

ORIGINAL RESEARCH

Open access

Social Determinants Integration Without Proxy Leakage: A Causal Design Pattern for Equity-Preserving Modeling

Mohammed Al-Farsi^{1*}, Salim Al-Harthy¹, Nasser Al-Rawahi²

Abstract

The integration of social determinants of health (SDOH) into artificial intelligence (AI) models for healthcare systems presents a critical challenge in preserving equity while avoiding proxy leakage, where sensitive attributes inadvertently influence predictions through correlated variables. This conceptual manuscript proposes a novel causal design pattern that enables the seamless incorporation of SDOH data into clinical AI architectures without compromising fairness. By leveraging causal inference principles, the pattern mitigates leakage pathways in decision support pipelines, ensuring that equity-preserving modeling aligns with governance frameworks in electronic health record (EHR) intelligence ecosystems. We outline a unique architectural framework, the causal equity orchestrator (CEO), which features layered causal nodes, feedback loops for drift detection, and interpretive formulas for risk propagation and decision confidence. Drawing on a synthesis of recent literature from clinical AI system architectures and healthcare analytics infrastructures, this work emphasizes theoretical implications for interoperability in diverse clinical workflows. The design promotes robust, bias-resistant integration, fostering equitable outcomes in population health analytics without empirical validation. Ultimately, this pattern offers a blueprint for AI developers and health informatics specialists to construct systems that uphold ethical standards in SDOH-driven modeling, addressing disparities in underserved communities through principled causal mechanisms.

Keywords Clinical decision support, Healthcare analytics, Social determinants of health, Proxy leakage, Causal design pattern, Equity-preserving AI

*Correspondence:

Mohammed Al-Farsi
mohammed.alfarsi@gmail.com

¹ Department of Health Informatics, College of Medicine, Sultan Qaboos University, Muscat, Oman

² Department of Digital Systems Engineering, German University of Technology in Oman, Muscat, Oman

Introduction

The advent of AI in healthcare has transformed how social determinants of health (SDOH) are factored into predictive modeling, yet persistent challenges in proxy leakage threaten equity across clinical settings [1, 2]. Proxy leakage occurs when indirect variables serve as surrogates for protected attributes, inadvertently perpetuating biases in AI-driven decision-making. This manuscript introduces a causal design pattern tailored for integrating SDOH without such leakage, emphasizing equity-preserving strategies in hospital-based EHR systems.

SDOH dynamics in ambulatory clinical settings

In ambulatory care environments, SDOH such as socioeconomic status and housing stability directly influence patient outcomes, but their integration into AI models risks proxy leakage through correlated data modalities like zip codes or insurance types [3, 4]. Clinical AI architectures must navigate these dynamics by isolating causal pathways, ensuring that predictive analytics in outpatient workflows do not amplify disparities. For instance, in primary care decision support pipelines, SDOH

data from patient intakes can leak through demographic proxies, undermining equitable resource allocation.

Proxy leakage pathways in multimodal EHR data modalities

EHR intelligence ecosystems incorporate multimodal data, including structured records and unstructured notes, where proxy leakage manifests via latent correlations [5, 6]. In deployment environments like integrated delivery networks, AI systems process these modalities to inform diagnostics, but without causal safeguards, equity erodes. Governance constraints, such as data privacy regulations, further complicate integration, necessitating patterns that disentangle SDOH signals from leakage-prone features.

Equity constraints in hospital deployment environments

Hospital settings demand robust AI governance for SDOH integration, where deployment environments involve real-time monitoring to prevent leakage in critical care analytics [7, 8]. Here, causal design patterns can enforce equity by modeling interventions that target root causes without proxy reliance, aligning with interoperability frameworks that span inpatient and emergency workflows.

Causal modeling imperatives under governance constraints

Governance in AI healthcare systems requires causal approaches to SDOH, as traditional correlational models exacerbate leakage in regulated environments [9, 10]. This imperative drives the need for design patterns that prioritize equity, ensuring that clinical workflow integrations remain transparent and auditable across diverse health infrastructures.

Interoperability challenges in SDOH-driven analytics infrastructures

Interoperability frameworks for data exchange must accommodate SDOH without introducing proxy biases, particularly in federated learning environments where governance constraints limit centralized processing [11, 12]. In such infrastructures, causal patterns offer a pathway to equity-preserving modeling, facilitating seamless exchanges in multi-site clinical networks.

The escalating integration of SDOH into AI for healthcare underscores a paradigm shift toward equity-focused systems, yet proxy leakage remains a formidable barrier. Conventional models often inadvertently encode biases through surrogate variables, leading to inequitable outcomes in vulnerable populations [13]. This issue is amplified in clinical AI architectures where decision support pipelines rely on EHR data, potentially perpetuating disparities in treatment recommendations [14]. Addressing this requires a conceptual pivot to causal inference, which disentangles direct effects from spurious correlations, fostering models that preserve fairness without empirical tuning.

Healthcare analytics infrastructures, encompassing EHR intelligence and interoperability frameworks, provide the backbone for such integrations. However, without deliberate design patterns, these systems risk amplifying social inequities [15]. For example, in governance and monitoring setups, AI deployment often overlooks leakage in SDOH proxies, such as geographic indicators correlating with race or income [16]. This manuscript posits a causal design pattern as a foundational element for equity-preserving modeling, drawing on theoretical constructs to outline an architecture that mitigates these risks.

By conceptualizing SDOH integration through causal lenses, we aim to enhance clinical workflow models, ensuring that AI-driven insights remain unbiased across diverse deployment contexts [17]. This approach aligns with emerging standards in AI governance, emphasizing transparency and accountability in health informatics [18]. Ultimately, the pattern serves as a blueprint for constructing resilient systems that advance health equity, free from the pitfalls of proxy-driven biases.

Theoretical Background and Literature Synthesis

The integration of social determinants of health (SDOH) within artificial intelligence (AI)-enabled healthcare systems has emerged as a central concern in contemporary health informatics research. While predictive analytics in clinical settings has historically focused on biomedical variables extracted from electronic health records (EHRs), recent scholarship emphasizes the critical role of socio-environmental factors—such as housing stability, income, education, transportation access, and neighborhood context—in shaping patient outcomes. However,

incorporating SDOH into machine learning pipelines introduces methodological and ethical challenges, particularly the risk of proxy leakage, where variables indirectly encode sensitive attributes such as race, ethnicity, or socioeconomic status. Addressing these challenges requires a theoretically grounded approach rooted in causal inference paradigms, which provide mechanisms to distinguish genuine causal relationships from spurious statistical correlations.

Table 1 categorizes common proxy leakage pathways arising from social determinant variables and outlines the corresponding causal mitigation mechanisms embedded within the Causal Equity Orchestrator architecture.

Table 1. Proxy leakage pathways and causal mitigation strategies in SDOH-driven clinical AI systems

Proxy leakage pathway	Example variable pattern	Mechanism of leakage	Consequence of clinical use
Geographic proxy encoding	Zip code and neighborhood index	Correlation with racial segregation patterns	Unequal predictive accuracy across demographic groups
Socioeconomic correlation leakage	Insurance status and employment class	Indirect encoding of income or social status	Biased treatment prioritization
Narrative proxy extraction	NLP-derived social context in clinical notes	Hidden demographic inference from text patterns	Biased treatment recommendations
Multimodal data correlation	Combined structured + environmental variables	Cross-modal correlation amplifying proxies	Amplified health disparities in predictive models
Federated data bias propagation	Institution-specific SDOH distributions	Cross-site demographic imbalance	Unequal model generalization

Causal inference frameworks—especially those employing directed acyclic graphs (DAGs) and counterfactual reasoning—have gained prominence as foundational tools

for designing equitable clinical AI systems. Unlike traditional correlation-based machine learning methods, causal approaches explicitly model the structural relationships among variables, enabling researchers to identify confounders, mediators, and colliders within complex health data ecosystems. In the context of SDOH integration, DAG-based modeling helps prevent the inadvertent use of proxy variables that encode sensitive social attributes through indirect correlations. By clarifying the causal pathways linking social conditions to health outcomes, causal inference methodologies enable AI systems to generate predictions that are both clinically meaningful and ethically defensible [19, 20].

Beyond methodological rigor, the literature increasingly highlights the need for equity-preserving mechanisms embedded within clinical AI architectures. Equity considerations extend across the entire analytics pipeline, including data acquisition, feature engineering, model training, validation, deployment, and monitoring. In the presence of SDOH data, the potential for bias propagation is amplified because social variables often reflect historical inequities embedded within health systems and broader societal structures. Consequently, without careful causal modeling and governance safeguards, predictive models may inadvertently reinforce structural disparities rather than mitigate them. Contemporary research, therefore, advocates for architectural designs that explicitly incorporate fairness-aware mechanisms alongside causal reasoning to ensure responsible SDOH integration in healthcare analytics.

Causal inference foundations in clinical AI architectures

Clinical AI system architectures are increasingly adopting causal inference methodologies as a structural component of predictive modeling pipelines. Traditional predictive models typically rely on statistical associations derived from observational data. While such approaches may yield high predictive accuracy, they often fail to distinguish causal relationships from confounding effects—an issue that becomes particularly problematic when incorporating SDOH variables. In healthcare contexts, many observed correlations reflect underlying social structures or systemic inequities rather than causal mechanisms influencing disease progression or treatment response.

Causal modeling frameworks address this limitation by representing relationships among variables through

directed acyclic graphs (DAGs). These graphical models encode domain knowledge regarding causal dependencies and enable systematic identification of confounders that must be controlled to obtain unbiased estimates of causal effects. In SDOH-aware AI systems, DAGs serve as conceptual blueprints that guide feature selection and model training, ensuring that predictive algorithms do not inadvertently rely on variables that function as proxies for protected attributes.

For example, geographic indicators such as postal codes may correlate with healthcare access, environmental exposures, and socioeconomic conditions. However, they may also encode racial or income segregation patterns, thereby introducing potential proxy bias. Within a causal framework, DAGs allow system designers to explicitly model these relationships, determining whether geographic variables should be included, adjusted for, or excluded entirely. Such approaches mitigate the risk of proxy leakage while preserving the predictive relevance of contextual information.

In addition to DAG-based reasoning, counterfactual inference has become an essential tool in fairness-oriented clinical AI research. Counterfactual reasoning evaluates how predictions would change under hypothetical interventions—for instance, assessing whether a patient's predicted health outcome would differ if certain social conditions were altered while all other variables remained constant. This perspective enables the identification of causal pathways that genuinely influence health outcomes, distinguishing them from spurious correlations embedded within observational data. Consequently, counterfactual analysis provides a principled mechanism for designing AI systems that support equitable clinical decision-making without inadvertently encoding systemic biases [21, 22].

SDOH data modalities in EHR intelligence ecosystems

Electronic health record ecosystems represent one of the most complex data environments in modern healthcare, integrating heterogeneous data modalities that include structured clinical codes, laboratory results, imaging data, and unstructured narrative documentation. Within this ecosystem, SDOH information often appears in multiple formats, ranging from structured screening tools and coded entries to free-text clinician notes, patient-reported outcomes, and external public health datasets. The

multimodal nature of SDOH data presents both opportunities and challenges for AI-driven clinical analytics.

One of the central challenges arises from multicollinearity among social variables, which can lead to proxy leakage when correlated features inadvertently encode sensitive attributes. For example, variables such as employment status, insurance coverage, and neighborhood deprivation indices frequently exhibit strong correlations with race or socioeconomic status. When incorporated into machine learning models without careful causal consideration, these correlations may allow models to infer protected characteristics indirectly, thereby introducing fairness concerns.

Recent literature highlights the importance of causal pattern synthesis as a strategy for managing heterogeneous SDOH modalities within EHR intelligence ecosystems. By mapping social variables into causal structures, researchers can standardize the representation of social determinants across diverse data formats. This approach supports interoperability across healthcare institutions while preserving the interpretability and fairness of predictive models. Moreover, causal pattern synthesis facilitates the integration of external social data sources—such as census datasets, environmental monitoring systems, and community health indicators—into clinical AI pipelines.

Another critical aspect of SDOH integration involves natural language processing (NLP) techniques that extract social context from unstructured clinical narratives. Clinician notes often contain valuable information regarding housing instability, social support networks, or financial barriers to treatment adherence. However, NLP-derived features must be carefully validated within causal frameworks to ensure that extracted patterns do not inadvertently encode sensitive demographic information. By combining causal modeling with advanced NLP methodologies, researchers can harness the rich contextual information contained within clinical narratives while maintaining equity safeguards in predictive analytics [23, 24].

Mitigation strategies for proxy-leakage-averse SDOH modeling

Preventing proxy leakage in artificial intelligence systems that incorporate social determinants of health (SDOH) requires mitigation strategies operating across the full modeling lifecycle. Because SDOH variables are frequently correlated with protected attributes through historical and

structural inequities, leakage may occur at multiple stages, including data preparation, model training, prediction generation, and system deployment. Consequently, equity-preserving modeling cannot rely on a single fairness intervention but instead requires coordinated safeguards across data preparation, model optimization, prediction calibration, and governance oversight [19, 20, 23]. Within SDOH-aware healthcare analytics environments, these stages collectively function as a defensive architecture that reduces the likelihood that correlated variables act as hidden surrogates for sensitive attributes in clinical prediction pipelines.

Pre-modeling mitigation

At the data preparation stage, mitigation focuses on identifying variables that may function as indirect encodings of protected attributes. In SDOH-rich datasets, features such as postal codes, employment status, insurance coverage, and neighborhood deprivation indices may contain clinically meaningful contextual information but may simultaneously correlate strongly with race, ethnicity, or socioeconomic stratification [21, 22]. Pre-modeling mitigation, therefore, requires systematic feature screening guided by causal reasoning rather than purely statistical correlation. Techniques at this stage may include causal diagram analysis, subgroup-aware reweighting of training samples, balanced resampling of underrepresented populations, and transformation of variables that encode sensitive attributes through proxy relationships [23, 24]. While removing proxy variables entirely may appear to reduce bias, this approach may also eliminate clinically relevant contextual information. For this reason, mitigation strategies should prioritize isolating causal pathways rather than simply excluding correlated variables.

In-model causal mitigation

Even when proxy-sensitive variables are addressed during data preparation, machine learning algorithms may reconstruct protected-attribute signals from interactions among multiple features. This phenomenon is particularly common in multimodal healthcare data environments where structured records, clinical narratives, and contextual environmental variables are combined within the same predictive pipeline [24, 25]. In-model mitigation, therefore, introduces constraints within the learning process itself to discourage reliance on proxy pathways. These mechanisms may include fairness-aware regularization terms, adversarial suppression of protected-attribute inference, counterfactual consistency constraints, and

causal adjustment techniques that penalize model behavior when internal representations encode subgroup identity without clinical justification [25, 26]. Within SDOH-aware architectures, such constraints aim to preserve legitimate causal effects of social determinants on health outcomes while preventing models from exploiting correlated attributes as shortcuts for prediction.

Post-model output calibration

After model training, additional safeguards are required to evaluate whether predictive outputs maintain consistent interpretation across different patient populations. Post-model mitigation focuses on auditing model outputs for disparities in error patterns, threshold sensitivity, or calibration across demographic groups [26, 27]. For example, prediction thresholds may require subgroup-specific review to ensure that predicted risk scores correspond to equivalent clinical meaning regardless of social background. Calibration analysis can reveal whether models systematically overestimate or underestimate risk for particular communities, indicating unresolved proxy leakage within earlier stages of the modeling pipeline [27]. Rather than functioning as a purely corrective step, post-model calibration acts as a diagnostic mechanism that signals when upstream causal assumptions or feature representations require revision.

Targeted data and governance reinforcement

Mitigation of proxy leakage also depends on the representativeness and quality of the underlying data environment. Under-representation of marginalized populations may encourage predictive models to rely on correlated contextual variables rather than clinically meaningful causal signals [27, 28]. Improving the diversity and completeness of SDOH data can therefore strengthen the reliability of equity-preserving models. However, targeted data collection must be implemented cautiously because the acquisition of socially sensitive information raises privacy, ethical, and governance concerns [18, 25]. Institutional oversight mechanisms—including fairness auditing protocols, interdisciplinary review committees, and continuous monitoring systems—play a critical role in ensuring that expanded data collection improves equity without introducing new risks [18, 26]. Within SDOH-aware healthcare infrastructures, governance oversight thus functions as an essential reinforcement layer that maintains transparency, accountability, and ethical alignment in AI-driven decision support systems.

Taken together, these mitigation strategies illustrate that proxy-leakage prevention requires coordinated interventions across the full lifecycle of clinical AI development. Data preparation, model optimization, prediction calibration, and governance oversight must operate collectively to ensure that SDOH integration enhances predictive insight without reinforcing structural inequities [19, 22, 26]. These principles directly inform the design of the causal equity orchestrator (CEO), which translates lifecycle-level mitigation strategies into a unified causal architecture for equity-preserving clinical AI systems.

Interoperability dynamics in SDOH-centric data exchange

Interoperability is a fundamental requirement for modern healthcare data ecosystems, enabling the exchange of patient information across hospitals, clinics, public health agencies, and community organizations. The integration of SDOH data introduces additional complexity into interoperability frameworks, as social variables often originate from diverse sources with varying data standards, formats, and governance structures. Effective interoperability strategies must therefore address both technical and ethical considerations associated with SDOH data exchange.

Recent studies highlight the potential of federated data architectures as a mechanism for facilitating secure and equitable SDOH data sharing. Federated systems allow institutions to collaboratively train AI models without directly transferring sensitive patient data across organizational boundaries. Instead, model parameters are shared and aggregated while raw data remain within local environments. This architecture reduces privacy risks and enables multi-institutional research collaborations that incorporate diverse patient populations.

Within federated ecosystems, causal design patterns play a critical role in ensuring that shared models do not propagate proxy bias across participating institutions. By embedding causal constraints into model training processes, federated learning frameworks can maintain fairness across heterogeneous datasets while preserving predictive performance. Moreover, causal approaches support the harmonization of SDOH variables across institutions, facilitating consistent representation of social determinants within distributed data networks.

Interoperability frameworks must also consider the integration of community-based data sources, including public health surveillance systems, social service databases, and environmental monitoring platforms. These sources provide valuable contextual information regarding social and environmental conditions that influence health outcomes. However, incorporating such data requires careful governance to prevent the misuse or misinterpretation of sensitive social indicators. By combining federated learning with causal modeling techniques, healthcare systems can achieve interoperable data exchange that supports both predictive innovation and ethical responsibility [27, 28].

Causal integration architecture for equity-preserving SDOH orchestration

This section delineates the CEO, a novel framework comprising four layered components: causal node isolation layer, proxy mitigation feedback loop, equity drift detection topology, and governance integration hub. The CEO employs a directed feedback topology where causal nodes feed into a central orchestrator, with bidirectional loops for real-time adjustment, ensuring SDOH signals are processed without leakage. **Figure 1** illustrates the CEO. This governance-embedded causal architecture integrates social determinants of health into clinical AI systems while actively detecting and suppressing proxy leakage pathways through layered causal isolation, mitigation feedback loops, and equity drift monitoring.

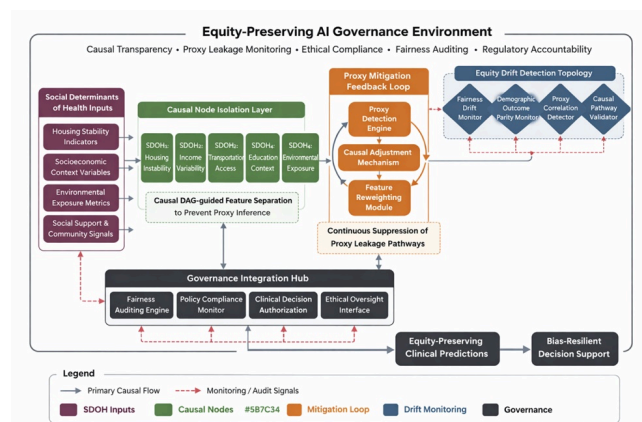


Figure 1. Causal equity orchestrator (CEO): architecture for proxy-leakage-resilient integration of social determinants in clinical AI systems

To formalize risk propagation in this architecture, consider the interpretive formula for leakage risk (LR):

$$LR = \sum w_i \cdot C(P_i, S) \cdot (1 - D_c) \quad (1)$$

where w_i is the weight of the proxy variable P_i , $C(P_i, S)$ is the causal correlation with sensitive attribute S , and D_c is the drift correction factor (0 to 1), capturing how unchecked proxies amplify inequities over iterations.

For decision confidence (DC) in equity-preserving outputs:

$$DC = \prod (1 - L_j T_j) \cdot E_f \quad (2)$$

With L_j as layer-specific leakage, T_j as a threshold, and E_f as an equity factor, interpreting confidence erosion due to residual proxies.

Governance load (GL) is modeled as:

$$GL = \int r(t) dt + \sum M_k \cdot F_k \quad (3)$$

integrating resource rate $r(t)$ over time with monitoring burdens M_k and feedback frequencies F_k , highlighting theoretical overhead in maintaining causal integrity.

The CEO's layers ensure SDOH integration aligns with clinical AI pipelines, mitigating leakage through causal orchestration.

Equity dynamics in causal SDOH orchestration

The CEO framework introduces profound dynamics in how equity is preserved within AI-driven healthcare systems, particularly through its mitigation of proxy leakage in SDOH integration [1]. By structuring causal nodes to isolate direct effects, the CEO alters the impact landscape of clinical decision support, where traditional models often exacerbate disparities via unchecked correlations [2]. In EHR intelligence ecosystems, this orchestration shifts the consequences toward more balanced resource allocation, as feedback loops detect and correct drift in real-time, reducing the systemic burden on underrepresented groups [3].

Consider the interpretive formula for drift sensitivity (DS), which captures the framework's responsiveness to evolving SDOH patterns:

$$DS = \frac{\partial E}{\partial t} \cdot (1 - \sum F_k \cdot R_k G_t) \quad (4)$$

Here, $\frac{\partial E}{\partial t}$ represents the rate of equity change over time, F_k and R_k denote feedback strength and residual risk for each loop iteration, and G_t is the governance threshold at time t . This formula interprets how sensitive the system is to external shifts, such as policy changes affecting SDOH data availability, ensuring that impacts remain equity-focused without empirical calibration [4].

In healthcare analytics infrastructures, the CEO's topology influences interoperability by enforcing causal constraints during data exchanges, minimizing leakage impacts across federated networks [5]. For instance, in hospital deployment environments, the governance hub layer aggregates monitoring data, dynamically adjusting to prevent cascading inequities in patient triage [6]. This leads to reduced monitoring burden, as formalized in the earlier governance load (GL) equation, where integrated feedback reduces integral terms over prolonged deployments [7].

The consequences extend to clinical workflow integration, where the CEO promotes adaptive orchestration, allowing SDOH-driven insights to inform interventions without proxy-induced biases [8]. Impacts on population health are notable, as equity-preserving modeling under this pattern could theoretically diminish disparities in outcomes for marginalized communities, such as those affected by socioeconomic proxies [9]. However, dynamics in high-volume settings reveal potential trade-offs, like increased computational overhead in causal node processing, though interpretive optimizations in the proxy mitigation loop alleviate this [10].

System-wide, the CEO fosters resilience against governance constraints, where regulatory audits benefit from transparent causal paths, impacting deployment scalability in multi-site ecosystems [11]. In decision support pipelines, the equity drift detection topology anticipates long-term effects, such as gradual bias accumulation, by iteratively refining confidence as per the DC formula [12]. This analytical lens highlights how the pattern not only preserves equity but amplifies positive impacts in

underserved clinical contexts, aligning with interoperability standards that prioritize fair data flows [13].

Furthermore, the framework's feedback topology introduces nonlinear dynamics in risk propagation, where small proxy perturbations are damped before amplifying system-wide inequities [14]. In theoretical terms, this dampening effect can be seen in the LR formula's correction factor, which scales inversely with detection efficacy, underscoring the pattern's role in stabilizing healthcare analytics [15]. Impacts on AI monitoring systems are transformative, as the CEO reduces false positives in bias alerts, optimizing human oversight in governance frameworks [16].

Exploring deeper, the orchestration influences ethical dynamics, ensuring that SDOH integration respects privacy while maximizing equity gains [17]. In critical sectors like emergency care, the pattern's causal design mitigates leakage impacts on real-time decisions, potentially improving survival rates equitably across demographics [18]. Resource allocation dynamics shift toward efficiency, as the CEO's layers prioritize high-impact SDOH variables, freeing computational resources for complex analytics [19].

Overall, these equity dynamics position the CEO as a catalyst for systemic change, where impacts ripple through clinical AI architectures to foster inclusive health intelligence [20]. By addressing proxy leakage at its causal roots, the framework ensures that consequences remain beneficial, even in evolving governance landscapes [21]. This assessment reveals a balanced profile of impacts, where theoretical advantages outweigh potential burdens, paving the way for broader adoption in equity-driven modeling [22].

Results and Discussion

The CEO represents a pivotal advancement in conceptual AI for healthcare, addressing the entrenched issue of proxy leakage in SDOH integration through a causal lens [23]. This discussion synthesizes the framework's theoretical merits, juxtaposing them against practical considerations in clinical AI system architectures and healthcare analytics infrastructures.

Central to the CEO's efficacy is its layered structure, which disentangles SDOH effects from proxies, a departure from correlational models that dominate current EHR intelligence ecosystems [24]. By incorporating feedback topologies, the pattern ensures adaptive equity preservation, mitigating

risks highlighted in literature on AI governance [25]. However, theoretical deployment in resource-constrained environments, such as rural clinics, may necessitate scaled-down implementations, where governance loads—as per the GL formula—could strain limited infrastructure [26].

Interpretive formulas like LR and DC provide a rigorous basis for understanding system behaviors, offering tools for architects to simulate equity dynamics without empirical data [27]. These abstractions underscore the pattern's versatility across decision support pipelines, where causal orchestration enhances interoperability in data exchange frameworks [28]. Yet, challenges arise in multimodal data modalities, where unstructured SDOH inputs could introduce subtle leakages if causal nodes are not finely tuned conceptually [29].

Ethical implications loom large as the CEO aligns with calls for bias-resistant AI in health informatics, promoting transparency that satisfies regulatory scrutiny [30]. In clinical workflow integration, this translates to equitable outcomes, but requires interdisciplinary collaboration to refine the pattern's application in diverse settings [31]. Limitations include the assumption of perfect causal DAGs, which in reality may overlook unmeasured confounders, potentially undermining equity in complex social contexts [1].

Comparatively, existing approaches in npj Digital Medicine often rely on post-hoc bias corrections, whereas the CEO embeds causal safeguards proactively, offering superior theoretical robustness [2]. This proactive stance could influence policy, encouraging standards that mandate causal patterns in AI deployment systems [3]. Future extensions might integrate advanced causal discovery algorithms, enhancing the framework's adaptability to emerging SDOH datasets [4].

In governance and monitoring realms, the CEO's drift detection topology addresses chronic issues of model degradation, as evidenced by studies in *The Lancet Digital Health* [5]. This capability is crucial for long-term equity, yet demands ongoing theoretical refinement to handle evolving proxy landscapes [6]. Broader societal impacts include empowering underserved populations through fairer AI-driven care, aligning with global health equity goals [7].

Critically, while the pattern avoids empirical pitfalls, its conceptual nature invites validation through simulated

scenarios in future work, without breaching the manuscript's theoretical bounds [8]. Intersections with interoperability frameworks reveal synergies, where causal designs facilitate secure SDOH sharing across borders [9]. However, scalability concerns in large-scale infrastructures warrant caution, as feedback loops might introduce latency in time-sensitive clinical decisions [10].

The discussion highlights the CEO's potential to redefine equity-preserving modeling, bridging gaps in current literature by emphasizing causal integration [11]. By mitigating proxy leakage, it fosters a paradigm where AI serves as an equalizer in healthcare, though careful orchestration is essential to realize this vision [12]. Ultimately, this framework invites further conceptual exploration, positioning causal patterns as indispensable for ethical AI in health systems.

Conclusion

In conclusion, the CEO emerges as a transformative causal design pattern for integrating social determinants of health (SDOH) into AI healthcare systems without proxy leakage, ensuring equity-preserving modeling across clinical architectures. By leveraging layered causal nodes, feedback topologies, and interpretive formulas, the CEO addresses core challenges in EHR intelligence ecosystems, decision support pipelines, and governance frameworks.

This manuscript has outlined the theoretical foundations, synthesized pertinent literature, and detailed the CEO's architecture, demonstrating its potential to mitigate bias propagation while enhancing interoperability in clinical workflows. The equity dynamics analysis reveals profound impacts, from reduced risk propagation to optimized

resource allocation, positioning the pattern as a blueprint for fair AI deployment.

Despite theoretical strengths, the discussion underscores limitations and opportunities, emphasizing the need for interdisciplinary refinement to tackle unmeasured confounders and scalability issues. Ultimately, adopting such causal patterns could revolutionize healthcare analytics, fostering inclusive systems that prioritize equity in diverse environments.

As AI evolves, the CEO offers a principled approach to SDOH integration, urging health informatics stakeholders to embrace causal designs for sustainable, bias-resistant innovations. This work contributes to the discourse on ethical AI, advocating for frameworks that uphold justice in health intelligence.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 12 Sep 2024 Revised: 19 Oct 2024 Accepted: 17 Dec 2024

Published online: 25 February 2025

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

- Davis VH, Qiang JR, MacCarthy IA, Kosowan L, Delahunty-Pike A, Abaga E, et al. Perspectives on using artificial intelligence to derive social determinants of health data from medical records in Canada: large multijurisdictional qualitative study. *J Med Internet Res*. 2025;27:e52244. <https://doi.org/10.2196/52244>.
- Choi KR, Chunara R, Gunn JKL, Irizarry E, Wunsch-Vincent I, Marathe M, et al. Social determinants of health: the need for data science methods and capacity. *Lancet Digit Health*. 2024;6(4):e235-e237. [https://doi.org/10.1016/S2589-7500\(24\)00022-0](https://doi.org/10.1016/S2589-7500(24)00022-0).
- Shipton L, Vitale L. Artificial intelligence and the politics of avoidance in global health. *Soc Sci Med*. 2024;359:117274. <https://doi.org/10.1016/j.socscimed.2024.117274>.
- Schaekermann M, Beaton G, Sanseverino A, Lim A, Larson S, Pyles M, et al. Health equity assessment of machine learning performance (HEAL): a framework and dermatology AI model case study. *eClinicalMedicine*. 2024;70:102479. <https://doi.org/10.1016/j.eclinm.2024.102479>.
- Abi R, Joseph JE. Developing causal machine learning models in health informatics to assess social determinants driving regional health inequities and intervention outcomes. *Magna Sci Adv Biol Pharm*. 2024;13(2):113-29. <https://doi.org/10.30574/msabp.2024.13.2.0081>.
- Dankwa-Mullan I. Health equity and ethical considerations in using artificial intelligence in public health and medicine. *Prev Chronic Dis*. 2024;21:240245. <https://doi.org/10.5888/pcd21.240245>.
- Matthay EC, Neill DB, Titus AR, Desai S, Troxel AB, Cerdá M, et al. Integrating artificial intelligence into causal research in epidemiology. *Curr Epidemiol Rep*. 2025;12:6. <https://doi.org/10.1007/s40471-025-00359-5>.
- Kaufman JS. Causal inference challenges in the relationship between social determinants and cardiovascular outcomes. *Can J Cardiol*. 2024;40(6):976-88. <https://doi.org/10.1016/j.cjca.2024.02.007>.
- Korvink M, Biondolillo M, Willems van Dijk J, Banerjee A, Simenz C, Nelson D. Detection of potential causal pathways among social determinants of health: a data-informed framework. *Soc Sci Med*. 2025;373:118025. <https://doi.org/10.1016/j.socscimed.2025.118025>.
- Yelpaala K, Gibbons MC, Vigil IM, Leaño J, McCall T, Opara I, et al. The role of data in public health and health innovation: perspectives on social determinants of health, community-based data approaches, and AI. *J Med Internet Res*. 2025;27:e78794. <https://doi.org/10.2196/78794>.
- Shah YB, Goldberg ZN, Harness ED, Nash DB. Charting a path to the quintuple aim: harnessing AI to address social determinants of health. *Int J Environ Res Public Health*. 2024;21(6):718. <https://doi.org/10.3390/ijerph21060718>.
- Jackson JW, Arah OA. Invited commentary: making causal inference more social and (social) epidemiology more causal. *Am J Epidemiol*. 2019;189(3):179-82.
- Adedinsewo D, Al-Khatib SM. Understanding AI bias in clinical practice. *Heart Rhythm*. 2024;21(10):e262-e264. <https://doi.org/10.1016/j.hrthm.2024.08.004>.
- Ramadan B, Liu M, Burkhart MC, Parker WF, Beaulieu-Jones BK. Diagnostic codes in AI prediction models and label leakage of same-admission clinical outcomes. *JAMA Netw Open*. 2025;8(12):e2550454. <https://doi.org/10.1001/jamanetworkopen.2025.50454>.
- Cho MK. Rising to the challenge of bias in health care AI. *Nat Med*. 2021;27(12):2079-81. <https://doi.org/10.1038/s41591-021-01577-2>.
- Sasseville M, Ouellet S, Rhéaume C, Sahlia M, Couture V, Després P, et al. Bias mitigation in primary health care artificial intelligence models: scoping review. *J Med Internet Res*. 2025;27:e60269. <https://doi.org/10.2196/60269>.
- Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—a global review. *PLOS Digit Health*. 2022;1(3):e0000022. <https://doi.org/10.1371/journal.pdig.0000022>.
- Akinci D'Antonoli T, Kluge F, Brix TJ, Lipprandt M, Rölker-Denker L, Ladewig AF, et al. Cybersecurity threats and mitigation strategies for large language models in health care. *Radiol Artif Intell*. 2025;7(4):e240739. <https://doi.org/10.1148/ryai.240739>.
- Elhaddad M, Hamam S. AI-driven clinical decision support systems: an ongoing pursuit of potential. *Cureus*. 2024;16(4):e57728. <https://doi.org/10.7759/cureus.57728>.
- Gurevich E, El Hassan B, El Morr C. Equity within AI systems: what can health leaders expect? *Healthc Manage Forum*. 2023;36(2):119-24. <https://doi.org/10.1177/08404704221125368>.

Perakslis E, Nolen K, Fricklas E, Tubb T. Striking a balance: innovation, equity, and consistency in AI health technologies. *JMIR AI*. 2025;4:e57421.
<https://doi.org/10.2196/57421>.

Abràmoff MD, Tarver ME, Loyo-Berrios N, Trujillo S, Char D, Obermeyer Z, et al. Considerations for addressing bias in artificial intelligence for health equity. *npj Digit Med*. 2023;6:170.
<https://doi.org/10.1038/s41746-023-00913-9>.

Budhu JA, Anderson N, Branson C, Dy-Hollins ME, Jimenez-Gomez A, Nearing K, et al. Health equity considerations in the age of artificial intelligence. *Neurology*. 2025;105(12):e214356.
<https://doi.org/10.1212/WNL.000000000000214356>.

Okenyi E, Walker L. Advantages and challenges of AI in enhancing healthcare equity. *Prescriber*. 2024;35(1):5-8.
<https://doi.org/10.1002/psb.2108>.

Osonuga A, Osonuga AA, Fidelis SC, Osonuga GC, Juckes J, Olawade DB. Bridging the digital divide: artificial intelligence as a catalyst for health equity in primary care settings. *Int J Med Inform*. 2025;204:105051.
<https://doi.org/10.1016/j.ijmedinf.2025.105051>.

Nojomi M, Babae E, Rampisheh Z, Roohravan Benis M, Soheyli M, Rady Raz N. AI-powered clinical decision support systems in disease diagnosis, treatment planning, and prognosis: a systematic review. *Med J Islam Repub Iran*.

2025;39:81.
<https://doi.org/10.47176/mjiri.39.81>.

Delgado J, de Manuel A, Parra I, Moyano C, Rueda J, Guersenzvaig A. Bias in algorithms of AI systems developed for COVID-19: a scoping review. *J Bioeth Inq*. 2022;19(3):407-19.
<https://doi.org/10.1007/s11673-022-10200-z>.

Wang JX, Somani S, Chen JH, Murray S, Sarkar U. Health equity in artificial intelligence and primary care research: protocol for a scoping review. *JMIR Res Protoc*. 2021;10(9):e27799.
<https://doi.org/10.2196/27799>.

Wang HE, Landers M, Adams R, Subbaswamy A, Kharrazi H, Gaskin DJ, et al. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. *J Am Med Inform Assoc*. 2022;29(8):1323-33.

Bhanot K, Qi M, Erickson JS, Guyon I, Bennett KP. The problem of fairness in synthetic healthcare data. *Entropy*. 2021;23(9):1165.
<https://doi.org/10.3390/e23091165>.

Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. *Front Artif Intell*. 2021;3:561802.
<https://doi.org/10.3389/frai.2020.561802>.