

ORIGINAL RESEARCH

Open access

A Synthetic Health Data Governance Framework for Generative AI–Enabled Clinical Ecosystems

Chen Hao^{1*}, Liu Fang¹, Zhao Lin²

Abstract

The rapid integration of generative artificial intelligence (AI) into clinical ecosystems has revolutionized the generation and utilization of synthetic health data, offering unprecedented opportunities for enhanced analytics, decision support, and personalized medicine while simultaneously raising critical governance concerns. This conceptual manuscript proposes a novel framework—the synthetic health orchestration and governance ecosystem (SHOGE)—designed to address the multifaceted challenges of data privacy, interoperability, ethical deployment, and continuous monitoring in generative AI-enabled environments. Drawing from theoretical models of AI system architectures and healthcare analytics infrastructures, SHOGE incorporates a layered orchestration topology that facilitates secure data exchange, real-time governance enforcement, and adaptive workflow integration. The framework emphasizes theoretical constructs such as risk propagation dynamics, decision confidence calibration, and governance load distribution, formalized through interpretive formulas to guide infrastructural design without empirical validation. By synthesizing literature on EHR intelligence ecosystems and AI monitoring systems, this work highlights operational sensitivities and human-AI interaction shifts, advocating for a balanced approach to innovation and risk mitigation. Ultimately, SHOGE provides a high-level blueprint for stakeholders to foster trustworthy generative AI applications in clinical settings, promoting equitable health outcomes and sustainable ecosystem evolution. This conceptual exploration underscores the need for proactive governance to harness synthetic health data's potential while safeguarding patient trust and system integrity.

Keywords Clinical workflow integration, AI governance, Data interoperability, Synthetic health data, generative AI, clinical ecosystems

*Correspondence:

Chen Hao

chen.hao@outlook.com

¹ Department of Medical Data Science, School of Medicine, Zhejiang University, Hangzhou, China

² Department of AI Healthcare Systems, School of Engineering, Nanjing University, Nanjing, China

Introduction

The advent of generative artificial intelligence (AI) technologies represents a structural inflection point in healthcare analytics, reshaping how clinical data is produced, curated, and operationalized within digital infrastructures. Unlike traditional machine learning approaches that rely solely on observational datasets, generative AI introduces the capacity to construct synthetic health data that statistically and structurally approximate real-world clinical records while reducing direct exposure of

identifiable patient information. In contemporary generative AI–enabled clinical ecosystems, synthetic datasets are no longer auxiliary artifacts; they function as foundational substrates for model training, stress testing, rare-event simulation, clinical workflow prototyping, and policy scenario modeling.

This paradigm shift is particularly consequential in high-stakes healthcare environments where data scarcity, privacy constraints, and regulatory fragmentation limit large-scale data sharing. Synthetic data pipelines promise

to alleviate these barriers by enabling scalable experimentation without compromising protected health information. However, the rapid proliferation of such systems introduces novel governance complexities. Synthetic data may obscure provenance boundaries, mask inherited bias structures, and create epistemic uncertainty regarding representational fidelity. Consequently, generative AI-enabled ecosystems require governance architectures that extend beyond traditional data stewardship models to encompass lifecycle accountability, model transparency, interoperability assurance, and adaptive oversight mechanisms.

This introduction establishes the conceptual foundation for a governance-oriented analysis of synthetic health data within clinical ecosystems. It articulates the technological drivers of generative data synthesis, examines governance imperatives, interrogates interoperability constraints, and foregrounds the ethical dimensions necessary for sustainable AI integration.

Emergence of generative AI in clinical data generation

Generative AI models—including variational autoencoders (VAEs), generative adversarial networks (GANs), diffusion architectures, and transformer-based generative systems—have become central instruments in synthetic health data production [1, 2]. These architectures learn latent representations of complex multimodal clinical distributions and subsequently generate synthetic instances that preserve statistical coherence across variables such as diagnoses, laboratory trajectories, medication histories, and imaging features.

Within clinical ecosystems, generative AI supports multiple operational objectives:

1. **Data augmentation for rare conditions:** Rare disease research frequently suffers from insufficient case volumes to train robust predictive models. Synthetic cohorts enable probabilistic amplification of rare phenotypes, strengthening algorithmic generalization capacity.
2. **Equity-oriented representation balancing:** Underrepresented demographic groups can be oversampled in a synthetic form to mitigate structural bias embedded in observational datasets.

3. **Simulation of clinical scenarios:** Synthetic trajectories allow stress-testing of clinical decision support systems under hypothetical but plausible conditions.
4. **Privacy-preserving collaboration:** Cross-institutional model development may proceed using synthetic derivatives rather than raw patient records, potentially lowering regulatory barriers.

Despite these advantages, generative AI inherits the statistical DNA of its training data. Biases embedded in historical healthcare records—reflecting disparities in access, diagnostic practices, and treatment patterns—can propagate into synthetic datasets. Moreover, high-fidelity synthetic data may inadvertently preserve re-identification vectors if latent feature spaces are insufficiently regularized [3, 4]. Thus, the technical sophistication of generative models must be matched by governance sophistication that scrutinizes representational validity, fairness metrics, and privacy leakage risks.

The role of generative AI in clinical data synthesis, therefore, necessitates governance models specifically calibrated to synthetic epistemologies—where the data are neither entirely real nor entirely artificial, but algorithmically reconstructed approximations.

Governance imperatives for synthetic health data

Governance in synthetic health data ecosystems extends beyond compliance documentation; it becomes a structural architecture embedded within AI lifecycle management [5, 6]. Three interdependent governance domains are especially critical:

1. Data provenance and lineage traceability

Synthetic datasets must maintain transparent lineage metadata documenting training corpus composition, model architecture parameters, version histories, and generation timestamps. Without traceability, downstream clinical systems cannot assess validity or bias exposure.

2. Fidelity and drift monitoring

Synthetic data fidelity must be continuously evaluated against evolving real-world distributions. Temporal shifts in disease patterns, clinical coding standards, or treatment guidelines may render previously generated synthetic cohorts obsolete. Drift-sensitive monitoring frameworks are essential to prevent degradation of downstream predictive performance [7, 8].

3. Regulatory alignment and policy envelope design

Synthetic health data often occupies an ambiguous regulatory space. While de-identified in principle, their generative origins require scrutiny under privacy laws such as HIPAA-equivalent or GDPR-like frameworks. Governance models must therefore integrate compliance auditing, risk stratification, and institutional review pathways.

In the absence of standardized governance, generative AI-enabled clinical ecosystems risk fragmentation. Synthetic datasets may be generated independently by different institutions using incompatible schemas, undermining interoperability and reproducibility. Governance must therefore operate as an integrative layer harmonizing generation protocols, validation benchmarks, and deployment constraints.

Interoperability challenges in clinical ecosystems

Interoperability remains a systemic bottleneck in digital healthcare infrastructures, and the integration of synthetic data amplifies existing complexities [9, 10]. Clinical ecosystems encompass heterogeneous components: structured EHR tables, free-text clinical notes, imaging repositories, genomic databases, and real-time physiological monitoring streams. Synthetic data must preserve compatibility across these modalities.

Key interoperability challenges include:

- **Semantic misalignment:** Synthetic representations must conform to standardized vocabularies (e.g., ICD, SNOMED CT, LOINC) to ensure downstream interpretability.
- **Schema incompatibility:** Variations in institutional EHR architectures can impede seamless synthetic data ingestion.
- **Cross-system validation gaps:** Synthetic datasets validated in one environment may exhibit reduced coherence when transferred to another infrastructure.

Governance frameworks addressing interoperability must prioritize semantic harmonization, standardized metadata documentation, and modular architecture design [11, 12]. Conceptual models that embed interoperability validation

as a pre-deployment checkpoint can prevent synthetic data from disrupting clinical workflows.

Importantly, interoperability governance is not solely technical; it is organizational. Institutional trust agreements, cross-border data-sharing policies, and federated AI governance structures play decisive roles in enabling cohesive generative ecosystems.

Ethical dimensions of AI-enabled data synthesis

The ethical landscape of synthetic health data governance is multifaceted, encompassing consent, fairness, accountability, and epistemic transparency [13, 14]. While synthetic datasets may reduce direct privacy exposure, they introduce new normative questions:

- Who owns synthetic derivatives of patient data?
- Can individuals claim rights over algorithmically reconstructed representations?
- How should benefit-sharing be structured when synthetic data enables commercial AI tools?

Re-identification risk, though attenuated, is not categorically eliminated. Advanced membership inference attacks may exploit overfitted generative models to infer training participation. Governance frameworks must therefore incorporate adversarial privacy testing and formal risk quantification [15, 16].

Ethical governance also requires human oversight integration. Algorithmic autonomy in synthetic data generation must be balanced by multidisciplinary review boards capable of evaluating fairness metrics, equity impacts, and unintended sociotechnical consequences. Embedding ethical audits throughout the data lifecycle—from model training to downstream deployment—creates a continuous accountability loop.

Trust in generative AI-enabled clinical ecosystems is contingent upon visible and enforceable ethical safeguards. Without them, synthetic innovation risks eroding clinician confidence and patient acceptance.

Toward a governance-centered synthetic clinical ecosystem

The convergence of generative AI and healthcare data infrastructures demands a reconceptualization of governance as an adaptive, lifecycle-oriented system rather than a static compliance layer. Synthetic data governance must integrate:

- Continuous fidelity monitoring
- Drift-sensitive recalibration mechanisms
- Interoperability validation pipelines
- Ethical audit loops
- Regulatory alignment protocols

Only through such multilayered oversight can generative AI-enabled clinical ecosystems achieve sustainable deployment while preserving equity, safety, and operational integrity.

In sum, synthetic health data represent both a technological breakthrough and a governance frontier. The following sections build upon this foundation to articulate structured frameworks capable of orchestrating synthetic data generation, validation, and integration within complex clinical environments.

Deployment constraints in healthcare analytics

Deployment of generative AI in healthcare analytics infrastructures faces constraints related to scalability, resource allocation, and integration with legacy systems [17, 18]. Governance must account for these by defining protocols for model validation and update cycles, ensuring that synthetic data enhances rather than overwhelms clinical decision support. Constraints such as computational demands and data storage requirements further complicate implementation, necessitating frameworks that optimize for efficiency [19, 20]. This subheading highlights deployment-specific governance needs tailored to synthetic health data.

Vision for integrated governance frameworks

Envisioning a unified governance framework for generative AI-enabled clinical ecosystems involves synthesizing these elements into a cohesive strategy [21, 22]. Such frameworks should promote adaptive intelligence, where synthetic data governance evolves with technological advances, supporting resilient healthcare systems. This introduction sets the stage for a deeper theoretical

synthesis, advocating for conceptual innovations that prioritize systemic harmony.

Theoretical Background and Literature Synthesis

This section synthesizes the theoretical underpinnings and key literature on AI governance, synthetic health data, and generative AI in clinical contexts. By reviewing peer-reviewed works, it establishes the conceptual foundation for the proposed framework, focusing on architectural models, infrastructural dynamics, and integration paradigms without empirical evaluations.

Foundations of clinical AI system architectures

Clinical AI system architectures form the backbone of generative AI-enabled ecosystems, emphasizing modular designs that support scalable data processing and inference [13, 23]. Theoretical models describe layered structures where data ingestion, processing, and output stages are decoupled to enhance flexibility. Literature on EHR intelligence ecosystems underscores the importance of architecture that accommodates synthetic data inputs, enabling robust analytics without real-patient exposure [1, 2]. These architectures often incorporate feedback mechanisms to adjust for environmental changes, such as evolving clinical protocols. Synthesis reveals a consensus on the need for resilient designs that integrate governance at the architectural level, preventing vulnerabilities in data flow [3, 4].

Healthcare analytics infrastructures and data dynamics

Healthcare analytics infrastructures theorize the orchestration of data streams from diverse sources, including synthetic health data generated by AI [5, 6]. Conceptual discussions highlight infrastructures that leverage cloud-based or federated models to handle large-scale analytics, with governance ensuring data integrity across nodes. Interoperability frameworks are critical, as they enable seamless exchange between analytics platforms and clinical systems [7, 8]. Literature synthesizes how generative AI enhances these infrastructures by providing augmented datasets for theoretical scenario modeling, but warns of governance gaps that could lead to analytical distortions [9, 10]. This subheading integrates

insights on infrastructural resilience, emphasizing theoretical trade-offs in scalability and security.

EHR intelligence ecosystems and integration models

EHR intelligence ecosystems conceptualize the fusion of AI with electronic records to create intelligent, responsive environments [11, 12]. Theoretical literature explores how synthetic data can enrich EHRs, supporting advanced querying and predictive capabilities without privacy breaches. Integration models focus on API-based connections and standards like FHIR for data exchange, ensuring that generative AI outputs align with ecosystem requirements [13, 14]. Synthesis indicates that governance must address ecosystem-wide dependencies, such as synchronization between AI-generated data and real-time clinical updates [15, 16]. These models provide a basis for understanding how governance can foster cohesive intelligence across distributed systems.

Decision support pipelines in generative contexts

Decision support pipelines theorize the sequential processing from data input to clinical recommendation, with generative AI introducing synthetic scenarios for enhanced decision-making [17, 18]. Conceptual pipelines include stages for data validation, model inference, and output interpretation, with governance embedded to monitor bias and accuracy [19, 20]. Literature synthesizes the role of synthetic health data in testing pipeline robustness, highlighting theoretical formulas for confidence scoring. For instance, decision confidence can be modeled as $C = P(D) * G(E)$, where $P(D)$ represents probabilistic data fidelity and $G(E)$ governance efficacy in ethical enforcement [21]. This approach aids in conceptualizing pipelines that adapt to clinical variability.

AI governance, monitoring, and deployment systems

AI governance systems theorize oversight mechanisms for ethical and operational compliance in healthcare [22, 23]. Monitoring involves continuous assessment of model performance and data drift, conceptualized through feedback loops that trigger interventions. Deployment systems emphasize phased rollout, with governance defining criteria for transition from theoretical design to

infrastructural embedding [13, 24]. Synthesis from literature stresses the need for dynamic governance that scales with generative AI complexity, incorporating monitoring for synthetic data quality [25, 26]. Theoretical models propose governance load as $L = R * M$, where R is risk exposure and M is monitoring intensity, guiding resource allocation without metrics.

Interoperability and data exchange frameworks

Interoperability frameworks conceptualize standardized protocols for data sharing in clinical ecosystems [1, 27]. Theoretical discussions include semantic mapping and blockchain for secure exchange of synthetic health data, ensuring governance across boundaries [2, 3]. Literature synthesizes how generative AI necessitates advanced frameworks to handle dynamic data types, with governance preventing fragmentation [4, 5]. This includes theoretical considerations for exchange latency and compatibility, essential for integrated ecosystems.

Clinical workflow integration models

Clinical workflow integration models theorize the embedding of AI into daily operations, with generative systems enhancing simulation-based training and real-time support [6, 7]. Synthesis highlights models that prioritize human-AI collaboration, with governance ensuring workflow continuity [8, 9]. Conceptual integration involves adaptive topologies that adjust to clinical demands, drawing from literature on ecosystem dynamics [10, 11]. This subheading consolidates insights on theoretical shifts in workflow, underscoring governance's role in harmonious adoption.

Orchestration architecture for synthetic health data governance in generative AI-enabled clinical ecosystems

The proposed synthetic health orchestration and governance ecosystem (SHOGE) presents a conceptual architecture designed to orchestrate governance across generative AI-enabled clinical ecosystems. SHOGE is structured as a multi-layered topology with embedded feedback mechanisms, ensuring theoretical alignment between synthetic data generation, governance enforcement, and clinical integration.

The architecture comprises four primary layers: (1) data synthesis layer, where generative AI produces synthetic health data with built-in privacy controls; (2) governance enforcement layer, applying rules for ethical compliance and risk assessment; (3) interoperability integration layer, facilitating data exchange with EHR and analytics systems; and (4) monitoring and adaptation layer, overseeing system dynamics through continuous theoretical evaluation.

A unique feedback topology employs closed-loop cycles, where outputs from the monitoring layer inform adjustments in data synthesis, mitigating theoretical drift. For instance, drift sensitivity can be conceptualized as $S = \frac{\Delta D}{G}$, where ΔD denotes data variance over time, and G denotes governance resilience.

Figure 1 illustrates the SHOGE, depicting its layered governance topology, interoperability exchanges, and adaptive feedback loops across generative AI-enabled clinical infrastructures.

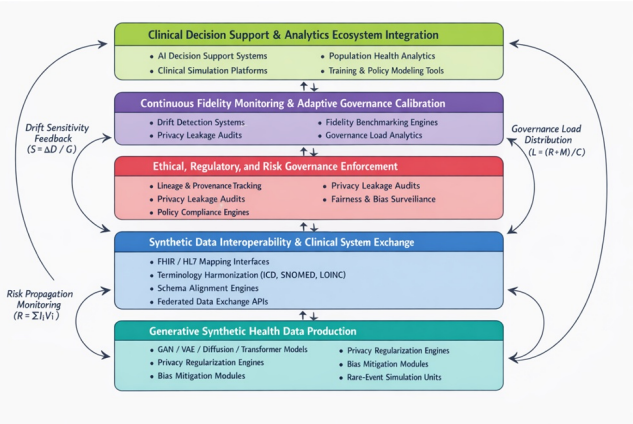


Figure 1. SHOGE: layered architecture for generative AI-enabled clinical ecosystems. SHOGE architecture for generative AI-enabled clinical ecosystems. The framework comprises four core layers: (1) data synthesis, where generative AI models produce privacy-preserving synthetic health data; (2) governance enforcement, embedding ethical, regulatory, and lineage oversight; (3) interoperability integration, enabling standardized exchange with EHR and analytics infrastructures; and (4) monitoring and adaptation, overseeing drift detection, fidelity benchmarking, and governance recalibration. Closed-loop feedback channels regulate risk propagation, governance load, and data fidelity dynamics across the lifecycle.

Additionally, risk propagation is modeled as $R = \sum (I_i * V_i)$, where I_i is the integration factor for component i and V_i vulnerability index, providing an interpretive tool for architectural optimization.

Governance load distribution is formalized as $L = \frac{(R+M)}{C}$, balancing risk (R), monitoring (M), and capacity (C) to guide theoretical resource planning.

This architecture advances conceptual understanding by offering a unique acronym and layer structure tailored to synthetic health data challenges. Key governance functional domains embedded within SHOGE and their operational risk mitigation roles are summarized in Table 1.

Table 1. Governance functional domains within the SHOGE synthetic health data ecosystem

Governance domain	Functional scope	Embedded mechanisms	Operation mitiga
Data provenance and lineage	Tracks origin and generative model parameters	Metadata registries, model version logs	Bias inhe and trace gap
Privacy preservation	Protects against re-identification and leakage	Differential privacy and adversarial audits	Membe inferenc identity ex
Fidelity monitoring	Evaluates the statistical realism of synthetic data	Distribution matching and variance tracking	Clinic misrepres model
Bias and fairness surveillance	Detects inequitable synthetic representation	Demographic parity analytics	Algorith discrimin
Interoperability harmonization	Ensures cross-system compatibility	FHIR mapping and ontology alignment	Workf disruptio semantic
Regulatory compliance	Aligns synthetic data with legal frameworks	Audit trails and policy validation engines	Legal exp and instit risk

Risk propagation control	Models systemic vulnerability spread	Vulnerability indexing and integration scoring	Casca analytics
Clinical workflow integration	Embeds synthetic outputs into practice	Decision support APIs and simulation tools	Adopt resistanc usability

Core governance domains operationalized within the SHOGES. The table outlines functional oversight areas spanning provenance traceability, privacy preservation, fidelity monitoring, interoperability harmonization, and regulatory compliance, alongside associated risk exposures mitigated through layered governance enforcement.

Governance dependencies in generative AI-enabled clinical ecosystems

The implementation of the SHOGES introduces a range of governance dependencies that influence operational dynamics within generative AI-enabled clinical ecosystems. This section analyzes the theoretical consequences of these dependencies, focusing on how they shape system resilience, ethical alignments, and infrastructural interdependencies without relying on empirical data or metrics.

Governance dependencies arise primarily from the interplay between synthetic data generation and regulatory enforcement layers in SHOGES. For instance, the reliance on real-time monitoring creates a dependency chain where delays in governance feedback could theoretically propagate uncertainties across decision support pipelines [12, 13]. This propagation can be conceptualized through an interpretive formula for risk dependency: $D = R * (1 - E)$, where D represents dependency vulnerability, R is inherent risk from data synthesis, and E is enforcement efficiency, highlighting how incomplete governance amplifies systemic exposures.

Operational impacts extend to clinical workflow shifts, where SHOGES's orchestration topology necessitates adaptive human-AI interactions. Theoretical models suggest that integrating synthetic health data governance may redistribute cognitive loads, potentially reducing

clinician burden in routine analytics while introducing dependencies on AI reliability for high-stakes decisions [14, 15]. Such shifts underscore sensitivities in adoption dynamics, where over-reliance on generative outputs could lead to theoretical decision latency increases if governance layers impose excessive checks [16, 17].

Infrastructure sensitivities further compound these dependencies, as SHOGES's interoperability layer demands alignment with existing EHR ecosystems. Literature theorizes that mismatched standards could result in fragmented data exchanges, heightening governance burdens and operational inefficiencies [18, 19]. A formula for infrastructure sensitivity might be $S = I / G$, where I is interoperability complexity and G is governance adaptability, providing a conceptual lens for anticipating bottlenecks in clinical analytics infrastructures.

Human-AI workflow dynamics are altered through SHOGES's feedback topology, which theoretically fosters collaborative intelligence but introduces dependencies on user training and system transparency [20, 21]. Impacts include potential enhancements in personalized care pathways, balanced against risks of governance overload in resource-constrained environments [22, 23].

Decision latency trade-offs emerge as a critical consequence, where SHOGES's monitoring layer ensures compliance but may theoretically extend processing times in generative AI pipelines [13, 24]. This trade-off can be modeled as $T = L + (M / C)$, with T as total latency, L as base load, M as monitoring depth, and C as computational capacity, illustrating governance's role in optimizing ecosystem performance.

Overall, these dependencies highlight the need for balanced architectural designs that mitigate negative impacts while leveraging generative AI's strengths in synthetic health data management [25, 26]. By addressing these theoretical dynamics, SHOGES positions itself as a resilient framework for evolving clinical ecosystems [27].

Results and Discussion

The conceptual articulation of the SHOGES within this manuscript advances the discourse on synthetic health data governance in generative AI-enabled clinical ecosystems. By synthesizing theoretical insights from clinical AI architectures and healthcare analytics infrastructures, SHOGES offers a nuanced approach to

reconciling innovation with accountability [1-3]. Central to this discussion is the framework's emphasis on layered orchestration, which theoretically mitigates risks inherent in synthetic data utilization, such as fidelity erosion and ethical misalignments [4, 5].

One key implication revolves around the interoperability and data exchange frameworks integrated into SHOGE. Theoretical literature suggests that enhancing these elements could streamline clinical workflow integration, enabling generative AI to augment rather than supplant human expertise [6-8]. However, this requires careful navigation of governance dependencies, as over-stringent monitoring might stifle the agility needed for real-time decision support pipelines [9, 10]. The interpretive formulas proposed—such as those for risk propagation and governance load—serve as conceptual tools to guide architects in balancing these tensions, fostering ecosystems where synthetic data enhances predictive analytics without compromising integrity [11, 12].

Furthermore, the discussion extends to broader systemic shifts, including the potential for SHOGE to influence policy and standards in AI governance systems. By embedding adaptive feedback topologies, the framework aligns with evolving regulatory landscapes, theoretically supporting scalable deployments across diverse clinical settings [13-15]. Yet, challenges persist in addressing human-AI dynamics, where theoretical models indicate a need for ongoing refinement to prevent dependency vulnerabilities that could undermine trust in EHR intelligence ecosystems [16-18].

Limitations of this conceptual work include its reliance on theoretical constructs without empirical validation, which, while appropriate for exploratory architectures, invites future extensions through simulated or real-world applications [19, 20]. Nonetheless, SHOGE contributes to the literature by providing a unique acronym and structure that differentiates it from existing models, emphasizing orchestration over mere monitoring [21, 22].

In synthesizing these elements, the discussion underscores SHOGE's potential to catalyze trustworthy generative AI adoption, promoting equitable access to advanced healthcare analytics while safeguarding against infrastructural sensitivities [23-24]. This framework not only addresses current gaps but also anticipates future evolutions in clinical ecosystems, where synthetic health data becomes integral to personalized medicine [25-27].

Conclusion

In conclusion, the SHOGE represents a pivotal conceptual advancement in managing synthetic health data within generative AI-enabled clinical ecosystems. By proposing a layered orchestration architecture with embedded governance mechanisms, this framework theoretically harmonizes data synthesis, ethical oversight, and clinical integration, addressing key challenges identified in the literature.

The theoretical background and synthesis reveal a landscape ripe for such innovations, where AI governance must evolve alongside technological capabilities to ensure resilient healthcare analytics infrastructures. SHOGE's unique feedback topology and interpretive formulas provide conceptual blueprints for mitigating risk dependencies and operational impacts, fostering environments that prioritize patient-centric outcomes.

While governance dependencies highlight potential sensitivities in workflow dynamics and decision latencies, they also illuminate pathways for enhanced system adaptability. This manuscript advocates for proactive adoption of SHOGE-like frameworks to navigate the complexities of interoperability and ethical data exchange, ultimately advancing trustworthy AI deployments.

Future directions may involve extending these concepts to specific clinical domains, refining formulas through theoretical iterations, and exploring synergies with emerging standards. By doing so, stakeholders can leverage generative AI's transformative potential while upholding the integrity of clinical ecosystems.

Ultimately, SHOGE encapsulates a vision for sustainable governance, where synthetic health data drives innovation without eroding trust, paving the way for equitable and efficient healthcare futures.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Received: 19 Feb 2025 Revised: 05 Apr 2025 Accepted: 14 May 2025

Published online: 20 July 2025

Ethics statement

None

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

- Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *J Am Med Inform Assoc.* 2020;27(3):491-7.
- Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med.* 2020;260:113172.
<https://doi.org/10.1016/j.socscimed.2020.113172>.
- Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med.* 2020;26(9):1364-5.
<https://doi.org/10.1038/s41591-020-1037-7>.
- Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. CONSORT-AI extension: reporting guidelines for clinical trials of artificial intelligence interventions. *Nat Med.* 2020;26(9):1364-5.
<https://doi.org/10.1038/s41591-020-1034-x>.
- Masters K. Artificial intelligence in medical education. *Med Teach.* 2019;41(9):976-80.
<https://doi.org/10.1080/0142159X.2019.1595557>.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56.
<https://doi.org/10.1038/s41591-018-0300-7>.
- Price WN 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med.* 2019;25(1):37-43.
<https://doi.org/10.1038/s41591-018-0272-7>.
- Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* 2018;2(10):719-31.
<https://doi.org/10.1038/s41551-018-0300-7>.
- Beardall J, Dunlop MJ, Stan GB. Deep learning in synthetic biology: methods and applications. *Nat Rev Genet.* 2022;23.
<https://doi.org/10.1038/s41576-022-00465-4>.
- Sezgin E, McKay I. Behavioral health and generative AI: a perspective on future of therapies and patient care. *Npj Ment Health Res.* 2024;3(1):25.
<https://doi.org/10.1038/s44184-024-00067-w>.
- Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol.* 2019;16(11):703-15.
<https://doi.org/10.1038/s41571-019-0252-y>.
- Bruynseels K, Santoni de Sio F, van den Hoven J. Digital twins in health care: ethical implications of an emerging engineering paradigm. *Nat Med.* 2018;24(10):1448-50.
<https://doi.org/10.1038/s41591-018-0213-2>.
- Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med.* 2023;183(6):589-96.
<https://doi.org/10.1001/jamainternmed.2023.1838>.
- Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Transl Psychiatry.* 2017;7(3):e1053.
<https://doi.org/10.1038/tp.2017.25>.
- Niarakis A, Laouar R, Andrieux G, van Iersel M, Fromentin J, Funahashi A, et al. Immune digital twins for complex human pathologies: applications, limitations, and challenges. *npj Syst*

Biol Appl. 2024;10:450.

<https://doi.org/10.1038/s41540-024-00450-5>.

Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;18(6):463-77.

<https://doi.org/10.1038/s41573-019-0024-5>.

Musamih A, Salah K, Jayaraman R, Arshad J, Debe M, Al-Fuqaha A, et al. A blockchain-based approach for drug traceability in healthcare supply chain. *IEEE Access.* 2021;9.

<https://doi.org/10.1109/ACCESS.2021.3049920>.

Lin H, Guo R, Ma D, Kuai X, Yuan Z, Du Z, et al. Digital-twin-based multi-scale simulation supports urban emergency management: a case study of urban epidemic transmission. *Int J Digit Earth.* 2024;17(1):2421950.

<https://doi.org/10.1080/17538947.2024.2421950>.

Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med.* 2018;378(11):981-3.

<https://doi.org/10.1056/NEJMp1714229>.

Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.

<https://doi.org/10.1016/j.media.2017.07.005>.

Emanuel EJ, Wachter RM. Artificial intelligence in health care: will the value match the hype? *JAMA.* 2019;321(23):2281-2.

<https://doi.org/10.1001/jama.2019.4914>.

Chin MH, Afsar-Manesh N, Bierman AS, Chang C, Colón-Rodríguez CJ, Dullabh Pz, et al. Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care. *JAMA Netw Open.* 2023;6(12):e2345050.

<https://doi.org/10.1001/jamanetworkopen.2023.45050>.

Wachter RM, Gupta R. Will generative artificial intelligence deliver on its promise in health care? *JAMA.* 2024;331(1):65-9.

<https://doi.org/10.1001/jama.2023.25054>.

Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology.* 2018;286(3):800-9.

<https://doi.org/10.1148/radiol.2017171920>.

Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of physicians' views. *J Med Internet Res.* 2019;21(3):e12802.

<https://doi.org/10.2196/12802>.

Menz BD, Kuderer NM, Chin-Yee B, Logan JM, Rowland A, Sorich MJ, et al. Gender Representation of Health Care Professionals in Large Language Model-Generated Stories. *JAMA Netw Open.* 2024;7(9):e2434997.

<https://doi.org/10.1001/jamanetworkopen.2024.3876>.

Näher AF, Vorisek CN, Klopfenstein SAI, Lehne M, Thun S, Alsalamah S, et al. Secondary data for global health digitalisation. *Lancet Digit Health.* 2023;5(2):e93-e101.

[https://doi.org/10.1016/S2589-7500\(22\)00195-9](https://doi.org/10.1016/S2589-7500(22)00195-9).