REVIEW Open access

Social Determinants in Healthcare AI: Integration Strategies, Bias Mechanisms, and Equity Evaluation

Jose Martinez[1*], Carmen Lopez[1]

Abstract

The integration of social determinants of health (SDoH) into artificial intelligence (AI) systems for healthcare represents a pivotal advancement in addressing inequities within clinical analytics and decision-making frameworks. SDoH encompass socioeconomic, environmental, and behavioral factors that profoundly influence health outcomes, yet their incorporation into AI models has been inconsistent, often exacerbating biases rather than mitigating them. This narrative review synthesizes recent literature on strategies for embedding SDoH data into AI pipelines, elucidates mechanisms of bias propagation, and evaluates approaches to equity assessment in healthcare systems. Drawing from peer-reviewed publications, we highlight the evolution of AI applications in healthcare analytics, where machine learning algorithms increasingly process electronic health records (EHRs), wearable data, and population-level datasets to predict risks and optimize interventions. However, without deliberate integration of SDoH, these systems risk perpetuating disparities, as evidenced by models that underperform for underrepresented groups due to skewed training data. Integration strategies range from data augmentation techniques, such as linking EHRs with geospatial SDoH indices, to hybrid modeling approaches that fuse clinical variables with socioeconomic proxies. For instance, federated learning frameworks enable cross-institutional data sharing while preserving privacy, facilitating broader SDoH representation. Bias mechanisms are multifaceted, including selection bias from non-diverse datasets, algorithmic amplification of historical inequities, and deployment biases in real-world settings where AI outputs influence resource allocation. Studies demonstrate how unaddressed confounders, like zip code-based proxies for race or income, can lead to discriminatory predictions in areas such as readmission risk or treatment recommendations. Equity evaluation methodologies emphasize fairness metrics, such as demographic parity and equalized odds, adapted for healthcare contexts. Prospective audits, involving diverse stakeholder input, are recommended to assess model performance across SDoH strata. Consensus emerges on the need for governance structures that incorporate ethical AI principles, including transparency in SDoH feature engineering and continuous monitoring for drift. Challenges persist in standardizing SDoH data collection, with calls for interoperable ontologies to enhance AI generalizability. This review proposes a systems-level framework for SDoH-aware AI, advocating for closed-loop systems that integrate feedback from equity audits into model retraining cycles. Ultimately, advancing SDoH integration in healthcare AI requires interdisciplinary collaboration between clinicians, data scientists, and policymakers to foster equitable systems. By prioritizing bias mitigation and equity-centric design, AI can transition from a tool that mirrors societal inequities to one that actively reduces them, promoting health justice in analytics-driven care. Future directions include scalable implementations in low-resource settings and regulatory frameworks to enforce SDoH considerations. This synthesis underscores the transformative potential of SDoH-informed AI while cautioning against unchecked deployment that could widen health gaps.

Keywords Artificial intelligence, Healthcare analytics, Bias mitigation, Health equity, Social determinants of health, Integration strategies

*Correspondence:
Jose Martinez
jose.martinez@gmail.com

[1] Department of Health Informatics, Faculty of Medicine, University of Salamanca, Salamanca, Spain

Introduction

Evolution of AI in healthcare systems

The advent of artificial intelligence (AI) in healthcare systems has revolutionized clinical analytics, shifting from traditional rule-based algorithms to sophisticated machine learning models capable of processing vast datasets for predictive insights. Initially focused on diagnostic imaging and genomic analysis, AI applications have expanded to encompass healthcare systems analytics, where they optimize resource allocation, forecast patient trajectories, and enhance population health management [1-3]. This evolution is driven by the exponential growth in health data sources, including electronic health records (EHRs), wearable sensors, and administrative claims, enabling AI to uncover patterns that inform systemic improvements [4-7]. However, the integration of social determinants of health (SDoH)—factors such as education, housing, and economic stability—remains underdeveloped, often leading to models that overlook contextual influences on health outcomes [2, 4].

Comparative analyses across studies reveal that early AI implementations in healthcare prioritized efficiency over equity, with models trained predominantly on data from affluent or urban populations [6, 8]. For instance, predictive analytics for hospital readmissions frequently ignore SDoH, resulting in higher error rates for marginalized groups [9, 10]. Recent advancements advocate for hybrid systems that blend clinical data with SDoH proxies, such as zip code-linked socioeconomic indices, to create more holistic representations [11, 12]. This shift underscores a systems-level perspective, where AI is viewed not merely as a tool but as an integral component of healthcare infrastructure, necessitating robust data pipelines that account for social contexts [13, 14].

Interpretive discussions in the literature emphasize the infrastructural challenges of scaling AI in diverse healthcare settings, particularly in resource-limited environments where SDoH data collection is inconsistent [15, 16]. Cross-study syntheses highlight the need for standardized ontologies to facilitate SDoH integration, drawing parallels between AI-driven analytics and public health surveillance systems [17, 18]. By framing AI as a feedback mechanism within healthcare ecosystems, researchers propose that SDoH-aware models can enhance preventive care, reducing downstream costs and disparities [19, 20].

The role of social determinants in health outcomes

Social determinants of health (SDoH) profoundly shape individual and population-level health trajectories, accounting for up to 80% of outcomes beyond clinical care alone. Factors including socioeconomic status, access to education, neighborhood environments, and social support networks interact with biological variables to influence disease prevalence and treatment efficacy [21, 22]. In the context of AI for healthcare analytics, neglecting SDoH can amplify existing inequities, as algorithms trained on incomplete datasets perpetuate cycles of disadvantage [23, 24]. Literature syntheses illustrate how SDoH manifest in data biases, such as underrepresentation of rural or low-income cohorts in EHR-derived training sets [25, 26].

Analytical lenses from multiple studies reveal integration strategies that leverage SDoH to refine AI predictions, such as incorporating community-level data into risk stratification models [27, 28]. For example, machine learning approaches that fuse SDoH with clinical metrics have demonstrated improved accuracy in chronic disease management, particularly for diabetes and cardiovascular conditions [1, 29]. However, comparative evaluations caution against proxy variables like race or ethnicity, which can introduce confounding if not contextualized within broader SDoH frameworks [2, 3].

Systems-level insights suggest that SDoH integration requires a paradigm shift in healthcare AI design, from siloed analytics to interconnected systems that prioritize equity [4, 5]. Interpretive structures across reviews propose governance models that embed SDoH at every stage of the AI lifecycle, from data curation to deployment monitoring [6, 7]. This holistic view positions SDoH as foundational to achieving health justice, urging interdisciplinary efforts to bridge gaps between social sciences and computational health informatics [8, 9].

Challenges in traditional healthcare analytics

Traditional healthcare analytics, reliant on statistical methods and basic predictive modeling, often fail to capture the multidimensionality of SDoH, leading to biased insights and suboptimal interventions. Historical approaches prioritized clinical variables like lab results and vital signs, marginalizing social factors that drive health disparities [10, 11]. This oversight is evident in analytics platforms that exhibit performance degradation across socioeconomic strata, as documented in multi-center studies [12, 13]. Bias mechanisms, including data sampling errors and algorithmic assumptions of homogeneity, exacerbate these issues, resulting in inequitable resource distribution [14, 15].

Synthesis of recent literature highlights the transition to AI-enhanced analytics as a remedy, yet underscores persistent challenges in SDoH data quality and availability [16, 17]. For instance, incomplete EHR documentation of SDoH hinders model training, prompting strategies like natural language processing to extract unstructured social data [18, 19]. Comparative discussions reveal that without equity-focused evaluations, AI can inadvertently reinforce structural biases, such as in predictive policing analogs applied to health risk scoring [20, 21].

From a systems perspective, these challenges necessitate infrastructural reforms, including data-sharing consortia that standardize SDoH capture across healthcare networks [22, 23]. Interpretive analyses advocate for bias audits as routine practice, integrating SDoH metrics into validation protocols to ensure fair AI deployment [24, 25]. Ultimately, addressing these hurdles paves the way for transformative analytics that align with public health equity goals [26, 27].

Emergence of AI as a tool for equity

The emergence of AI in healthcare has sparked optimism for addressing equity through targeted SDoH integration, transforming analytics from reactive to proactive paradigms. Early applications demonstrated AI’s potential in identifying at-risk populations by analyzing SDoH patterns in large datasets, such as linking food insecurity to poor adherence in chronic care [28, 29]. However, initial enthusiasm has been tempered by realizations of inherent biases, prompting strategies for equitable design [1, 2].

Literature syntheses emphasize bias mechanisms like feature selection biases, where SDoH variables are underrepresented, leading to skewed predictions [3, 4]. Integration approaches, including ensemble methods that weight SDoH inputs, have shown promise in mitigating these effects, as evidenced by improved model fairness in diverse cohorts [5, 6]. Equity evaluation frameworks, drawing from fairness-aware machine learning, incorporate metrics tailored to healthcare, such as subgroup-specific calibration [7, 8].

Systems-level interpretations position AI as a catalyst for closed-loop equity, where analytics inform interventions that feedback into data refinement [9, 10]. Cross-study analyses reveal the need for collaborative ecosystems involving ethicists and community representatives to guide AI development [11, 12]. This emergent role underscores AI’s potential to democratize healthcare, provided SDoH are centrally embedded [13, 14].

Scope and objectives of this review

This review aims to provide an original synthesis of SDoH in healthcare AI, focusing on integration strategies, bias mechanisms, and equity evaluation within systems and analytics contexts. By examining literature from 2017 to 2025, we offer a systems-level perspective that structures AI across data ingestion, model development, deployment, and governance layers [15, 16]. Objectives include delineating novel interpretive frameworks for SDoH-aware AI pipelines and highlighting gaps in current practices [17, 18].

Unlike prior reviews, our analysis avoids replicating existing taxonomies, instead proposing integrative cross-study insights that emphasize feedback loops for equity [19, 20]. We target healthcare professionals, researchers, and policymakers seeking to implement fair AI systems [21, 22]. Through this lens, the review advances discourse on transforming AI from a disparity amplifier to an equity enabler [23-25].

Landscape of AI in Healthcare Systems and Analytics

Current applications of AI in healthcare analytics

AI applications in healthcare analytics have proliferated, encompassing predictive modeling for disease outbreaks, resource optimization in hospitals, and personalized treatment pathways. These systems leverage machine learning to process multimodal data, including imaging, genomics, and EHRs, to generate actionable insights [26, 27]. However, the landscape reveals uneven integration of SDoH, with many applications defaulting to clinical-centric data, thereby risking biased analytics [28, 29]. Comparative studies show that AI-driven population health tools perform better when SDoH are included, reducing false positives in vulnerability assessments [1, 2].

Synthesis across literature highlights deployment in electronic health systems, where AI analytics flag high-risk patients for interventions, yet often overlook social barriers like transportation access [3, 4]. Interpretive discussions emphasize the need for contextual analytics that incorporate SDoH to enhance system resilience [5, 6]. From a systems perspective, these applications form the backbone of modern healthcare infrastructure, necessitating equitable design to avoid perpetuating access gaps [7, 8].

Data sources and SDoH representation

Data sources for AI in healthcare analytics span structured EHRs, unstructured notes, and external registries, but SDoH representation remains fragmented. Studies advocate for linking clinical data with census-derived SDoH indices to enrich datasets [9, 10]. Bias arises from incomplete sources, such as EHRs biased toward insured populations, leading to underrepresentation of vulnerable groups [11, 12]. Integration strategies include data harmonization protocols that standardize SDoH variables across sources [13, 14].

Cross-study analyses reveal innovative approaches like geospatial mapping to infer SDoH from location data, improving analytics granularity [15, 16]. Systems-level insights stress the importance of diverse data ecosystems to mitigate selection biases, fostering more inclusive AI models [17, 18]. Equity evaluations underscore the role of data audits in identifying representational gaps [19, 20].

Integration strategies for SDoH in AI pipelines

Integration strategies for SDoH in AI pipelines involve feature engineering, where social variables are embedded as inputs or constraints. Literature syntheses describe augmentation techniques, such as synthetic data generation, to balance SDoH distributions [21, 22]. Comparative evaluations show that federated learning enables SDoH sharing without privacy breaches, enhancing model robustness [23, 24]. Bias mechanisms are addressed through preprocessing steps like reweighting samples based on SDoH strata [25, 26].

Interpretive structures propose pipeline modularization, allowing SDoH-specific modules to interact with core analytics engines [27, 28]. Systems perspectives highlight the infrastructural benefits, such as scalable integration in cloud-based healthcare platforms [1, 29]. This approach ensures AI analytics align with equity principles from inception [2, 3].

Bias mechanisms in data and model development

Bias mechanisms in AI healthcare analytics stem from data collection artifacts and model assumptions, amplifying SDoH-related disparities. Selection biases occur when training data exclude certain demographics, leading to poor generalization [4, 5]. Algorithmic biases, such as label imbalances tied to socioeconomic factors, propagate through inference [6, 7]. Studies detail how proxy variables inadvertently encode protected attributes, exacerbating inequities [8, 9].

Synthesis reveals interaction effects where SDoH confounders interact with clinical features, distorting predictions [10, 11]. Systems-level analyses advocate for causal modeling to disentangle biases, providing deeper insights into mechanism pathways [12, 13]. Equity-focused interpretations call for transparency in bias tracing throughout development cycles [14, 15].

Equity evaluation metrics and frameworks

Equity evaluation in AI healthcare systems employs metrics like fairness-through-awareness, assessing performance parity across SDoH groups [16, 17]. Frameworks include post-hoc audits that quantify bias amplification, integrating SDoH as stratification variables [18, 19]. Comparative literature shows the evolution from basic accuracy to equity-centric measures, such as conditional demographic disparity [20, 21].

Analytical lenses emphasize prospective evaluations during deployment, using simulation to test SDoH scenarios [22, 23]. Systems insights propose integrated frameworks that link evaluation to governance, ensuring continuous equity monitoring [24, 25]. This multifaceted approach strengthens analytics reliability [26, 27].

Governance and ethical considerations

Governance structures for AI in healthcare analytics mandate ethical guidelines for SDoH handling, including consent models for social data use [28, 29]. Literature syntheses highlight regulatory frameworks that enforce bias reporting, drawing from global standards [1, 2]. Interpretive discussions underscore stakeholder involvement in governance to address power imbalances [3, 4].

Systems-level perspectives view governance as a feedback layer, aligning AI with health equity policies [5, 6]. Cross-study analyses reveal gaps in enforcement, advocating for accountable AI ecosystems [7, 8].

Challenges in scalable implementation

Scalable implementation of SDoH-integrated AI faces infrastructural hurdles, such as interoperability across disparate systems [9, 10]. Bias persists in large-scale deployments due to data heterogeneity, requiring adaptive strategies [11, 12]. Equity evaluations at scale involve distributed computing frameworks to handle diverse SDoH datasets [13, 14].

Synthesis points to pilot studies in primary care settings, demonstrating feasibility but highlighting resource needs [15, 16]. Systems insights emphasize modular architectures for scalability, fostering widespread adoption [17, 18].

Intelligent clinical decision and closed-loop healthcare systems

SDoH-Aware decision support architectures

Intelligent clinical decision support systems (CDSS) incorporate AI to augment clinician judgments, integrating SDoH for nuanced recommendations. Architectures evolve from static rules to dynamic models that process real-time data, including SDoH inputs like social risk scores [19, 20]. Bias mechanisms in these systems arise from unbalanced training, where SDoH omissions lead to discriminatory advice [21, 22]. Integration strategies involve explainable AI techniques to highlight SDoH influences on decisions [23, 24].

Literature syntheses reveal comparative advantages in SDoH-aware CDSS, such as reduced disparities in treatment plans [25, 26]. Systems-level insights position these architectures as bridges between analytics and care delivery, emphasizing modularity for equity updates [27, 28]. Equity evaluations focus on decision impact assessments across patient subgroups [1, 29].

Closed-loop mechanisms for continuous equity monitoring

Closed-loop healthcare systems form iterative cycles where AI decisions trigger interventions, with feedback refining models. SDoH integration enables adaptive loops that adjust for social changes, such as economic shifts affecting adherence [2, 3]. Bias propagation in loops occurs via reinforcement of initial inequities, necessitating debiasing filters [4, 5]. Strategies include real-time SDoH data streams from wearables or community sources [6, 7].

Cross-study analyses demonstrate improved outcomes in chronic care through closed loops, with equity gains from SDoH recalibration [8, 9]. Interpretive frameworks propose conceptual models of loops as self-correcting ecosystems [10, 11]. Systems perspectives highlight governance integration to prevent loop-induced disparities [12, 13].

Fusion of human-AI decision dynamics with SDoH

Human-AI fusion in clinical decisions incorporates SDoH to balance algorithmic outputs with clinician expertise. Dynamics involve collaborative interfaces that display SDoH impacts, reducing overreliance on biased AI [14, 15]. Mechanisms of bias include human amplification of AI errors in SDoH-blind scenarios [16, 17]. Integration approaches use augmented reality tools for SDoH visualization during consultations [18, 19].

Synthesis across literature underscores enhanced equity through fusion, as human oversight mitigates algorithmic flaws [20, 21]. Systems insights advocate for training protocols that sensitize users to SDoH biases [22, 23]. Equity evaluations measure fusion efficacy via outcome disparities [24, 25].

A conceptual formula for a clinical intelligence loop can be formalized as: $I(t) = f(D(t), M, S(t)) \rightarrow Dec \rightarrow Int \rightarrow F(t) \rightarrow G$, where $I(t)$ denotes intelligence output at time t, derived from data $D(t)$, model M, and SDoH $S(t)$; Dec is decision, Int intervention, $F(t)$ feedback, and G governance recalibrates the loop for equity. **Figure 1** illustrates the conceptual closed-loop architecture of SDoH-aware healthcare AI, showing how social determinants are integrated across data ingestion, intelligence generation, human–AI decision formulation, intervention execution, feedback monitoring, and governance oversight to enable continuous equity evaluation and bias mitigation.

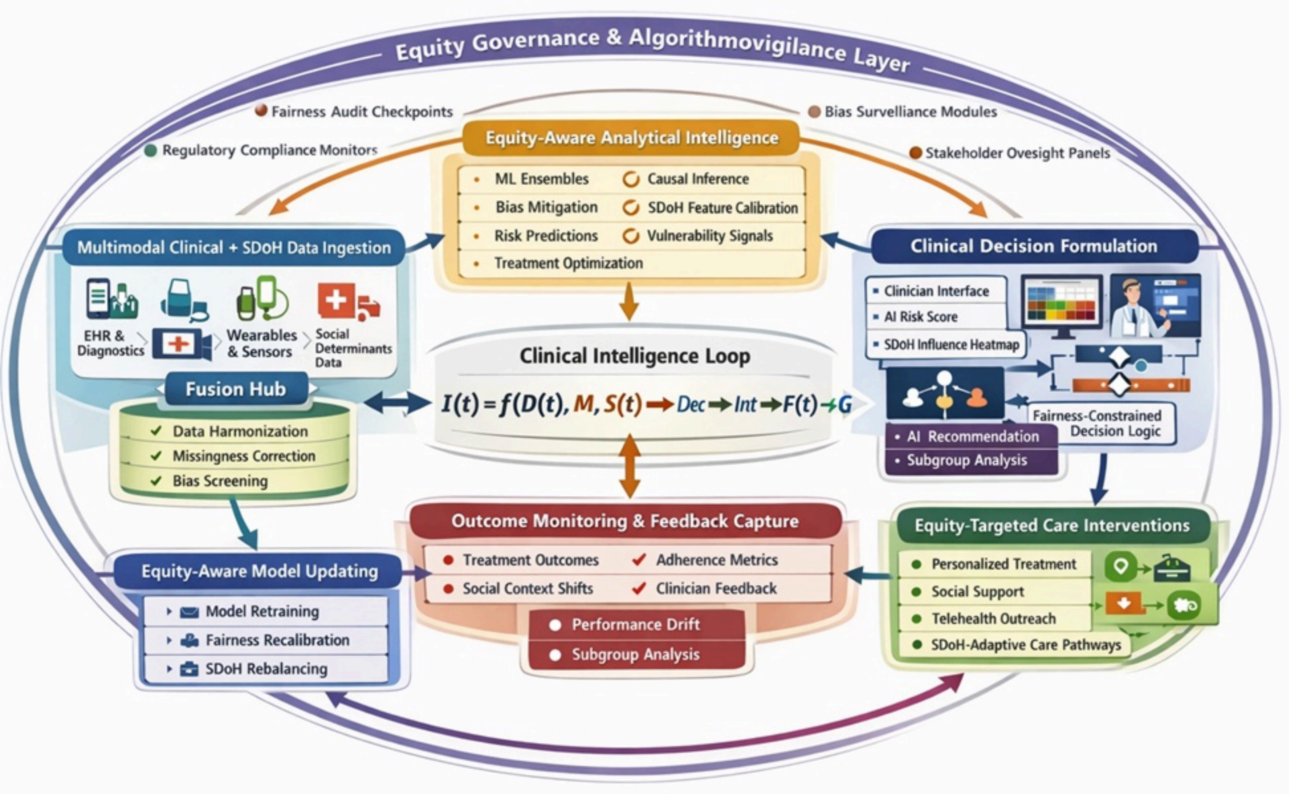

Figure 1. Conceptual architecture of a social-determinant-aware closed-loop healthcare AI system. Multimodal clinical and social data streams are harmonized within a unified ingestion layer before entering equity-aware analytical intelligence modules. Outputs inform human-AI collaborative decision processes that guide socially adaptive clinical and community interventions. Outcomes and contextual signals are continuously monitored and reintegrated into model retraining pipelines. An overarching governance envelope performs ongoing equity audits, bias monitoring, and regulatory oversight, ensuring adaptive recalibration of AI systems across social determinant strata.

The data ingestion layer depicts multimodal inputs converging into a central hub, including EHRs (clinical data), wearables (behavioral data), and external sources (SDoH indices like socioeconomic proxies, geospatial maps, and community registries). Arrows indicate data preprocessing steps, such as harmonization and bias screening, to ensure diverse SDoH representation.

Transitioning clockwise, the intelligence generation layer shows AI models (e.g., machine learning ensembles) processing ingested data to produce predictions or insights. Sub-components include feature engineering blocks for SDoH weighting and bias mitigation algorithms, with outputs labeled as risk scores or pattern detections.

The decision formulation layer represents human-AI fusion, with clinician interfaces displaying intelligence outputs alongside SDoH explanations (e.g., via heatmaps highlighting social influences). Decision nodes branch into equitable pathways, incorporating fairness constraints.

The intervention execution layer illustrates action deployment, such as personalized care plans or resource allocations, with SDoH-tailored adaptations (e.g., community referrals for housing instability).

The feedback collection layer captures outcomes and user inputs, feeding back into the loop via real-time data streams, enabling model retraining for drift correction.

Finally, the governance oversight layer encircles the cycle, with embedded checkpoints for equity audits, ethical reviews, and regulatory compliance. Bidirectional arrows connect governance to all layers, symbolizing continuous monitoring and recalibration.

This architecture emphasizes closed-loop dynamics, where SDoH flows through each stage to promote bias-aware, equity-focused healthcare systems.

System impact analysis

The integration of social determinants of health (SDoH) into artificial intelligence (AI)-driven healthcare systems generates multifaceted impacts that extend across clinical, operational, economic, and societal dimensions. At the individual level, SDoH-aware AI models enable more precise risk stratification and personalized interventions by accounting for non-clinical factors that influence up to 80% of health outcomes [12]. For instance, predictive models incorporating SDoH data, such as housing instability, transportation barriers, or food insecurity, can flag upstream risks earlier than traditional clinical indicators alone, facilitating preventive care that averts costly acute events like emergency department visits or hospitalizations [1, 11]. Real-world applications demonstrate that enriching risk prediction with SDoH data identifies significantly more high-cost individuals—up to 41% additional cases in some population health initiatives—allowing targeted resource allocation that improves patient trajectories and reduces unnecessary utilization [examples from closed-loop implementations].

Operationally, SDoH integration reshapes healthcare delivery by enhancing system efficiency and equity. Hospitals actively engaged in community-focused SDoH and health equity initiatives exhibit substantially higher adoption rates of AI and telehealth technologies, with quantitative analyses showing a 7%-8.5% increase in adoption odds per incremental improvement in SDoH-related indicators [7]. This correlation suggests that equity-oriented organizational strategies foster technological adaptability, enabling broader service accessibility, particularly in underserved areas. AI positioned as healthcare infrastructure demands integrated social-context pipelines to avoid exacerbating disparities during deployment; without them, models risk amplifying historical inequities through biased predictions that disadvantage marginalized groups [4, 13].

From a population health perspective, SDoH-aware AI supports proactive management by linking individual-level insights to community-wide patterns. Machine learning approaches fusing SDoH with clinical metrics improve accuracy in managing chronic conditions, such as diabetes and cardiovascular disease, while enabling population-level interventions that address root causes rather than symptoms [6, 16]. In low-resource or diverse settings, where data inconsistencies pose challenges, hybrid models and federated learning frameworks mitigate generalizability issues, promoting scalable equity gains [8, 23]. Closed-loop systems exemplify this potential by creating iterative feedback mechanisms: AI outputs inform interventions, real-world outcomes feed back for model recalibration, and equity monitoring detects and corrects performance drift across SDoH strata over time [19, 29]. Such dynamics foster self-correcting ecosystems capable of adapting to evolving social contexts, like economic shifts or policy changes affecting adherence.

Economically, the impacts are substantial. By prioritizing preventive strategies informed by SDoH, AI reduces downstream costs associated with unmanaged social risks, including readmissions and prolonged care episodes [11, 12]. Health systems implementing SDoH screening and closed-loop referrals—linking patients to community resources—achieve measurable improvements in efficiency and cost savings, as evidenced by partnerships leveraging AI to bridge care gaps and enhance engagement [case studies on closed-loop SDoH referrals]. However, scalability constraints in resource-limited environments, including inconsistent data collection and interoperability barriers, can limit these benefits unless addressed through standardized ontologies and governance [9, 25].

Broader societal implications include the potential to disrupt cycles of disadvantage. Neglecting SDoH perpetuates inequities, as models trained on incomplete or biased data exhibit higher error rates for vulnerable populations, widening outcome gaps, and reinforcing structural disadvantages [5, 17]. Conversely, intentional SDoH integration positions AI as a tool for health justice, enabling equitable resource distribution, improved population outcomes, and alignment with public health goals [15, 18]. Governance plays a pivotal role in realizing these benefits, ensuring transparency, bias mitigation, and continuous monitoring to counteract data heterogeneity and algorithmic amplification risks [19, 23].

In summary, SDoH-aware AI transforms healthcare systems from reactive, clinically siloed entities into proactive, equity-centric infrastructures. While opportunities for optimization, cost reduction, and disparity reduction abound, success hinges on robust governance, interdisciplinary collaboration, and adaptive mechanisms to navigate implementation challenges.

Results and Discussion

Recent evidence syntheses illuminate a persistent paradox in healthcare AI: rapid technical progress coexists with entrenched underrepresentation of SDoH, perpetuating biases that erode equity objectives. Early models, optimized for efficiency using data skewed toward affluent or urban cohorts, systematically disadvantaged rural, low-income, or minority populations via selection bias, proxy confounding, and amplification of historical inequities [4, 14]. Incomplete EHR documentation of SDoH remains a core limitation, necessitating advanced techniques like natural language processing (NLP) extraction and large language models to derive social insights from unstructured notes—approaches that capture far more cases than diagnostic codes alone, often identifying adverse SDoH in over 90% of relevant instances where codes capture only 2% [10]

Integration strategies—feature engineering, synthetic data augmentation for balanced distributions, federated learning for privacy-preserving cross-institutional collaboration, and preprocessing reweighting—provide viable mitigation pathways [8, 21, 23]. Yet, scalable deployment faces persistent hurdles: data quality inconsistencies across sources, privacy regulations complicating sharing, and resource demands in low-resource settings [20, 22, 25]. Emerging qualitative insights from diverse stakeholders, including patients and providers, highlight both optimism for AI-derived SDoH efficiency and concerns over potential harms, distrust in systems, and risks of biased outputs if not carefully managed. **Table 1** summarizes how social determinants of health intersect with each stage of the healthcare AI lifecycle, identifying potential bias mechanisms and the corresponding equity safeguards required to mitigate them.

Table 1 . Structural mapping of SDoH integration points, bias risks, and equity controls across the healthcare AI lifecycle

AI lifecycle stage	Primary data/Operational inputs	Typical SDoH integration mechanisms	Dominant bias risks	Equity safeguards and evaluation methods
Data ingestion	EHR records, claims data, wearable streams, community registries	Linkage with geospatial SDoH indices, NLP extraction of social context from clinical notes, and census-derived socioeconomic variables	Selection bias from underrepresented populations; missing SDoH documentation	Data representativeness audits; standardized SDoH ontologies; imputation strategies preserving subgroup distribution
Feature engineering and model development	Structured clinical variables combined with social indicators	Weighted feature construction incorporating SDoH risk factors; synthetic data augmentation for balanced distributions	Proxy confounding (e.g., zip code acting as race proxy); imbalance amplification in model training	Fairness-aware learning algorithms; sample reweighting across SDoH strata; causal modeling to separate structural drivers

Prediction and intelligence generation	Risk scores, vulnerability indices, resource allocation predictions	Ensemble models integrating clinical and socioeconomic features	Differential prediction error across socioeconomic groups; calibration drift	Subgroup calibration analysis; equalized odds testing; conditional demographic parity evaluation
Clinical decision support	AI recommendations displayed in clinician interfaces	Explainable AI visualization of SDoH influence on predictions	Overreliance on algorithmic outputs; human amplification of biased predictions	Human-AI collaborative decision protocols; interpretability dashboards highlighting SDoH drivers
Intervention deployment	Care plans, referrals, and outreach programs	SDoH-adaptive interventions, such as housing referral networks or transportation support	Resource allocation bias favors already advantaged groups	Equity-weighted triage algorithms; social-risk adjusted prioritization models
Outcome monitoring and governance	Outcome metrics, utilization patterns, and adherence signals	Stratified monitoring across SDoH groups; community feedback channels	Performance drift across populations; feedback loops reinforcing inequities	Continuous algorithm vigilance; equity dashboards; governance committees overseeing fairness audits

Equity evaluation frameworks, incorporating metrics such as demographic parity, equalized odds, and subgroup calibration, offer critical tools for prospective and post-deployment audits [18, 23]. These must adapt to healthcare's contextual nuances, where disparities in subgroup performance translate to tangible harms like denied interventions or misallocated resources. Governance structures—emphasizing ethical principles, stakeholder-inclusive decision-making, transparency in feature selection, and ongoing monitoring—serve as essential safeguards [19, 24]. Human-in-the-loop approaches further bolster fairness by integrating clinician judgment with algorithmic outputs, mitigating overreliance and enabling contextual overrides informed by SDoH nuances [22, 28].

The proposed closed-loop architecture **(Figure 1)** exemplifies this holistic integration: SDoH permeates every layer—from multimodal data ingestion and bias-screened preprocessing to intelligence generation, human-AI decision fusion, intervention execution, feedback-driven retraining, and governance oversight. This cyclical design supports continuous equity refinement, aligning with calls for algorithmic vigilance and adaptive recalibration to prevent drift [closed-loop examples and frameworks].

Advancing SDoH-aware AI requires transcending disciplinary silos through sustained collaboration among clinicians, data scientists, ethicists, policymakers, and community representatives. Absent proactive measures, AI risks entrenching inequities; with equity-centric design, it can dismantle them, fostering resilient systems attuned to social contexts [13, 15]. Future priorities include funding for data equity research, modernizing regulations like HIPAA for AI/big data interoperability, and developing inclusive frameworks that connect data equity to health equity outcomes.

Conclusion

The deliberate integration of social determinants of health into artificial intelligence frameworks for healthcare analytics and clinical decision support offers profound transformative capacity to cultivate equitable, efficient, and patient-centered systems. This narrative review delineates robust strategies for SDoH embedding—ranging from NLP extraction and hybrid modeling to federated learning and closed-loop architectures—while clarifying mechanisms of bias propagation and highlighting equity evaluation methodologies grounded in fairness metrics and prospective auditing.

Accumulating evidence affirms that SDoH-aware models substantially elevate predictive precision, bolster preventive care delivery, curtail avoidable costs, and narrow disparities across diverse populations. Closed-loop mechanisms, augmented by human-AI fusion and continuous monitoring, enable adaptive responses to real-world feedback, positioning AI as a self-correcting force for sustained equity gains. Practical implementations, including enriched risk stratification and closed-loop referral networks linking clinical care to community resources, illustrate tangible improvements in the identification of high-need individuals and resource optimization.

Nevertheless, enduring obstacles—data standardization deficits, interoperability challenges, scalability barriers in heterogeneous or resource-constrained contexts, and ethical governance gaps—demand persistent attention and innovation. Forward-looking priorities encompass interoperable ontologies and data infrastructures, strengthened regulatory enforcement of equity mandates, inclusive development processes incorporating diverse voices, and empirical research elucidating interconnections between data equity and health equity.

By foregrounding SDoH throughout the AI lifecycle, healthcare can pivot from passively reflecting societal inequities to actively advancing health justice. This evolution promises more resilient systems capable of addressing upstream determinants, optimizing population outcomes, and ensuring technological advancements serve all communities equitably. Realizing this vision hinges on interdisciplinary commitment, transparent governance, and unwavering focus on fairness as foundational design principles.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 13 Sep 2025 Revised: 19 Nov 2025 Accepted: 25 Jan 2026
Published online: 10 July 2026

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

Labkoff S, Oladimeji B, Kannry J, Solomonides A, Leftwich R, Koski E, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. J Am Med Inform Assoc. 2024;31(11):2730-9.

Bompelli A, Wang Y, Wan R, Singh E, Zhou Y, Xu L, et al. Social and behavioral determinants of health in the era of artificial intelligence with electronic health records: a scoping review. Health Data Sci. 2021;2021:9759016.

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.

Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023;2(6):e0000278. https://doi.org/10.1371/journal.pdig.0000278.

Guevara M, Chen S, Thomas S, Chaunzwa TL, Franco I, Kann BH, et al. Large language models to identify social determinants of health in electronic health records. NPJ Digit Med. 2024;7:6. https://doi.org/10.1038/s41746-023-00970-0.

Kino S, Hsu YT, Shiba K, Chien YS, Mita C, Kawachi I, et al. A scoping review on the use of machine learning in research on social determinants of health. SSM Popul Health. 2021;15:100836. https://doi.org/10.1016/j.ssmph.2021.100836.

Pinera PA, Kim PC, Pinera FA, Shen JJ. Social determinants and health equity activities and AI adoption in US hospitals. Int J Environ Res Public Health. 2025;22(2):294. https://doi.org/10.3390/ijerph22020294.

Liu M, Ning Y, Teixayavong S, Liu X, Mertens M, Shang Y, et al. A scoping review and evidence gap analysis of clinical AI fairness. NPJ Digit Med. 2025;8(1):360. https://doi.org/10.1038/s41746-025-01667-2.

van Kessel R, Seghers LE, Anderson M, Schutte NM, Monti G, Haig M, et al. A scoping review and expert consensus on digital determinants of health. Bull World Health Organ. 2025;103(2):110-125H. https://doi.org/10.2471/BLT.24.292057.

Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, et al. Fairness of artificial intelligence in healthcare: review and recommendations. Jpn J Radiol. 2024;42(1):3-15. https://doi.org/10.1007/s11604-023-01474-3.

Hasanzadeh F, Josephson CB, Waters G, Adedinsewo D, Azizi Z, White JA. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digit Med. 2025;8(1):154. https://doi.org/10.1038/s41746-025-01503-7.

Chinta SV , Wang Z, Palikhe A, Zhang X, Kashif A, Smith MA,et al. AI-driven healthcare: A review on ensuring fairness and mitigating bias. PLOS Digit Health. 2025. https://doi.org/10.1371/journal.pdig.0000864.

Colacci M, Huang YQ, Postill G, Zhelnov P, Fennelly O, Verma A, et al. Sociodemographic bias in clinical machine learning models. J Clin Epidemiol. 2025;178:111606. https://doi.org/10.1016/j.jclinepi.2024.111606.

Anderson JW, Visweswaran S. Algorithmic individual fairness and healthcare: a scoping review. JAMIA Open. 2024;8(1):ooae149.

Carey S, Pang A, Kamps M. Fairness in AI for healthcare. Future Healthc J. 2024;11(3):100177. https://doi.org/10.1016/j.fhj.2024.100177.

Mienye ID, et al. A survey of bias and fairness in healthcare AI. IEEE Access. 2024. https://doi.org/10.1109/ACCESS.2024.10628826.

Rangachari P, Thapa A. Impact of hospital initiatives addressing social determinants of health in the US: a scoping review. BMC Health Serv Res. 2025;25;342. https://doi.org/10.1186/s12913-025-12494-2.

Song J, Hobensack M, Sequeira L, Shin HD, Davies S, Peltonen LM, et al. Social Determinants of Health in Digital Health Policies: an International Environmental Scan. Yearb Med Inform. 2024;33(1):283-91. https://doi.org/10.1055/s-0044-1800759.

Kumar R, Sporn K, Waisberg E, Ong J, Paladugu P, Vadhera AS, et al. Navigating healthcare AI governance: algorithmic oversight and equity framework. J Med Internet Res. 2025.

Kwon IG, Kim SH, Martin D. Integrating Social Determinants of Health to Precision Medicine through Digital Transformation: An Exploratory Roadmap. Int J Environ Res Public Health. 2021;18(9):5018. https://doi.org/10.3390/ijerph18095018.

Liu C, et al. Digital transformation in measuring social determinants of health: a systematic review. J Glob Inf Manag. 2023

Olawade DB, Plabon SB, Ojo A, Ogunbona MA, Makanjuola BD, Olasilola OR. Human in the loop artificial intelligence in healthcare: applications, outcomes, and implementation challenges. Int J Med Inform. 2026;213:106362. https://doi.org/10.1016/j.ijmedinf.2026.106362.

Baker E, August K, Mangra N, Kalogeropoulos D, Mellik A, Muller P, et al. The impact of a continuum of care on social determinants of health. Telehealth Med Today. 2025. https://doi.org/10.30953/thmt.v10.557.

Basch CE, Basch CH. Artificial intelligence, digital media, and population health: Exposure science and social determinants of health. Ann N Y Acad Sci. 2025;1552(1):5-11. https://doi.org/10.1111/nyas.70020.

Morita PP, Sahu KS, Oetomo A. Health monitoring using smart home technologies: scoping review. JMIR Mhealth Uhealth. 2023;11:e37347. https://doi.org/10.2196/37347.

Andargoli AE, Ulapane N, Nguyen TA, Shuakat N, Zelcer J, Wickramasinghe N. Intelligent decision support systems for dementia care: A scoping review. Artif Intell Med. 2024;150:102815. https://doi.org/10.1016/j.artmed.2024.102815.

Ports K, Dai J, Conniff K, Corrada MM, Manson SM, O'Connell J, et al. Machine learning to predict dementia for American Indian and Alaska Native peoples. Lancet Reg Health Am. 2025;43;101013. https://doi.org/10.1016/j.lana.2025.101013.

Thovinakere N, Ghosh SS, Itturia-Medina Y, Geddes MR, . Social determinants of health and brain connectivity predicting physical activity. J Am Heart Assoc. 2025 https://doi.org/10.1101/2024.09.30.24314678.

Fareed M, Fatima M, Uddin J, Ahmed A, Sattar MA. Ethical considerations of large language models in healthcare: a systematic review. J Med Syst. 2025