

Article

Explainable Artificial Intelligence in Rehabilitation Nursing: A Sociotechnical Framework for Human-Centered Clinical Decision Support

Filipe P. Ramos ^{1,2,*} , Arnaldo Santos ^{1,3} , Tania Rocha ^{3,4} , Fernando Petronilho ^{5,6}  and Rui Pereira ^{5,6} 

¹ Department of Sciences and Technology, Universidade Aberta, Rua da Escola Politécnica, 147, 1269-001 Lisboa, Portugal; arnaldo.santos@uab.pt

² INESC TEC—Institute for Systems and Computer Engineering, Technology and Science, Centre HumanISE, Universidade Aberta, Rua da Escola Politécnica, 147, 1269-001 Lisboa, Portugal

³ INESC TEC—Institute for Systems and Computer Engineering, Technology and Science, Centre Digimedia, Universidade Aberta, Rua da Escola Politécnica, 147, 1269-001 Lisboa, Portugal; trocha@utad.pt

⁴ Department of Engineering, School of Science and Technology, University of Trás-os-Montes and Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal

⁵ School of Nursing, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal; fpetronilho@ese.uminho.pt (F.P.); ruipereira@ese.uminho.pt (R.P.)

⁶ Unidade de Investigação em Ciências da Saúde: Enfermagem, Health Sciences Research Unit: Nursing, University of Coimbra, Rua Doutor José Alberto Reis, 3000-232 Coimbra, Portugal

* Correspondence: filipe.p.ramos@inesctec.pt

Abstract

Healthcare systems are complex adaptive environments in which clinical work, digital technologies, and organizational routines interact continuously, often challenging the integration of artificial intelligence (AI) into everyday practice. Although explainable AI (xAI) has been proposed to address concerns related to algorithmic opacity and professional trust, explainability is still frequently approached. Grounded in General Systems Theory, sociotechnical systems theory, and complexity science, this study conceptualizes explainability as an emergent system-level property of healthcare systems. Using Design Science Research as a systems-oriented inquiry methodology, a human-centered conceptual framework for AI-supported clinical decision-making was developed through iterative cycles of problem framing, design, demonstration, and evaluation. The framework was explored in rehabilitation nursing, a domain characterized by multidimensional patient data, longitudinal decision processes, and close professional–patient interaction. Iterative engagement with Rehabilitation Nursing Specialists informed design principles related to user participation, contextualized explanations, and workflow alignment. An exploratory evaluation with 144 Specialists assessed perceived usefulness, comprehensibility of explanations, and acceptance of AI-supported recommendations in realistic scenarios. The findings indicate that explainability is experienced not as a property of the algorithm alone, but as an outcome emerging from interactions between AI behavior, human interpretation, and organizational context. The framework shows potential to meaningfully support clinical decision-making in Rehabilitation Nursing by providing contextually aligned, human-centered explanatory outputs.

Keywords: explainable artificial intelligence; human-centered design; sociotechnical systems; rehabilitation nursing; clinical decision-making; interactive decision support systems; complexity science; design science research



Academic Editor: Peter Tsasis

Received: 22 March 2026

Revised: 15 May 2026

Accepted: 19 June 2026

Published: 2 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Healthcare systems are increasingly understood as complex adaptive systems in which multiple actors interact, decisions are distributed, and outcomes emerge through nonlinear dynamics. Clinical practice does not unfold in isolation; it unfolds with technology, within routines, professional relationships, organizational constraints, and evolving regulatory frameworks. In such environments, system performance depends less on individual components and more on how those components interact over time. As a result, interventions that focus narrowly on technical optimization often struggle to achieve sustained integration or meaningful impact in real-world healthcare settings.

The rapid incorporation of digital technologies into healthcare, particularly AI-based clinical decision support systems, illustrates these systemic challenges. While AI holds promise for supporting clinical reasoning, managing complex patient data, and optimizing care pathways, its introduction frequently produces unintended consequences, including workflow disruption, professional resistance, and erosion of trust. These effects are particularly evident in care contexts such as rehabilitation nursing, where decision-making is cognitively demanding, longitudinal, and strongly context-dependent. From a systemic perspective, such outcomes usually reflect misalignments among technological artifacts, professional sense-making, and the organizational context, rather than limitations of the algorithms alone.

Explainable artificial intelligence (xAI) has emerged as a response to concerns regarding opacity and accountability in AI-supported decision-making. Much of the existing literature, however, conceptualizes explainability primarily as a technical property of algorithms, emphasizing interpretability methods, visualizations, or post hoc explanation techniques [1].

Recent research in explainable AI in healthcare has highlighted the need to move beyond algorithm-centric interpretability toward approaches that consider the broader system, including human, organizational, and contextual dimensions [2,3].

While these contributions are important, they often overlook how explanations are interpreted, negotiated, and used in practice. In clinical settings, explanations are not evaluated solely by their technical correctness, but by whether they support action, understanding, and trust. This shifts the focus from explaining models in isolation to understanding the conditions under which explanations become usable and meaningful within complex sociotechnical systems.

Despite long-standing insights from systems theory and sociotechnical research, their application to explainable AI remains limited. Conceptual frameworks that explicitly treat explainability as a system-level phenomenon, emerging from interactions between algorithmic behavior, human cognition, professional practice, and organizational context, remain scarce.

Addressing this gap is essential for both systems research and the effective integration of AI into healthcare practice. In response, this study adopts a sociotechnical systems perspective grounded in General Systems Theory, sociotechnical systems theory, and complexity science. From this standpoint, explainability is not treated as a standalone technical attribute, but as an emergent property of the sociotechnical system that arises when AI behavior, explanatory representations, user interpretation, and organizational conditions are coherently aligned. Recent evidence reinforces this gap, indicating that current explainable AI approaches in clinical decision support still face unresolved challenges related to integration, usability, and real-world applicability [4].

Methodologically, the study develops the proposed conceptual framework using Design Science Research (DSR) as a system-oriented inquiry approach. Rather than treating DSR solely as a method for artifact construction, it is used here as a form of systemic

intervention, enabling iterative cycles of problem framing, design, demonstration, and evaluation within a complex adaptive environment. Through these cycles, feedback from the system informs successive refinements, reflecting the nonlinear and evolutionary nature of sociotechnical change. This approach allows the study to simultaneously generate theoretically grounded insights and explore their practical implications within a real-world healthcare setting.

The proposed conceptual framework is explored within the domain of rehabilitation nursing, which serves as an empirical context to illustrate and examine the systemic dynamics of explainable AI integration. Rehabilitation nursing is characterized by multidimensional patient information, longitudinal decision processes, and close interaction between professionals and patients, making it a particularly suitable context for investigating explainability as a sociotechnical phenomenon. Importantly, rehabilitation nursing is not positioned as the primary object of theoretical contribution, but as a field of application through which systems-level principles can be examined and articulated.

This article contributes in three ways. First, it advances systems theory by conceptualizing explainable AI as an emergent property of socio-technical healthcare systems. Second, it proposes a systems-based, human-centered conceptual framework to guide the design and evaluation of AI-assisted decision-making in complex clinical settings. Third, it demonstrates how Design Science Research can be operationalized as a systemic research methodology to study and shape human–AI interaction in healthcare systems. Together, these contributions aim to support both theoretical development in systems research and a more effective, context-sensitive integration of AI into healthcare practice.

2. Theoretical Framework

2.1. *Systems Theory and Complexity in Health Care*

The foundations of systems thinking were established by Bertalanffy's General Systems Theory (TGS) [5], which proposes the understanding of phenomena as integrated wholes, rather than mere sets of isolated parts. According to Bertalanffy [5], a system is defined by the holism and interdependence of its components, as well as by their interactions with the surrounding environment. A central feature of systems is the emergence of properties, that is, attributes of the system that cannot be deduced solely from the properties of individual components.

The application of the concepts in the context of health care reveals that hospitals and medical care networks function as complex adaptive systems. In these systems, health professionals, patients, technologies, equipment, and clinical routines interact in a nonlinear way [6,7]. In these types of systems, small changes can have amplified and unpredictable consequences, and overall behavior is not easily predictable from fixed rules. For example, the introduction of a new technological tool in a hospital can have different impacts on teams, leading to informal adaptations in the workflow or unanticipated side effects. This phenomenon is consistent with the adaptive and evolutionary nature of these systems [8]. For this reason, studies of AI in healthcare benefit from a systemic perspective that recognizes technology as part of a broader clinical ecosystem. Instead, technology becomes part of a broader clinical ecosystem, which is influenced by multiple contextual factors.

2.2. *Sociotechnical Systems and Human Factors*

In the twentieth century, Eric Trist and colleagues [9] introduced the concept of sociotechnical systems, demonstrating in classical studies of work organization that the optimal performance of a system requires the joint optimization of its social (human) and technical components. Subsequently, Albert Cherns [10] formulated principles of

socio-technical design, such as the active participation of users in the design process, the sharing of decision-making power, and the mutual adaptation between technologies and work processes. These principles aim to ensure that the introduction of new technologies effectively improves the system without adverse effects on people. These ideas remain fully relevant in the digital age, as sociotechnical systems research emphasizes the importance of considering interactions between users, tasks, and technologies from the outset. In health settings, Sittig et al. [11] proposed an eight-dimensional sociotechnical model to assess the introduction of clinical information systems.

This model covers components such as hardware and software, clinical content, human–computer interface, workflow processes, clinical team, internal organizational environment, external characteristics (regulations, market), and monitoring/evaluation systems. The main lesson drawn from these models is that focusing exclusively on technology, without proper adaptation of the social and organizational context, rarely produces the desired results. For example, the implementation of a clinical decision support system without adequate training for health professionals or without the proper adaptation of work protocols can result in a suboptimal use of technology or even in shortcuts and procedural deviations [12]. Consequently, a truly systemic analysis of any health innovation, such as the introduction of Explainable Artificial Intelligence tools, must consider this complex web of socio-technical interdependencies. In the case under study, this means recognizing that the performance and usefulness of the proposed framework will be conditioned by factors such as nurses' adherence, the institution's policies, the ease of integration into the rehabilitation workflow, and the support of top management.

2.3. Explainable AI and Contextual Explainability in Healthcare

xAI has emerged in the last decade as a vital domain for ensuring transparency and interpretability in machine learning algorithms [13]. In the context of healthcare systems, AI is of particular relevance, as AI-assisted clinical decisions depend on the trust of professionals and patient acceptance. These elements are closely associated with the ability to understand the recommendations generated by the system. Explainability can be understood as a qualitative characteristic of the AI system that allows it to articulate, in an intelligible way to human beings, the reasons underlying its predictions or suggestions. However, interdisciplinary research has shown that making models “explainable” is a challenge that transcends merely technical issues, involving the understanding of how professionals interpret and use explanations in practice [13]. In particular, explainability should be contextual and user-centered, i.e., explanations should be tailored to the professional's profile (e.g., level of experience, specialty) and the ongoing clinical situation [14]. This view of contextualized explainability is aligned with the notion that effective transparency is an emergent property of the entire sociotechnical system, fundamentally depending on the alignment between algorithm behavior, human interpretation, and organizational context [15]. For example, an extensive and technical explanation may be valuable to an experienced researcher or physician in an audit situation.

Still, it would be counterproductive to a nurse in the emergency department who needs quick, practical guidance [15]. Therefore, xAI systems in the field of health should be flexible to adapt the depth and form of their explanations to the circumstances. In critical scenarios, such as emergencies, they should enable the presentation of succinct and action-directed justifications, while in rehabilitation or case management contexts, they can provide more detailed clarifications, integrating information on the patient's evolution and clinical foundations. Importantly, explainability only generates value if users trust the system. Trust is built not only through transparency, but also through perceived clinical relevance, consistency with professional knowledge, and experiential

validation. In this sense, involving end-users in the design of explanations is essential to ensure that AI-generated knowledge representations align with professional reasoning and contextual needs.

To synthesize the theoretical elements discussed, Table 1 summarizes the main conceptual frameworks that support the present study, highlighting their key contributions and the way they inform the development of our framework.

Table 1. Theoretical foundations and their contributions to the conceptualization of the proposed sociotechnical framework for explainable AI in rehabilitation nursing.

Theoretical Concept	Key Contributions	Relevance to This Study
General Systems Theory [5]	Systems as integrated wholes; interdependence of components; input-processing-output; Emerging Properties.	The holistic approach of the study is oriented, with the framework being analyzed as an integral part of a complex health system, with emerging behaviors (e.g., trust) that arise from the interactions between AI and professionals.
Adaptive complex systems [6,7]	Multi-agent health systems; nonlinear interactions; unpredictable behavior; need for iterative and flexible change approaches.	The present research justifies the use of iterative methods (DSR) and contextual interventions. The implementation of the framework, therefore, requires continuous monitoring and constant adaptation to the real clinical environment, thus recognizing the unpredictability and variability of the rehabilitation context.
Sociotechnical Systems	Sociotechnical Systems Co-optimization of social and technical systems; user participation; mutual adaptation technology-organization; design centered on people and the context of Use.	It is imperative to involve rehabilitation nurses in the framework design process. The integration of the technological solution into existing processes and culture is essential to prevent resistance, workarounds, or adoption failures resulting from sociotechnical misalignment.
Contextual Explainability (xAI) [13–15]	AI explanations should be understandable, relevant, and tailored to the user and the scenario; Transparency is a multidimensional concept (technical, cognitive, organizational).	The present study aims to design a framework using explainable Artificial Intelligence (AI) with different levels and formats of explanation (visual graphics, user-directed natural language), adjustable to the nurse's experience and the stage of the clinical process, to maximize practical utility and trust in the system.

2.4. Integrated Theoretical Synthesis

Taken together, these theoretical perspectives converge toward a unified system-based view in which explainable artificial intelligence is not conceptualized as a standalone technological feature, but as an emergent sociotechnical phenomenon. General Systems Theory provides the ontological foundation for understanding healthcare as an interconnected whole; complexity science explains the nonlinear, adaptive dynamics through which AI interventions unfold; and sociotechnical systems theory offers design principles for aligning technological artifacts with human, professional, and organizational subsystems. This integrated perspective establishes the conceptual basis for the framework proposed in this study, positioning explainability as a system-level outcome arising from coherent interactions between algorithms, users, and contextual conditions, rather than as an intrinsic property of AI models alone.

3. Methodology

3.1. Design Science Research as a Systems-Oriented Inquiry Approach

Building on the conceptual framework and exploratory evaluation presented above, this section discusses the findings from a systems-oriented perspective. This study adopts DSR as a system-oriented inquiry methodology, aligned with sociotechnical systems theory and complexity science. Rather than treating DSR solely as an artifact-building approach, it is employed here as a form of systemic intervention, enabling iterative exploration and refinement within a complex adaptive healthcare environment. DSR is particularly suitable for this purpose, as it explicitly integrates theory-informed design, contextualized intervention, and reflective evaluation through iterative feedback cycles [16,17].

Consistent with systems thinking, the research is grounded in the assumption that knowledge about explainable AI emerges through interaction with the sociotechnical system, rather than from detached observation. In this sense, DSR is conceptually aligned with action research and Action Design Research, which emphasize co-creation, contextual sensitivity, and learning through intervention in real-world systems [18,19]. This positioning allows the study to address the inherent complexity, non-linearity, and adaptivity of healthcare systems.

3.2. Design Requirements

Based on the literature review and consultations with rehabilitation nursing specialists and domain experts in technological fields, a set of design requirements was identified to guide the development of the proposed framework, as summarized in Table 2.

Table 2. Design requirements for the proposed human-centered sociotechnical explainable AI framework.

Requirement	Description	Source
R1	The framework must provide access to relevant clinical and contextual data to support AI-based decision-making.	Literature + Clinical context
R2	The system must generate reliable AI-based predictions that support clinical reasoning without replacing professional judgment.	Literature
R3	The framework must ensure explainability through transparent and interpretable outputs, including justification of AI recommendations.	Literature + Expert input
R4	The system must support human interpretation by aligning AI explanations with clinical reasoning processes and user expertise.	Expert input
R5	The framework must facilitate clinical decision-making processes, including assessment, care planning, and intervention adjustment.	Clinical context
R6	The system must consider organizational context and workflow integration, including interprofessional collaboration and institutional constraints.	Literature + Expert input
R7	The framework must support continuous system adaptation through feedback mechanisms reflecting real-world clinical Use.	DSR principles

3.3. Iterative Cycles of Systemic Design and Evaluation

The study was conducted through iterative cycles of design, demonstration, and evaluation, reflecting the dynamics of complex adaptive systems. Each cycle functioned as a feedback loop, in which insights generated from interaction with the system informed subsequent refinements of the conceptual framework. Initial problem framing and requirement elicitation were informed by a review of the literature and by engagement with

rehabilitation nursing specialists, enabling the identification of systemic challenges related to explainability, trust, and decision-making in rehabilitation contexts.

Based on these requirements, a conceptual framework for human-centered explainable AI was developed, explicitly incorporating sociotechnical design principles such as user participation, contextualized explanations, and alignment with clinical workflows. Rather than aiming for technical optimization, the framework was designed to explore how explainability could emerge from coherent interactions between AI behavior, human interpretation, and organizational context.

To provide a clearer representation of the proposed approach, Figure 1 presents the human-centered sociotechnical explainable AI framework. The model highlights the interaction between technical components, human interpretation, and organizational context in supporting clinical decision-making processes in rehabilitation nursing.

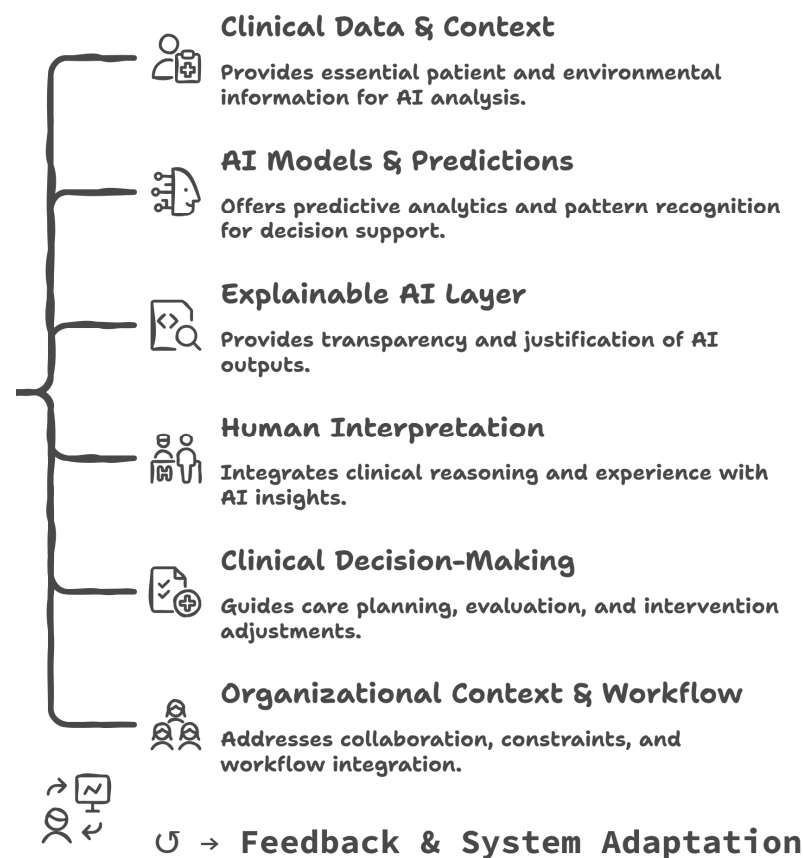


Figure 1. Human-centered sociotechnical explainable AI framework for clinical decision support in rehabilitation nursing.

The framework illustrates the interactions between clinical data, AI models, explainable mechanisms, human interpretation, and organizational context in supporting clinical decision-making.

The feedback loop represents continuous system adaptation based on clinical use and contextual factors.

3.4. Demonstration and Exploratory System-Level Evaluation

The framework was subsequently demonstrated and explored using realistic rehabilitation scenarios derived from post-stroke care pathways. These scenarios were used to support an exploratory, formative evaluation focused on system-level properties, including perceived usefulness, comprehensibility of explanations, trust, and alignment with professional reasoning. A panel of 144 rehabilitation nursing specialists from multiple

institutions participated in this phase, interacting with the framework through a simulated decision-support interface.

Importantly, this evaluation did not aim to validate clinical effectiveness or algorithmic performance. Instead, it sought to examine how explainability was experienced within the sociotechnical system, capturing perceptions related to human–AI interaction, cognitive support, and organizational fit. Quantitative questionnaire data and qualitative feedback were used to inform iterative refinement of the framework and to reflect on explainability as an emergent system-level outcome.

3.5. Ethical Considerations and AI Use Transparency

All study procedures complied with ethical and legal requirements. The participants provided their informed consent. In accordance with MDPI's guidelines on AI use, the authors declare that generative AI tools were used exclusively for linguistic refinement and proofreading. All conceptualization, theoretical framing, methodological design, data analysis, and interpretation were conducted by the authors, ensuring the scientific integrity of the work.

4. Results

4.1. Sample Characterization

The sample consisted of 144 Specialist Nurses in Rehabilitation Nursing (EEER), mostly female (70.9%), with a mean age of 45.6 years ($SD = 7.929$). The average length of professional practice was 22.79 years ($SD = 7.775$), while the specific experience as an EEER had a mean of 11.4 years ($SD = 6.523$), as illustrated in Figure 2.

Sex

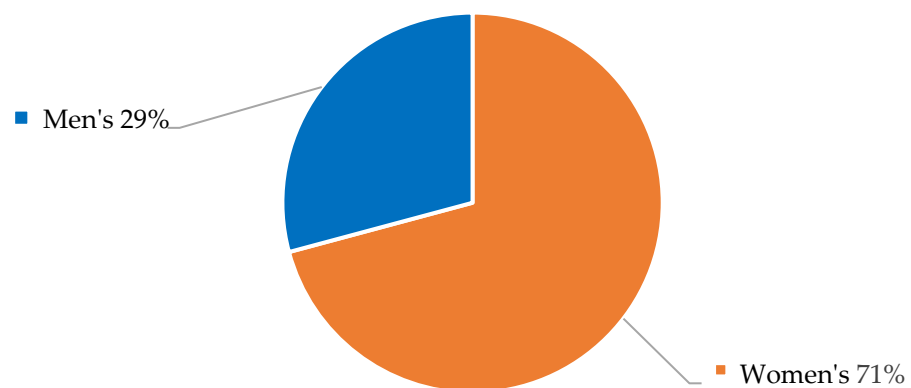


Figure 2. Gender distribution of the rehabilitation nursing specialists.

In terms of academic qualifications, 56.9% had a bachelor's degree, 40.3% a master's degree, and 2.8% a doctorate, as illustrated in Figure 3.

The clinical context of exercise was diverse, with differentiated care (57.6%) and primary health care (32.6%) standing out. The geographical distribution proportionally reflected the organization of the Portuguese health system, with a predominance of the North region (66.7%).

This composition confers robustness and representativeness on the evaluation carried out.

Academic Qualifications

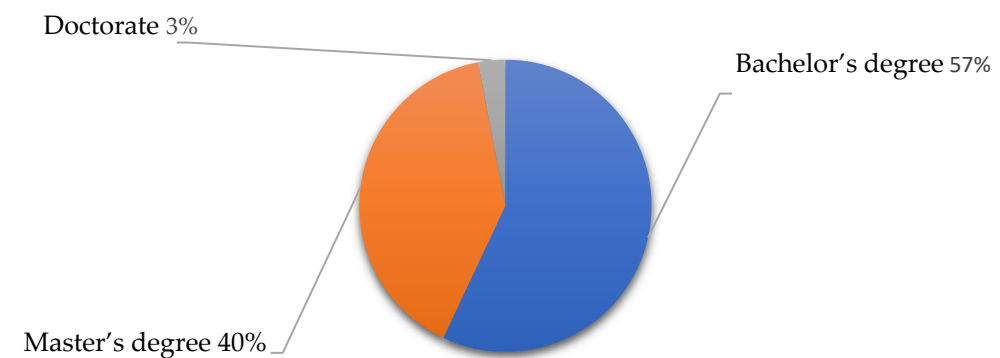


Figure 3. Distribution of participants according to academic qualifications.

4.2. Global EEER Opinion on the Use of xAI in Clinical Decision Making

The overall perception of the usefulness of xAI as a decision support tool was clearly positive, with an overall average of 3.89 on a Likert scale from 1 to 5, ranging from “agree” to “strongly agree”. This result shows a broad acceptance of technology and recognizes the value of explainability in contexts of high clinical complexity.

4.3. Evaluation of the Seven Dimensions Analyzed

As detailed in Table 3, the evaluation of the seven functional dimensions revealed a consistently positive assessment, with mean scores ranging from 3.75 to 4.13 on a 5-point Likert scale. The standard deviations (SDs), ranging from 0.68 to 0.91, indicate a strong consensus among the 144 specialists, with relatively low variance across all domains. The highest agreement was observed in ‘Research, Training, and Care Design’ ($M = 4.13$; $SD = 0.75$), suggesting that the framework is perceived as a robust tool for evidence-based practice and Cognitive support in rehabilitation nursing. Conversely, organizational domains such as ‘Quality Management’ and ‘Continuity of Care’ ($M = 3.75$) showed slightly higher variance, reflecting the inherent complexity of institutional workflows.

Table 3. Psychometric properties and descriptive analysis of the xAI framework ($n = 144$).

Dimension (Items)	Mean (M)	Std. Dev (SD)	Cronbach's α
1. Evaluation and Diagnosis (3)	4.02	0.8	0.92
2. Planning and Execution (6)	3.93	0.75	0.92
3. Results and Final Evaluation (9)	3.94	0.68	0.94
4. Quality Management (7)	3.75	0.76	0.92
5. Communication & Interprofessional Coord. (2)	3.78	0.82	0.80
6. Continuity of Care (2)	3.75	0.91	0.87
7. Research, Training, and Care Design (2)	4.13	0.75	0.89
Global Scale (31 items)	3.89	0.7	0.98

The internal consistency of the evaluation instrument was assessed using Cronbach's alpha (α). The global scale (31 items) demonstrated excellent reliability ($\alpha = 0.98$), with all seven dimensions yielding coefficients between 0.80 and 0.94, confirming high internal consistency across all theoretical domains.

4.4. Highest Appraisal: Cognitive and Clinical Support

The highest levels of acceptance were observed in the dimensions of Research, Training, and Care Design ($M = 4.13$; $SD = 0.75$) and Evaluation and Diagnosis ($M = 4.02$; $SD = 0.80$).

These results suggest that Rehabilitation Nursing Specialists perceive xAI as a robust tool for mediating complex cognitive tasks, such as critical thinking and longitudinal care planning. In this context, the explanatory nature of the framework is experienced as a reinforcement of professional autonomy and a facilitator of evidence-based practice, rather than a replacement for clinical judgment. The comparatively low standard deviations in these domains reflect a strong professional consensus: xAI provides meaningful cognitive offloading during high-stakes clinical reasoning.

4.5. Intermediate Appraisal: Operational Implementation

Dimensions related to Results and Evaluation ($M = 3.94$; $SD = 0.68$) and Planning and Intervention ($M = 3.93$; $SD = 0.75$) received intermediate appraisals. While these scores remain firmly positive, the slightly lower intensity of agreement compared to the cognitive domains suggests that the utility of xAI is perceived as more impactful during the analysis and decision-making phases than in the direct, operational execution of care. From a systems perspective, this indicates that while the framework effectively supports the sense-making process, its integration into the physical and procedural performance of nursing interventions may be moderated by deeply ingrained manual routines and existing technical limitations at the bedside.

4.6. Lower Appraisal: Organizational and Interprofessional Complexity

The dimensions with the lowest—yet still positive—appraisals were Communication and Interprofessional Coordination ($M = 3.78$; $SD = 0.82$), Continuity of Care ($M = 3.75$; $SD = 0.91$), and Quality Management ($M = 3.75$; $SD = 0.76$). These areas involve multiple stakeholders and complex institutional interdependencies, which participants implicitly identified as potential bottlenecks for xAI integration. The higher variance ($SD = 0.91$) observed in Continuity of Care reflects the diverse organizational maturity and varying interoperability challenges across different clinical settings. These findings underscore a critical design requirement for future systems: the need for collaborative explainability, where xAI outputs must be tailored not only to the individual nurse but also to the shared requirements of the interdisciplinary team and the broader institutional workflow.

4.7. Interpretative Synthesis of the Results

The empirical data demonstrate that the initial iteration of the xAI framework achieved high levels of systemic acceptance among the 144 specialists, as evidenced by a global mean of 3.89 ($SD = 0.70$) and an excellent internal consistency ($\alpha = 0.98$). This robust statistical support confirms that xAI is perceived as a highly viable intervention in rehabilitation care, particularly for cognitively demanding tasks with significant clinical impact. The most positive appraisals were observed in dimensions such as Research, Training, and Care Design ($M = 4.13$) and Evaluation and Diagnosis ($M = 4.02$), suggesting that the framework effectively supports professional autonomy and evidence-based practice. Conversely, organizational and interprofessional domains—such as Continuity of Care ($M = 3.75$) and Quality Management ($M = 3.75$)—emerged as priority areas for sociotechnical optimization. The higher variability observed in these dimensions (SD up to 0.91) is not merely a statistical artifact, but a direct reflection of the structural and regional differences in clinical contexts, a phenomenon characteristic of complex adaptive systems. These findings highlight a critical sociotechnical gap: while the framework provides strong cognitive support for the individual clinician, its broader scalability depends on deeper integration with institutional workflows and interoperable systems. Even so, the consistent pattern of positive agreement across all 31 items reinforces the framework's stability and its potential to serve as a human-centered mediator in complex clinical decision-making.

5. Discussion

5.1. Explainability as a System-Level Outcome

The findings of this study support the view that explainable artificial intelligence in healthcare should be understood as a sociotechnical and systemic phenomenon rather than as a purely technical challenge. When interpreted through systems theory and complexity science, perceived usefulness, acceptance, and trust appear not as isolated user responses, but as system-level outcomes shaped by interactions between AI behavior, human sense-making, and organizational context. In this study, explainability became meaningful when explanations were perceived as clinically relevant, cognitively accessible, and aligned with professional reasoning and workflow structures. This pattern is consistent with recent MDPI evidence indicating that the acceptance and perceived usefulness of xAI-based clinical decision support systems depend not only on algorithmic accuracy but also on usability, workflow integration, and the alignment of explanations with clinical practice [20]. From a systems perspective, the positive reception of the proposed framework by Rehabilitation Nursing Specialists does not indicate the success of a technological artifact per se, but rather a systemic fit between the designed intervention and the existing clinical ecosystem. This finding reinforces the argument that explainability cannot be reduced to interpretability techniques alone but must be understood as a relational and situated phenomenon shaped by sociotechnical interactions.

From a systems perspective, the positive reception of the proposed framework by rehabilitation nursing specialists does not indicate the success of a technological artifact per se, but rather a systemic fit between the designed intervention and the existing clinical ecosystem. This finding is consistent with sociotechnical systems theory, which emphasizes that system performance and sustainability depend on the co-optimization of social and technical subsystems [9,10]. In this study, explainability emerged as a meaningful system-level outcome when explanations were contextually relevant, cognitively accessible, and aligned with professional reasoning and workflow structures.

But by the coherence of the overall sociotechnical configuration. This reinforces the argument that explainability cannot be reduced to interpretability techniques alone but must be understood as a relational and situated phenomenon. Such an interpretation extends prior work in xAI by explicitly embedding explainability within systems theory, thereby addressing a gap identified in both systems science and human–AI interaction research.

5.2. Trust, Emergence, and Human–AI Complementarity

Importantly, trust in AI-supported recommendations was not driven solely by transparency of algorithmic logic. Instead, trust emerged from the coherence of the broader sociotechnical configuration in which the AI system was embedded. Participants consistently described AI-generated explanations as cognitive support rather than authoritative prescriptions. From a systems perspective, this behavior reflects adaptive regulation, in which professionals retain judgment while selectively integrating technological input. This pattern suggests a complementary human–AI relationship, in which explainability supports reflection, learning, and sense-making without displacing human agency.

This finding is particularly relevant in light of concerns regarding automation bias and over-reliance on AI. Rather than observing such effects, the system configuration explored in this study appeared to foster a complementary human–AI relationship, in which explainability supported reflection, learning, and sense-making. This balance suggests that when sociotechnical design principles are appropriately applied, explainable AI can enhance system resilience by reinforcing, rather than undermining, professional expertise.

Moreover, the iterative refinement of the framework through feedback loops aligns with the behavior of complex adaptive systems, where interventions are continuously

adjusted in response to local conditions and emergent effects. This reinforces the suitability of Design Science Research as a systemic inquiry methodology, capable of accommodating nonlinear dynamics and evolving system states.

5.3. Design Science Research as Systemic Intervention

Beyond its empirical findings, the study also offers methodological insights into how Design Science Research can be enacted within complex sociotechnical systems. The iterative refinement of the proposed framework reflects the behavior of complex adaptive systems, where interventions evolve through feedback and local adaptation rather than linear implementation. In this respect, Design Science Research proved suitable as a systems-oriented inquiry methodology, enabling theory-informed exploration while remaining sensitive to contextual dynamics. Rather than seeking technical optimization, the framework was refined to explore how explainability could emerge through coherent interactions among technological, human, and organizational elements.

First, it advances the conceptualization of explainable AI by framing explainability as an emergent sociotechnical property, rather than as an intrinsic attribute of AI models. This perspective extends General Systems Theory and sociotechnical systems theory into the domain of AI-supported decision making.

Second, the proposed framework operationalizes systems-based design principles, such as user participation, contextual alignment, and iterative adaptation, within a contemporary digital health context. In doing so, it demonstrates how abstract systems concepts can be translated into actionable design guidance without resorting to reductionist or technology-centric solutions.

Third, the study illustrates how Design Science Research can be enacted as a systemic intervention methodology, supporting theory-informed exploration and learning within complex adaptive systems. This methodological contribution is particularly relevant for systems scholars seeking to study and influence real-world sociotechnical environments.

5.4. Implications for Systems Theory and Healthcare Practice

From a theoretical standpoint, this study contributes to systems research by conceptualizing explainable AI as an emergent sociotechnical property rather than an intrinsic feature of AI models. It extends systems theory and sociotechnical design principles into the domain of AI-supported clinical decision making and demonstrates how abstract systems concepts can inform actionable design guidance. At the practical level, the findings suggest that organizations aiming to integrate AI into healthcare should adopt a holistic systems-oriented approach, giving equal attention to sociotechnical alignment, professional sense-making, and organizational readiness, rather than focusing exclusively on algorithmic performance.

In the context of rehabilitation nursing, the framework highlights how explainable AI can support longitudinal and multidimensional decision processes without compromising the central role of the nurse. By embedding explanations within professional reasoning and clinical context, AI-supported systems may contribute to learning, consistency, and reflective practice, while preserving human accountability and ethical responsibility.

6. Limitations and Future Research

This study has several limitations that should be considered when interpreting the findings. The exploratory evaluation focused on perceived usefulness, trust, and system alignment within simulated clinical scenarios, rather than on measurable clinical outcomes. As such, the results should be interpreted as insights into system-level dynamics, rather than as evidence of clinical effectiveness.

Firstly, the evaluation is based on the perceptions of Specialist Nurses in Rehabilitation Nursing (EER), collected through a questionnaire. As in any study relying on participant perceptions, there is a possibility of social desirability bias, where responses may reflect what is perceived as professionally appropriate rather than actual practice.

Secondly, the demonstration of the framework was conducted using simulated and controlled scenarios. While this approach allows for systematic initial evaluation, it does not replace experimentation in real clinical settings. Therefore, the findings do not allow for conclusions regarding direct impact on the quality of care, but rather on perceived usefulness, clarity, and alignment with professional reasoning.

Additionally, although the sample of 144 participants provides a relevant size and some geographic diversity, all participants belong to the Portuguese context. This may limit the transferability of the results to other healthcare systems with different organizational structures, resources, or levels of digital maturity.

Finally, version 1.0 of the framework has not yet been integrated with real clinical information systems. This limits the evaluation of aspects such as interoperability, real-time performance, and adaptation to complex organizational workflows. These aspects are critical for future development and practical implementation.

Future research should focus on longitudinal and real-world implementations of explainable AI frameworks to examine how system-level properties evolve and how continuous adaptation unfolds in operational environments. Further studies should also explore scalability, interoperability with existing health information systems, and the transferability of the proposed framework to other clinical domains.

These limitations do not compromise the validity of the findings but provide important directions for future research and for the progressive development of the framework.

7. Conclusions

Version 1.0 of the human-centered explainable AI (xAI) framework suggests feasibility, relevance, and potential to support clinical decision-making in Rehabilitation Nursing. The evaluation conducted with 144 Rehabilitation Nursing Specialists revealed a globally positive perception ($M = 3.90/5$), with greater appreciation in cognitively demanding domains such as “Research, Training and Care Design” and “Evaluation and Diagnosis”. These findings indicate that xAI is perceived not as an isolated technical feature, but as a meaningful support tool that can reinforce professional reasoning, promote reflective practice, and enhance the clarity and transparency of AI-supported recommendations.

Although organizational and interprofessional dimensions also received positive evaluations, they revealed comparatively lower values, highlighting persistent challenges related to workflow integration, coordination across professional groups, and the complexity of institutional contexts. These aspects reinforce the understanding that the alignment between AI behavior, human interpretation, and organizational structures shapes explainability.

Future development of the framework should therefore prioritize three complementary directions:

- (i) refining explanatory mechanisms better to accommodate different professional profiles and levels of expertise;
- (ii) strengthening the availability of transparent, auditable documentation, including version histories, decision criteria, and underlying reasoning;
- (iii) promoting effective integration into real-world institutional workflows, supported by interoperable APIs and mechanisms for ongoing monitoring of performance and equity.

These findings should be interpreted in light of the study’s scope, which is based on a conceptual framework and perception-based evaluation rather than on evidence of clinical effectiveness in real-world settings.

Overall, the framework shows promising potential to support clinical decision-making in Rehabilitation Nursing by providing contextually aligned, human-centered explanatory outputs that complement, rather than replace, professional judgment.

Author Contributions: Conceptualization, F.P.R., F.P. and A.S.; methodology, F.P.R., A.S. and T.R.; validation, A.S., T.R., F.P. and R.P.; formal analysis, F.P.R.; investigation, F.P.R.; resources, F.P., A.S. and F.P.R.; data curation, F.P.R.; writing—original draft preparation, F.P.R.; writing—review and editing, F.P.R., A.S., T.R., F.P. and R.P.; visualization, F.P.R.; supervision, F.P. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki. No approval by the Institutional Ethics Committee was necessary, as all data were collected anonymously from capable, consenting adults. The data are not traceable to participating individuals. The procedure complies with the General Data Protection Regulation (GDPR). Institutional Review Board approval was not required for this study owing to its observational design; Moreover, neither disease course nor patient treatment was involved in or affected by the conduct of this study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Acknowledgments: The authors would like to thank the rehabilitation nursing specialists who participated in the study for their valuable contributions.

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

References

1. Kostopoulos, G.; Davrazos, G.; Kotsiantis, S. Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. *Electronics* **2024**, *13*, 2842. [\[CrossRef\]](#)
2. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
3. McCradden, M.D.; Joshi, S.; Anderson, J.A.; London, A.J. A normative framework for artificial intelligence as a sociotechnical system in healthcare. *Patterns* **2023**, *4*, 100864. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Salimparsa, M.; Sedig, K.; Lizotte, D.; Abdullah, S.; Chalabianloo, N.; Muanda, F. Explainable AI for clinical decision support systems: Literature review, key gaps, and research synthesis. *Informatics* **2025**, *12*, 119. [\[CrossRef\]](#)
5. von Bertalanffy, L. *General System Theory: Foundations, Development, Applications*; George Braziller: New York, NY, USA, 1968.
6. Plsek, P.E.; Greenhalgh, T. Complexity science: The challenge of complexity in health care. *BMJ* **2001**, *323*, 625–628. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Braithwaite, J.; Churrua, K.; Long, J.C.; Ellis, L.A.; Herkes, J. When complexity science meets implementation science: A theoretical and empirical analysis of systems change. *BMC Med.* **2018**, *16*, 63. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Begun, J.W.; Zimmerman, B.; Dooley, K.J. Health care organizations are complex adaptive systems. In *Advances in Health Care Organization Theory*; Mick, S.M., Wyttenbach, M.E., Eds.; Jossey-Bass: San Francisco, CA, USA, 2003; pp. 253–288.
9. Trist, E.L.; Bamforth, K.W. Some social and psychological consequences of the longwall method of coal-getting. *Hum. Relat.* **1951**, *4*, 3–38. [\[CrossRef\]](#)
10. Cherna, A. The principles of sociotechnical design. *Hum. Relat.* **1976**, *29*, 783–792. [\[CrossRef\]](#)
11. Sittig, D.F.; Singh, H. A new socio-technical model for studying health information technology in complex adaptive healthcare systems. *Qual. Saf. Health Care* **2010**, *19*, i68–i74. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Eason, K. Afterword: The past, present, and future of sociotechnical systems theory. *Appl. Ergon.* **2014**, *45*, 213–220. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ehsan, U.; Saha, K.; De Choudhury, M.; Riedl, M.O. Charting the sociotechnical gap in explainable AI: A framework to address the gap in xAI. *Proc. ACM Hum.-Comput. Interact.* **2023**, *7*, 34. [\[CrossRef\]](#)

15. Rätz, T.; Pahud de Mortanges, A.; Reyes, M. Explainable AI in medicine: Challenges of integrating xAI into the future clinical routine. *Front. Radiol.* **2025**, *5*, 1627169. [[CrossRef](#)] [[PubMed](#)]
16. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. *MIS Q.* **2004**, *28*, 75–105. [[CrossRef](#)]
17. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A Design Science Research Methodology for Information Systems Research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77. [[CrossRef](#)]
18. Checkland, P.; Holwell, S. Action research: Its nature and validity. *Syst. Pract. Action Res.* **1998**, *11*, 9–21. [[CrossRef](#)]
19. Sein, M.K.; Henfridsson, O.; Purao, S.; Rossi, M.; Lindgren, R. Action design research. *MIS Q.* **2011**, *35*, 37–56. [[CrossRef](#)]
20. Abbas, Q.; Jeong, W.; Lee, S.W. Explainable AI in clinical decision support systems: A meta-analysis of methods, applications, and usability challenges. *Healthcare* **2025**, *13*, 2154. [[CrossRef](#)] [[PubMed](#)]

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.