International Journal of Medical Informatics 203 (2025) 106015

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Toward responsible AI governance: Balancing multi-stakeholder perspectives on AI in healthcare

Leon Rozenblit [a,b,c], Amy Price [d,e], Anthony Solomonides [f], Amanda L. Joseph [b,g,h], Eileen Koski [i], Gyana Srivastava [b], Steven Labkoff [j,b], David Bray [k,l], Monica Lopez-Gonzalez [m], Reva Singh [n], Dave deBronkart [o,p,q], Paul J. Barr [d], Peter Szolovits [r], Kiran Dattani [s], Charles Jaffe [t], Douglas Fridsma [u], Russell Baris [v], Russell Leftwich [w], Robert Stolper [x], Mark G. Weiner [y], Nuria Pastor [z], Unai Sanchez Luque [z], Baihan Lin [aa], Tien Thi Thuy Bui [b,ab], Bilikis Oladimeji [ac], Tayler Williams [n], Gretchen Purcell Jackson [ad,ae], Pei-Yun Sabrina Hsueh [af], Yuri Quintana [b,h,ag,*]

[a] Q.E.D. Institute, New Haven, CT, USA
[b] Division of Clinical Informatics, Beth Israel Deaconess Medical Center, Boston, MA, USA
[c] Yale School of Management, New Haven, CT, USA
[d] The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
[e] BMJ, London, UK
[f] Research Institute, Endeavor Health, Evanston, IL, USA
[g] School of Health Information Science, University of Victoria, Victoria, BC, Canada
[h] Homewood Research Institute, Guelph, Ontario, Canada
[i] IBM Research, Yorktown Heights, NY, USA
[j] Quantori, Boston, MA, USA
[k] Loomis Innovation Council, Stimson Center, Washington, DC, USA
[l] Business Executives for National Security, Washington, DC, USA
[m] Cognitive Insights for Artificial Intelligence, Baltimore, MD, USA
[n] American Medical Informatics Association, Washington, DC, USA
[o] Society for Participatory Medicine, Boston, MA, USA
[p] HL7® Patient Empowerment Workgroup, USA
[q] The BMJ's Patient Advisory Panel, London, UK
[r] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[s] Google Cloud Healthcare and Life Sciences, New York, NY, USA
[t] Health Level 7 International, Ann Arbor, MI, USA
[u] Health Universe, San Francisco, CA, USA
[v] eLumindata, Westport, CT, USA
[w] Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA
[x] IQVIA, Boston, MA, USA
[y] Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA
[z] HumanITcare, Barcelona, Spain
[aa] Departments of Neuroscience and Systems Biology, Columbia University Irving Medical Center, New York, NY, USA
[ab] Massachusetts College of Pharmacy and Health Sciences, Boston, MA, USA
[ac] UnitedHealth Group, Minnesota, MN, USA
[ad] Intuitive Surgical, Sunnyvale, CA, USA
[ae] Pediatrics and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
[af] Pfizer Inc., New York, NY, USA
[ag] Department of Medicine, Harvard Medical School, Boston, MA, USA

* Corresponding author.
E-mail address: yquintan@bidmc.harvard.edu (Y. Quintana).

https://doi.org/10.1016/j.ijmedinf.2025.106015
Received 10 November 2024; Received in revised form 6 June 2025; Accepted 12 June 2025
Available online 19 June 2025

ARTICLE INFO

Keywords:
AI governance
Healthcare ethics
Multi-stakeholder framework
Clinical decision support
Responsible innovation
Real World Research
Generative AI
Large Language Models
CoProduction
Public and Patient Involvement
PPI
RWE
CDCS
Cosumer Involvement
Consumer Health

ABSTRACT

Introduction: The rapid integration of artificial intelligence (AI) into healthcare presents significant governance challenges, requiring balanced approaches that safeguard safety, efficacy, equity, and trust (SEET). This study proposes a cognitive framework to guide AI governance, addressing tradeoffs between speed, scope, and capability.
Objective: To develop a structured governance model that harmonizes stakeholder perspectives, focusing on multi-dimensional challenges and ethical principles essential for AI in healthcare.
Methods: A multidisciplinary team convened at the *Blueprints for Trust* conference, organized by the American Medical Informatics Association (AMIA), and the Division of Clinical Informatics at Beth Israel Deaconess Medical Center. Following extensive discussions with 190 participants across sectors, three governance models were identified to address specific domains: (1) Clinical Decision Support (CDS), (2) Real-World Evidence (RWE), (3) Consumer Health (CH).
Results: Three governance models emerged, tailored to CDS, RWE, and CH domains. Key recommendations include establishing a Health AI Consumer Consortium for patient-centered oversight, initiating voluntary accreditation and certification frameworks, and piloting risk-level-based standards. These models balance rapid adaptation with SEET-focused safeguards through transparency, inclusivity, and ongoing learning.
Conclusion: A proactive, constraint-based governance framework is critical for responsible AI integration in healthcare. This structured, multi-stakeholder approach provides a roadmap for ethical, transparent governance that can evolve with technological advancements, enhancing trust and safety in healthcare AI applications.

1. Introduction

Responsible governance of artificial intelligence (AI) in healthcare must balance rapid innovation with introducing complementary safeguards to ensure safety, efficacy, equity, and trust (SEET). This paper proposes a structured approach to healthcare-specific governance models, such as voluntary certifications, domain-specific oversight bodies, and participatory design frameworks. These models operate at the intermediate level—above individual institutions but below national regulation—where sector-specific norms can shape AI deployment. Rapid deployment and iteration favor lightweight frameworks, prioritizing cooperation over regulation. However, appropriate controls and public trust depend on transparent disclosure and independent auditing procedures. Anchoring governance with ethical principles that empower patients and consumers can promote human-centered design (HCD) and prudent clinical integration.

Drawing from four months of structured deliberation with 190 experts, our core recommendations include a patient-led Health AI Consumer Consortium (HAIC2) for representation and oversight, accrediting cross-disciplinary educational programs, and instituting stepwise accountability mechanisms for AI in healthcare systems stratified by risk level. Success will require a multifaceted approach, including commitment across stakeholders to instantiate governance through participatory standards development, pilots of novel certification methods, gradual accumulation of experience, and consensus. With concerted near-term efforts centered on transparency and stakeholder values alignment, AI's vast potential can be harnessed responsibly to enable more equitable access and optimal patient and provider outcomes.

Several initiatives aspire to guide the development of AI governance in general, and healthcare AI in particular [1,2]. The direction of our work builds on methods-first proposals, such as Embi's "algorithmovigilance," which refers to the systematic monitoring and evaluation of algorithmic tools to ensure they are continually updated based on real-world performance data [3]. Industry-wide implementation guidelines, similar to those developed by the Coalition for Health AI (CHAI), are downstream of methods-first efforts [4,5]. This paper develops the overarching framework, while parallel papers explore detailed governance use cases in the domains of Clinical Decision Support (CDS), Real-World Evidence (RWE), and Consumer Health (CH) [1,2].

2. Methods

The aim of the consensus-building process was to generate pragmatic, stakeholder-informed governance models tailored to three domains—CDS, RWE, and CH—where governance challenges are most pressing. These domains were selected based on pre-conference scoping reviews and stakeholder input identifying them as high-priority areas for intervention. The broadly inclusive *Blueprints for Trust* conference [3] convened a multidisciplinary group of experts on September 2023 at Harvard Medical school to provide diverse perspectives on ensuring AI's safe and equitable use in healthcare. The American Medical Informatics Association (AMIA) and the Division of Clinical Informatics (DCI) at Beth Israel Deaconess Medical Center (BIDMC) led in-depth discussions on the topic of AI governance. The group included 140 online members and 50 in-person members from industry, academia, the informatics community, patients, advocates, clinicians, scientists, and regulators. This article is an output of of four months of comprehensive discussion, iteration, and evaluation with a multifaceted team working toward consensus building through discussions [1,2,4]. Our deliberations produced three separate but complementary governance models with different requirements for three domains: (1) Clinical Decision Support (CDS), (2) Real World Evidence (RWE), and (3) Consumer Health (CH). Moreover, each domain was shown to have different governance requirements.

The consensus-building process spanned four months and included multiple delivery mediums such as: educational webinars, a three-day conference on location at Harvard Medical School, followed by structured deliberative sessions using the following methods:

2.1. Research timeline and documentation

The four-month timeline encompassed multiple iterative stages:

1. **Pre-Conference Phase (Weeks 1–4):** A series of educational webinars introduced key governance challenges and foundational AI ethics principles.
2. **Conference (Week 5):** A hybrid (in-person and virtual) event convened 50 in-person and 140 online participants, facilitating structured panel discussions, breakout sessions, and workshops.
3. **Post-Conference Deliberation (Weeks 6–16):** Meeting transcripts and session summaries were analyzed using natural language processing (NLP) to identify key themes. Transcripts from breakout and plenary sessions were analyzed using NLP tools. Techniques included keyword extraction, clustering algorithms to detect dominant themes, and sentiment analysis to gauge consensus levels. Manual validation by human analysts was used to refine and interpret NLP-generated themes. Findings were discussed and iteratively refined

through robust biweekly virtual discussions, and cross-functional stakeholder feedback was incorporated to reach consensus.

4. **Finalization Phase (Weeks 17–18):** The resulting framework underwent a final review by external domain experts in AI ethics, clinical governance, and informatics.

All meeting minutes, discussion summaries, and thematic analyses were documented for transparency, and summaries are available upon request.

2.2. Participant selection and representation

Participants (n = 190) were identified through a structured selection process led by AMIA and the DCI. The criteria for inclusion included domain expertise in AI governance, CDS, RWE, and CH. To ensure diversity, selection was stratified across key stakeholder groups, including:

- Clinicians and Informatics Researchers (35 %)
- AI Scientists, Solution Architects, and Industry Leaders (30 %)
- Legal and Policy Experts (15 %)
- Patient and Consumer Advocates (20 %)

Demographic diversity was considered with respect to professional background and gender. However, we didn't consider race and ethnicity, as we were unlikely to achieve representativeness with a convenience sample of N = 190.

2.3. Consensus-building process

Consensus in our work emerged through an iterative, deliberative process rather than through formal voting mechanisms. The process followed these structured steps:

1. **Initial Framework Development:** Facilitated small-group discussions generated preliminary governance models.
2. **Iterative Refinement:** Draft recommendations were discussed and refined in follow-up meetings. NLP of meeting recordings helped identify recurring themes and areas of general agreement. The NLP outputs were synthesized into draft recommendation documents. These were reviewed during biweekly meetings, where each proposal was debated and revised. Consensus was reached through group convergence, and areas of disagreement were explicitly documented.
3. **Voting and Dispute Resolution:**
   - Consensus was considered reached when discussion participants demonstrated general agreement on key principles and recommendations through consistent articulation across meetings.
   - Areas of disagreement were specifically noted in discussions, with multiple perspectives incorporated into draft documents to ensure comprehensive representation.
   - Meeting facilitators encouraged inclusive participation and ensured all stakeholder viewpoints received consideration. Facilitators were charged with neutrality and were not to represent the viewpoints of any institution.
   - Final synthesis documents were circulated for comment, with feedback incorporated until general agreement was achieved on the core frameworks presented in this paper.

Outcomes: The final consensus resulted in the delineation of general constraints on AI governance as well as special considerations for each of the three distinct governance models tailored to the needs of: (1) AI in CDS [2], (2) AI in RWE, and (3) AI in CH [1]. Each tailored model provides structured recommendations for ensuring ethical, transparent, and accountable AI deployment in healthcare.

3. Results

3.1. The challenges of governing AI in healthcare

The governance of AI in healthcare poses unique challenges because it must allow for rapid innovation in a vibrant and chaotic ecosystem while promoting SEET. The dimensions of speed, breadth, and capability were recurrent themes in participant discussions. These were identified through NLP clustering and validated through manual coding. Stakeholders often described tradeoffs in terms of delivery timelines, scope of applicability, and the rigor of oversight, leading us to formalize these as the primary axes of our governance-selection framework.

Developing appropriate governance models requires a wise allocation of limited resources to balance the three dimensions of governance: (a) speed – how quickly the governance model can be implemented and adapted to future changes in technologies or contexts, (b) breadth – how much scope, in terms of functions, markets, vendors, and user classes will the governance model cover, and (c) capability – how well the governance model will achieve the desired standards for SEET. Some application domains, like CDS, and some modalities, like machine learning (ML), pose especially acute governance challenges, the former because it has the potential to directly impact patient outcomes and clinical systems, the latter because opacity undermines trust.

Amid these challenges, the emergence of generative AI (GenAI) has further shaped the AI-in-healthcare landscape. GenAI technology, known for its capacity to generate content ranging from textual information to medical imagery, holds considerable potential in healthcare applications. GenAI could expedite tasks such as medical image analysis, drug discovery, and even clinical note generation. [4] However, its rapid evolution introduces unique governance, ethical, and safety considerations.

Because AI in healthcare, powered by GenAI, is moving at an unprecedented pace, any governance model must be developed and implemented to match the speed of innovation. The urgency to settle on governance designs for the many facets of AI are tempered by concerns that selected designs must neither hinder innovation nor compromise safety. They must be resilient enough to respond to unanticipated changes yet capable enough to ensure efficacious patient-centric outcomes.

Despite AI's novel and exciting potential, development across its varied modalities should adopt a pragmatic and balanced perspective. A multidisciplinary approach integrating HCD principles encompassing a broad spectrum of stakeholders (i.e., patients, advocates, providers, developers) from the outset is vital. Keeping the human in the loop promotes the development and design of SEET-oriented technologies [5].

While perspectives must be broad, the focus of governance design must be narrowly drawn around specific application domains—in this instance, CDS, RWE, and CH. This is because stakeholder incentives, legal contexts, and ethical implications differ across domains. For example, CDS directly affects clinical decision-making and demands expert medical validation, while CH involves personal autonomy and user-driven oversight. Coherent governance models can best be designed within domains encompassing similar stakeholder relationships, constraints, and incentives.

Further, a cognitive framework (Fig. 1) for selecting governance models must begin with acknowledging the constraints on model selection and then develop three layers of analysis: (a) the structural characteristics of desirable governance models, (b) the processes and methods for developing and implementing appropriate models, and (c) the content of desirable governance models.

3.2. Constraints-based cognitive framework for selecting governance models

Information technology (IT) project managers and leaders know the

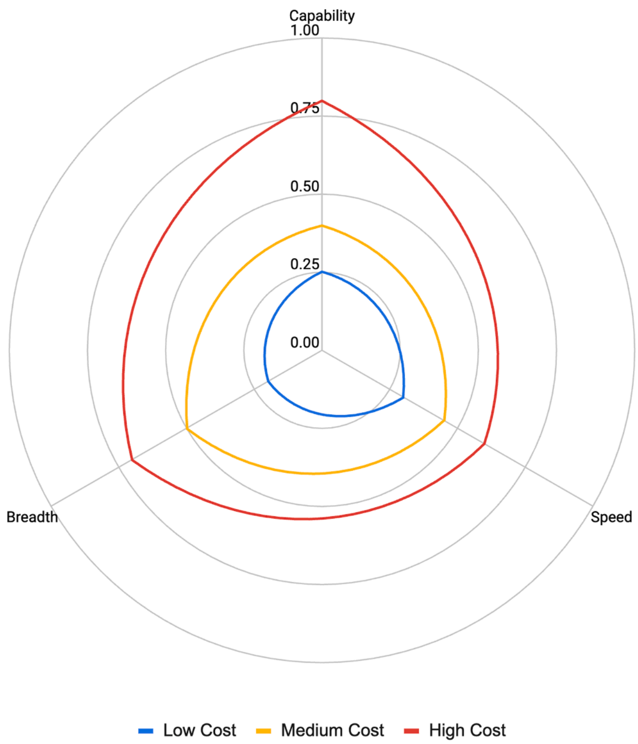

Fig. 1. The Triple-Constraint Framework for Governance Model Selection.

adage: "Good, fast, cheap. Pick two." Underpinning the saying is the concept of "the triple constraint" on the selection of project plans: for any specified budget level, you can control two out of three dimensions of project delivery: time, scope, and quality; the third is then determined by your other choices [6]. An analogous concept applies to the selection of governance models. Given a resource level, the selection of a governance model represents a set of design tradeoffs between the three dimensions of governance: (1) **speed** of governance framework implementation and adaptation, (2) **breadth** of the proposed governance, and (3) **capability** of the governance model (Fig. 1). This mental model outlines a constraint-based cognitive framework.

We define a "cognitive framework" as a conceptual model used to guide decision-makers' reasoning towards the best-fit governance approach. By "constraints-based cognitive framework," we mean one that explicates the tradeoffs between model attributes under resource constraints. We will first discuss the general constraints on selecting governance models and then describe three subsequent framework layers:

(a) the structural characteristics, (b) the methods of development, and (c) the preliminary content recommendations.

Fig. 1, the Triple-Constraint Framework for Governance Model Selection, illustrates the selection space for governance models with the dimensions of speed, breadth, and capability represented as polar coordinates. This triad is inspired by the 'triple constraint' in project management [9], where tradeoffs between time, scope, and quality must be managed. Resource levels are represented by the maximum total area of each triangular shape. Fig. 1 illustrates graphically the tradeoffs inherent in governance model selection under resource constraints: a model that optimizes for broad coverage must yield on speed or capability, optimizing for speed means reduced breadth or capability, and optimizing for capability means less speed or breadth.

Levels of Capability, Speed, and Breadth Achievable at Each Resource Level range from 0 to 1. Maximums are triple-constrained within the boundary set by resources available to develop and implement a particular governance model; for each resource level, the total area must remain constant so that an increase on one axis requires a decrease on at least one other [7].

(1) "Speed" in governance models is analogous to "time" in project management. It refers to how quickly the model can be implemented and how quickly it can adapt to changes. Note that "speed" in this context doesn't refer to how quickly the model makes governance decisions, no more than the "time" in IT project management refers to how quickly the delivered software will run. The run-time characteristics of the software are part of its quality. Analogously, the governance throughput of the selected model – the speed with which it can make decisions and the number of decisions it can handle per unit of time – is part of its "capability."
(2) "Breadth" in governance models is analogous to "scope" in project management. Seeking to govern broadly means covering a large fraction of the possible use cases in a domain. Narrow governance focuses on a small fraction of the most essential use cases (e.g., the most common or those posing the highest risk of harm). Broad governance may consume resources on handling low-value use cases, while narrow governance may leave rare but undesirable occurrences ungoverned.
(3) Lastly, "capability" in governance models is analogous to "quality" in project management. With AI in healthcare, our capability metric is the model's ability to deliver the desired level of SEET, but it also includes fostering innovation by providing rapid decisions and minimal burden. Since higher capability consumes more resources (i.e., requiring greater time from the governing body to review each rule or case), finding an appropriate level requires deliberation and negotiation among stakeholders about the necessary tradeoffs. This is the reason that early multi-stakeholder engagement is essential as was done throughout the DCI Network AI streams (i.e., CDS, RWE, CH, Governance). Critical stakeholders, like patients and providers, must be given a strong voice early in any governance design process, as they are, symbiotically, the ultimate end-user of healthcare services and delivery.

AI governance models in healthcare must arrive quickly, adapt rapidly (speed), and foster an environment for rapid innovation and growth while improving SEET (capability). To achieve this at a particular level of available funding (resource level), they must limit use cases covered by the model (breadth).

The rapid emergence of AI-powered symptom checker applications (apps) and chatbots available directly to healthcare consumers offers an illustrative case study [8]. Apps can provide initial triage advice based on reported symptoms quickly and conveniently. However, without SEET oversight, concerns remain. A high-speed but low-capability governance approach might rapidly approve new symptom app submissions without deep review [9]. This enables unhindered innovation but risks harm from flawed advice. In contrast, a high-capability, low-speed approach might require a rigorous approval process for new apps that would take several years to design and implement. This would ensure very high levels of safety at the end but would halt the delivery of innovative products during development. Unless very well-resourced, this may further impose an unacceptable burden on innovation afterward. A compromise solution might be to limit breadth: applying slow-to-deploy governance only to those apps that inherently pose serious risks while leaving others either ungoverned or only governed via quick-to-deploy models.

While no governance model can optimize for all three constraints simultaneously (Fig. 1), a coherent selection framework should catalyze a set of tradeoffs that best fit the practical situations they are attempting to govern. Speed is an especially acute concern with GenAI at the time of writing. Innovation with this modality is relentless; any governance approach that takes years to deliver or adapt will miss the mark. Of course, as the triple constraint implies, optimizing for speed involves tradeoffs. A model may arrive too fast for stakeholders to learn and adjust, for a full range of voices to be included in deliberation, or for results to be adequately evaluated. Appropriate selection will balance rapid delivery with adequate planning, deliberation, and evaluation.

Importantly, resource contention continues when making choices within each dimension. For example, deciding how to deliver high **capability** may involve further tradeoffs between governance throughput – how quickly the model can produce decisions – and decision alignment with governance goals (with SEET, in the case of AI in healthcare). To return to the case study, a high-alignment, low-throughput model might require lengthy and ongoing AI app-specific validation studies. In addition to slowing down potentially useful products, such an approach may be prohibitively expensive to all but the best-funded vendors. It may encourage them to form close relationships with regulators, leading to regulatory capture and ultimately hinder innovation. An intermediate-throughput approach might instead require apps to submit a standardized transparency report on data sources, intended use cases, and test results. Apps could launch immediately upon submitting the report but would be required to display usage guidance. A safety review board would retrospectively audit apps and flag those with mismatches between transparency reports and performance. Selecting such an intermediate-throughput approach might strike a pragmatic balance between continuous evolution and adequate safeguards.

4. Recommendations

The recommendations presented were derived from analysis of themes emerging from the use case discussions in the CDS, RWE, and CH working groups. Each domain-specific working group drafted its own recommendations, which were then harmonized and refined by the steering group using the constraint-based framework. Informed by the constraint-based cognitive framework, we can proceed through three layers of analysis to frame our recommendations: (a) the structural characteristics of appropriate governance models, (b) the methods for developing appropriate governance models, (c) preliminary recommendations about the content of plausible governance models for AI in healthcare.

Table 1 shows how each level of recommendation interacts with the proposed application domain of governance. Cross-cutting

Table 1
Structural characteristics, development methods, and preliminary content recommendations for AI governance models by use-case domain.

Gov. Models	(a) Structural Characteristics	(b) Methods for Developing and Implementing	(c) Preliminary Content Recommendations
SHARED / CROSS-CUTTING	In environments of rapid change, speed and adaptability are the driving priorities. Thus, appropriate models must: ▪ In contrast to a more rigid regulatory approach, adopt minimally restrictive approaches that favor cooperation over regulation, modeled on examples such as Leadership in Energy and Environmental Design (LEED) or Underwriters Laboratories (UL). ▪ Be built on a resilient and adaptable foundation to handle the many “unknown unknowns” that are bound to arise. ▪ Be aligned with the needs of specific use cases to be sufficiently narrow and to avoid scope creep. ▪ Integrate diverse perspectives into a harmonious voice, like themes in a symphony.	Methods for model development must also meet speed and adaptability constraints and limit breadth to sufficiently narrow domains. They must incorporate ethical principles to ensure that AI in healthcare aligns with the requirements of fairness, transparency, and patient empowerment. They entail: ▪ Engagement with all stakeholders (AI designers and vendors, patients, consumers, providers, payers, clinicians, and scientists) [10]. ▪ A multidisciplinary approach that brings together informatics, clinical knowledge, behavioral sciences, information technology, and regulatory expertise [11]. ▪ Empowering the voice of the patient or consumer and creating institutional mechanisms that channel their values and preferences to decision-makers [12,13]. ▪ Encourage organizational design that manages the risk of future regulatory capture by industry or other stakeholders [14]. (See expanded discussion in text.)	▪ Establish a multi-stakeholder Health AI Consumer Consortium (see discussion below). ▪ Require high transparency and accountability in AI systems, informing users about potential biases and uncertainties. ▪ Invest in educational programs for healthcare providers and patients to promote understanding and effective use of AI. ▪ Facilitate voluntary adoption of AI accreditation and certification in healthcare by industry, following successful models like LEED, UL, and Energy Star, avoiding the need for heavy regulatory involvement.
Unique to CDS	Use by clinicians in patient care could lead to care and outcomes based on unreliable decision- making if AI phenomena are not better understood. Because CDS can be directly involved with the delivery of care, it promises medical benefits but poses an acute and immediate risk of actual harm to patients. Use is already embedded within complex institutional, professional, and regulatory contexts.	CDS use cases have more focused needs that may require the consensus of local or domain- specific expertise to judge “*what good looks like.*” You may need medical experts with specialty training to inform AI model training.	Clinician end-users need trustworthy systems, and patients need to trust the recommendations provided through CDS.
Unique to RWD/ RWE	The term “Real-world data” (RWD) refers to data gathered during routine healthcare interactions instead of data collected in the context of a research protocol. RWD encompasses a wide variety of information, including electronic health records, insurance claims, diagnostic test results, and patient- generated data from surveys and wearable or implanted devices. Analysis of RWD can provide RWE to advance research, support healthcare decision-making and innovation, and inform public policy. The data quality used for training may be a major challenge in this domain.	RWE is used in a wide variety of healthcare applications. Evaluating population-level effects of medications is one of the best known. However, RWE is also used in predictive modeling for risk and disease progression, clinical trial enrichment or outcomes, assessing and addressing disparities, improving diagnostics, and countless other applications. Decisions must be made by multi- stakeholder communities that include health outcomes researchers, biostatisticians, payors, and patients.	RWD/RWE has a profound effect on model output, so it is essential to ensure that it is fit for purpose, fully documented, as representative as possible of the target use case and population, and that appropriate bias detection and mitigation strategies have been deployed.
Unique to CH	Consumer values and preferences shape usage and policy, not vice versa. Potential queries may be autonomous and individualized.	HCD, from initiation through evolution, is essential for achieving SEET, delivering value, and constructing space for autonomy. Consumers must be empowered to negotiate for their preferences.	A funded Health AI Consumer Consortium must validate products, catalyze patient empowerment, and build literacy.

considerations are in the top row. The next three rows show what is specific or unique to each governed domain.

We expand on the shared/cross-cutting recommendations below. The respective working groups in separate papers detail recommendations for consumer health [1] and clinical decision support [2].

Health AI Consumer Consortium (HAIC2). One critical element for effective AI governance is the establishment of a multi-stakeholder consortium that directly represents patient interests, affords patients an influential voice in steering policies and best practices, and provides continuous learning about AI governance. We propose an independent non-profit entity that can operate as a "Health AI Consumer Consortium." This Consortium would have a mixed board structure encompassing patients, patient advocates, ethicists, clinicians, legal experts, Life Sciences and IT industry leaders, technology leaders, and AI developers. Voting rules would favor control by the patients. Broad consumer participation would be encouraged through tools like digital town halls and localized chapters. With these elements, HAIC2 can act as an influential convening body for developing patient-centered governance models, generating multi-stakeholder consensus, co-developing AI standards, coordinating education campaigns, and providing independent consumer oversight to align innovation with public interests. The Consortium would also identify shared values and concerns around the use of AI in medicine and develop governance principles, guidelines, pilot projects, and other concrete steps reflecting input from all groups. Finally, HAIC2 will develop and maintain a feedback loop to continually reassess and update guidelines based on real-world learning. Throughout, HAIC2 must balance innovation-promoting flexibility with trust- preserving safety and transparency. We present a plan centered on HAIC2 that optimizes for rapid delivery of value in a separate paper covering AI in Consumer Health [1].

Industry cooperation and voluntary accreditation/certification. Industry-led oversight bodies like Leadership in Energy and Environmental Design (LEED), Underwriters Laboratories (UL), and Energy Star demonstrate effective voluntary frameworks for driving standards adoption outside formal regulation. All of these organizations convene diverse stakeholder groups, including competitors, to shape consensus around guidelines and best practices. They incentivize adoption through non-financial rewards like certification status signaling social responsibility or simply through the inherent benefits like energy savings from optimization. Additionally, independence from specific commercial interests lends credibility. Though participation requires some investment, costs are modest compared to alternatives. Continuous learning comes through published product ratings, new use-case expansion, and periodic revisions to standards. Core strengths revolve around industry cooperation in pre-competitive zones. These groups focus on common objectives and consensus before seeking regulatory ceilings, so market forces bolster ethical practices. Their successes lend insights as potential models for coalescing around governance in emerging domains like AI in healthcare.

Transparency, Accountability, and Education. We strongly encourage approaches that provide high levels of transparency and accountability, informing end users about potential biases and uncertainties linked to the system's advice and predictions. We propose industry investment and governmental funding for educational programs for upskilling/training healthcare providers and patients/consumers. Training should teach users to identify accredited/certified AI tools trained on data that is findable, accessible, interoperable, reusable, and trusted (FAIR+) [15–17]. In addition, the educational programs themselves may be accredited and certified through the same industry-led oversight mechanism described above and overseen by the Consumer Consortium. Voluntary accreditation and certification can minimize regulatory delays and overhead for educational programs, technology, and transparency artifacts. For clinical knowledge, medical specialty societies may be the best curators of relevant expertise and may need representation on each clinical use case for education, transparency, usability, and application functions.

4.1. Special challenge: Regulatory capture

Precautions are required to prevent regulatory capture by industry and other influential stakeholders [14]. Regulatory capture occurs when regulatory agencies become dominated by and advance the commercial interests of the industry they oversee over the public interest. Precautions are vital for AI governance bodies to retain independence and objectivity. Some precautions include:

(1) Establish a board structure that prevents any single private sector entity or block from controlling decision-making.
(2) Prohibit direct financial relationships between the governance body and member companies by funding operations from pooled membership dues or independent grants.
(3) Encourage accountability, making all policies and debates publicly accessible rather than permitting closed-door meetings between regulators and corporate lobbyists.
(4) Set explicit term limits for board members and rotating seats to prevent dominance by any permanent in-group membership.

4.2. Governance

To ensure the Health AI Consumer Consortium (HAIC2) is effectively led and governed, the selection process for its leadership will prioritize individuals with a combination of lived patient experience and domain expertise in areas such as digital health, patient advocacy, health policy, or AI ethics. While every individual is a patient, leadership roles within HAIC2 will require demonstrable experience in engaging with healthcare systems, participating in patient-centered research, or contributing to healthcare governance discussions. Building upon the Stakeholder Action and Motivation Incentive Alignment Map (SAM-I-AM) framework we previously developed, HAIC2 will serve as the central mechanism for aligning multi- stakeholder interests while preserving patient-centricity. To maintain credibility and balance, the HAIC2.

leadership selection process will be overseen by an Implementation Steering Committee (ISC) composed of representatives from patient organizations, clinicians, ethicists, informaticians, and industry stakeholders. This committee will establish eligibility criteria for leadership positions, vet nominees, and oversee transparent elections or appointments. The consortium leader will report to an advisory board composed of multi-stakeholder representatives to ensure alignment with public interest and ethical AI governance. To guarantee HAIC2′s effectiveness, clear performance metrics—such as participation in policy recommendations, engagement with regulatory bodies, and the integration of patient feedback into AI governance models—will be established. Additionally, regular audits and an annual impact assessment report will evaluate the consortium's progress and adherence to its objectives, ensuring that HAIC2 remains a transparent and accountable entity in AI governance.

5. Discussion

This paper makes several distinct contributions to the discourse on AI governance amidst a burgeoning AI consumer healthcare technology landscape. First, it offers a novel conceptual framework for analyzing tradeoffs under constraints when selecting governance models for specific AI use contexts. It begins by defining the governance goal as SEET plus innovation, identifying the dimensions – speed, breadth, and capability- of the space from which appropriate governance models can be selected, and proposing three analysis steps for reaching structured recommendations about targeted governance.

Second, the analysis suggests differentiating governance approaches by three key spheres: CDS, RWE, and CH tools. This contrasts with the dominant trend for global guidance that attempts to cover all of AI in Healthcare. Explicitly acknowledging unique constraints, incentives, and stakeholder relationships within each subdomain enables crafting governance aligned with their relevant particulars.

Third, our approach advances the state-of-the-art by generating a consensus perspective through a more collaborative and inclusive process: convening a multi-stakeholder group encompassing an unprecedented diversity of cross-sector perspectives. As a result, our preliminary content recommendations emphasize including often neglected consumer voices as essential partners, allowing them to shape governance models toward their needs, values, and preferences through formal representation and oversight mechanisms. Contributions from ethicists, clinicians, and technology leaders further enhance safety, transparency, and accountability. This proposed approach balances responsive oversight that engenders trust and measured innovation that expands healthcare access. Core principles like HCD and participatory standards development, along with a vision for voluntary accreditation/certification, carry this work beyond previous policy recommendations and approaches.

6. Conclusion

Selecting appropriate governance modes for AI in healthcare is essential to continued progress. However, a better framework is required for conceptualizing the tradeoffs among speed, breadth, and capability for trustworthy innovation that safeguards patients. We propose a constraint-based cognitive framework for balancing multi-stakeholder perspectives and offer a three-layer analysis for reaching structured recommendations. Our analysis suggests models should be organized by application domain, negotiated through multi-stakeholder engagement, centered on empowered patient voices, focused on transparency and accountability (initially through voluntary accreditation and certification), and emphasize continuous learning.

While directionally promising, the ideas about model content require refinement. The next steps include formally convening a representative assembly of stakeholders, devising transparent accreditation standards across risk levels, and piloting voluntary certification processes. Gradually nurturing these structures over the next few years can facilitate the orderly emergence of proactive governance rather than reactive policymaking.

The following changes would help achieve more responsible governance of AI in healthcare: (1) the framework for evaluating governance models for tradeoffs between speed, breadth, and capability is well-understood by stakeholder organizations; (2) appropriate models are designed, evaluated, and selected by multiple organizations, and their use becomes routine; (3) CDS, RWE and CH domains select differentiated governance models that fit the domain constraints, incentives, and stakeholder relationships; (4) the selected models are adaptable and resilient to change, as well as empirically sound, well-documented, and continually updated; and (5) the transparency artifacts generated in response to the models are useful and dynamic.

We hope this paper generates timely conversations about appropriate models, how we create them, what they must include, and the boundaries of their operative domains. We encourage all stakeholders, from AI developers and industry leaders to healthcare policymakers and end-users, to engage in the ongoing conversation surrounding AI governance in healthcare. We must collectively take action to pave the way toward a responsible future where AI improves healthcare delivery while safeguarding patient well-being and nurturing the blueprint for trust.

CRediT authorship contribution statement

Leon Rozenblit: Writing – review & editing, Writing – original draft, Investigation, Methodology, Conceptualization. **Amy Price:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Anthony Solomonides:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Amanda L. Joseph:** Investigation, Methodology, Writing – original draft, Writing – review & editing. **Eileen Koski:** Writing – review & editing, Methodology, Investigation. **Gyana Srivastava:** Writing – review & editing, Investigation, Formal analysis. **Steven Labkoff:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **David Bray:** Writing – review & editing, Methodology, Investigation. **Monica Lopez-Gonzalez:** Writing – review & editing, Methodology, Investigation. **Reva Singh:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Dave deBronkart:** Writing – review & editing, Investigation. **Paul J. Barr:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Peter Szolovits:** Writing – review & editing, Investigation. **Kiran Dattani:** Writing – review & editing, Methodology, Investigation. **Charles Jaffe:** Writing – review & editing, Conceptualization. **Douglas Fridsma:** Writing – review & editing, Investigation. **Russell Baris:** Writing – review & editing, Investigation. **Russell Leftwich:** Writing – review & editing, Methodology, Investigation. **Robert Stolper:** Writing – review & editing, Investigation. **Mark G. Weiner:** Writing – review & editing, Methodology, Investigation. **Nuria Pastor:** Writing – review & editing, Conceptualization. **Unai Sanchez Luque:** Writing – review & editing. **Baihan Lin:** Writing – review & editing. **Tien Thi Thuy Bui:** Writing – review & editing, Investigation, Conceptualization. **Bilikis Oladimeji:** Writing – review & editing, Investigation. **Tayler Williams:** Writing – review & editing. **Gretchen Purcell Jackson:** Writing – review & editing. **Pei-Yun Sabrina Hsueh:** Writing – review & editing. **Yuri Quintana:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Steven Labkoff is an employee of Quantori, LLC. Eileen Koski is employed by and owns stock in IBM. David Bray is the Principal and CEO at LeadDoAdapt, consultant to Health Level Seven's AI Taskforce, and is a periodic keynote speaker on Healthcare, AI, and related issues. Douglas Fridsma is an employee of Health Universe. Russell Leftwich is an employee of InterSystems Corporation. Bilikis Oladimeji is an employee and own stocks/stock options of UnitedHealth Group. Gretchen Purcell Jackson is an employee of Intuitive Surgical. Gretchen owns stock and/or stock options for IBM, Kyndryl, and Intuitive Surgical. Gretchen was the Past President and Past Board Chair of the American Medical Informatics Association. Pei-Yun Sabrina Hsueh is an employee of Pfizer Inc. Sabrina owns stock and/or stock options for IBM, Bayesian Health, and Pfizer. Sabrina is on the Practitioners Board of Association for Computing Machinery (ACM). The authors had no other conflicts or competing interests to disclose. All opinions are the authors' and not the view of my employer or affiliated organizations.

Acknowledgments

We would like to thank Dr. David Bray, Dr. Tiffani J. Bright, and Dr. Isaac Kohane, for their keynote talks at the *Blueprints for Trust* conference [4] conference. We thank Dr. David Bray for his initial input in early discussions on the framing of this paper. We would like to thank Rabbi Daniel Cohen and Reverend Greg Doll for their webinar on Bioethics and AI, Dr. Luke Sato, Dr. Paul R. DeMuro, and Kenneth E. White, JD for their webinar on AI in Healthcare: Risk Management, Trust, and Liability, Dr. Amy Price and Dave deBronkart (ePatient Dave) for their webinar on AI in Healthcare: The Patient Perspective, and Dr. Anne-Marie Meyer, Mark Shapiro, and Rob Stolper for their webinar: AI in Healthcare: Real World Data Generation And The Regulatory Perspective. Lastly, we would like to thank Dr. Charles Safran for his input on later versions of the paper as well as his guidance at the DCI Network. This study received no funding.

Code availability

No computer code was produced or analyzed for this article.

Data availability

All data used in this article has been reported in this paper. There is no other data associated with this paper. No datasets were generated or analyzed during the current study.

References

[1] L. Rozenblit, A. Price, A. Solomonides, A.L. Joseph, G. Srivastava, S. Labkoff, D. deBronkart, R. Singh, K. Dattani, M. Lopez-Gonzalez, P.J. Barr, E. Koski, B. Lin, E. Cheung, M.G. Weiner, T. Williams, T.T. Thuy Bui, Y. Quintana, Towards a Multi-Stakeholder process for developing responsible AI governance in consumer health, Int. J. Med. Inf. 195 (2025) 105713, https://doi.org/10.1016/j.ijmedinf.2024.105713.

[2] S. Labkoff, B. Oladimeji, J. Kannry, A. Solomonides, R. Leftwich, E. Koski, A. L. Joseph, M. Lopez-Gonzalez, L.A. Fleisher, K. Nolen, S. Dutta, D.R. Levy, A. Price, P.J. Barr, J.D. Hron, B. Lin, G. Srivastava, N. Pastor, U.S. Luque, T.T.T. Bui, R. Singh, T. Williams, M.G. Weiner, T. Naumann, D.F. Sittig, G.P. Jackson, Y. Quintana, Toward a responsible future: Recommendations for AI-enabled clinical decision support, J. Am. Med. Inform. Assoc. 31 (11) (2024 Nov 1) 2730–2739, https://doi.org/10.1093/jamia/ocae209. PMID: 39325508; PMCID: PMC11491642.

[3] DCI Network. Blueprints for Trust: Best practices and Regulatory Pathways for ethical AI in healthcare, September 21-22, 2023, Harvard Faculty Club, Harvard University, Boston, MA, USA. Retrieved January 22, 2024. Available from: https://www.dcinetwork.org/events/192.

[4] D.C.I. Network, Recordings of session from blueprints for trust: Best practices and Regulatory Pathways for ethical AI in healthcare conference, Available from: Retrieved March 8 (2024) https://www.dcinetwork.org/node/313.

[5] R. Iniesta, The human role to guarantee an ethical AI in healthcare: a five-facts approach, AI Ethics (2023), https://doi.org/10.1007/s43681-023-00353-x.

[6] C. Van Wyngaard, Theory of the triple constraint — A conceptual review, in: 2012 IEEE International Conference on Industrial Engineering and Engineering Management, 2012, pp. 1991–1997, https://doi.org/10.1109/IEEM.2012.6838095.

[7] L. Rozenblit, A constraints-based cognitive framework for selecting governance models. [Accessed on 2024- 02-01] Available from: https://open.substack.com/pub/knowledgerefinery/p/a-constraints-based-cognitive- framework.

[8] C.J. Wiedermann, A. Mahlknecht, G. Piccoliori, A. Engl, Redesigning primary care: The emergence of artificial- intelligence-driven symptom diagnostic tools, J. Personalized Medi. 13 (9) (2023) 1379, https://doi.org/10.3390/jpm13091379.

[9] Y. Quintana, S. Lu, G. Srivastava, S. Kennedy, Bridging the gap: A comprehensive review and public consultation on what is needed to ensure the quality, safety, transparency, and effectiveness of apps for mental health and substance use health, Homewood Res. Inst. 2024 (2024), https://doi.org/10.13140/RG.2.2.15300.6464. Available at https://hriresearch.com/bridging-the-gap-report-scc/.

[10] F. Delgado, S. Yang, M. Madaio, Q. Yang, Stakeholder participation in AI: Beyond“ Add Diverse Stakeholders and Stir” (2021), arXiv preprint arXiv:2111.01122.

[11] S.W. Smye, A.F. Frangi, Interdisciplinary research: Shaping the healthcare of the future, Future Healthc J. 8 (2) (2021 Jul) e218–e223, https://doi.org/10.7861/fhj.2021-0025. PMID: 34286188; PMCID: PMC8285142.

[12] P.B. Bhattad, L. Pacifico, Empowering patients: Promoting patient education and health literacy, Cureus 14 (7) (2022 Jul 27) e27336, https://doi.org/10.7759/cureus. 27336. PMID: 36043002; PMCID: PMC9411825.

[13] The Power of the Patient Voice: How Health Care Organizations Empower Patients and Improve Care Delivery. Available from: https://cssjs.nejm.org/landing-page/cj-ebook-2021/The-Power-of-the-Patient-Voice.pdf.

[14] O. Chisholm, H. Critchley, Future directions in regulatory affairs, Front. Med. (Lausanne). 9 (9) (2023 Jan) 1082384, https://doi.org/10.3389/fmed.2022.1082384. PMID: 36698838; PMCID: PMC9868628.

[15] M. Wilkinson, M. Dumontier, I. Aalbersberg, et al., The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (2016) 160018, https://doi.org/10.1038/sdata.2016.18.

[16] J.E. Richardson, B. Middleton, J.E. Platt, B.H. Blumenfeld, Building and maintaining trust in clinical decision support: Recommendations from the Patient-Centered CDS Learning Network, Learn. Health. Syst. 4 (2) (2019 Dec 11) e10208, https://doi.org/10.1002/lrh2. 10208. PMID: 32313835; PMCID: PMC7156865.

[17] D. Lin, J. Crabtree, I. Dillo, et al., The TRUST principles for digital repositories, Sci. Data 7 (2020) 144, https://doi.org/10.1038/s41597-020-0486-7.