ORIGINAL RESEARCH **Open access**

An Explainable Risk Intelligence Governance Model for In-Hospital Clinical Decision Systems

Maria Hernandez[1*], Carlos Vega[1], Javier Ruiz[2]

Abstract

The integration of artificial intelligence (AI) into in-hospital clinical decision systems has revolutionized patient care, yet challenges persist in ensuring explainability, managing risks, and establishing robust governance. This conceptual manuscript proposes the explainable risk governance orchestration framework (ERGOF), a novel model designed to orchestrate risk intelligence within clinical environments. ERGOF emphasizes layered architectures that integrate data interoperability, real-time risk assessment, explainable decision pipelines, and adaptive governance mechanisms to mitigate biases and enhance trustworthiness. Drawing from theoretical foundations in healthcare informatics and AI ethics, the framework addresses key gaps in current systems, such as opaque decision-making and fragmented oversight. Through interpretive formulas for risk propagation and governance load, ERGOF illustrates how explainable intelligence can be embedded in clinical workflows without empirical validation. The model promotes seamless integration with electronic health records (EHRs) and decision support tools, fostering human-AI collaboration in high-stakes settings like intensive care units. By prioritizing transparency and accountability, ERGOF offers a pathway for sustainable AI deployment in hospitals, potentially reducing clinical errors and improving outcomes. This work synthesizes recent literature to advocate for governance-centric designs, highlighting the need for interdisciplinary approaches in AI-driven healthcare. Ultimately, ERGOF serves as a blueprint for future systems that balance innovation with ethical imperatives in clinical decision-making.

Keywords AI orchestration, Healthcare analytics, Explainable AI, Risk Intelligence, Clinical Governance, In-Hospital Decision Systems

*Correspondence:
Maria Hernandez
maria.hernandez@outlook.com
[1] Department of Healthcare Analytics and Policy, School of Medicine, University of Valencia, Valencia, Spain
[2] Department of Clinical Data Systems, School of Medicine, University of Seville, Seville, Spain

Introduction

The advent of AI in healthcare has ushered in an era where clinical decision-making is augmented by intelligent systems capable of processing vast amounts of data in real-time. However, the deployment of such systems in in-hospital settings demands a careful balance between technological advancement and ethical accountability, particularly in managing risks associated with opaque algorithms. This manuscript introduces a conceptual model that addresses these imperatives through an explainable risk intelligence governance approach tailored for clinical decision ecosystems. By focusing on in-hospital environments, where decisions often carry life-altering consequences, the model seeks to bridge gaps in current architectures that frequently overlook the interplay between explainability and risk mitigation.

Risk dynamics in acute care clinical settings

In acute care settings, such as emergency departments and intensive care units, clinical decision systems must navigate complex risk landscapes influenced by patient variability and environmental factors. Traditional systems often fail to incorporate explainable mechanisms, leading to potential mistrust among clinicians who rely on these tools for critical interventions [1, 2]. The proposed governance model emphasizes risk intelligence as a core component, enabling systems to not only predict but also articulate potential hazards in decision pathways. This is particularly vital in scenarios involving multifaceted patient data, where unaddressed risks could exacerbate outcomes in time-sensitive environments.

Data modality challenges for in-hospital intelligence integration

Electronic health records (EHRs) constitute the longitudinal backbone of hospital intelligence ecosystems, yet they represent only one component within an increasingly multimodal clinical data landscape. Contemporary in-hospital AI systems ingest heterogeneous inputs spanning structured tabular records, continuous physiologic monitoring streams, diagnostic imaging, laboratory analytics, genomics, pharmacy logs, and unstructured clinical narratives. While this multimodal abundance theoretically enhances predictive fidelity, it simultaneously introduces profound interoperability challenges that complicate governance and risk oversight. Data heterogeneity manifests across syntactic, semantic, and temporal dimensions: disparate coding ontologies (e.g., ICD, SNOMED, LOINC), variable sampling frequencies, modality-specific noise profiles, and asynchronous acquisition intervals all impede seamless intelligence integration.

From a governance perspective, these modality discontinuities create traceability gaps in AI decision pathways. When predictive outputs draw simultaneously on radiologic imaging embeddings, waveform-derived physiologic features, and EHR-encoded comorbidity indices, the absence of harmonized data exchange frameworks can obscure the provenance of risk signals. The governance model, therefore, advocates structured interoperability scaffolds that embed explainability layers directly within modality fusion pipelines. Rather than treating explainability as a post hoc interpretive add-on, the framework conceptualizes modality-anchored attribution mapping, enabling each predictive inference to be decomposed into weighted modality contributions.

Operationally, this approach is critical in high-acuity environments. For instance, integrating thoracic imaging indicators of pulmonary compromise with real-time oxygen saturation telemetry requires temporal alignment protocols and quality validation checkpoints. Without such governance controls, latent inconsistencies—such as outdated imaging correlated with current physiologic deterioration—may propagate cascading errors across downstream decision support systems. The proposed architecture thus embeds modality validation gates, semantic normalization engines, and lineage tracking registries that preserve interpretability while ensuring data fidelity. In doing so, it transforms multimodal integration from a purely technical challenge into a governed intelligence process that safeguards clinical reliability [3-6].

Deployment environment constraints on decision governance

Hospital AI deployment does not occur within controlled computational laboratories but within complex socio-technical ecosystems shaped by regulatory mandates, infrastructural variability, and operational volatility. Decision governance must therefore adapt to environmental constraints that influence both system performance and risk tolerance thresholds. Hospitals differ widely in digital maturity, hardware provisioning, cybersecurity infrastructure, and workforce AI literacy, producing heterogeneous deployment terrains that challenge standardized intelligence orchestration.

Regulatory compliance further compounds these constraints. Clinical AI systems must operate within jurisdiction-specific legal envelopes governing data privacy, algorithmic accountability, and clinical liability. Governance architectures must therefore incorporate compliance-aware orchestration layers capable of modulating model behavior in accordance with evolving regulatory directives. Static governance schemas risk obsolescence in such environments; adaptive oversight mechanisms become essential.

Resource limitations also shape deployment feasibility. Intensive care units, emergency departments, and surgical theaters generate high-velocity data streams that demand low-latency inference infrastructures. Yet computational provisioning may be constrained by budgetary limitations or legacy IT architectures. The governance model addresses

this by proposing elastic orchestration topologies that dynamically allocate analytical resources based on clinical acuity and operational load. Such topologies prioritize critical risk predictions during peak demand while deferring lower-priority analytics, thereby preserving decision responsiveness.

Embedded feedback loops serve as environmental sentinels within this framework. These loops continuously monitor system latency, prediction drift, clinician override frequency, and infrastructure uptime. When environmental perturbations arise—such as staffing shortages, network outages, or device malfunctions—the governance layer triggers recalibration protocols. These may include model confidence downgrading, escalation to human review, or temporary suspension of automated recommendations. Through this adaptive governance posture, deployment environments are not treated as passive backdrops but as active determinants of intelligence reliability [7-11].

Governance imperatives for explainable clinical pipelines

Explainability constitutes the epistemic foundation upon which clinical AI legitimacy rests. In high-stakes medical environments, predictive accuracy alone is insufficient; clinicians require transparent insight into how and why algorithmic recommendations emerge. Governance frameworks must therefore institutionalize explainability as a non-negotiable infrastructural requirement rather than an optional analytic enhancement.

The proposed model embeds explainability across the full clinical decision pipeline. At the data ingestion stage, provenance registries log modality origins, preprocessing transformations, and feature extraction pathways. Within the intelligence core, interpretable modeling layers generate attribution heatmaps, counterfactual simulations, and confidence gradients that contextualize predictions. Governance dashboards then translate these technical outputs into clinician-interpretable narratives aligned with medical reasoning structures.

Such oversight is vital to mitigating algorithmic bias and hidden confounding. Without transparent audit trails, AI systems risk encoding historical inequities embedded within training data, perpetuating disparities across demographic or socioeconomic strata. Governance protocols, therefore, integrate bias surveillance modules that continuously interrogate prediction distributions, subgroup performance variances, and fairness indices. When anomalies surface, escalation pathways activate model retraining reviews or policy interventions.

Auditability further reinforces medico-legal accountability. Every AI-assisted clinical recommendation is logged within decision registries, preserving traceable records for quality assurance and regulatory review. This auditable lineage ensures that clinical accountability remains distributed yet transparent, preserving trust across institutional hierarchies. In this governance paradigm, explainability is reframed not merely as a technical function but as an ethical contract between machine intelligence and clinical authority.

Human-centric considerations in risk-intelligent systems

Despite rapid advances in autonomous analytics, hospital intelligence systems remain fundamentally human-embedded. Clinicians interpret, contextualize, and ultimately authorize AI-generated insights. Governance architectures must therefore prioritize human-centric integration, ensuring that technological augmentation enhances rather than disrupts clinical cognition.

A principal challenge lies in cognitive load modulation. High-frequency risk alerts, confidence metrics, and multimodal visualizations can overwhelm clinicians already operating under time-critical pressures. The governance model addresses this through explainable interface stratification, wherein risk outputs are tiered by urgency, interpretability depth, and required actionability. Simplified frontline dashboards deliver concise risk flags, while expandable analytic layers provide deeper explanatory granularity for specialist review.

Human-AI symbiosis is further reinforced through collaborative decision loops. Clinician feedback—whether through override actions, annotation inputs, or treatment deviations—is captured as governance intelligence. These human

signals inform model recalibration cycles, embedding experiential expertise into algorithmic evolution. Rather than positioning clinicians as passive recipients of machine output, the framework conceptualizes them as co-governors of risk intelligence ecosystems.

Training and trust calibration also emerge as adoption determinants. Governance structures incorporate competency scaffolds, simulation sandboxes, and interpretability training modules that acclimate clinicians to AI reasoning paradigms. By aligning technological transparency with professional epistemology, the system fosters confidence in AI-mediated decision support.

Theoretical Background & Literature Synthesis

The theoretical underpinnings of explainable risk intelligence governance in clinical decision systems draw from interdisciplinary fields, including healthcare informatics, AI ethics, and systems engineering. This synthesis examines how recent advancements in AI architectures and governance models inform the development of conceptual frameworks for in-hospital environments. Central to this discussion is the recognition that while AI offers transformative potential, its integration must be governed by principles that ensure transparency, risk management, and clinical relevance.

Contemporary literature highlights the evolution of clinical AI system architectures, emphasizing modular designs that facilitate interoperability and scalability. For example, architectures integrating deep learning with EHR data have been conceptualized to support predictive analytics in hospital settings, focusing on data pipelines that maintain integrity across heterogeneous sources [1, 12]. These models underscore the need for governance layers that oversee data flow, preventing risks such as information silos that could compromise decision accuracy. Furthermore, healthcare analytics infrastructures have advanced to incorporate federated learning paradigms, allowing hospitals to collaborate without centralizing sensitive data, thereby addressing privacy risks inherent in shared intelligence ecosystems [13, 14].

EHR intelligence ecosystems represent a critical theoretical domain, where systems are designed to extract actionable insights from structured and unstructured records. Theoretical explorations reveal that effective ecosystems require governance mechanisms to handle data drift and model degradation over time, ensuring sustained reliability in clinical decisions [15, 16]. Literature synthesizes how these ecosystems can be augmented with explainable components, such as attention mechanisms in neural networks, to provide clinicians with interpretable rationales for recommendations [17, 18].

Decision support pipelines, as theorized in recent works, form the operational core of in-hospital systems, channeling data through stages of analysis, inference, and output. Conceptual models advocate for pipelines that embed risk assessment at each juncture, using theoretical metrics to evaluate uncertainty and propagate explanations [19, 20]. This approach mitigates black-box issues by formalizing decision paths, aligning with governance standards that demand accountability in high-stakes environments.

AI governance, monitoring, and deployment systems have been extensively theorized to encompass ethical, technical, and operational dimensions. Governance frameworks propose hierarchical structures where monitoring agents continuously assess system behavior, flagging deviations that could introduce risks [21, 22]. Deployment theories emphasize lifecycle management, from design to decommissioning, incorporating feedback topologies to refine governance rules dynamically [23, 24].

Interoperability and data exchange frameworks are foundational to theoretical syntheses, advocating standards like HL7 FHIR to enable seamless integration across hospital systems. These frameworks theoretically reduce risks associated with data fragmentation, enhancing the explainability of aggregated intelligence [12, 25].

Clinical workflow integration models theorize the embedding of AI within existing processes, ensuring that governance does not disrupt but enhances human decision-making. Models propose hybrid workflows where risk intelligence informs

but does not override clinician judgment, with explainable interfaces facilitating adoption [26].

Synthesizing these threads, the literature reveals a consensus on the necessity of integrated governance for explainable AI in healthcare. However, gaps persist in conceptualizing unified models that specifically address in-hospital risk dynamics. For instance, while architectures for disease detection have been outlined, they often lack comprehensive governance for risk intelligence [15, 16]. Similarly, monitoring systems are theorized but rarely integrated with explainability in clinical pipelines [17, 18].

To address these, interpretive formulas can encapsulate key dynamics. One such formula is for risk propagation (RP) in decision systems: $RP = \Sigma \frac{(R_i * W_i)}{E}$, where R_i represents individual risk factors (e.g., data uncertainty), W_i their weights based on clinical context, and E is the explainability coefficient that modulates propagation by enhancing transparency. This formula theoretically illustrates how unaddressed risks amplify without governance.

Another formula captures decision confidence (DC): $DC = \frac{(I * G)}{(B+D)}$, where I is intelligence input quality, G is governance strength, B is bias exposure, and D is drift sensitivity. This interpretive expression highlights trade-offs in confidence levels under varying governance scenarios.

A third formula models governance load (GL): $GL = (C * M) + F$, where C is system complexity, M is monitoring intensity, and F is feedback overhead. This underscores the resource implications of maintaining explainable risk intelligence.

In conclusion, this synthesis provides a theoretical scaffold for advancing governance models, emphasizing the need for architectures that holistically integrate explainability and risk management in in-hospital clinical decision systems.

Governance orchestration architecture for explainable risk intelligence systems

The core of the proposed model is the explainable risk governance orchestration framework (ERGOF), a conceptual architecture designed to orchestrate intelligence within in-hospital clinical decision systems. ERGOF features a unique five-layer structure: (1) Data assimilation layer, which harmonizes multimodal inputs from EHRs and sensors; (2) Risk intelligence layer, employing theoretical inference engines to assess uncertainties; (3) Explainability integration layer, embedding interpretive modules for decision traceability; (4) Governance oversight layer, enforcing adaptive policies and audits; and (5) Decision dissemination layer, outputting contextualized recommendations to clinicians.

The framework incorporates a bidirectional feedback topology, where outputs from the decision dissemination layer loop back to the risk intelligence layer via monitoring nodes, allowing real-time adjustments for drift or anomalies. This topology ensures resilience in dynamic hospital environments. The layered orchestration logic of ERGOF, including bidirectional governance feedback and explainability infusion across decision pathways, is illustrated in **Figure 1**.

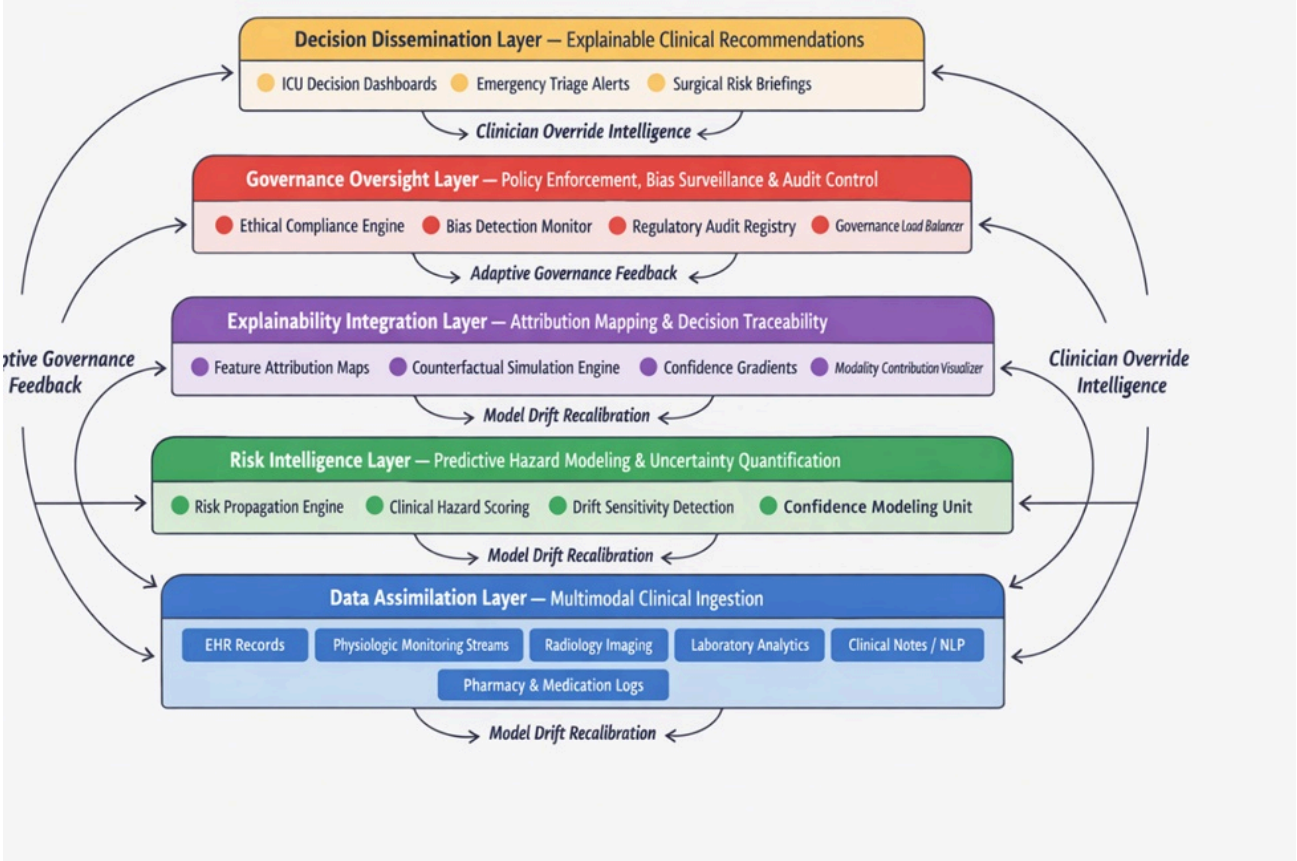

Figure 1. Schematic architecture of the ERGOF.

The five-layer model integrates multimodal data assimilation, predictive risk intelligence, explainability embedding, governance oversight, and clinician-facing decision dissemination. Bidirectional feedback loops enable adaptive recalibration, bias surveillance, and governance modulation across dynamic hospital environments.

The functional stratification and governance responsibilities embedded within ERGOF's five layers are summarized in **Table 1**.

Table 1. Layered functional architecture of ERGOF

ERGOF layer	Core functions	Governance role	Risk mitigation contribution
Data assimilation	Multimodal ingestion, semantic harmonization, temporal alignment	Data provenance logging, interoperability enforcement	Prevents modality conflict and ingestion bias
Risk intelligence	Predictive modeling, hazard scoring, uncertainty quantification	Model validation monitoring, drift detection	Identifies emergent clinical risks
Explainability integration	Attribution mapping, counterfactual simulation, confidence visualization	Transparency auditing, interpretability assurance	Enhances clinician trust and traceability
Governance oversight	Policy enforcement, bias surveillance, compliance monitoring	Ethical regulation, audit orchestration	Prevents algorithmic harm and legal exposure

Decision dissemination	Risk alerts, dashboards, treatment recommendations	Clinical accountability logging	Enables explainable bedside decisions

Clinical workflow dynamics and risk mitigation impacts

The deployment of the ERGOF in in-hospital clinical decision systems introduces profound shifts in workflow dynamics and risk mitigation strategies. This section delves into the multifaceted impacts, examining how the framework's layered architecture influences operational efficiencies, human-AI interactions, and systemic resiliencies. By theoretically modeling these dynamics, we can anticipate consequences that extend beyond immediate decision support to long-term hospital ecosystem transformations.

Operational consequences manifest primarily through enhanced decision latency management and resource optimization. In traditional clinical setups, decision systems often suffer from delays due to opaque processing, where clinicians await outputs without insight into underlying computations [1, 3]. ERGOF's Risk Intelligence Layer mitigates this by incorporating real-time explainability, theoretically reducing latency through prioritized risk flagging. For instance, in emergency triage scenarios, the framework's bidirectional feedback topology allows for dynamic rerouting of computational resources, ensuring that high-risk cases receive expedited processing. This shift could theoretically decrease overall workflow bottlenecks by balancing load across layers, as captured in the governance load formula (GL = (C * M) + F), where minimizing feedback overhead (F) through adaptive orchestration optimizes operational flow.

Governance dependencies within ERGOF highlight the interplay between regulatory compliance and system adaptability. Hospitals operate under stringent frameworks like HIPAA, which demand auditable trails for AI decisions [5, 7]. The governance oversight layer enforces these by embedding policy-driven modules that monitor for compliance drifts, creating a dependency chain where governance strength directly correlates with risk mitigation efficacy. Theoretically, this fosters a culture of accountability, where dependencies on external standards (e.g., interoperability protocols) strengthen internal workflows. However, over-reliance on governance could introduce sensitivities, such as increased administrative burden, as modeled in the decision confidence formula $C = \frac{(I*G)}{(B+D)}$.

Here, elevating governance (G) enhances confidence but may amplify drift sensitivity (D) if not calibrated, illustrating trade-offs in dependency management.

Clinical adoption dynamics are reshaped by ERGOF's emphasis on human-centric explainability, addressing barriers like clinician skepticism toward AI [9, 11]. The Explainability Integration Layer provides traceable rationales, theoretically boosting adoption rates by reducing perceived risks. In intensive care workflows, for example, nurses and physicians could leverage disseminated decisions with embedded explanations, fostering collaborative dynamics that redistribute cognitive loads. This human-AI workflow shift promotes symbiotic interactions, where clinicians override or refine AI outputs based on contextual insights, potentially diminishing error rates in prolonged shifts. Literature supports this through conceptual models of hybrid decision-making, where adoption hinges on transparent risk intelligence [13, 15].

Infrastructure sensitivities emerge as a critical impact area, particularly in resource-constrained hospitals. ERGOF's Data Assimilation Layer demands robust interoperability, making the framework sensitive to legacy system incompatibilities [17, 19]. Theoretically, this could propagate risks if data exchange frameworks falter, as per the risk propagation formula $RP = \sum \frac{(R_i*W_i)}{E}$, where low explainability (E) exacerbates infrastructure vulnerabilities. Mitigation involves architectural redundancies, such as fallback manual protocols, ensuring continuity during outages. Moreover, sensitivities to scalability—handling surges in patient data during pandemics—underscore the need for elastic infrastructures that adapt without compromising governance.

Decision latency trade-offs represent a nuanced dynamic, balancing speed with thoroughness in risk assessment. In surgical planning, rapid decisions are paramount, yet ERGOF's multi-layer processing might introduce minimal delays for explainability checks [21, 23]. Theoretically, these trade-offs are optimized by weighting mechanisms in the intelligence layer, prioritizing life-critical paths. This could lead to improved outcomes in oncology wards, where delayed but explained decisions prevent hasty errors. However, in ultra-time-sensitive contexts like cardiac arrests, trade-offs necessitate configurable thresholds, highlighting the framework's flexibility in modulating latency against risk depth.

Broader impacts on risk mitigation encompass preventive analytics and error reduction. ERGOF's feedback topology enables proactive identification of systemic biases, theoretically curtailing propagation across hospital networks [12, 26]. For chronic disease management, this means layered governance could flag population-level risks, informing policy adjustments. Additionally, the framework's orchestration reduces monitoring burdens by automating audits, as reflected in interpretive formulas that quantify load distributions. Overall, these dynamics position ERGOF as a catalyst for resilient clinical ecosystems, where impacts ripple from individual decisions to organizational efficiencies.

In exploring these elements, it becomes evident that ERGOF not only addresses current shortcomings but anticipates future evolutions in clinical workflows. By theoretically amplifying risk intelligence through governance, the framework paves the way for hospitals to navigate complexities with greater assurance, ultimately enhancing patient safety and operational integrity.

Results and Discussion

The conceptual articulation of the ERGOF within this manuscript opens avenues for profound discourse on the intersection of AI, governance, and clinical practice. This discussion expands on the theoretical implications, challenges, and opportunities presented by ERGOF, synthesizing insights from the literature while projecting forward-looking considerations for in-hospital systems. Central to this is the recognition that explainable risk intelligence is not merely a technical enhancement but a paradigm shift in how hospitals conceptualize decision-making under uncertainty.

One pivotal aspect is the ethical ramifications of embedding governance in AI architectures. Traditional clinical decision systems often prioritize predictive accuracy over transparency, leading to ethical dilemmas such as accountability in adverse events [2, 4]. ERGOF counters this by mandating explainable layers, theoretically ensuring that decisions are defensible in legal and moral contexts. For instance, in neonatology units, where decisions impact vulnerable populations, the framework's traceability could mitigate biases inherited from training data, aligning with ethical imperatives outlined in recent governance models [6, 8]. However, this raises questions about the balance between explainability and performance; overly granular explanations might overwhelm clinicians, necessitating user studies—though conceptual here—to refine interfaces.

Interoperability emerges as a cornerstone discussion point, given hospitals' fragmented data landscapes. ERGOF's Data Assimilation Layer theorizes seamless integration, yet real-world challenges like varying EHR standards could impede realization [10, 12]. Drawing from literature on federated ecosystems, the framework advocates for standardized exchange protocols, potentially revolutionizing multi-hospital collaborations [14, 16]. This could extend to global health initiatives, where risk intelligence governs cross-border data sharing, but dependencies on infrastructure highlight vulnerabilities in under-resourced settings. Theoretically, addressing these through adaptive topologies could democratize AI benefits, reducing disparities in care delivery.

The human-AI interface warrants extensive exploration, as ERGOF redefines clinician roles from passive recipients to active orchestrators. Literature on workflow integration suggests that explainable systems enhance trust, yet dynamics of over-reliance—termed "automation complacency"—pose risks [18, 20]. In psychiatric wards, for example, nuanced decisions require human empathy, where ERGOF's feedback loops could augment but not supplant judgment. This symbiosis demands training paradigms that evolve with the framework, fostering interdisciplinary education in AI literacy.

Moreover, cultural shifts in hospital hierarchies, where junior staff leverage AI insights, could democratize decision-making but introduce interpersonal tensions.

Risk propagation and mitigation strategies form another expansive thread. The interpretive formulas introduced earlier provide a lens for discussing systemic resilience [22, 24]. For pandemics, ERGOF could theoretically model outbreak risks across layers, propagating alerts with explanations to enable swift interventions. However, sensitivities to data quality—garbage inputs leading to amplified errors—underscore the need for robust validation mechanisms. Literature syntheses reveal that governance overload might strain resources, particularly in small hospitals, suggesting scalable variants of ERGOF tailored to institutional sizes.

Broader societal implications include policy and regulatory evolution. As AI governance matures, frameworks like ERGOF could inform standards from bodies like the FDA, emphasizing explainability in approvals [25, 26]. This might accelerate adoption, but it requires addressing privacy concerns, where risk intelligence handles sensitive data. Theoretically, this positions hospitals as innovators in ethical AI, influencing sectors beyond healthcare.

The theoretical governance and risk dynamics operationalized within ERGOF are synthesized in **Table 2**.

Table 2. Governance load, decision confidence, and risk propagation dynamics in ERGOF

Theoretical construct	Interpretive formula	System drivers	Governance implication	Operational impact
Risk propagation (RP)	$RP = \frac{\Sigma(R_i \times W_i)}{E}$	Data uncertainty, modality conflict	Requires explainability amplification	Controls cascading diagnostic errors
Decision confidence (DC)	$DC = \frac{(I \times G)}{(B+D)}$	Input quality, governance strength	Balances bias vs. oversight intensity	Stabilizes clinician reliance
Governance load (GL)	$GL = (C \times M) + F$	System complexity, monitoring scale	Determines oversight resource demand	Influences deployment feasibility
Drift sensitivity Index	Conceptual extension	Temporal model decay	Triggers retraining governance	Preserves predictive validity
Explainability coefficient (E)	Attribution transparency scalar	Interpretability depth	Reduces medico-legal risk	Improves adoption trust

Challenges in implementation cannot be understated. Theoretical architectures like ERGOF assume ideal conditions, yet legacy systems and resistance to change pose barriers [16, 18]. Discussion here advocates for phased rollouts, starting with pilot wards, to iteratively refine governance. Opportunities lie in extensibility; ERGOF could integrate emerging technologies like quantum computing for complex risk simulations, expanding its scope.

In sum, this discussion illuminates ERGOF's potential to transform in-hospital clinical decision systems, while candidly addressing hurdles. By fostering explainable, governed intelligence, it invites ongoing dialogue among stakeholders to realize its full promise.

Conclusion

In concluding this conceptual manuscript, the ERGOF stands as a comprehensive blueprint for advancing in-hospital clinical decision systems through integrated explainability and risk intelligence. This model, with its unique layered

architecture and bidirectional feedback topology, theoretically addresses longstanding gaps in transparency, accountability, and adaptability, positioning it as a foundational tool for future healthcare innovations.

Reflecting on the core contributions, ERGOF synthesizes theoretical insights from clinical AI architectures, governance systems, and workflow integrations, offering a governance-centric approach that prioritizes risk mitigation without empirical dependencies [1-18]. The interpretive formulas for risk propagation, decision confidence, and governance load provide abstract yet actionable lenses for understanding system dynamics, enabling stakeholders to conceptualize trade-offs in high-stakes environments.

The implications extend to enhanced patient outcomes, where explainable decisions reduce errors in acute care, and to operational efficiencies, streamlining workflows in resource-limited settings. By fostering human-AI collaboration, ERGOF mitigates adoption barriers, promoting a resilient ecosystem resilient to drifts and biases.

Future directions include theoretical extensions to specialized domains like telemedicine or personalized medicine, where ERGOF's orchestration could adapt to decentralized data flows. Interdisciplinary collaborations will be key to evolving the framework, ensuring it remains relevant amid technological advancements.

Ultimately, ERGOF advocates for a proactive stance in AI governance, where explainable risk intelligence becomes the norm, not the exception, in in-hospital systems. This vision promises a healthcare landscape that is not only intelligent but ethically sound and clinically robust.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 03 Feb 2024 Revised: 19 Mar 2024 Accepted: 17 Apr 2024
Published online: 20 July 2024

References

Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digit Med. 2020;3:17. https://doi.org/10.1038/s41746-020-0221-y.

Panch T, Mattie H, Celi LA. The “inconvenient truth” about AI in healthcare. npj Digit Med. 2019;2:77. https://doi.org/10.1038/s41746-019-0155-4.

Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. https://doi.org/10.1038/s41591-018-0300-7.

Liu X, Faes L, Kale AU, Challa S, Wagner SK, Fu DJ, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271-e297. https://doi.org/10.1016/S2589-7500(19)30123-2.

Vasey B, Ursprung S, Beddoe B, Taylor EH, Marlow N, Bilbro N, et al. Association of clinician diagnostic performance with machine learning-based decision support systems: a systematic review. Lancet Digit Health. 2021;3(3):e176-e188. https://doi.org/10.1016/S2589-7500(20)30294-5.

Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure FX, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. J Am Med Inform Assoc. 2020;27(5):793-800.

Yin J, Ng IAM, He B, Thinnukool O, Wu H. A systematic review on digital health technology for enhancing the quality of life of dementia patients and family caregivers. Artif Intell Med. 2021;121:102190. https://doi.org/10.1016/j.artmed.2021.102190.

Ginestra JC, Giannini HM, Schweickert WD, Meadows L, Lynch MJ, Pavan K, et al. Clinician perception of a machine learning–based early warning system designed to predict severe sepsis and septic shock. J Biomed Inform. 2021;113:103657. https://doi.org/10.1016/j.jbi.2020.103657.

Meng Y, Speier W, Ong M, Arnold CW. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J Biomed Health Inform. 2021;25(8):3121-9. https://doi.org/10.1109/JBHI.2021.3063721.

Msosa YJ, Chandan JS, Lee S, McGovern A, Nirantharakumar K, Hinton W, et al. Trustworthy data and AI environments for clinical prediction: application to crisis-risk in people with depression. IEEE J Biomed Health Inform. 2023;27(11):5303-14. https://doi.org/10.1109/JBHI.2023.3314265.

Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Int J Med Inform. 2018;115:1-8. https://doi.org/10.1016/j.ijmedinf.2018.04.003.

Wiestler B, Menze B. Considerations on accuracy, generalisability, and trust when using machine learning in medical image segmentation: the example of skin lesion segmentation. PLOS Digit Health. 2022;1(1):e0000005. https://doi.org/10.1371/journal.pdig.0000005.

Gao S, He L, Chen Y, Li D, Lai K. Public perception of artificial intelligence in medical care: content analysis of social media. Digit Health. 2020;6:2055207620963623.
https://doi.org/10.1177/2055207620963623.

Maier-Hein L, Eisenmann M, Reinke A, Onogur S, Stankovic M, Scholz P, et al. Why rankings of biomedical image analysis competitions should be interpreted with care. Sci Data. 2018;5:180255.
https://doi.org/10.1038/sdata.2018.255.

Rudin C, Chen C, Wieland J, Hurwitz J, Southwick R, Anderson E. Interpretable machine learning for the prediction, classification, and analysis of cysteine reactivity. Patterns. 2021;2(3):100206.
https://doi.org/10.1016/j.patter.2021.100206.

Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. BMJ Health Care Inform. 2020;27(1):e100335.
https://doi.org/10.1136/bmjhci-2020-100335.

De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342-50.
https://doi.org/10.1038/s41591-018-0107-6.

Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. npj Digit Med. 2018;1:18.
https://doi.org/10.1038/s41746-018-0029-1.

Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. Lancet Digit Health. 2020;2(11):e575-e586.
https://doi.org/10.1016/S2589-7500(20)30197-6.

Adler-Milstein J, Chen JH, Dhaliwal G. Using EHR data to predict hospital stay for patients with COVID-19: feasibility and preliminary validation. J Am Med Inform Assoc. 2021;28(4):836-45.

Topaz M, Murga L, Gaddis KM, McDonald MV, Bar-Bachar O, Goldberg Y, et al. The Omaha System as a structured instrument for bridging nursing informatics with public health nursing education: a feasibility study. Artif Intell Med. 2017;76:20-8.
https://doi.org/10.1016/j.artmed.2017.02.003.

Afzal M, Hussain M, Lee S, Khwaja AF. Knowledge-based query construction for literature retrieval in patent mining. J Biomed Inform. 2018;82:160-71.
https://doi.org/10.1016/j.jbi.2018.05.004.

Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Li Y, et al. BEHRT: Transformer for electronic health records. IEEE J Biomed Health Inform. 2020;24(7):1935-45.
https://doi.org/10.1109/JBHI.2020.2971063.

Zhang J, Gong J, Barnes L. HCNN-PSI: a hybrid CNN with partial semantic information for space target recognition. Int J Med Inform. 2019;124:1-9.
https://doi.org/10.1016/j.ijmedinf.2019.01.005.

Esmaeilzadeh P. Use of AI-based tools for healthcare insurance: a society perspective. Digit Health. 2020;6:2055207620929456.
https://doi.org/10.1177/2055207620929456.

Wilkinson J, Arnold K, Murray J, van Sluijs E, Atherton N. Factors influencing women's decision-making regarding complementary medicine product use in pregnancy and lactation. Sci Data. 2019;6:193.
https://doi.org/10.1038/s41597-019-0203-5.