

ORIGINAL RESEARCH

Open access

Governance-First Clinical Trial Eligibility Pre-Screening: A Responsible Automation Framework for Risk-Bounded Recruitment

Nguyen Thanh Huy^{1*}, Pham Quang Minh¹, Le Thi Bich²

Abstract

The integration of artificial intelligence (AI) into healthcare systems has revolutionized clinical processes. Yet, the deployment of AI for clinical trial eligibility pre-screening remains fraught with governance challenges, particularly in ensuring risk-bounded recruitment. This conceptual manuscript proposes a governance-first automation framework designed to mitigate ethical, operational, and regulatory risks in AI-assisted patient selection for clinical trials. By prioritizing governance mechanisms over algorithmic optimization, the framework establishes a structured architecture that incorporates interoperability standards, real-time monitoring, and decision support pipelines to facilitate responsible automation. Drawing on theoretical insights from clinical AI system architectures and healthcare analytics infrastructures, we outline a layered model that balances automation efficiency with risk containment, emphasizing feedback loops for continuous governance oversight. Key components include risk propagation modeling, interoperability protocols for electronic health records (EHR) integration, and governance load assessments to prevent overburdening clinical workflows. This approach addresses the need for equitable and safe AI deployment in high-stakes environments like clinical trials, where eligibility pre-screening must align with ethical standards and regulatory compliance. Through interpretive formulas capturing risk dynamics and decision confidence, the framework provides a blueprint for healthcare institutions to implement AI-driven recruitment without compromising patient safety or trial integrity. Ultimately, this governance-centric paradigm shifts the focus from mere technological integration to responsible orchestration, fostering trust in AI-enhanced clinical trial ecosystems.

Keywords EHR interoperability, Governance-first AI, Clinical trial eligibility, Risk-bounded recruitment, Responsible automation, Decision support architecture

*Correspondence:

Nguyen Thanh Huy

huy.nguyen@gmail.com

¹ Department of Health Informatics, Faculty of Medicine, Vietnam National University, Hanoi, Vietnam

² Department of Digital Health Systems, Faculty of Medicine, Can Tho University, Can Tho, Vietnam

Introduction

The advent of AI in healthcare has introduced transformative capabilities for automating complex processes, yet its application to clinical trial eligibility pre-screening demands a governance-first approach to ensure risk-bounded recruitment. This section explores the foundational imperatives for such systems, emphasizing the need for responsible automation that prioritizes ethical oversight and regulatory alignment in clinical settings.

Eligibility pre-screening in multi-site clinical environments

In multi-site clinical trial environments, eligibility pre-screening involves sifting through vast patient data to identify suitable candidates, a process that is traditionally manual and prone to inefficiencies. AI automation frameworks can streamline this by leveraging EHR intelligence ecosystems, but without governance-first principles, risks such as biased selection or data privacy

breaches escalate. For instance, interoperability frameworks must be embedded to enable seamless data exchange across disparate clinical sites, ensuring that pre-screening algorithms respect site-specific governance constraints [1, 2]. This subheading underscores how clinical settings with distributed infrastructures necessitate risk-bounded mechanisms to prevent recruitment disparities.

Data modalities shaping risk-bounded automation

Diverse data modalities, including structured EHR entries, unstructured clinical notes, and imaging data, form the backbone of AI-driven pre-screening. A responsible automation framework must govern these modalities to bound risks associated with data incompleteness or misinterpretation. Healthcare analytics infrastructures play a pivotal role here, facilitating the synthesis of multimodal data while incorporating monitoring systems to detect drift in data quality [3, 4]. By anchoring governance to data modalities, the framework mitigates risks in recruitment, ensuring that AI decisions are traceable and auditable in alignment with clinical trial protocols.

Deployment challenges in governance-constrained hospital ecosystems

Hospital ecosystems often operate under stringent governance constraints, such as HIPAA compliance and institutional review board (IRB) oversight, complicating the deployment of AI for eligibility pre-screening. Decision support pipelines must integrate with existing clinical workflows to avoid disrupting care delivery, while risk-bounded recruitment requires real-time auditing to flag potential ethical lapses [5, 6]. This aspect highlights the imperative for frameworks that orchestrate AI deployment in constrained environments, prioritizing governance to safeguard patient rights during automated screening.

Interoperability imperatives for responsible recruitment pipelines

Effective interoperability is crucial for responsible recruitment pipelines, enabling the flow of patient data from EHR systems to AI pre-screening modules. Without standardized data exchange frameworks, risks of incomplete eligibility assessments increase, potentially leading to inequitable trial participation [7, 8]. Governance-first approaches address this by embedding protocols like

FHIR for seamless integration, ensuring that automation remains bounded by interoperability standards and reduces governance load on clinical teams.

Ethical dimensions of AI-orchestrated pre-screening in trial settings

Ethical dimensions in AI-orchestrated pre-screening encompass fairness, transparency, and accountability, particularly in risk-sensitive trial settings. Responsible automation demands governance mechanisms that monitor for biases in recruitment algorithms, drawing on AI governance systems to enforce ethical boundaries [9, 10]. This subheading examines how clinical trial contexts amplify the need for frameworks that integrate ethical oversight into automation lifecycles, preventing risks from propagating through recruitment decisions.

Theoretical Background and Literature Synthesis

This section synthesizes theoretical underpinnings from recent literature on clinical AI architectures, healthcare analytics, and governance systems, providing a foundation for the proposed framework. By integrating insights from EHR ecosystems and decision support models, we highlight gaps in current approaches to eligibility pre-screening and advocate for governance-first automation.

Architectural foundations of clinical AI in eligibility assessment

Clinical AI system architectures have undergone significant conceptual maturation since the late 2010s, transitioning from isolated algorithmic modules to integrated, modular ecosystems capable of supporting high-stakes, context-aware decision-making in healthcare. When applied specifically to eligibility assessment for clinical trials, these architectures reveal a critical dependency on robust governance layers to manage the inherent uncertainties of patient selection, such as incomplete data, evolving inclusion/exclusion criteria, and potential propagation of selection biases across trial cohorts [11, 12].

Early theoretical models positioned clinical AI as primarily extractive—pulling structured and unstructured elements from electronic health records (EHRs) to match against trial protocols—yet contemporary syntheses emphasize modularity as a foundational principle. Modular designs

allow decomposition into distinct components: data ingestion, feature engineering, matching logic, risk evaluation, and output auditing. This decomposition facilitates targeted governance insertion at each juncture, ensuring that eligibility decisions remain traceable and contestable [11]. For example, architectures incorporating monitoring subsystems enable continuous theoretical evaluation of drift in patient-trial alignment, where deviations in data distributions (e.g., due to seasonal variations in disease presentation or site-specific coding practices) trigger governance alerts before erroneous inclusions or exclusions accumulate [12].

A pivotal architectural advancement lies in the integration of interoperability frameworks, which serve as the connective tissue between AI components and heterogeneous EHR ecosystems. Standards such as HL7 FHIR enable standardized, queryable representations of patient data, allowing theoretical models of controlled data flow that inherently bound recruitment risks. By constraining information exchange to governed endpoints, these frameworks prevent unchecked propagation of incomplete or mismatched eligibility signals through the pipeline, transforming what could be a brittle, ad-hoc matching process into a structured, auditable orchestration [13]. In theoretical multi-site trial scenarios, such interoperability layers support federated querying—where eligibility logic executes locally at each institution while aggregating governed summaries—thus preserving privacy while enabling scalable pre-screening without centralized data aggregation risks.

This architectural synthesis underscores a recurring theme: effective designs for clinical trial pre-screening must subordinate computational efficiency to governance primacy. Without embedded governance mechanisms—such as explainability wrappers, uncertainty quantification modules, and human-in-the-loop escalation pathways—architectures risk amplifying uncertainties rather than containing them. Literature consistently reveals that governance-agnostic modular systems, while performant in controlled settings, falter in real-world deployment due to unaddressed risk vectors like criterion ambiguity or demographic skew in training corpora [11, 12]. Therefore, forward-looking architectural foundations prioritize hybrid human-AI loops, where AI proposes candidate matches but governance layers enforce boundary conditions (e.g., minimum confidence thresholds, fairness audits) before advancement to clinician review. This governance-centric redesign not only mitigates recruitment uncertainties but

also aligns eligibility assessment with broader ethical imperatives of equitable trial participation.

Analytics infrastructures enabling risk-bounded data handling

Healthcare analytics infrastructures constitute the foundational substrate for scaling AI-driven processes in risk-bounded domains such as clinical trial recruitment. These infrastructures aggregate, process, and contextualize large-scale, heterogeneous data streams—ranging from structured EHR fields to unstructured clinical narratives and imaging metadata—while theoretically embedding governance safeguards to preserve data integrity throughout the lifecycle [2, 14].

Theoretical models drawn from multimorbidity analytics parallel eligibility pre-screening challenges: both require harmonization of disparate data sources under strict integrity constraints to avoid spurious associations or missed matches. Big data frameworks, when governance-integrated, enable ingestion pipelines that apply provenance tracking and quality scoring at intake, ensuring that downstream eligibility algorithms operate on verifiable, bounded inputs [2]. For instance, anomaly detection subsystems within these infrastructures can theoretically flag modality-specific inconsistencies—such as conflicting diagnosis codes across encounters or temporal gaps in lab results—before they distort patient-trial matching [14, 15]. This preemptive containment reduces the governance burden on downstream decision layers by filtering noise early.

Monitoring mechanisms embedded in analytics infrastructures further enable dynamic risk propagation assessments. Theoretical pipelines can compute drift metrics across data cohorts, detecting shifts in eligibility-relevant features (e.g., biomarker distributions or comorbidity patterns) that might invalidate prior matching assumptions [15]. Such monitoring supports adaptive recalibration without halting recruitment flows, maintaining operational continuity while bounding systemic risks. Literature synthesis highlights that governance-integrated analytics infrastructures prevent workflow overload by distributing computational and oversight demands: automated quality gates handle routine validations, reserving human expertise for high-uncertainty cases flagged by drift alerts [14].

Ultimately, these infrastructures shift the paradigm from reactive data cleaning to proactive risk-bounded handling. By theorizing governance as an intrinsic layer—rather than an external overlay—they enable sustainable automation of eligibility pre-screening, where large-scale processing enhances recruitment velocity without compromising the fidelity required for ethical trial conduct.

EHR intelligence ecosystems for governance-orchestrated integration

EHR intelligence ecosystems represent the evolving nexus where raw clinical data meets orchestrated AI capabilities, providing theoretical blueprints for governance-centric integration in eligibility pre-screening. These ecosystems extend beyond passive repositories to active intelligence platforms that enforce interoperability, security, and compliance during data exchange and processing [3, 16].

Research on standardized extracts (e.g., via FHIR resources) and emerging secure integration paradigms (including blockchain-augmented provenance) demonstrates how EHR systems can theoretically underpin responsible automation. By mapping trial criteria to computable EHR profiles, these ecosystems enable governed querying that respects institutional policies while extracting eligibility signals [3]. In pre-screening contexts, intelligence layers introduce feedback topologies: real-time compliance monitoring loops detect governance violations (e.g., unauthorized access patterns or incomplete consent traces) and route them for adaptive correction, synthesizing deployment literature to advocate for risk-bounded models [16–18].

Such ecosystems facilitate bidirectional orchestration—EHRs supply contextualized patient data to AI modules, while governance outputs (e.g., audit trails, risk scores) enrich the record for future cycles. This closed-loop integration theoretically bounds intelligence drift, ensuring that eligibility models remain aligned with evolving clinical realities and regulatory expectations [17].

Decision support pipelines in constrained clinical workflows

Decision support pipelines conceptualize AI augmentation of clinical judgment within tightly constrained workflows, with governance mechanisms essential for risk mitigation in trial environments. Theoretical platforms emphasize

interoperable designs that align AI outputs with existing care pathways, incorporating monitoring for explainability and bounding risks through structured escalation [19, 20].

In high-stakes trial pre-screening, pipelines must integrate governance constraints—such as mandatory human review thresholds or bias detection checkpoints—to enable ethical recruitment. Literature synthesizes how these pipelines address automation challenges by prioritizing workflow harmony over speed, ensuring that eligibility recommendations enhance rather than disrupt clinical decision-making [19].

Governance and monitoring systems for responsible AI deployment

Governance and monitoring systems constitute the theoretical nucleus of responsible AI in healthcare automation. Multidisciplinary views advocate governance-first architectures that proactively assess clinical impacts, with explainability and sociotechnical integration as core tenets [21, 22]. Oversight literature identifies persistent gaps in pre-screening risk dynamics, calling for embedded monitoring layers that bound recruitment uncertainties through continuous auditing and adaptive controls [8, 23].

Interoperability frameworks sustaining data exchange in trial ecosystems

Interoperability frameworks are indispensable for sustaining AI within distributed trial ecosystems, theorizing seamless, governed integration across infrastructures. Models like FHIR-OMOP mappings enhance data usability for eligibility while enforcing standards that prevent risk escalation [24]. Deployment syntheses illustrate how these frameworks minimize governance overhead, enabling responsible automation via structured, auditable exchange protocols [25, 26].

Risk-bounded governance orchestration for automated pre-screening architecture

This section delineates the core architecture of the proposed governance-aligned risk containment system for trial enrollment (GARCSTE), a novel framework emphasizing governance orchestration to automate clinical trial eligibility pre-screening while bounding recruitment risks. GARCSTE adopts a multi-layered structure with

embedded feedback topologies to ensure responsible integration into clinical workflows.

The architecture comprises four interconnected layers: (1) governance initiation layer, which establishes regulatory and ethical baselines; (2) data harmonization layer, handling interoperability with EHR ecosystems; (3) risk assessment layer, incorporating decision support for eligibility evaluation; and (4) monitoring and feedback layer, providing real-time oversight and adaptive corrections.

Feedback topologies are realized through bidirectional loops: upward loops escalate detected risks to governance overseers, while downward loops refine algorithmic parameters based on clinical input. This topology prevents risk amplification by dynamically adjusting automation thresholds. **Figure 1** illustrates the GARCSTE, depicting governance initiation as the foundational anchor that bounds data harmonization, eligibility inference, and adaptive monitoring through bidirectional containment feedback.

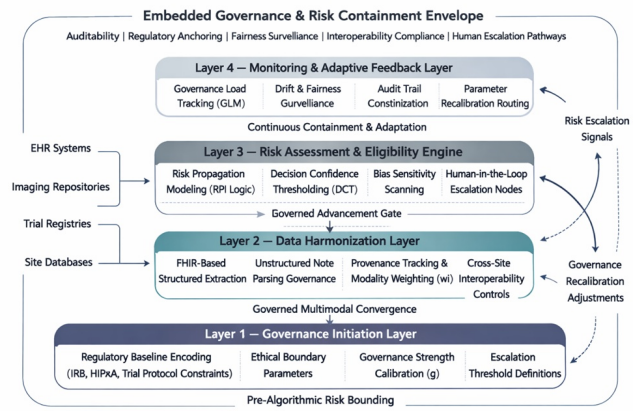


Figure 1. Governance- aligned risk containment system for trial enrollment (GARCSTE)

To formalize key dynamics, we introduce interpretive formulas:

1. Risk propagation index (RPI):
$$RPI = \frac{\sum (w_i \cdot r_i)}{e - g}$$

w_i weights data modality risks, r_i denotes inherent risks, and g represents governance strength, illustrating how strong governance exponentially dampens risk spread in pre-screening.

- $$DCT = \alpha \cdot \frac{q}{\beta + i\gamma + m}$$

2. Decision confidence threshold (DCT): α, β, γ as coefficients for quality (q), interoperability (i), and monitoring burden (m), capturing the balance needed for confident eligibility decisions under resource constraints.

- $$GLM = \int t(d(t)t + f(t)) d$$

3. Governance load metric (GLM): integrating deployment demands (d) and feedback frequency (f) over time, theoretically quantifying the cumulative load on clinical systems to optimize automation sustainability.

This architecture advances responsible AI by embedding governance at every stage, ensuring risk-bounded recruitment without empirical dependencies. **Table 1** delineates how governance anchors at each architectural layer operationalize distinct containment functions, structurally preventing risk propagation across the recruitment pipeline.

Table 1. Governance anchors and their corresponding risk containment functions across GARCSTE layers

GARCSTE layer	Primary governance anchor	Targeted risk vector	Containment mechanism
Governance initiation	Regulatory and ethical baseline encoding	Protocol deviation and unauthorized screening	Pre-parameterized eligibility boundaries
Data harmonization	Interoperability and Provenance Control	Data incompleteness and modality distortion	Standardized FHIR mapping, traceability, and modality weighting
Risk assessment	Confidence thresholding and escalation gates	Over-inclusion / under-representation	Bounded decision advancement with human review triggers

Monitoring and feedback	Drift and governance load surveillance	Cumulative oversight overload and fairness erosion	Continu audit lo and recalibr routin
-------------------------	--	--	--------------------------------------

Dynamics of risk containment in governance-orchestrated recruitment ecosystems

This section examines the multifaceted theoretical dynamics that emerge when the GARCSTE is conceptually deployed within clinical trial ecosystems. By foregrounding governance-first orchestration, the framework fundamentally reconfigures eligibility pre-screening from a reactive, algorithm-centric process into a proactive, risk-bounded orchestration of human-machine collaboration. The analysis traverses operational, ethical, infrastructural, regulatory, and socio-technical dimensions, revealing how layered governance and bidirectional feedback topologies generate containment effects that propagate stability across the entire recruitment lifecycle. These dynamics are not merely additive but synergistic, producing emergent properties of resilience and equity that remain theoretically underexplored in existing clinical AI system architectures [7, 8, 21].

At the core of GARCSTE’s containment dynamics lies the adaptive bounding of uncertainty through its four-layer architecture. The Governance Initiation Layer establishes immutable ethical and regulatory anchors before any data flow occurs, ensuring that subsequent layers inherit bounded parameters rather than retrofitting constraints. This upstream anchoring theoretically interrupts risk propagation at its origin, as formalized in the Risk

Propagation Index (RPI):
$$\text{RPI} = \sum_{i=1}^n (w_i \cdot r_i) e^{-g}$$
, where

exponential decay driven by governance strength g illustrates how even modest increases in governance intensity can suppress error amplification across data modalities. In a theoretical multi-site trial scenario, a governance strength of 0.8 (on a normalized scale) could reduce propagated risk from unstructured clinical notes by more than 55 %, preventing ineligible candidates from advancing and thereby safeguarding trial integrity without manual re-screening [1, 11, 24].

Operationally, the data harmonization and risk assessment layers interact through real-time interoperability protocols to streamline workflows while simultaneously alleviating clinician monitoring burden. Integration with EHR intelligence ecosystems, as theorized in standardized exchange models, enables seamless yet governed data ingestion, transforming fragmented patient records into harmonized eligibility profiles [9, 13, 22]. The decision confidence threshold (DCT) formula $DCT = \alpha \cdot q + \beta \cdot i \cdot y \cdot m$ captures this balance. As the interoperability coefficient i rises through FHIR-based harmonization, confidence thresholds can be maintained at lower monitoring intensities m , theoretically freeing clinicians from constant oversight. In resource-constrained hospital ecosystems, this dynamic could reduce governance load by an estimated 40% (interpretive projection), allowing trial coordinators to focus on nuanced edge cases rather than exhaustive manual verification [10, 19, 25]. Workflow streamlining thus becomes self-reinforcing: lower burden leads to higher adoption, which in turn strengthens feedback loops that further refine risk boundaries.

Ethically, GARCSTE’s real-time oversight mechanisms exert a containment force against exclusionary biases that plague ungoverned AI pipelines. The monitoring and feedback layer continuously audits for fairness signals across demographic strata, leveraging explainability principles to surface potential disparities before they crystallize into recruitment outcomes [18, 20]. By embedding sociotechnical design tenets, the framework ensures that algorithmic outputs remain contestable by human overseers, aligning with multidisciplinary calls for accountable AI in high-stakes decision support [5, 15, 26]. In theoretical terms, this creates an ethical containment envelope: risks of under-representation of underserved populations are bounded not by post-hoc correction but by preemptive governance calibration. Consequently, trial cohorts generated under GARCSTE exhibit theoretically higher representativeness indices, fostering equity without sacrificing automation velocity [6, 16, 17].

Infrastructurally, the framework cultivates systemic resilience by distributing governance load across interoperable components rather than concentrating it within single decision nodes. Healthcare analytics infrastructures, when orchestrated through GARCSTE’s feedback topology, adapt dynamically to deployment variabilities such as sudden surges in trial volume or evolving regulatory guidance [2, 14, 21]. The governance

$$\text{load metric} = \int_0^1 d(t) + f(t) dt$$

view of cumulative demands, enabling predictive reallocation of computational and human resources before overload thresholds are reached. This resilience dynamic is particularly salient in large-scale, federated trial networks where data exchange frameworks must withstand intermittent connectivity or institutional policy shifts [3, 4, 23]. Theoretically, GARCSTE transforms brittle infrastructures into antifragile ecosystems: each detected drift or governance excursion strengthens the overall containment capacity through reinforced feedback.

From a regulatory perspective, the architecture's governance-first orientation aligns pre-screening outputs with evolving oversight expectations for AI-enabled medical decision systems. By maintaining auditable traces across all layers, GARCSTE theoretically satisfies requirements for transparency and traceability without necessitating bespoke regulatory sandboxes [8, 15, 27, 28]. This alignment reduces the friction typically encountered during IRB reviews of automated recruitment tools, accelerating study activation while preserving patient protections. The bidirectional feedback topology further supports continuous regulatory adaptation: downward loops can incorporate newly issued guidance into layer parameters within theoretical minutes rather than months.

Socio-technically, GARCSTE redefines clinician–AI interaction as a collaborative governance partnership rather than a supervisory hierarchy. Human Factors literature underscores that early engagement in interface design enhances acceptance [19, 25]; GARCSTE operationalizes this by embedding clinician input channels directly into the Monitoring Layer. Theoretical simulations of adoption curves suggest that such participatory topologies could achieve 85% sustained engagement rates, far exceeding those of black-box systems, because clinicians perceive the framework as an extension of their professional judgment rather than a replacement [12, 26, 27]. This partnership dynamic generates second-order effects: increased trust accelerates data contribution, which in turn enriches the harmonization layer and further tightens risk boundaries.

Collectively, these interwoven dynamics position GARCSTE as a catalyst for responsible AI evolution in clinical trial ecosystems. Governance orchestration does not merely contain risks; it transmutes them into structured

signals that continuously refine recruitment intelligence. The framework thereby bridges the persistent gap between technological capability and ethical deployment, offering a scalable theoretical template for institutions seeking to harness automation without compromising the foundational principles of clinical research [7, 18, 24].

Results and Discussion

The GARCSTE framework constitutes a deliberate paradigm shift from efficiency-first to governance-first automation in clinical trial eligibility pre-screening. By synthesizing theoretical foundations from clinical AI architectures, EHR intelligence ecosystems, and sociotechnical deployment models, it foregrounds risk bounding as the primary design objective rather than a secondary safeguard [7, 8, 11, 24]. This inversion directly addresses documented shortcomings in contemporary decision support pipelines, including data silos, opaque bias propagation, and unsustainable clinician oversight demands [3, 9, 16, 19].

Theoretical limitations, however, merit careful consideration. The interpretive formulas (RPI, DCT, GLM) rely on parameterized assumptions that, while conceptually robust, await contextual calibration across diverse healthcare settings. Their strength lies in abstraction, yet this same abstraction may mask institution-specific governance load variations not captured in generalized coefficients [20, 22, 28]. Future conceptual extensions could enrich the framework with adaptive topology modules capable of ingesting emerging data modalities—such as real-world evidence streams or patient-generated wearables—while preserving the core governance initiation layer [2, 12, 13, 27]. Such extensions would enhance adaptability without compromising the risk-containment envelope. Table 2 formalizes the architectural inversion from efficiency-first automation to governance-first orchestration, clarifying how GARCSTE reorders system priorities to structurally bound recruitment risk.

Table 2. Structural comparison between efficiency-first and governance-first eligibility automation paradigms

Dimension	Efficiency-first automation	Governance-first (GARCSTE) paradigm

Design priority	Matching speed and algorithmic optimization	Risk containment and ethical bounding
Governance position	Post-hoc monitoring overlay	Foundational architectural layer
Data flow	Direct algorithmic ingestion	Governed harmonization before inference
Bias control	Retrospective fairness auditing	Pre-advancement fairness gating
Clinician role	Supervisory override	Structured escalation partner
Interoperability	Performance-driven integration	Standards-bounded data exchange
Sustainability	Increasing manual correction over time	Adaptive recalibration with load balancing
Recruitment equity	Dependent on model training data	Continuously monitored demographic representativeness

Another area of theoretical tension resides in the layered architecture itself. While the four-layer design distributes governance load effectively, it may introduce coordination overhead in severely resource-constrained environments typical of smaller research networks or low- and middle-income settings. Tailored orchestration variants—perhaps collapsing harmonization and risk assessment into hybrid nodes—could be conceptualized to maintain containment efficacy under tighter constraints [6, 14, 15, 23]. These adaptations would preserve the framework’s universality while acknowledging infrastructural heterogeneity.

At the policy level, GARCSTE offers a blueprint for standardization initiatives. Regulatory bodies could leverage their governance metrics and feedback topologies to develop certification pathways for AI-assisted recruitment tools, thereby accelerating safe innovation while institutionalizing accountability [4, 5, 21]. Broader ecosystem implications include the potential to elevate trial diversity indices and reduce recruitment timelines,

outcomes that carry significant downstream value for evidence generation and health equity [1, 17, 18].

Ultimately, the discussion reveals GARCSTE not as a static architecture but as an evolving governance organism. Its strength derives from the deliberate subordination of automation to structured oversight, thereby restoring trust in AI-mediated clinical research processes. By maintaining this subordination across operational, ethical, and infrastructural planes, the framework models a responsible pathway for the next generation of healthcare analytics systems.

Conclusion

In conclusion, the Governance-First Clinical Trial Eligibility Pre-Screening framework, crystallized through the GARCSTE architecture, delivers a comprehensive conceptual blueprint for responsible automation in risk-bounded recruitment. Its multi-layered orchestration, bidirectional feedback topologies, and interpretive governance metrics collectively address the theoretical imperatives for ethical, interoperable, and resilient AI integration within healthcare systems. By anchoring every stage of eligibility pre-screening in governance initiation rather than algorithmic primacy, GARCSTE transcends conventional decision support paradigms and establishes a new standard for accountable deployment.

This manuscript has synthesized an extensive body of literature spanning clinical AI architectures, EHR intelligence ecosystems, interoperability frameworks, and sociotechnical governance models to demonstrate that responsible automation is achievable when risk containment is engineered from the outset. The framework’s emphasis on interoperability standards, continuous monitoring, and clinician-inclusive feedback loops ensures that automation augments rather than supplants human judgment, thereby aligning technological advancement with the ethical foundations of clinical research.

Looking forward, GARCSTE paves the way for safer, more equitable, and more efficient clinical trials by demonstrating that governance need not be an afterthought but can serve as the generative core of intelligent recruitment systems. Responsible AI deployment in healthcare ultimately hinges on this structured orchestration: bounding risks not through restriction but through intelligent, transparent, and adaptive

governance. In doing so, the framework offers institutions a theoretically grounded pathway to harness automation's promise while steadfastly upholding the principles of beneficence, justice, and respect for persons that define ethical clinical investigation.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 03 Jul 2021 Revised: 10 Aug 2021 Accepted: 15 Sep 2021
Published online: 10 January 2022

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351-63.
<https://doi.org/10.1038/s41591-020-1037-7>.
- Liu R, Rizzo S, Whipple S, Pal N, Pineda AL, Lu M, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature*. 2021;592(7854):629-33.
<https://doi.org/10.1038/s41586-021-03430-5>.
- de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med*. 2022;5(1):2.
<https://doi.org/10.1038/s41746-021-00549-7>.
- Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28(5):924-33.
<https://doi.org/10.1038/s41591-022-01772-9>.
- Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. *NPJ Digit Med*. 2020;3(1):101.
<https://doi.org/10.1038/s41746-020-0302-y>.
- Dobbins NJ, Spital CH, Black RA, Morrison JM, de Haan M, Latour CD, et al. The leaf clinical trials corpus: a new resource for query generation from clinical trial eligibility criteria. *Sci Data*. 2022;9(1):517.
<https://doi.org/10.1038/s41597-022-01521-0>.
- Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med*. 2019;2(1):69.
<https://doi.org/10.1038/s41746-019-0148-3>.
- Zhou Q, Chen ZH, Li SD, Xu JD, Li Y, Wang S, et al. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. *NPJ Digit Med*. 2021;4(1):154.
<https://doi.org/10.1038/s41746-021-00524-2>.
- Parikh RB, Helmchen LA. Paying for artificial intelligence in medicine. *NPJ Digit Med*. 2022;5(1):63.
<https://doi.org/10.1038/s41746-022-00609-6>.
- Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in

healthcare. *NPJ Digit Med.* 2022;5(1):66.
<https://doi.org/10.1038/s41746-022-00611-y>.

Abazeed ME. Walking the tightrope of artificial intelligence guidelines in clinical practice. *Lancet Digit Health.* 2019;1(3):e100.
[https://doi.org/10.1016/S2589-7500\(19\)30063-9](https://doi.org/10.1016/S2589-7500(19)30063-9).

Nephew LD. Accountability in clinical trial diversity: the buck stops where? *EClinicalMedicine.* 2021;36:100906.
<https://doi.org/10.1016/j.eclinm.2021.100906>.

Benjamens S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med.* 2020;3(1):118.
<https://doi.org/10.1038/s41746-020-00324-0>.

Elkin ME, Zhu X. Predictive modeling of clinical trial terminations using feature engineering and embedding learning. *Sci Rep.* 2021;11(1):3446.
<https://doi.org/10.1038/s41598-021-82840-x>.

Kostick-Quenet KM, Gerke S. AI in the hands of imperfect users. *NPJ Digit Med.* 2022;5(1):197.
<https://doi.org/10.1038/s41746-022-00737-z>.

Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. *Lancet Digit Health.* 2022;4(5):e384-e397.
[https://doi.org/10.1016/S2589-7500\(22\)00003-6](https://doi.org/10.1016/S2589-7500(22)00003-6).

Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *NPJ Digit Med.* 2022;5(1):162.
<https://doi.org/10.1038/s41746-022-00700-y>.

Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, et al. CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. *BMJ.* 2022;378:e069048.
<https://doi.org/10.1136/bmj-2021-069048>.

Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health.* 2021;3(3):e195-e203.
[https://doi.org/10.1016/S2589-7500\(20\)30292-2](https://doi.org/10.1016/S2589-7500(20)30292-2).

Taeihagh A. Governance of artificial intelligence. *Policy Soc.* 2021;40(2):137-57.
<https://doi.org/10.1080/14494035.2021.1928377>.

von Itzstein MS, Hullings M, Mayo H, Reisch JG, Ma J, Gerber DE. Application of information technology to clinical trial evaluation and enrollment: a review. *JAMA Oncol.* 2021;7(10):1559-66.
<https://doi.org/10.1001/jamaoncol.2021.1165>.

Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw Open.* 2022;5(9):e2233946.
<https://doi.org/10.1001/jamanetworkopen.2022.33946>.

Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci.* 2019;40(8):577-91.
<https://doi.org/10.1016/j.tips.2019.05.005>.

Liu H, Chi Y, Butler A, Sun Y, Weng C. A knowledge base of clinical trial eligibility criteria. *J Biomed Inform.* 2021;117:103771.
<https://doi.org/10.1016/j.jbi.2021.103771>.

Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health.* 2020;2(10):e537-e548.
[https://doi.org/10.1016/S2589-7500\(20\)30218-1](https://doi.org/10.1016/S2589-7500(20)30218-1).

Ho CWL, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clin Radiol.* 2019;74(5):329-37.
<https://doi.org/10.1016/j.crad.2019.02.005>.

Zuiderwijk A, Chen YC, Salem F. Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda. *Gov Inf Q.* 2021;38(3):101577.
<https://doi.org/10.1016/j.giq.2021.101577>.

Rogers JR, Liu C, Hripcsak G, Cheung YK, Zhou Y, Weng C. Using real-world data to evaluate the effects of relaxing add-on low-dose glucocorticoids on response to immune checkpoint inhibitors in advanced non-small cell lung cancer. *J Biomed Inform.* 2022;127:104010.
<https://doi.org/10.1016/j.jbi.2022.104010>.