. Table 11 summarizes the validation experiments and their ordering-stability outcomes across the four tested axes.

Table 11. Shock–Recovery Landmarks in S4 from Weekly Tracking.

Landmark	PXI Evidence	EXI Evidence	Approximate Tick/Week	Interpretation
Initial disruption	69.88 → 68.18	65.10 → 62.35	Tick 1/Week 0.1	Stress produces an immediate but limited early decline
Recovery reinforcement begins	Intensify factor first increases from 1.00 to 1.15	Same pattern	Tick 57/Week 8.1	The architecture shifts from absorption to reinforced correction
Sustained cross-over over S3	S4 PXI remains above S3 from tick 59 onward	S4 EXI remains above S3 from tick 58 onward	Ticks 58–59/Weeks 8.3–8.4	S4 no longer only recovers; it overtakes the non-stress full-governance trajectory
First post-shock plateau	PXI fluctuates around ~80.78–80.98	EXI fluctuates around ~77.73–78.02	Ticks 80–100/Weeks 11.4–14.3	A stable post-shock operating band emerges before later reinforcement steps
Later reinforcement phases	Intensify factor rises again at ticks 113, 169, and 225	Same pattern	Weeks 16.1, 24.1, and 32.1	S4 continues adaptive amplification through staged reinforcement
Terminal state	Final PXI = 88.02	Final EXI = 85.36	Tick 364/Week 52	S4 ends at the highest terminal level among all scenarios

Note. This table is derived directly from the uploaded weekly tracking raw file rather than from terminal snapshots alone. The earlier phrasing “stable by around week 12” should be interpreted more precisely as the emergence of a first post-shock plateau around weeks 11–14, not as the final equilibrium of the entire annual trajectory.

The Tukey HSD results also supported the ordering of $S1 < S2 < S3$ for both PXI and EXI. However, these statistical results should be read as supplementary evidence reinforcing that the scenario gradient is not a chance pattern, rather than as a substitute for the architecture-level contribution [16]. In particular, Table 12 is based on the seed-robustness dataset, whose analytic purpose differs from that of the main 1000-run terminal snapshot table; this distinction should therefore be maintained in interpreting the results. The corresponding ANOVA and Tukey HSD statistics for the seed-robustness dataset are reported in Table 13.

Table 12. Validation Experiments Summary: Ordering Stability across Seeds and Parameter Changes.

Validation File	Tested Conditions	Ordering Stability	Interpretation
Robustness_SeedCheck (v3)	30 seeds	30/30 runs preserved $S1 < S2 < S3$	Scenario gradient remains stable under seed perturbation
Sensitivity_Alpha (v2)	5 alpha levels	5/5 settings preserved $S1 < S2 < S3$	Results remain stable under moderate learning-rate variation

Sensitivity_Agents (v3)	5 scale conditions	5/5 settings preserved S1 < S2 < S3	Scenario ordering remains robust under agent-scale change
Sensitivity_InteractProb (v3)	20 interaction-probability combinations	20/20 settings preserved S1 < S2 < S3	Relationship-density variation does not overturn the main pattern

Note. The purpose of the validation set is not to reproduce identical point estimates under every parameterization, but to test whether the direction and ordering of the governance gradient remain intact across seed and parameter changes.

Table 13. Seed-Robustness ANOVA Summary for PXI and EXI.

Measure/Scenario	Mean $\pm$ SD	95% CI	ANOVA/Tukey
PXI—S1	58.26 $\pm$ 0.08	[58.23, 58.29]	F(2,87) = 477,528.23, p < 0.001; Tukey S1 < S2 < S3
PXI—S2	65.60 $\pm$ 0.09	[65.57, 65.63]	
PXI—S3	78.44 $\pm$ 0.08	[78.41, 78.46]	
EXI—S1	52.09 $\pm$ 0.15	[52.04, 52.14]	F(2,87) = 151,107.39, p < 0.001; Tukey S1 < S2 < S3
EXI—S2	61.35 $\pm$ 0.17	[61.29, 61.41]	
EXI—S3	74.62 $\pm$ 0.16	[74.56, 74.67]	

Note. sim-seed 1–30, n = 30 per scenario; Robustness_SeedCheck dataset, NetLogo 7.0.3). The existing ANOVA/Tukey results are retained and augmented with SD and 95% CI. The small standard deviations (PXI SD $\approx$ 0.08–0.09; EXI SD $\approx$ 0.15–0.17) and narrow, fully non-overlapping confidence intervals indicate that the ordering S1 < S2 < S3 is highly stable across random seeds. Note: this seed-robustness dataset is analytically distinct from the main 1000-run terminal snapshot (Table 9), which is why the S3 values here (78.44/74.62) differ slightly from the main-experiment values (78.29/74.50).

Accordingly, the conclusion of this section should be stated in a restrained manner. Within the scope of this study, the main pattern was broadly preserved across changes in seed and core parameters. However, this does not mean that the same numerical values and trajectories would be reproduced across all hospitals, all regulatory environments, or all operational rule sets. The significance of this validation does not lie in providing a complete proof of generalization. More precisely, it lies in demonstrating that the proposed HCEE-based healthcare AI governance architecture exhibited directional robustness within the implemented simulation space. This supports, without exaggeration, the claim that the preceding scenario comparison and stress-response interpretation are not accidental products of a single run. On this basis, the next chapter discusses what these validation results imply for the theoretical, design-oriented, and practical implications of the HCEE-based architecture.

6. Discussion

The findings of this study reaffirm that an approach that understands healthcare AI governance merely as a matter of normative oversight may not sufficiently explain the actual operational context. Existing discussions of healthcare AI governance have made important contributions by articulating principles such as safety, accountability, transparency, fairness, and human oversight, but they have been relatively less specific about what signals are detected under changing field conditions, through what pathways such signals are interpreted, and how they are translated into operational adjustment [4,7,71]. This study reformulated precisely this point as a systems problem. In other words, it argued that the core issue in healthcare AI governance lies not so much in the addition of principles themselves, but in whether the experience signals generated during operations are actually sensed, interpreted, and connected to adjustment and learning

through an operable feedback structure. This perspective leads healthcare AI governance to be understood not as a static oversight device, but as the operating architecture of an adaptive socio-technical system.

At the center of this reformulation lies a shift in the status of experience. In prior discussions, patient experience and employee experience have largely been treated as post hoc outcome indicators used to assess acceptance or satisfaction after implementation [29,66]. However, in actual healthcare settings, experiential changes such as patient anxiety and distrust, a perceived lack of explanation, staff workflow friction and coordination fatigue, and a reduced sense of control may appear before technical failure is formally identified [15,72]. This study interpreted such changes not as merely incidental reactions, but as operational signals that reveal misalignment within the system and the early need for intervention. The design logic of HCEE, which repositions PXI and EXI as governance inputs, is based precisely on this shift. In this respect, the present study does not simply reiterate the importance of human-centeredness, but can be understood as proposing an architecture-level framework that incorporates human experience into the actual judgment and control circuits of governance.

The scenario comparison results suggest that this design logic does not remain at the level of a conceptual proposal alone, but may appear as a system-level difference. The consistently observed ordering across S1, S2, and S3 indicates that differences in healthcare AI governance are difficult to reduce simply to the question of “whether governance exists or not” [16,40]. Rather, what matters is the extent to which experience signals are embedded within the structures of sensing, control, and learning. Following the activation logic defined in Section 3.1, these structural differences produced differences in both terminal outcomes and temporal paths, and the consistent pattern of $S1 < S2 < S3$ suggests that the maturity of experience-based governance may be associated with system performance.

Particular attention should be paid here to the meaning of S2. Although S2 showed improved results compared to S1, it also revealed clear limitations relative to S3. This does not mean that the sensing and visibility of experience signals are themselves meaningless. On the contrary, it indicates that some degree of improvement is possible even through sensing and visualization alone [2,16]. At the same time, however, it also indicates that this alone is not sufficient. Simply observing experience information does not immediately adjust the system, and unless the sensed signals are connected to interpretation, rule activation, control lever adjustment, and subsequent learning, a considerable portion of operational instability may remain. It is precisely at this point that the present study presents the message that “sensing alone is not enough” in a more academically restrained form. That is, the core of experience-based governance lies not in measurement itself, but in whether measured experience information is incorporated into a control pathway that actually changes the operational state.

The results of S3 extend this discussion one step further. The advantages of the full HCEE condition were not limited to higher terminal values, but also appeared in dynamic characteristics such as the buffering of the initial decline, a faster rebound, and stabilization at a higher level. This suggests that healthcare AI governance should not be understood only in terms of differences in performance levels, but also in terms of how the system adjusts and stabilizes itself over time [73,74]. In other words, what this study demonstrates is not merely that experience-based governance “produced” better outcomes, but that it was able to form a more adjustable and more responsive operational trajectory. In the language of Systems, this may be read as a result indicating that the success or failure of governance depends less on the mere existence of norms than on the operability of feedback, adaptation, and stabilization [74].

The interpretation of S4 requires greater caution. As noted in Section 5.2, S4 is a stress-test extension of the same architecture rather than a further maturity stage; it is therefore not appropriate to interpret its high terminal state as a positive effect of stress itself. Therefore, it is not appropriate to interpret the high terminal state of S4 as a positive effect of stress itself [73,74]. A more defensible interpretation is that, within the rule structure of this study, closed-loop governance was able to re-read experience signals after the shock, readjust the intensity of intervention, and reconfigure the subsequent trajectory. In this respect, S4 expands the meaning of healthcare AI governance beyond a mere improvement in the level of control to the issue of resilience capacity, that is, the capacity to restore functional alignment and re-stabilize after disruption. In particular, the sequential appearance of recovery, re-stabilization, and subsequent reinforcement after the initial loss suggests that a functioning closed-loop structure may serve not simply as a defensive mechanism, but as an adaptive operational order [73–75].

These findings clarify the theoretical contributions of this study more explicitly [73]. First, the study shifts healthcare AI governance from static oversight to an adaptive socio-technical operating architecture. This is an attempt to move away from viewing governance as a set of norms and inspections and to reconstruct it instead as a design problem that responds structurally to changing operational realities. Second, the study repositions patient experience and employee experience from downstream outcomes to governance inputs. This is meaningful in that it does not treat experience merely as an ethical supplement to human-centered discourse, but as a system variable that drives actual judgment and control. Third, the study proposes, through the four-layer closed-loop structure of Policy–Governance–Control–Experience, the relationship among policy, judgment, adjustment, and experience as a single architecture-level model. These contributions are significant in that they reinterpret the principle–practice gap in healthcare AI governance as the “absence of an operational architecture” and convert that gap into the problem of a designable structure [7,15,40].

The practical implications of this study are also clear. For healthcare institutions to implement AI governance in a substantive way, it is not sufficient merely to declare principles or strengthen ex post inspection [4,40]. Hospitals need to treat experience signals such as PXI and EXI not as peripheral reference materials, but as operational signals, and to gradually design a structure that connects them to threshold monitoring, trigger activation, escalation logic, control lever adjustment, and learning update. What matters here is not simply measuring experience more frequently, but jointly designing under what conditions such experience information is judged to represent a problem, through what level of review it is processed, through which operational levers it is adjusted, and in what way it is reflected in subsequent learning. In this sense, the maturity of hospital AI governance may depend less on the number of rules or the strength of declarations than on the ability to actually operate meaningful feedback loops. The results of this study suggest that the focus of healthcare AI governance should shift from “whether principles are in place” to “whether feedback loops actually operate” [16,40].

At the same time, the scope of interpretation of this study should be understood within clear limitations. This study is based on simulation-based validation, and the thresholds, trigger conditions, and operating parameters were set at a level intended to examine structural operability rather than to replicate the reality of any specific healthcare institution exactly. Therefore, it is not appropriate to generalize the results of this study uniformly to all healthcare institutions or to interpret them as indicating that identical values and trajectories would be reproduced in the field. In addition, the operationalization of PXI and EXI is also a design choice intended to convert the theoretical logic of this study into a testable form; accordingly, actual field application would require site-specific calibration that considers job composition, patient-group

characteristics, organizational scale, regulatory environment, and the feasibility of data collection. Furthermore, the robustness demonstrated in this study constitutes supporting evidence for consistency at the directional level, not a complete proof of universal predictive power. Therefore, the contribution of this study is best understood as mechanism-level explanatory evidence showing what structural differences experience-based closed-loop governance can produce.

Nevertheless, this study offers a significant turning point for research on healthcare AI governance. As AI rapidly diffuses across multiple operational touchpoints in hospitals, it is no longer sufficient to treat governance only as a matter of adoption or regulatory compliance. What is actually required is a feedback-based governance architecture that continuously reads experience signals from the field, connects them to operational adjustment and learning, and can re-establish stabilization even after disruption.

Through HCEE, this study proposes precisely this point at the design level and examines its structural possibility through simulation. Accordingly, the implications of this Discussion are relatively clear. Healthcare AI governance is difficult to explain sufficiently through the language of static oversight alone, and it needs to be re-understood in the language of an adaptive closed-loop system operating around experience signals. This discussion may lead naturally into the next chapter, the conclusion, where the study can more concisely summarize how it redefined healthcare AI governance and what its theoretical and practical implications are.

7. Conclusions

This study reconceptualized healthcare AI governance not as a static problem of checking the normative consistency of principles or the mere presence of oversight mechanisms, but as a problem of an adaptive system that detects experiential signals under changing operational conditions and connects them to pathways of judgment, adjustment, and learning. While prior discussions of healthcare AI governance have made significant contributions by establishing principles such as safety, accountability, transparency, and fairness, this study focused on how such principles can be translated into an operable order through a concrete structure within the actual operating processes of healthcare organizations. From this perspective, the core task of healthcare AI governance lies not in adding more principles per se but in building a feedback structure capable of interpreting signals generated during operations and linking them to subsequent adjustments.

In response to this problem, this study proposed HCEE (Human-Centered Experience Engineering) as an architecture-level design framework for healthcare AI governance. The core of HCEE lies not in confining patient experience and employee experience to post hoc indicators of satisfaction or acceptance but in repositioning them, through the Patient Experience Index (PXI) and Employee Experience Index (EXI), as governance input signals that reveal the state of the system, failures of alignment, and the need for intervention. On this basis, this study conceptualized healthcare AI governance as a closed-loop structure in which the four layers of Policy, Governance, Control, and Experience are mutually interconnected, and sought to interpret experience not as peripheral reference material but as a central input to the operation of governance.

The simulation results suggest that this reconceptualization has a certain degree of explanatory power. In the comparison of governance maturity, a consistent ordering of S1, S2, and S3 was observed, and the condition with a fully closed-loop structure exhibited higher terminal performance and more stable operational trajectories. In addition, under the stress-test condition, the same closed-loop structure demonstrated the possibility of establishing a path of recovery and re-stabilization after exogenous disruption. These findings suggest that the effectiveness of healthcare AI governance may depend not

simply on whether principles are formally present, but on the extent to which experiential signals are substantively connected to the structures of sensing, control, and learning. In other words, the findings of this study indicate that the practical success or failure of healthcare AI governance may depend less on the declaration of governance itself than on the operability of the feedback loop.

In this regard, the contributions of this study may be summarized as follows. First, it shifts healthcare AI governance from the perspective of static oversight to that of an adaptive socio-technical operating architecture. Second, by repositioning patient and employee experience not as downstream outcomes but as governance inputs, it extends the status of experience into a central variable of operational design. Third, it proposes a four-layer closed-loop structure linking policy, governance, control, and experience and presents it in a form that can be examined through simulation-based validation. These contributions indicate that healthcare organizations need to design AI governance beyond the level of declarative principles, as an operational structure in which sensing, judgment, adjustment, and learning are iteratively connected.

Of course, the findings of this study should be understood within the scope of simulation-based validation, and actual field application will require contextual calibration for individual institutions as well as additional empirical examination. Nevertheless, this study clarifies that the core of healthcare AI governance should be placed not in the enumeration of principles but in the design and operational feasibility of a workable feedback structure. Ultimately, what matters is not merely declaring the existence of governance, but whether that governance is structured so that it can operate repeatedly in actual healthcare settings. In this sense, this study proposes experience-based closed-loop governance as a meaningful approach to reconfiguring healthcare AI governance.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; software, M.K.; validation, M.K.; formal analysis, M.K.; investigation, M.K.; data curation, M.K.; writing—original draft preparation, M.K.; writing—review and editing, M.K. and J.C.; visualization, M.K.; supervision, J.C.; project administration, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study is based entirely on agent-based simulation and did not involve human participants or animals; no real patient or staff data were collected or used.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulation data and code generated and analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HCEE	Human-Centered Experience Engineering
PXI	Patient Experience Index
EXI	Employee Experience Index
TC	Trigger-Control
AAL	AI Autonomy Level
HOI	Human Oversight Intensity

TR	Trust Recalibration
IF	Information Flow
ABM	Agent-Based Modeling
HCAI	Human-Centered Artificial Intelligence
AX	AI Transformation

Appendix A. Detailed Operating Specification of the HCEE-Based Healthcare AI Governance Architecture

This appendix provides supplementary detail for architecture-level elements referred to in the main text but not presented there in full. Its purpose is to improve interpretability and reproducibility without interrupting the main argumentative flow of the paper. In particular, Table A1 summarizes how leverage points are distributed across the four-layer HCEE architecture, whereas Table A2 presents the trigger-control rulebook that specifies how experiential signals are translated into adaptive governance responses within the closed-loop structure.

Appendix A.1

Table A1 provides a supplementary mapping of leverage points across the four-layer HCEE architecture. Its purpose is to clarify that intervention depth is not concentrated in a single decision point, but distributed across the Policy, Governance, Control, and Experience layers with different functional roles. The table should therefore be read as an analytical mechanism map showing where information flows, rule structures, authority allocation, learning processes, and threshold-buffer arrangements operate within the architecture. It is included to support interpretation of the main text rather than to present a separate empirical result.

Table A1. Leverage Point Mapping Across the Four-Layer HCEE Architecture.

Leverage Dimension	Policy Layer	Governance Layer	Control Layer	Experience Layer
Information Flows	Defines what categories of experience-related information must be recognized within the governance boundary.	Routes PXI and EXI signals into trigger evaluation, review, and escalation processes.	Returns current implementation status and lever-adjustment outcomes to the upper layers.	Generates patient- and employee-experience signals that initiate upward feedback.
Rules and Intervention Criteria	Establishes high-level boundary conditions related to safety, accountability, fairness, and human oversight.	Interprets experience signals through trigger rules, escalation logic, and review criteria.	Translates governance direction into operational lever adjustment, including AAL, HOI, TR, and IF.	Supplies the observed conditions under which intervention criteria become relevant.
Authority and Escalation	Retains authority over higher-order constraints and exceptional review conditions.	Determines whether detected deterioration requires oversight review, selective adjustment, or escalation.	Executes permitted interventions and reports residual issues or implementation effects.	Accumulates field-level consequences that may call for further governance attention.
Delays and Learning	Reviews persistent system-level concerns and adjusts higher-	Interprets repeated warning patterns and incorporates them into subsequent	Recalibrates intervention intensity on the basis of repeated signals and	Reflects lagged effects of prior interventions in

	order priorities when needed.	review and learning cycles.	prior adjustment outcomes.	subsequent PXI and EXI trajectories.
Thresholds and Buffers	Defines acceptable operating boundaries within which lower-level adjustment may occur.	Interprets sustained decline, cross-domain imbalance, persistence, and stress cues as governance-relevant signals.	Applies threshold-based activation and stabilizing control logic to avoid overreaction to short-term fluctuation.	Generates the underlying variation that is subsequently filtered, interpreted, and fed back into governance.

Note. Table A1 is intended as an architecture-level mechanism map. It summarizes where different forms of leverage are located within the HCEE framework and how they contribute to closed-loop adjustment. The table supports interpretation of the architecture presented in the main text and should not be read as a separate cell-by-cell empirical validation table. System-level implications are examined through the scenario-based results reported in the Results section.

Appendix A.2

Table A2 provides the full trigger-control rulebook referred to in the main text. Its purpose is to clarify how changes in patient and employee experience are translated into governance-relevant signals and subsequently into adaptive control responses within the HCEE closed-loop architecture. The table summarizes the trigger conditions, their governance interpretation, the principal control direction, and the intended system-level effect. It is included to make the operational logic of the architecture more explicit and should be read as a design-oriented rule summary rather than as a separate empirical result.

Table A2. Full Trigger-Control Rulebook for the HCEE Closed-Loop Governance Architecture.

Trigger Code	Trigger Condition	Governance Interpretation	Primary Control Direction	Intended System Effect
TC-01	PXI < 70 after smoothing and rolling-window evaluation	Patient-side experience deterioration requiring corrective attention	Increase trust-repair effort and, where necessary, strengthen oversight and information support	Reduce patient-side dissatisfaction and support recovery of service experience
TC-02	EXI < 65 after smoothing and rolling-window evaluation	Employee-side experience deterioration indicating operational strain or coordination burden	Increase oversight intensity and coordination support; adjust autonomy if required	Reduce staff-side strain and support restoration of operational stability
TC-03	$ PXI - EXI > 15$ after smoothing and rolling-window evaluation	Cross-domain imbalance between patient and employee experience requiring rebalancing	Strengthen review and information-flow adjustment to rebalance system conditions across domains	Limit divergence between PXI and EXI and improve cross-domain alignment
TC-04	Persistent warning condition over time	Sustained deterioration suggesting structural rather than temporary instability	Escalate intervention intensity and maintain adjustment until the warning condition is moderated	Prevent prolonged instability and support re-stabilization of the operating state

TC-05	Stress cue or disruption-related instability signal	Elevated risk condition requiring broader adaptive response across the closed loop	Reconfigure lever combination as needed across autonomy, oversight, trust repair, and information flow	Support recovery under stress and restore a stable operating trajectory
-------	---	--	--	---

Note. In the simulation, trigger detection is based on rolling-window monitoring with signal smoothing. For TC-01 to TC-03, the main text specifies a 7-day rolling average with exponential smoothing ($\alpha = 0.3$). Table A2 summarizes the operating logic through which experiential deterioration is converted into adaptive governance response. The table should therefore be interpreted as a rule specification embedded in the architecture, while validation of system effects is reported through the scenario-based and robustness analyses in the Section 5.

References

1. Poon, E.G.; Lemak, C.H.; Rojas, J.C.; Guptill, J.; Classen, D. Adoption of artificial intelligence in healthcare: Survey of health system priorities, successes, and challenges. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 1093–1100.
2. Hassan, M.; Kushniruk, A.; Borycki, E. Barriers to and facilitators of artificial intelligence adoption in health care: Scoping review. *JMIR Hum. Factors* **2024**, *11*, e48633.
3. Morley, J.; Machado, C.C.; Burr, C.; Cows, J.; Joshi, I.; Taddeo, M.; Floridi, L. The ethics of AI in health care: A mapping review. *Soc. Sci. Med.* **2020**, *260*, 113172.
4. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*; World Health Organization: Geneva, Switzerland, 2024.
5. Mennella, C.; Maniscalco, U.; De Pietro, G.; Esposito, M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon* **2024**, *10*, e26297.
6. Ibáñez, J.C.; Olmeda, M.V. Operationalising AI ethics: How are companies bridging the gap between practice and principles? An exploratory study. *AI Soc.* **2022**, *37*, 1663–1687.
7. Hassan, M.; Borycki, E.M.; Kushniruk, A.W. Artificial intelligence governance framework for healthcare. *Healthc. Manag. Forum* **2025**, *38*, 125–130.
8. Hussein, R.; Zink, A.; Ramadan, B.; Howard, F.M.; Hightower, M.; Shah, S.; Beaulieu-Jones, B.K. Advancing healthcare AI governance through a comprehensive maturity model based on systematic review. *npj Digit. Med.* **2026**, *9*, 236.
9. Rahimi, A.K.; Pienaar, O.; Ghadimi, M.; Canfell, O.J.; Pole, J.D.; Shrapnel, S.; van der Vegt, A.H.; Sullivan, C. Implementing AI in hospitals to achieve a learning health system: Systematic review of current enablers and barriers. *J. Med. Internet Res.* **2024**, *26*, e49655.
10. Kim, J.Y.; Hasan, A.; Kueper, J.; Tang, T.; Hayes, C.; Fine, B.; Balu, S.; Sendak, M. Establishing organizational AI governance in healthcare: A case study in Canada. *npj Digit. Med.* **2025**, *8*, 522. <https://doi.org/10.1038/s41746-025-01909-3>.
11. Moy, S.; Irannejad, M.; Manning, S.J.; Farahani, M.; Ahmed, Y.; Gao, E.; Prabhune, R.; Lorenz, S.; Mirza, R.; Klinger, C. Patient perspectives on the use of artificial intelligence in health care: A scoping review. *J. Patient-Centered Res. Rev.* **2024**, *11*, 51–62.
12. Ayorinde, A.; Mensah, D.O.; Walsh, J.; Ghosh, I.; Ibrahim, S.A.; Hogg, J.; Peek, N.; Griffiths, F. Health care professionals' experience of using AI: Systematic review with narrative synthesis. *J. Med. Internet Res.* **2024**, *26*, e55766.
13. Fazakarley, C.-A.; Breen, M.; Thompson, B.; Leeson, P.; Williamson, V. Beliefs, experiences and concerns of using artificial intelligence in healthcare: A qualitative synthesis. *Digit. Health* **2024**, *10*, 20552076241230075.
14. Li, L.Z.; Yang, P.; Singer, S.J.; Pfeffer, J.; Mathur, M.B.; Shanafelt, T. Nurse burnout and patient safety, satisfaction, and quality of care: A systematic review and meta-analysis. *JAMA Netw. Open* **2024**, *7*, e2443059.
15. van Leersum, C.M.; Maathuis, C. Human centred explainable AI decision-making in healthcare. *J. Responsible Technol.* **2025**, *21*, 100108.
16. Wells, B.J.; Nguyen, H.M.; McWilliams, A.; Pallini, M.; Bovi, A.; Kuzma, A.; Kramer, J.; Chou, S.-H.; Hetherington, T.; Corn, P.; et al. A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. *npj Digit. Med.* **2025**, *8*, 514. <https://doi.org/10.1038/s41746-025-01900-y>.
17. Hwang, G.; Hejazian, S.S.; Vafaei Sadr, A.; Wagner, J.K.; Hao, P.; Vemuri, A.; Kawamura, Y.; Nawab, K.; Hijjawi, S.; Zand, R. Translating AI to the Bedside with Physician Buy-In: Recommendations from a Meta-Analysis and Systematic Review of the Literature. *Bioengineering* **2025**, *12*, 1363.
18. Katonai, G.; Arvai, N.; Mesko, B. AI and primary care: Scoping review. *J. Med. Internet Res.* **2025**, *27*, e65950.

19. Nair, M.; Nygren, J.; Nilsen, P.; Gama, F.; Neher, M.; Larsson, I.; Svedberg, P. Critical activities for successful implementation and adoption of AI in healthcare: Towards a process framework for healthcare organizations. *Front. Digit. Health* **2025**, *7*, 1550459.
20. Adnan, H.S.; Shidani, A.; Clifton, L.; Bankhead, C.R.; Perera-Salazar, R. Implementation framework for AI deployment at scale in healthcare systems. *iScience* **2025**, *28*, 112406.
21. Chan, A.; Rahimi-Ardabili, H.; Rogers, W.A.; Coiera, E. The real-world impact of artificial intelligence ethics frameworks across a decade in healthcare: A scoping review. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 1767–1777.
22. Shiferaw, K.B.; Roloff, M.; Balaur, I.; Welter, D.; Waltemath, D.; Zeleke, A.A. Guidelines and standard frameworks for artificial intelligence in medicine: A systematic review. *JAMIA Open* **2025**, *8*, ooae155.
23. Zabala, G.; Pruitt, Z.M.; Fairbanks, R.J.; Ratwani, R. Assessing the Readiness of Health Care Organizations for Safe AI Integration: Perspectives from Quality and Safety Leaders. *J. Patient Saf.* **2026**, *22*, 168–172.
24. Nouis, S.C.; Uren, V.; Jariwala, S. Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: A qualitative study of healthcare professionals' perspectives in the UK. *BMC Med. Ethics* **2025**, *26*, 89.
25. Olawade, D.B.; Weerasinghe, K.; Teke, J.; Msiska, M.; Boussios, S.; Hatzidimitriadou, E. Evaluating AI adoption in healthcare: Insights from the information governance professionals in the United Kingdom. *Int. J. Med. Inform.* **2025**, *199*, 105909.
26. Elgin, C.Y.; Elgin, C. Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: A qualitative study of healthcare professionals' perspectives. *BMC Med. Ethics* **2024**, *25*, 148.
27. Osnat, B. Patient perspectives on artificial intelligence in healthcare: A global scoping review of benefits, ethical concerns, and implementation strategies. *Int. J. Med. Inform.* **2025**, *203*, 106007.
28. Foresman, G.; Biro, J.; Tran, A.; MacRae, K.; Kazi, S.; Schubel, L.; Visconti, A.; Gallagher, W.; Smith, K.M.; Giardina, T. Patient perspectives on artificial intelligence in health care: Focus group study for diagnostic communication and tool implementation. *J. Particip. Med.* **2025**, *17*, e69564.
29. Tun, H.M.; Rahman, H.A.; Naing, L.; Malik, O.A. Trust in Artificial Intelligence-Based Clinical Decision Support Systems Among Health Care Workers: Systematic Review. *J. Med. Internet Res.* **2025**, *27*, e69678. <https://doi.org/10.2196/69678>.
30. Mikkonen, K.; Tuunainen, S.; Oikarinen, A.; Jansson, M.; Woo, B.; Zhou, W.; Tam, W.; Tuomikoski, A.M.; Kaakinen, P.; Juntunen, J. Artificial Intelligence Technologies Supporting Nurses' Clinical Decision-Making: A Systematic Review. *J. Clin. Nurs.* **2026**, *35*, 1525–1540.
31. Berger, S.A.; Håland, E.; Solbjør, M. Patient Perspectives on Trust in Artificial Intelligence-Powered Tools in Prostate Cancer Diagnostics. *Qual. Health Res.* **2025**, *36*, 276–288. <https://doi.org/10.1177/10497323251387545>.
32. Chandrasekaran, R.; Moustakas, E. Patient attitudes toward ambient artificial intelligence scribes in clinical care: Insights from a cross-sectional study. *J. Am. Med. Inform. Assoc.* **2026**, *33*, 263–272.
33. Lawrence, K.; Kuram, V.S.; Levine, D.L.; Sharif, S.; Polet, C.; Malhotra, K.; Owens, K. Informed Consent for Ambient Documentation Using Generative AI in Ambulatory Care. *JAMA Netw. Open* **2025**, *8*, e2522400. <https://doi.org/10.1001/jamanetworkopen.2025.22400>.
34. Rosenbacke, R.; Melhus, Å.; McKee, M.; Stuckler, D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: Systematic review. *JMIR AI* **2024**, *3*, e53207.
35. Kelly, A.; Bhardwaj, N.; Sainte-Marie, T.T.H.; Van de Ven, P.; Melia, R.; Williams, J.E.; Mathiasen, K.; Nielsen, A.S. Investigating How Clinicians Form Trust in an AI-Based Mental Health Model: Qualitative Case Study. *JMIR Hum. Factors* **2025**, *12*, e79658.
36. Omori, M.; Basnayake, P.; Frazer, H.M.; Keogh, L.; Kunicki, K.; Lippey, J.F. Trust in AI is a “fluid process”: Building trust of AI through clinicians' needs in the BreastScreen Victoria Program—A qualitative study. *Qual. Health Res.* **2025**, *36*, 262–275. <https://doi.org/10.1177/10497323251351521>.
37. Ahmad, O.; Mason, S.; Stanley, S.; Nwosu, A.C. Exploring Perspectives of Health Care Professionals on AI in Palliative Care: Qualitative Interview Study. *JMIR Hum. Factors* **2025**, *12*, e79514.
38. Martínez-Martínez, H.; Martínez-Alfonso, J.; Sánchez-Rojo-Huertas, B.; Reynolds-Cortez, V.; Turégano-Chumillas, A.; Meseguer-Ruiz, V.A.; Cekrezi, S.; Martínez-Vizcaino, V. Perceptions of, barriers to, and facilitators of the use of AI in primary care: Systematic review of qualitative studies. *J. Med. Internet Res.* **2025**, *27*, e71186.
39. Khan, M.M.; Shah, N.; Shaikh, N.; Thabet, A.; Belkhair, S. Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges. *Int. J. Med. Inform.* **2025**, *195*, 105780.
40. Lekadir, K.; Frangi, A.F.; Porras, A.R.; Glocker, B.; Cintas, C.; Langlotz, C.P.; Weicken, E.; Asselbergs, F.W.; Prior, F.; Collins, G.S. FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* **2025**, *388*, e081554.

41. Shevtsova, D.; Ahmed, A.; Boot, I.W.; Sanges, C.; Hudecek, M.; Jacobs, J.J.; Hort, S.; Vrijhoef, H.J. Trust in and acceptance of artificial intelligence applications in medicine: Mixed methods study. *JMIR Hum. Factors* **2024**, *11*, e47031.
42. McCradden, M.D.; Thai, K.; Assadi, A.; Tonekaboni, S.; Stedman, I.; Joshi, S.; Zhang, M.; Chevalier, F.; Goldenberg, A. What makes a ‘good’ decision with artificial intelligence? A grounded theory study in paediatric care. *BMJ Evid.-Based. Med.* **2025**, *30*, 183–193. <https://doi.org/10.1136/bmjebm-2024-112919>.
43. Kim, J.Y.; Hasan, A.; Balu, S.; Sendak, M. People process technology and operations framework for establishing AI governance in healthcare organizations. *npj Digit. Med.* **2026**, *9*, 224.
44. Tsoi, A.H.; Gartner, G.; Cotten, S.W.; Kim, J.; Nazarian, J.; Thomas, J.; McSwain, S.D.; Ahmadi-Moosavi, R.; Rimal, R. Establishing and implementing a responsible artificial intelligence framework: A 1-year review. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 1778–1784.
45. Beger, J. Not someone, but something: Rethinking trust in the age of medical AI. *Eur. J. Radiol. Artif. Intell.* **2025**, *3*, 100038.
46. Maimaitiaili, M.; Jiamaliding, Y.; Dai, G.; Xiao, H.; Kuerbanjiang, W.; Yi, Y. Artificial intelligence platform architecture for hospital systems: Systematic review. *J. Med. Internet Res.* **2025**, *27*, e79788.
47. Patil, S.V.; Myers, C.G.; Dai, T. Protecting clinical value judgment in the age of AI. *npj Digit. Med.* **2026**, *9*, 269.
48. Nong, P.; Maurer, E.; Dwivedi, R. The urgency of centering safety-net organizations in AI governance. *npj Digit. Med.* **2025**, *8*, 117.
49. Jacob, C.; Brasier, N.; Laurenzi, E.; Heuss, S.; Mougiakakou, S.-G.; Cöltekin, A.; Peter, M.K. AI for IMPACTS framework for evaluating the long-term real-world impacts of AI-powered clinician tools: Systematic review and narrative synthesis. *J. Med. Internet Res.* **2025**, *27*, e67485.
50. Stroud, A.M.; Minter, S.A.; Zhu, X.; Ridgeway, J.L.; Miller, J.E.; Barry, B.A. Patient information needs for transparent and trustworthy cardiovascular artificial intelligence: A qualitative study. *PLoS Digit. Health* **2025**, *4*, e0000826.
51. Alelyani, T. A validated framework for responsible AI in healthcare autonomous systems. *Sci. Rep.* **2025**, *15*, 44432.
52. Templin, T.; Fort, S.; Padmanabham, P.; Seshadri, P.; Rimal, R.; Oliva, J.; Lich, K.H.; Sylvia, S.; Sinnott-Armstrong, N. Framework for bias evaluation in large language models in healthcare settings. *npj Digit. Med.* **2025**, *8*, 414.
53. Hong, C.; Chowdhury, A.; Sorrentino, A.D.; Wang, H.; Agrawal, M.; Bedoya, A.; Bessias, S.; Economou-Zavlanos, N.J.; Wong, I.; Pean, C. Application of unified health large language model evaluation framework to In-Basket message replies: Bridging qualitative and quantitative assessments. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 626–637.
54. You, J.G.; Hernandez-Boussard, T.; Pfeffer, M.A.; Landman, A.; Mishuris, R.G. Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. *npj Digit. Med.* **2025**, *8*, 107.
55. Boudierhem, R. Shaping the future of AI in healthcare through ethics and governance. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 416.
56. Richter, F.; Holmes, E.; Richter, F.; Guttmann, K.; Duong, S.Q.; Gangadharan, S.; Schadt, E.E.; Salmasian, H.; Gelb, B.D.; Glicksberg, B.S. Toward governance of artificial intelligence in pediatric healthcare. *npj Digit. Med.* **2025**, *8*, 636.
57. Biro, J.; Jabbarpour, Y.; Ratwani, R. A new approach to artificial intelligence safety in primary care. *Lancet Prim. Care* **2026**, *2*, 100115.
58. Rosenthal, J.T.; Beecy, A.; Sabuncu, M.R. Rethinking clinical trials for medical AI with dynamic deployments of adaptive systems. *npj Digit. Med.* **2025**, *8*, 252.
59. Jackson, N.J.; Brown, K.E.; Miller, R.; Murrow, M.; Cauley, M.R.; Collins, B.X.; Novak, L.L.; Benda, N.C.; Ancker, J.S. Factors influencing the effectiveness of artificial intelligence-assisted decision-making in medicine: A scoping review. *J. Am. Med. Inform. Assoc.* **2026**, *33*, 1054–1064.
60. Saenz, A.D.; McCoy, T.; Mantha, A.B.; Martin, R.; Damiano, R.; Adair, D.; Heaney, D.; Sisodia, R.; Park, L.; Forsberg, R.; et al. Establishing responsible use of AI guidelines: A comprehensive case study for healthcare institutions. *npj Digit. Med.* **2024**, *7*, 348. <https://doi.org/10.1038/s41746-024-01300-8>.
61. Solaiman, B.; Mekki, Y.M.; Qadir, J.; Ghaly, M.; Abdelkareem, M.; Al-Ansari, A. A “True Lifecycle Approach” towards governing healthcare AI with the GCC as a global governance model. *npj Digit. Med.* **2025**, *8*, 337. <https://doi.org/10.1038/s41746-025-01614-1>.
62. Collins, B.X.; Bélisle-Pipon, J.-C.; Evans, B.J.; Ferryman, K.; Jiang, X.; Nebeker, C.; Novak, L.; Roberts, K.; Were, M.; Yin, Z. Addressing ethical issues in healthcare artificial intelligence using a lifecycle-informed process. *JAMIA Open* **2024**, *7*, o0ae108.
63. Stogiannos, N.; O’Regan, T.; Scurr, E.; Litosseliti, L.; Pogose, M.; Harvey, H.; Kumar, A.; Malik, R.; Barnes, A.; McEntee, M.F. Lessons on AI implementation from senior clinical practitioners: An exploratory qualitative study in medical imaging and radiotherapy in the UK. *J. Med. Imaging Radiat. Sci.* **2025**, *56*, 101797.

64. Starke, G.; Gille, F.; Termine, A.; Aquino, Y.S.J.; Chavarriaga, R.; Ferrario, A.; Hastings, J.; Jongsma, K.; Kellmeyer, P.; Kulynych, B. Finding consensus on trust in AI in health care: Recommendations from a panel of international experts. *J. Med. Internet Res.* **2025**, *27*, e56306.
65. Bodnari, A.; Travis, J. Scaling enterprise AI in healthcare: The role of governance in risk mitigation frameworks. *npj Digit. Med.* **2025**, *8*, 272.
66. Gundlack, J.; Thiel, C.; Negash, S.; Buch, C.; Apfelbacher, T.; Denny, K.; Christoph, J.; Mikolajczyk, R.; Unverzagt, S.; Frese, T. Patients' Perceptions of Artificial Intelligence Acceptance, Challenges, and Use in Medical Care: Qualitative Study. *J. Med. Internet Res.* **2025**, *27*, e70487. <https://doi.org/10.2196/70487>.
67. Chen, C.; Cui, Z. Impact of AI-Assisted Diagnosis on American Patients' Trust in and Intention to Seek Help from Health Care Professionals: Randomized, Web-Based Survey Experiment. *J. Med. Internet Res.* **2025**, *27*, e66083. <https://doi.org/10.2196/66083>.
68. Abdelwanis, M.; Simsekler, M.C.E.; Gabor, A.F.; Sleptchenko, A.; Omar, M. Artificial intelligence adoption challenges from healthcare providers' perspectives: A comprehensive review and future directions. *Saf. Sci.* **2026**, *193*, 107028.
69. Kandaswamy, S.; Muthu, N.; Braykov, N.; Carter, R.; Blanco, R.; Bui, T.; Orenstein, E.; Mai, M. Human performance evaluation of a pediatric artificial intelligence sepsis model. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 1552–1561. <https://doi.org/10.1093/jamia/ocaf106>.
70. Jain, S.S.; Goto, S.; Hall, J.L.; Khan, S.S.; MacRae, C.A.; Ofori, C.; Pegus, C.; Pencina, M.; Peterson, E.D.; Schwamm, L.H. Pragmatic Approaches to the Evaluation and Monitoring of Artificial Intelligence in Health Care: A Science Advisory from the American Heart Association. *Circulation* **2025**, *152*, e433–e442.
71. Bailo, P.; Nittari, G.; Pesel, G.; Basello, E.; Spasari, T.; Ricci, G. Governing Healthcare AI in the Real World: How Fairness, Transparency, and Human Oversight Can Coexist: A Narrative Review. *Sci* **2026**, *8*, 36.
72. Mertz, M.; Toskovich, K.; Shields, G.; Attema, G.; Dumond, J.; Cameron, E. Exploring trust factors in AI-healthcare integration: A rapid review. *Front. Artif. Intell.* **2025**, *8*, 1658510.
73. Salwei, M.E.; Carayon, P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. *J. Cogn. Eng. Decis. Mak.* **2022**, *16*, 194–206.
74. Matthieu, V.; Alexei, S. The resilience of complex sociotechnical systems: A meta-review of conceptualisations. *Systems* **2026**, *14*, 71.
75. Sáez, C.; Ferri, P.; García-Gómez, J.M. Resilient artificial intelligence in health: Synthesis and research agenda toward next-generation trustworthy clinical decision support. *J. Med. Internet Res.* **2024**, *26*, e50295.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.