Applying the Deep Learning–Sector–Governance (DLSG) Framework to the U.S. Healthcare System: Opportunities, Deployment Pathways, and Policy-Aligned Evaluation

Yisong Chen[1]*, Chuqing Zhao[2], Yishan Zhong[3],
Ziyu Wang[4], Jiazhao Shi[5], Wenjia Zheng[6]

[1] College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, United States
[2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States
[3]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, United States
[4] Independent Researcher, Washington, DC 20004, United States
[5] Tandon School of Engineering, New York University, Brooklyn, NY, United States
[6] Computer and Information Science Department, Fordham University, Bronx, NY, United States
*Email: ychen841@gatech.edu

Abstract

The United States healthcare system faces persistent challenges in cost, access, quality, and administrative burden. Although deep learning (DL) has achieved strong performance in prediction and representation learning, real-world impact remains constrained by data heterogeneity, workflow integration, and regulatory and governance requirements. This paper adapts the Deep Learning–Sector–Governance (DLSG) framework to the U.S. healthcare context, positioning it as a deployment blueprint that aligns model design, sector workflows, and governance controls. We examine applications spanning clinical decision support, imaging triage, operations, administrative automation, and payment integrity, and outline evaluation strategies that prioritize safety, calibration, and policy-aligned outcomes over standalone accuracy metrics. Given the scale of documented improper payments and administrative burden in U.S. healthcare, governance-aligned efficiency and integrity gains could plausibly translate into about $30–60 billion in annual system-level savings, depending on adoption depth and sustained operational alignment. We argue that the primary contribution of DLSG is to improve in predictive accuracy and to increase the likelihood that algorithmic advances translate into measurable improvements in patient outcomes and system efficiency under real regulatory and operational constraints. In addition, this framework is in alignment with national health objectives such as those articulated in Healthy People 2030.

Keywords: Deep Learning in Healthcare; AI Governance; Clinical Decision Support; Payment Integrity; Health System Evaluation; Digital Health Deployment

Article History:
Received November 05, 2025
Revised January 02, 2026
Accepted March 18, 2026
Available Online March 30, 2026

Applying the Deep Learning–Sector–Governance (DLSG) Framework to the U.S. Healthcare System: Opportunities, Deployment Pathways, and Policy-Aligned Evaluation

1. Introduction

The U.S. healthcare system has high spending, uneven outcomes, and fragmented care delivery. It also has large administrative overhead. Providers, payers, and public agencies are now trying to use machine learning (ML) and deep learning (DL) to improve care, reduce avoidable utilization, and limit fraud, waste, and abuse. At the same time, healthcare is high-stakes. Because the model is tightly integrated in the clinical workflow, errors will negatively impact the safety of the patients. On top of that, regulatory and ethical constraints can limit the healthcare data used as the model input. [Liao et al., 2025; Chen et al., 2025; Chen et al., 2026; Wiens et al., 2019; Wong et al., 2021; Topol, 2019; Murff et al., 2011]

U.S. health care spending reached $4.5 trillion dollars in 2022, and it was 17.3% of GDP (down from 19.5% in 2020 and 18.2% in 2021). In FY2023, CMS reported estimated improper payments of $31.2B (Medicare FFS), $16.6B (Medicare Part C), $3.4B (Medicare Part D), $50.3B (Medicaid), and $2.1B (CHIP). For Medicaid, CMS noted that 82% of improper payments were driven by insufficient documentation. The HHS OIG/DOJ HCFAC program also reported that civil health care fraud settlements and judgments under the False Claims Act exceeded $1.8 billion, and that more than $3.4 billion was returned to the Federal Government or paid to private persons in FY2023. These numbers show why evaluation should go beyond model accuracy. For DLSG-aligned systems, we focus on measurable outcomes such as cost and administrative throughput, time-to-care, and payment integrity, while still accounting for safety and compliance [Himmelstein et al., 2020]. [Bao et al., 2023; Tamo et al., 2023; Zhong et al., 2025].

In 2022, total health spending was $4.5 Trillion and 92% of the population was insured (26.6M uninsured). These baselines support evaluations that include access, cost, and administrative burden, not only predictive accuracy of the model on paper [CMS, 2022; Lin, 2025; Esteva et al., 2017; Gulshan et al., 2016; Wynants et al., 2020; Rajkomar et al., 2019].

The Deep Learning–Sector–Governance (DLSG) framework is a sector-aware way to move from algorithm ideas to compliant and scalable deployment [Chen et al., 2025]. DLSG highlights three parts that should be designed together: (1) the DL method (model architecture, training setup, and uncertainty handling), (2) the sector context (how data is generated, operational constraints, and incentives), and (3) governance (privacy, security, auditing, regulation, and accountability). In this paper, we apply DLSG to U.S. healthcare. We argue that many failures in healthcare AI are not only accuracy problems. They are integration and governance problems that block deployment even when a model performs well offline. [Rajpurkar et al., 2018]

Contributions:

1) We define a DLSG mapping for U.S. healthcare and translate sector and governance needs into concrete design controls.

ISSN-3067-7505
See: https://inatgi.in/index.php/jtis/index for more information. DOI: 10.63646/jtis.2026.040103

2) We review high-impact applications in provider and payer settings and explain how DLSG affects model choice and deployment decisions.

3) We propose evaluation and monitoring criteria that emphasize clinical utility, safety, fairness, and operational value. [Obermeyer et al., 2019; Wolff et al., 2019]

4) We present six real-world case studies (with a payer/insurance focus) and summarize DLSG-aligned lessons for implementation.

2. Background: DLSG and U.S. Healthcare Constraints

The Deep Learning–Sector–Governance (DLSG) framework conceptualizes AI deployment in regulated environments as the interaction of three mutually constraining domains: algorithmic capability, sector-specific operational structure, and governance requirements. In this framework, model development is considered alongside operational and regulatory constraints, because the predictive performance, workflow integration, and regulatory accountability are co-determined for deployment success.

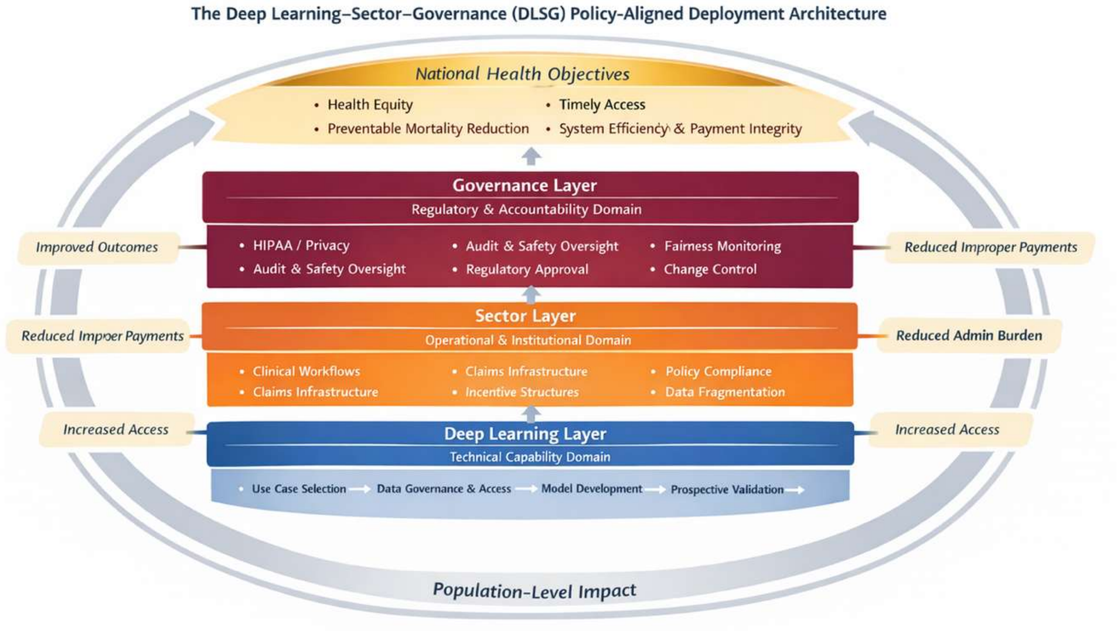

Figure 1. DLSG mapping from deep learning components to sector constraints, governance, and outcomes.

At the technical level, deep learning encompasses model architecture, representation learning strategies, training data regimes, robustness, calibration, and uncertainty estimation. These design choices shape predictive capacity but do not, by themselves, ensure safe or scalable deployment. At the sector level, AI systems must operate within domain-specific data generation processes, documentation practices, workflow constraints, and incentive structures. In healthcare, this includes Electronic Health Record (EHR) structures, billing systems, institutional variation, and resource limitations. At the governance level, deployment is shaped by legal and regulatory requirements, privacy and security controls, auditability standards, and accountability mechanisms that govern how model outputs may influence decisions. [Johnson et al., 2016; Levit et al., 1990; Habib, 2025]

In real world, these domains interact closely pre- and post- development. For example, a high-capacity transformer trained on clinical notes may improve discrimination for risk prediction, yet governance constraints may restrict data movement or impose documentation standards, while sector constraints may require low-latency inference within an EHR interface. Under DLSG, model design, operational integration, and compliance obligations are not sequential layers but interacting constraints that jointly determine feasibility and impact. Figure 1 illustrates this mapping from deep learning components through sector realities and governance controls to measurable system-level outcomes.

The U.S. healthcare system has special characteristics that make this integration particularly salient. Healthcare data are heterogeneous, spanning claims, structured EHR fields, clinical narratives, imaging, laboratory values, wearable signals, and social determinants of health. Interoperability tends to be incomplete, documentation practices can vary across institutions, and many labels are administratively defined rather than clinically precise. For example, diagnosis codes often reflect billing requirements rather than ground truth disease states. These factors introduce missingness, measurement error, and institutional variability that directly affect model design and evaluation. [Richman et al., 2022; Tseng et al., 2018; Bodenheimer, 2005]

Deployment must also respect workflow constraints. Clinicians operate under time pressure, attention is limited, and alert fatigue is well documented. Explainability requirements differ by use case, and model outputs must integrate into existing care pathways without increasing cognitive or administrative burden. These operational conditions reinforce the DLSG premise that predictive accuracy alone is insufficient.

Insurance and payer environments introduce additional complexities. Although claims data are standardized through coding systems such as ICD-10-CM, CPT/HCPCS, and NDC, they are optimized for billing rather than clinical representation. Claims exhibit time lags, periodic code-set revisions, benefit design effects, and policy rule overlays that complicate causal interpretation. so, payer-oriented models must be evaluated on discrimination metrics and on downstream operational outcomes, that includes reduced improper payments, improved adjudication speed, and measurable member-level impact. [Bodenheimer, 2005]

Governance considerations are central to this deployment landscape. In the United States, privacy and security expectations are shaped by HIPAA and its Security Rule [U.S. Department of Health and Human Services (HHS)], and strengthened through the HITECH Act [U.S. Congress, 2009]. For some intended use, regulatory oversight by the Food and Drug Administration may apply to software functioning as a medical device (SaMD) [FDA 2018a. FDA, 2018b, FDA, 2021]. Beyond statutory requirements, healthcare organizations maintain internal governance processes governing model approval, version control, monitoring, incident response, and audit logging. Under DLSG, governance is embedded within the lifecycle from use-case definition through post-deployment monitoring, rather than treated as an external constraint applied after model development.

3. From Application to Policy-Aligned Impact: Operationalizing DLSG in U.S. Healthcare

Clinical decision support and risk prediction represent some of the most developed applications of deep learning in healthcare. Representative use cases include early warning for sepsis, prediction of readmission risk, clinical deterioration modeling, and care pathway optimization [Rajkomar et al., 2018]. Within a DLSG framework, model design must align with the structure of EHR data. This often favors sequence-based architectures and calibrated uncertainty estimates that support safe escalation. On the other hand, the selection of prediction targets should be directly tied to actionable clinical interventions rather than abstract risk scores. Deployment so requires clearly defined human-in-the-loop policies, documentation standards, and systematic drift-monitoring to ensure ongoing safety and reliability.

From a payer perspective, prediction targets must align with operational levers such as case management enrollment, post-discharge follow-up, and medication adherence outreach. DLSG emphasizes that intervention protocols should be specified before model deployment so that success is measured in terms of improved outcomes resulting from paired interventions, rather than solely by improvements in AUROC or other offline metrics.

Deep learning also demonstrates strong potential in imaging-centered workflows, particularly in radiology and pathology triage, where models can prioritize suspicious studies, support quality control, and automate quantitative measurements [De Fauw et al., 2018; Li et al., 2021; Cohen et al., 2017]. In this setting, DLSG requires robust vision models that account for domain shift across scanners and institutions, careful integration into PACS and notification workflows, and governance mechanisms that preserve auditability and validate performance across sites and subpopulations. Because imaging outputs may influence time-sensitive decisions, escalation pathways and continuous monitoring are essential components of safe deployment.

Beyond direct clinical diagnosis, hospitals can use DL and ML systems to forecast bed demand, optimize staffing allocation, and manage operating room scheduling. Although these applications often carry lower direct regulatory risk, they can have large operational impact. DLSG highlights the importance of aligning predictive outputs with concrete operational levers such as discharge planning, staffing adjustments, and throughput management, while also monitoring for unintended inequities in resource allocation across patient populations.

Administrative functions—that includes coding, billing, documentation, and prior authorization—represent another high-impact domain. Administrative burden remains a major driver of healthcare costs, and natural language processing systems can assist with documentation summarization, coding recommendations, denial management, and prior authorization preparation. Under DLSG principles, generated outputs should be retrieval-grounded and traceable, with strict controls over protected health information access. Traceability must extend to documenting that sources informed each generated output, how it was reviewed by humans, and how it influenced downstream decisions, thereby maintaining accountability across administrative workflows.

Fraud, waste, and abuse detection—particularly in claims-based environments—remains one of the most policy-aligned applications of deep learning. Claims anomaly detection systems can identify suspicious billing patterns, upcoding behavior, or provider-level fraud risks. Because such signals are sparse and often adversarial, DLSG recommends imbalance-aware training strategies and graph or sequence modeling of claims trajectories. Decision thresholds must be calibrated to investigative capacity, and explainability and auditability are critical, given that model outputs may trigger claim denial, recoupment, or formal investigation. Public-sector payers have already documented large-scale analytics programs aimed at fraud prevention and improper payment reduction [Centers for Medicare & Medicaid Services (CMS). Government Accountability Office (GAO). HHS Office of Inspector General (OIG)], reinforcing the importance of governance-aligned deployment in this domain.

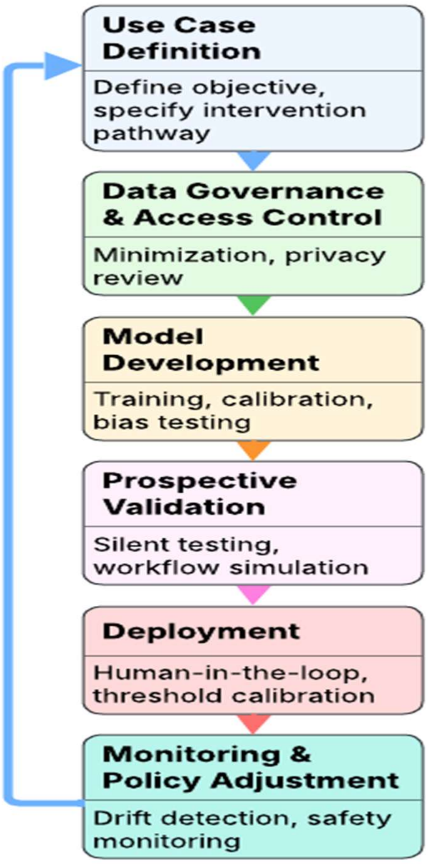

Figure 2. DLSG Deployment Lifecycle

Translating these application domains into durable system impact requires a structured deployment architecture. Under DLSG, deployment is conceptualized not as a one-time model release but as a continuous, governance-aware lifecycle. A practical DLSG-aligned system begins with data ingestion governed by minimization principles, that includes purpose limitation and least-privilege access. Feature and embedding storage must be subject to strict access controls and retention policies. Inference services require versioning and rollback capabilities to ensure seamless updates. Continuous monitoring should track predictive performance, calibration, drift, bias, latency, and safety-sensitive signals. Human review interfaces remain essential for high-stakes decisions, particularly where outputs influence payment adjudication and investigative actions. Comprehensive audit logging and incident response protocols ensure traceability and institutional accountability.

In payer workflows, additional architectural requirements arise. Policy rule versioning is necessary to accommodate changes in benefit design, reimbursement rules, and coverage criteria. Code-set updates, that includes ICD and CPT revisions, must be tracked to prevent unintended degradation caused by shifting labeling conventions. End-to-end traceability is required so that model scores can be linked directly to adjudication downstream financial or clinical consequences. These sector-specific constraints illustrate how deployment architecture is co-determined by technical design, operational realities, and governance requirements.

Evaluation under DLSG extends beyond traditional offline discrimination metrics. While measures such as AUROC provide an initial assessment of predictive capacity, healthcare deployment requires a broader set of endpoints that connect algorithmic outputs to operational and policy outcomes. DLSG-aligned evaluation incorporates payment-integrity and administrative-burden indicators, that includes review routing yield, documentation completeness, and appeal cycle time. Outcome-linked measures—such as avoided improper payments, reduced time-to-treatment, prevention of avoidable denials, and reduced avoidable utilization—provide stronger evidence of real-world impact. Simultaneously, governance-linked measures, that includes incident rates, drift triggers, auditability compliance, and fairness monitoring, are necessary to ensure sustained alignment with regulatory and institutional standards.

Beyond discrimination metrics, healthcare evaluation should assess calibration and decision-curve analysis such as time-to-treatment and claims turnaround time. Prospective validation is also critical before model outputs influence clinical or financial decisions. Post-deployment monitoring, supported by formal change-management processes, ensures that updates remain transparent and accountable.

In payer contexts specifically, evaluation endpoints include reduction in improper payments and improved recovery yield net of investigation cost [Government Accountability Office (GAO), 2024. HS Office of Inspector General (OIG), 2024], reduced turnaround time for claims adjudication without and increases denial error, decreased avoidable utilization when paired with care management interventions, and fairness monitoring across plan types, geographies, and vulnerable subpopulations. By linking predictive systems to measurable operational, financial, and population-level outcomes, DLSG reframes evaluation from isolated model performance toward policy-aligned system impact.

4. Real-World Case Studies (10+), With DLSG Lessons (Focus: Medical + Medical Insurance)

This section condenses the prior set of twelve brief examples into six richer, end-to-end cases. Each case includes (i) a short context, (ii) how DLSG shapes the technical + organizational design, and (iii) forward-looking recommendations for scaling impact in U.S. healthcare (Medicare/Medicaid and commercial).

Case A— Payment integrity at scale (Medicare + commercial): fraud, waste, and abuse (FWA) risk scoring

Brief: Payers process massive claim volumes under tight timeliness constraints. FWA signals are sparse, adversarial, and change over time. false positives create provider abrasion and downstream administrative load.

DLSG application: (DL) use calibrated risk scoring + anomaly detection with human-in-the-loop labeling. (Sector) align features and labels to claim-adjudication rules, coding practices, and provider network structure. (Governance) ensure audit trails, explainable factors for adverse actions, privacy/security controls, and post-deployment drift monitoring.

Recommendations: use risk-based routing (auto-pay vs targeted review), maintain a formal model change-control board, and track end outcomes (recoveries, avoided improper payments, reviewer throughput) rather than offline AUROC alone.

Case B— Prior authorization (PA) modernization: risk-based automation instead of blanket automation

Brief: PA is a major source of friction and delays. Naïve automation can amplify inequities or deny necessary care. purely manual workflows do not scale.

DLSG application: (DL) triage requests into ‘low-risk’ vs ‘needs clinical review’ using uncertainty estimates. (Sector) encode guideline logic and benefit design. (Governance) implement appealability, clinician override, and monitoring for disparate impact.

Recommendations: publish a measurement plan (turnaround time, approval/denial parity, downstream utilization), run silent trials before go-live, and keep a policy layer that is separable from the model.

Case C— Denials & appeals operations: NLP to reduce administrative burden

Brief: Denials and appeals generate high administrative cost for providers and payers. Many cases are ‘documentation-completeness’ issues rather than true medical-necessity disputes.

DLSG application: (DL) document classification + information extraction to pre-fill packets and flag missing elements. (Sector) integrate with EHR/RCM tooling and payer portals. (Governance) maintain provenance for extracted fields and limit model hallucination via constrained extraction.

Recommendations: prioritize high-volume denial reasons, measure staff minutes saved, and create feedback loops from appeal outcomes to improve extraction rules/models.

Case D— Imaging triage with regulatory constraints: stroke LVO / critical findings routing

Brief: Time-sensitive imaging (e.g., stroke) benefits from rapid triage, but errors can be catastrophic. FDA-cleared workflows exist, yet deployment success depends on site integration and continuous monitoring.

DLSG application: (DL) robust models + calibration with site-specific validation. (Sector) align to radiology and ED workflows (notification routing, escalation paths). (Governance) safety monitoring, incident reporting, and version control across sites.

Recommendations: require prospective 'silent mode' evaluation, define alert thresholds per site capacity, and monitor alert fatigue and time-to-treatment metrics.

Case E— Sepsis/clinical deterioration alerts: the 'workflow + governance' failure mode

Brief: Sepsis early-warning is a classic high-stakes prediction problem. public controversies highlight that poor calibration, unclear labeling, and weak governance can undermine trust.

DLSG application: (DL) emphasize calibration, temporal validation, and missingness handling. (Sector) pair prediction with actionable intervention protocols. (Governance) transparent validation reports, bias checks, and clear accountability when models change.

Recommendations: treat the model as part of a socio-technical system—evaluate whether interventions triggered by alerts improve outcomes. publish model factsheets and run periodic re-validation.

Case F— Ambient clinical documentation: efficiency gains with privacy + safety constraints

Brief: Automated note generation can reduce clinician time on documentation, but introduces privacy risk and clinical safety risk if generated text is wrong.

DLSG application: (DL) constrained generation + attribution, with human verification. (Sector) integrate into note-sign workflows and specialty templates. (Governance) strict PHI handling, access control, retention policy, and monitoring for hallucinations/unsafe suggestions.

Recommendations: start with low-risk sections (e.g., HPI summarization), require explicit clinician attestation, and measure time saved plus downstream coding quality/denial rates.

5. IMPACT ESTIMATION

While DLSG is a deployment and governance framework rather than a single predictive intervention, its potential economic impact is best understood when we examine the scale of addressable cost pools in U.S. healthcare. Publicly reported data indicate that annual improper payments across major federal programs exceed $100 billion, and administrative activities are frequently estimated to account for 15–25% of total healthcare spending, with total national health expenditures reaching about $4.5 trillion in 2022[CMS, 2022. CMS,

2023. OIG, 2024]. These figures define the structural domains in that DLSG-aligned deployment may plausibly generate measurable savings.

Reductions in improper payments—5–10%—could save several billion dollars annually. Similarly, if governance-aligned AI systems reduced only 1–2% of administrative overhead through improved prior authorization routing, denial management, documentation support, and claims processing efficiency, the resulting savings would likely fall in the tens of billions of dollars per year. Under conservative assumptions, system-wide impact may so plausibly reach about $10–20 billion annually. Broader and mature deployment across both public and commercial payers could raise this range further, potentially into the $30–60 billion range, depending on adoption depth and sustained operational alignment.

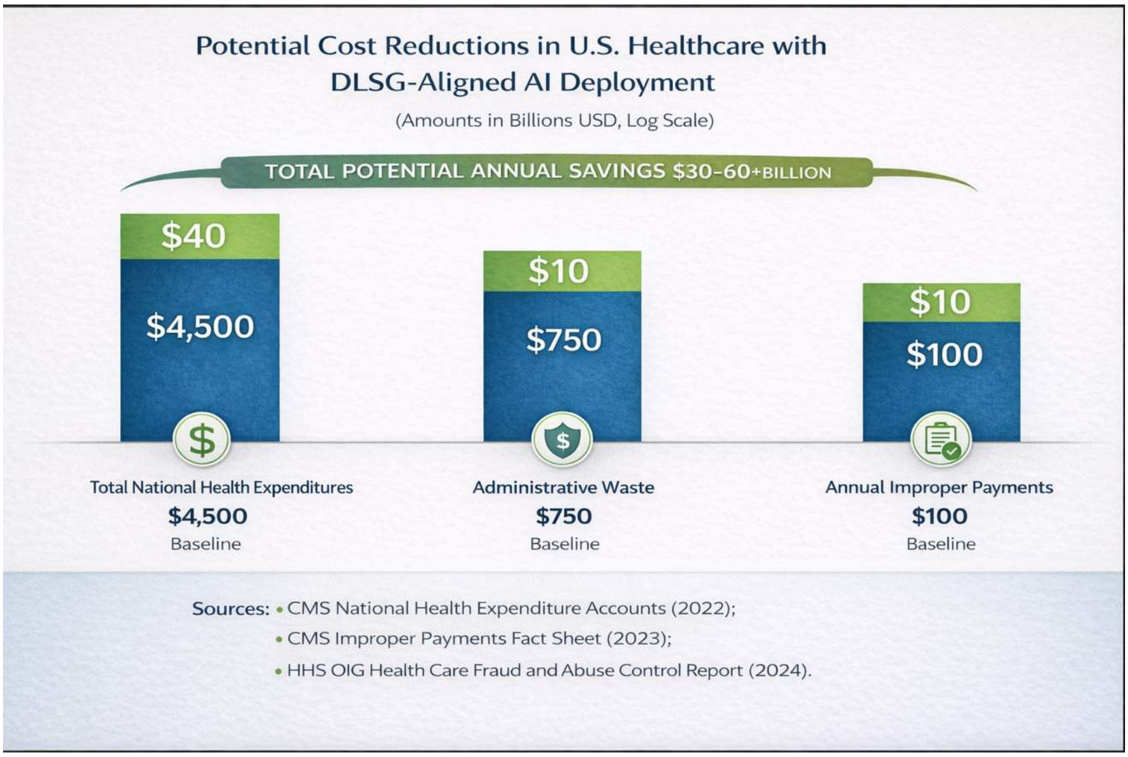

Figure 3. Potential Cost Reductions in U.S. Healthcare with DLSG-Aligned AI Development

These estimates do not assume dramatic automation or workforce displacement. Rather, they reflect incremental efficiency and integrity gains made durable through calibration, workflow integration, auditability, and monitoring—features central to the DLSG framework., DLSG's economic contribution lies less in and increases model accuracy and more in and increases the probability that predictive improvements translate into realized system-level value.

These figures should be interpreted as directional rather than causal estimates. However, given the scale of documented improper payments and administrative burden, even small percentage improvements enabled by governance-aligned deployment could save material national impact.

6. Alignment With Healthy People 2030 National Objectives

The Deep Learning–Sector–Governance (DLSG) framework is aligned closely with national public health priorities articulated in the Healthy People 2030. An initiative led by the U.S. Department of Health and Human Services, Healthy People 2030 define measurable objectives dedicated to advancing health equity, expanding access to high-quality care, and optimizing health system performance. Although DLSG is primarily a deployment and governance framework, its focus on policy-aligned evaluation inherently situates technical AI implementation within this broader national agenda.

A central feature of Healthy People 2030 is its focus on reducing disparities across socioeconomic, geographic, and demographic dimensions [HHS, 2024]. DLSG contributes to this objective by requiring subgroup validation, fairness monitoring, and drift detection as part of routine model governance. In high-stakes settings such as prior authorization, risk prediction, and payment integrity analytics, performance variation across plan types and vulnerable populations can produce downstream inequities. By embedding equity-sensitive evaluation directly into deployment pipelines, DLSG transforms fairness from a post-hoc analytic concern to a continuous operational requirement, thereby aligning technical deployment with national equity goals.

Healthy People 2030 also prioritize timely access to care and the mitigation of preventable morbidity and mortality. DLSG-aligned systems—such as risk-based routing in utilization management—are evaluated not solely on predictive discrimination but on their effect on intervention effectiveness. This emphasis on operational endpoints connects model performance to real-world health outcomes, consistent with national objectives targeting preventable hospitalizations, emergency response timeliness, and care coordination.

Furthermore, Healthy People 2030 underscore the importance of high-quality, efficient health services. Administrative burden, documentation gaps, and improper payments represent structural inefficiencies that indirectly affect patient access and resource allocation. By promoting calibrated risk routing, auditability, and workflow-aware automation, DLSG is able improve in system efficiency. These improvements, while often framed in economic terms, also serve broader public health objectives by reallocating resources for direct patient care and strengthening institutional accountability.

DLSG does not redefine national health goals. rather, it offers an implementation lens through that AI deployment can be evaluated against already established public policy targets. By explicitly linking evaluation endpoints—such as reduced improper payments—to nationally defined objectives, DLSG reframes model success as contribution to measurable population-level impact., the framework provides a bridge between algorithmic innovation and the strategic priorities articulated in Healthy People 2030.

7. Challenges And Future Directions

Despite growing enthusiasm for artificial intelligence in healthcare, large structural and institutional barriers continue to constrain real-world impact. A central challenge lies in the fragmentation of data systems across providers, payers, and public agencies. The current HER systems are often siloed and governed by distinct access policies. Site variation further complicates deployment. For example, local clinical norms differ across institutions could limit model portability and and increases the risk of performance degradation when systems are transferred across settings. In parallel, incentives between payers and providers are not always aligned. While payers may prioritize cost containment and payment integrity, providers often prioritize throughput, quality metrics, and patient satisfaction. Models that optimize one objective without accounting for institutional incentive structures risk generating friction and unintended consequences.

Governance constraints introduce an additional layer of complexity. Privacy and security compliance requirements demand strict controls over protected health information, and often limits data movement and model retraining strategies. Vendor risk management processes require contractual clarity and ongoing monitoring of third-party systems. As AI systems increasingly generate recommendations or draft clinical and administrative content, accountability for automated outputs becomes a critical concern. Institutions must determine how responsibility is allocated when model-generated suggestions influence treatment decisions or denial actions. Safe handling of generated content—that includes mitigation of hallucinations—requires formal oversight mechanisms that extend beyond traditional model validation.

Experience across early deployments suggests that successful implementation rarely begins with full automation. Instead, high-impact systems typically start as decision support or risk-based routing tools, augmenting rather than replacing human judgment. Automation expands gradually as governance evidence accumulates and institutional trust increases. Prospective validation strategies, that includes silent-mode or shadow deployments in that predictions are generated but not acted upon, allow organizations to assess calibration and subgroup performance before exposing patients or financial processes to risk. Documentation practices are equally important. Model cards, change logs, and explicit statements of intended use can help clarify operational boundaries and known failure modes, supporting both internal governance and external auditability [Mitchell et al., 2019].

Privacy-preserving technical approaches represent another recurring mitigation pattern. De-identification and federated learning architectures can enable cross-site collaboration while limiting centralized data exposure [Nguyen et al., 2021]. However, these approaches must be paired with rigorous evaluation to ensure that privacy constraints do not inadvertently degrade performance or amplify disparities. In high-stakes contexts, technical safeguards alone are insufficient, and they must be embedded within institutional processes for model lifecycle management.

Looking forward, several directions appear particularly promising. Multimodal foundation models may improve representation learning across heterogeneous inputs. Advances in uncertainty estimation and calibration can strengthen trustworthiness in risk-sensitive applications such as triage and utilization management. Governance tooling is also evolving, with and increases emphasis on standardized reporting templates, automated audit trails, and continuous fairness monitoring embedded directly within production systems. For payer organizations, the highest-impact work will likely couple technical innovation with policy-aligned evaluation frameworks, explicitly measuring net improper-payment reduction, member-level outcomes, turnaround time, and access-related harms., the next phase of progress may depend less on incremental gains in predictive accuracy and more on strengthening the institutional architectures that allow AI systems to operate safely, equitably, and accountably at scale.

8. Conclusion

Adapting the Deep Learning–Sector–Governance (DLSG) framework to the U.S. healthcare landscape underscores a critical reality: successful AI deployment is not merely a function of predictive accuracy. Rather, it is contingent upon how effectively models are embedded within complex clinical and administrative workflows. In a high-stakes, regulated environment, algorithmic performance transcends statistical metrics, requiring evaluation that accounts for operational constraints, human agency, and institutional accountability [Tomašev et al., 2019].

By harmonizing deep learning architecture with sector-specific data realities, DLSG operationalizes a pragmatic pathway for translating algorithmic innovation into systemic impact. Across domains ranging from imaging triage to payment integrity, the framework posits that calibration, auditability, and continuous monitoring are the foundations of durable value. Ultimately, this approach reframes success, moving beyond isolated performance indicators toward sustained gains in healthcare access, safety, efficiency, and equity.

DLSG does not propose new health policy objectives. rather, it offers an implementation lens through that AI systems can be evaluated against existing national priorities, that includes those articulated in Healthy People 2030. By grounding evaluation in policy-aligned endpoints—such as reduced administrative burden, improved timeliness of care, equitable performance across populations, and strengthened payment integrity—the framework increases the likelihood that AI deployment contributes meaningfully to population-level health outcomes.

This work argues that the primary contribution of DLSG lies in and increases the probability that advances in deep learning translate into real-world impact under regulatory and operational constraints. As healthcare organizations continue to invest in AI capabilities, frameworks that integrate technical performance with sector context and governance will be essential to achieving high-impact, trustworthy, and policy-aligned deployment at scale.

References

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digital Medicine, 1(1). https://doi.org/10.1038/s41746-018-0040-6

American Medical Association. (2023). 2023 AMA prior authorization physician survey.

Bao, H., Deng, J., Xing, S., Zhong, Y., Shi, W., Marteau, B., Das, B., Shehata, B., Deshpande, S., & Wang, M. D. (2023). Rare heart transplant rejection classification using diffusion-based synthetic image augmentation. In 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 1–4). https://doi.org/10.1109/bhi58575.2023.10313377

Bodenheimer, T. (2005). High and rising health care costs. Part 1: Seeking an explanation. Annals of Internal Medicine, 142(10), 847–854. https://doi.org/10.7326/0003-4819-142-10-200505170-00010

Bodenheimer, T. (2005). High and rising health care costs. Part 2: Technologic innovation. Annals of Internal Medicine, 142(11), 932–937. https://doi.org/10.7326/0003-4819-142-11-200506070-00012

Centers for Disease Control and Prevention. (2022). CDC guideline for prescribing opioids for chronic pain. https://www.cdc.gov

Centers for Medicare & Medicaid Services. (2022). National health expenditures 2022 highlights. https://www.cms.gov/newsroom/fact-sheets/national-health-expenditures-2022-highlights

Centers for Medicare & Medicaid Services. (2023). Fiscal year 2023 improper payments fact sheet. https://www.cms.gov/newsroom/fact-sheets/fiscal-year-2023-improper-payments-fact-sheet

Centers for Medicare & Medicaid Services. (n.d.). Fraud prevention system (FPS). https://www.cms.gov

Chen, Y., Gao, Y., Yu, S., Zhao, C., & Lu, Y. (2026). Large language models in mental health: A systematic review of applications, innovations, and ethical challenges. Journal of Industrial Integration and Management, 1–33. https://doi.org/10.1142/s2424862225300042

Chen, Y., Zhao, C., & Gao, Y. (2025). Blockchain applications in health insurance: A review of applications, challenges, and prospects. Frontiers in Blockchain, 8. https://doi.org/10.3389/fbloc.2025.1699290

Chen, Y., Zhao, C., & Nie, C. (2025). Health insurance fraud detection: The role of feature engineering and preprocessing techniques. In Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence (pp. 858–862). https://doi.org/10.1145/3745238.3745373

Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2025). Deep learning in financial fraud detection: Innovations, challenges, and applications. Data Science and Management. https://doi.org/10.1016/j.dsm.2025.08.002

Chen, Z., Zhao, C., Zhao, M., Zhan, Q., Dong, Y., Zhao, S., Yu, S., Yao, C., Xie, Y., Dong, Z., Sun, Q., Liu, Y., Yu, C. H., Yang, H., & Wang, G. (2025). Rethinking subjective trust in LLM: Actualizing tangibility from uncertainty. TechRxiv. 10.36227/techrxiv.176463545.53813864/v1

Cohen, J. P., Irvin, J., Ding, D., Kelly, K., & Rajpurkar, P. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv. https://arxiv.org/abs/1711.05225

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine, 24(9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

Food and Drug Administration. (2018). FDA permits marketing of first medical device to use artificial intelligence to detect diabetic retinopathy. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-first-medical-device-use-artificial-intelligence-detect-diabetic

Food and Drug Administration. (2018). Software as a medical device (SaMD). https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd

Food and Drug Administration. (2021). Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-samd

Food and Drug Administration. (2025). 510(k) summary for Viz.ai LVO detection.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402. https://doi.org/10.1001/jama.2016.17216

Habib, A. R. (2025). Health care administrative costs. JAMA, 334(1), 89. https://doi.org/10.1001/jama.2025.3761

Himmelstein, D. U., Campbell, T., & Woolhandler, S. (2020). Health care administrative costs in the United States and Canada, 2017. Annals of Internal Medicine, 172(2), 134–142. https://doi.org/10.7326/m19-2818

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3(1). https://doi.org/10.1038/sdata.2016.35

Levit, K. R., Freeland, M. S., & Waldo, D. R. (1990). National health care spending trends: 1988. Health Affairs, 9(2), 171–184. https://doi.org/10.1377/hlthaff.9.2.171

Li, E., Banerjee, S. S., Huang, S., Iyer, R. K., & Chen, D. (2021). Improved GPU implementations of the Pair-HMM forward algorithm for DNA sequence alignment. In Proceedings of the IEEE 39th International Conference on Computer Design (ICCD) (pp. 299–306). https://doi.org/10.1109/ICCD53106.2021.00055

Liao, R., Zhao, C., Li, J., Feng, W., Lyu, Y., Chen, B., & Yang, H. (2025). CATP: Cross-attention token pruning for accuracy-preserved multimodal model inference. In 2025

IEEE Conference on Artificial Intelligence (CAI) (pp. 1100–1104). https://doi.org/10.1109/cai64502.2025.00191

Lin, Z. (2025). Tax share analysis and prediction of kernel extreme learning machine optimized by vector weighted average algorithm. Advances in Economics, Management and Political Sciences, 194(1), 1–8. https://doi.org/10.54254/2754-1169/2025.24147

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from electronic health records. Scientific Reports, 6. https://doi.org/10.1038/srep26094

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229). https://doi.org/10.1145/3287560.3287596

Murff, H. J., FitzHenry, F., Matheny, M. E., Gentry, N., Kotter, K. L., Crimin, K., Dittus, R. S., Rosen, A. K., Elkin, P. L., Brown, S. H., & Speroff, T. (2011). Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA, 306(8). https://doi.org/10.1001/jama.2011.1204

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for Internet of Things: A comprehensive survey. IEEE Communications Surveys & Tutorials, 23(3), 1622–1658. https://doi.org/10.1109/COMST.2021.3075439

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1). https://doi.org/10.1038/s41746-018-0029-1

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C. P., et al. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Medicine, 15(11), e1002686. https://doi.org/10.1371/journal.pmed.1002686

Richman, B. D., Kaplan, R. S., Kohli, J., Purcell, D., Shah, M., Bonfrer, I., Golden, B., Hannam, R., Mitchell, W., Cehic, D., et al. (2022). Billing and insurance–related administrative costs: A cross-national analysis. Health Affairs, 41(8), 1098–1106. https://doi.org/10.1377/hlthaff.2022.00241

Tamo, J. B., Nnamdi, M. C., Lesbats, L., Shi, W., Zhong, Y., & Wang, M. D. (2023). Uncertainty-aware ensemble learning models for out-of-distribution medical imaging analysis. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 4243–4250). https://doi.org/10.1109/bibm58861.2023.10385418

Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S., Protsyuk, I., et al. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572(7767), 116–119. https://doi.org/10.1038/s41586-019-1390-1

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Tseng, P., Kaplan, R. S., Richman, B. D., Shah, M. A., & Schulman, K. A. (2018). Administrative costs associated with physician billing and insurance-related activities at an academic health care system. JAMA, 319(7), 691. https://doi.org/10.1001/jama.2017.19148

U.S. Congress. (2009). Health Information Technology for Economic and Clinical Health (HITECH) Act.

U.S. Department of Health and Human Services. (2013). HIPAA administrative simplification regulations: 45 CFR Parts 160, 162, and 164. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2024). Explore the new Healthy People 2030 leading health indicators infographics. https://odphp.health.gov/news/202412/explore-new-healthy-people-2030-leading-health-indicators-infographics

U.S. Department of Health and Human Services, Office of Inspector General. (2024). Health care fraud and abuse control program report, fiscal year 2023. https://oig.hhs.gov/reports

U.S. Department of Health and Human Services, Office of Inspector General. (2025). Medicaid program integrity and improper payments.

U.S. Government Accountability Office. (2024, April 16). Medicare and Medicaid: Additional actions needed to enhance program integrity and save billions (GAO-24-107487). https://www.gao.gov/products/gao-24-107487

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. Medical Decision Making, 26(6), 565–574. https://doi.org/10.1177/0272989X06295361

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., et al. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/m18-1376

Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., Phillips, M., Konye, J., Penoza, C., Ghous, M., & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Internal Medicine. https://doi.org/10.1001/jamainternmed.2021.2626

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Albu, E., Arshi, B., Bellou, V., Bonten, M. M. J., et al. (2020). Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. BMJ, m1328. https://doi.org/10.1136/bmj.m1328

Zhong, Y., Shi, W., Tamo, J. B., Nnamdi, M. C., Yuan, Y., & Wang, M. D. (2025). Multi-agent LLM reasoning for clinical procedure sequencing from high-granularity EHR data. In Proceedings of the 16th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 1–12). https://doi.org/10.1145/3765612.3767238

Zhong, Y., Xu, J., Marteau, B., Tan, S. Q. Y., & Wang, M. D. (2025). Anatomically constrained near-eye gaze tracking and movement classification using transformer. In IEEE International Conference on Healthcare Informatics (ICHI) (pp. 280–287). https://doi.org/10.1109/ichi64645.2025.00040