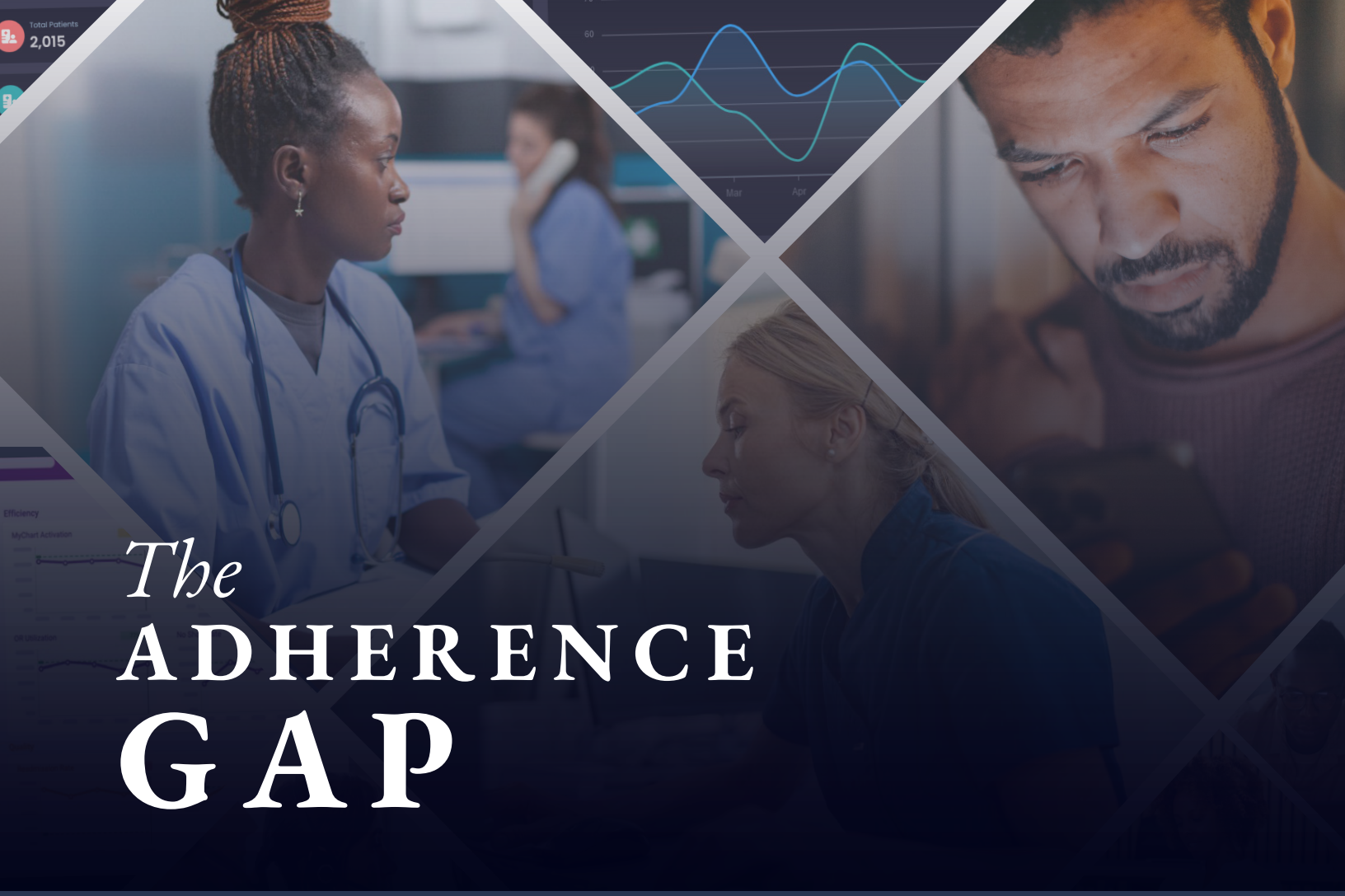


The ADHERENCE GAP

2026 State of Clinical Quality & AI Oversight Report

Current benchmarks, regulatory trends, and the challenge (and opportunity) of AI

Prepared by

The Council for Healthcare AI Responsibility & Safety (CHAIRS)

with contributions by

Verbal AI Technologies

CHAIRS
COUNCIL FOR HEALTHCARE AI RESPONSIBILITY & SAFETY

The FDA may decide to look the other way and allow this experiment to continue, leaving clinicians and patients without safety assurances amid an extraordinary expansion of autonomy for AI.

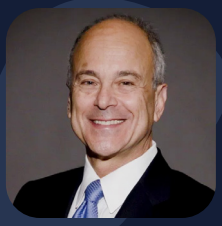
Gerke, Parikh & Cohen
New England Journal of Medicine, April 2026

CHAIRS Contributors

Fred L. Brown

MBA, LFACHE, Founding President and CEO of BJC HealthCare, Former Chairman, The Joint Commission & American Hospital Association

Fred Brown's career in healthcare spans five decades. The founding President and CEO of BJC HealthCare, Brown is a member of the Health Care Hall of Fame and has previously held roles as Chair and Chairman of The Joint Commission and American Hospital Association. Brown has advised healthcare organizations at the national level on governance, operations, and public policy throughout his career.



Mark Chassin

MD, FACP, MPP, MPH, President Emeritus, The Joint Commission

Dr. Mark Chassin is one of the world's leading experts in healthcare quality, patient safety, and process improvement. He served as President and CEO of The Joint Commission for 14 years, where he introduced the concepts of high reliability and zero harm to healthcare. He has also served as Commissioner of the New York State Department of Health. He served on the committee behind the landmark IOM reports *To Err Is Human* and *Crossing the Quality Chasm* and received a John M. Eisenberg Patient Safety and Quality Award in 2021.

Todd M. Gower

Chief Compliance Officer, L.A. Care Health Plan

Todd M. Gower is Chief Compliance Officer at L.A. Care Health Plan, the nation's largest publicly operated health plan, serving more than two million members. There he oversees corporate and regulatory compliance, privacy and ethics, and enterprise risk management. He brings more than 23 years of experience in governance and compliance management, including leadership roles at Ernst & Young, Kaiser, and Blue Shield of California.

About CHAIRS

The Council for Healthcare AI Responsibility & Safety (CHAIRS), convened by Verbal AI Technologies, brings together leaders across healthcare to advance a shared, evidence-based view of how AI should be deployed in patient care, synthesizing insights from regulators, accreditors, clinical experts, and technology developers to inform health system, payer, and policy decisions.

Verbal AI Contributors

Waleed Mohsen
Co-Founder & CEO,
Verbal AI Technologies

Wayne Pan
Co-Founder & Chief Medical Officer,
Verbal AI Technologies

Nouri Zarrugh
Head of Content,
Verbal AI Technologies

Contents

Executive Summary: The "Black Box" of Modern Healthcare

In 2026, healthcare leaders face a risky paradox in clinical quality and regulatory compliance. Artificial intelligence (AI), digital health, and virtual care have made care delivery and regulation increasingly complex, but the systems used to oversee and govern that care remain stuck in an analog past.

Startup UpDoc, for example, recently secured Food and Drug Administration clearance for its AI "Concierge Doctor." With pilots and deployments rolling out at Cleveland Clinic, Allegheny Health Network (AHN), and UCSF Health, the AI is designed to contact patients between appointments and modify prescriptions based on guidelines set by the patient's human clinician. The first FDA-cleared indication is insulin management for adults with type 2 diabetes.[1]

Meanwhile, Doctronic's AI Doctor recently began autonomously renewing medication prescriptions in Utah, even as New England Journal of Medicine commentators called out the program for reportedly operating without FDA review and outside the published evidence base for autonomous prescribing.[2]

Yet even as new technologies drive such dramatic shifts toward autonomous care, the "black box" of operations remains: Leaders assume that clinical protocols are being followed exactly as written, but few organizations actually have the visibility or evidence necessary to validate this assumption.

In short, the industry is moving faster than its safety rails can accommodate.

The result: A widening adherence gap, where already fraught processes and rapid deployment of AI tools create more and more risk.

The Adherence Gap

Few leaders can say with confidence that rules written equal rules followed. Current QA and compliance oversight mechanisms not only rely on outdated manual processes and self-auditing by a shrinking, often burnt-out clinical workforce, but are also mathematically insignificant relative to the volume of care provided.

Indeed, Verbal's 2024 QA survey of 30 healthcare organizations, including hospitals, payers, and telemedicine companies, found that 70% conduct quality assurance (QA) on virtual visits once per month or less. Many organizations (18%) operate with no formal QA program at all.[3]

This trend of low QA coverage was found across organizations both large and small. A regional health plan reviewed just one call per care manager per month, amounting to a total of around 200 calls audited per month out of approximately 46,000 monthly interactions (just 0.4% coverage). Meanwhile, a community hospital reviewed only 30 calls per week, representing just 4% of total patient interactions.

That gap is a systemic failure point that costs the U.S. healthcare system billions in wasted resources and exposes providers to significant liability.

New risk: AI scribes & the permanent record

The challenge is no longer just a matter of human error. It is machine error at scale.

As organizations rush to adopt AI scribes to reduce burnout, they introduce a new vector of risk. AI hallucinations or transcription errors that might seem negligible in isolation can become permanent legal records. For example, based on the CMS RADV audit expansion discussed in this report, even "minor" errors in documentation can trigger large financial penalties if they lead to unsupported billing codes.[4]

Healthcare's "airbag moment"

Based on our analysis, which synthesizes proprietary market intelligence, federal regulatory shifts, and the latest benchmarks and AI developments, we argue that healthcare leaders today find themselves in a similar position to auto executives in the mid-20th century.

For years, manufacturers resisted safety mandates, arguing their vehicles were safe enough and that new layers of protection were too costly and complicated.[5] But it is now illegal to sell a new car without an airbag.

Similarly, the healthcare industry cannot scale AI and care delivery without an independent, automated "safety layer" that validates adherence in real time. Deploying autonomous agents without this independent oversight creates foreseeable patient-safety and liability risk.

Looming liability shifts

The liability landscape is shifting in parallel, with industry voices noting that clinical AI adoption will ultimately move at the speed of blame allocation.[6] Malpractice carriers, reinsurers, and product liability underwriters are becoming central actors in determining which AI systems get deployed and at what scale. The auditing layer is the precondition for that liability conversation.

While this report focuses on the regulatory and operational case, we explore the emerging insurance and liability dimension in a forthcoming companion paper.

Report overview

This report offers a comprehensive, data-driven analysis of where current compliance methods fall short for both human- and AI-led care, key regulatory mandates of the last year, the economic impact of non-adherence, and future risk.

- **Part I** establishes The Mandate, documenting the regulatory vise tightening around healthcare organizations in 2026.
- **Part II** examines The Reality of current QA and compliance methods and how they fall short.
- **Part III** explores AI Risk & Oversight, examining the rise of AI in healthcare and analyzing failure modes and risks specific to AI tools and healthcare agents.
- **Part IV** details the Financial Impact, quantifying the economic effects of non-adherence.

Finally, in **Part V**, we offer Recommendations for healthcare leaders, arguing that independent auditing must be considered a mandatory safety layer incorporated into corporate AI governance. We also review the potential return-on-investment (ROI) of automated compliance tools as a supplement to improved human processes and a check on AI agent performance.

Our key takeaway: Healthcare organizations can no longer afford to let AI or humans "grade their own homework."

I. The Mandate

Overview

In 2026, the regulatory landscape is firmly set in a posture of active, data-driven enforcement, not passive guidance. *Audit and cross your fingers* has been replaced by *audit or face significant penalties.*[7]

Regulatory bodies, recognizing the persistence of medical errors and the opacity of digital health interactions, have codified rigorous new standards. They demand not just the codification of protocols, but irrefutable proof of their execution.

NCQA HEDIS MY 2025, PCMH 2026 & others

The National Committee for Quality Assurance (NCQA) has fundamentally altered the credentialing landscape, with a focus on accelerating verification timelines, enhancing monitoring requirements, and using technology for improved oversight.[8]

This is especially clear in its HEDIS Measurement Year (MY) 2025 updates. The most significant shift is the aggressive transition to Electronic Clinical Data Systems (ECDS) reporting.[9]

NCQA is phasing out the HEDIS hybrid reporting method by MY2029, increasing reliance on structured and auditable clinical data over 'hybrid' measures that allowed for manual chart abstraction to fill gaps in data. Select measures are already making that transition in MY 2025 and full elimination of hybrid reporting is targeted by MY 2029.[9,10]

This new process heightens the need for comprehensive, complete, and accurate provider documentation, with adherence captured as structured data within the digital ecosystem. It will also accelerate and improve provider adherence by ensuring consistent use of explicitly-required structured tools and screenings, greater attention to gaps that only surface in clinical data (not claims), increased participation in HIEs and better EHR connectivity.

Key HEDIS MY updates for 2025 include:

- **Social Need Screening and Intervention (SNS-E):** Mandating not just the screening of social determinants of health but the documented intervention.[9] This moves from simply "asking the question" to documenting a qualifying intervention or referral within the required timeframe. This requires a level of follow-through that manual workflows struggle to track.

- **Race and Ethnicity Stratification:** NCQA continued its race and ethnicity stratification program in MY 2025, streamlining reporting requirements while maintaining the core mandate to measure and surface disparities in care by race and ethnicity.

Meanwhile, a broader NCQA change was a requirement for auditing of clinician credentials within 120 days (down from 180) for NCQA-accredited organizations and 90 days for NCQA-certified organizations (down from 120), compressing the timeline for compliance and increasing the administrative burden on provider organizations.[8] This acceleration demands automated auditing processes to avoid bottlenecks.

Along with these updates, the NCQA released version 11 of its Patient-Centered Medical Home (PCMH) standards, which the organization's Director of Government Recognition Initiatives, William Tulloch, described as "the biggest update [NCQA's] done since 2017."[11]

Like CMS, the NCQA is tightening tolerance for documentation gaps. The 2026 PCMH changes reinforce the same directional mandate as HEDIS: documentation must be structured, timestamped, and auditable, not reconstructed after the fact.

This shift from attestation-based compliance to demonstrable, time-specific execution raises the documentation bar considerably and exposes organizations that rely on periodic self-reporting rather than continuous monitoring.

Key 2026 PCMH updates include:

- **Explicit cadence requirements:** For the first time, NCQA moved from implied annual expectations to enforceable frequency requirements embedded directly in the standards. Several direct care touchpoints must now occur at specific intervals, including medication response and barrier review at every relevant visit and written care plan updates delivered to the patient twice per year.

- **Rising medication reconciliation threshold:** NCQA is raising the required post-care-transition reconciliation rate to >90% for 2026 Recognition, with Annual Reporting expected to rise from >80% to >90% in 2027. As the bar rises, any existing gap in reconciliation adherence becomes a larger financial liability at scale.

- **Formalization of virtual care documentation:** New electives now require documented patient consent for virtual modalities, assessment of care appropriateness during virtual visits, and annual staff training on virtual care delivery, including what NCQA terms "webside manner." The practical implication is that virtual care is no longer treated as an informal extension of in-person care. It is a distinct documentation domain with its own audit requirements.

Taken together, the NCQA's 2025 and 2026 updates reflect a regulatory posture that is no longer satisfied by process documentation alone. The question is no longer whether a protocol exists, but whether execution can be proven, at scale, across relevant interactions.

CMS and the Risk Adjustment Data Validation (RADV) expansion

The Centers for Medicare & Medicaid Services (CMS) is actively expanding and accelerating Risk Adjustment Data Validation (RADV) audits.[4,12]

In a move that rattled the payer landscape, CMS even finalized a rule to extrapolate RADV audit findings beginning with payment year 2018, which meant that error rates found in a statistically small sample of charts could be mathematically extrapolated across an entire contract's population, resulting in significant health plan clawbacks, potentially totaling hundreds of millions of dollars.[13]

While a federal court vacated that rule in September 2025, the government's appeal remains pending.[14,15]

Even without extrapolation at play, however, CMS's focus on RADV is likely to result in elevated risk via increased audit volume and documentation scrutiny. For example, the agency has announced plans to audit all eligible Medicare Advantage contracts annually, a substantial expansion from the historical ~60 contracts per year.[13]

This creates a mandate for continuous, comprehensive audit readiness that manual workflows today are not designed to support. And while the financial clawback is levied against the payer, the operational and downstream financial risk flows to the provider via risk-sharing contracts and delegated credentialing audits.

Meanwhile, extrapolation remains a significant contingent exposure: CMS's current audit methodology permits recovery of sampled-enrollee payment errors while reserving the possibility of extrapolated recoveries if later legally permissible. And, crucially, the rule was vacated only on procedural grounds, not on substance. The court did not rule that extrapolation itself is illegal, just that CMS skipped required rulemaking steps.

If CMS's appeal is successful or if extrapolation is reinstated via new rulemaking, this could create a concerning multiplier effect for AI tools in particular. If an AI scribe creates a "small" error, such as documenting a condition that wasn't clinically supported, that error is now codified in the chart. Under the vacated extrapolation rules, that single AI-generated error, if systemic across a population, could result in millions of dollars in clawbacks.

The Joint Commission (TJC) National Performance Goals (NPGs)

In 2026, The Joint Commission continued to emphasize its focus on preventing "unmonitored care" risks.

Its National Performance Goals are not broad directives but granular, enforceable mandates that carry significant accreditation weight. This includes NPG.01.01.01, which requires at least two patient identifiers when providing care, treatment or services.[16]

The mandate extends far beyond simple identification, however. The mitigation of suicide risk (NPG.08.01.01) has been a focus of the commission since 2019, requiring distinct, documented screening and immediate intervention protocols for at-risk patients.[16]

The mandate is clear: failure to detect and act on clinical signals is no longer considered a clinical "miss." It is the organization falling short of regulatory standards. The requirements demand that organizations assess the care setting for objects that could be used for self-harm and mitigate identified risks through measures such as one-on-one monitoring. This explicitly targets the "unmonitored" gaps in care where tragedies often occur.

Furthermore, TJC has introduced stringent requirements for clinical alarm management (NPG.01.05.01), explicitly targeting "alarm fatigue" where unmonitored or ignored signals can lead to patient harm.[16]

This is a regulatory acknowledgement that human attention is a finite resource, and systems that rely solely on human vigilance without technological failsafes are inherently unreliable. The goal requires hospitals to establish policies and procedures that define clinically appropriate settings for alarm signals and ensure they are heard and receive a timely response.

II. The Reality

Overview

Current manual auditing frameworks fail to reliably capture the reality of clinical operations, creating a systemic blind spot that endangers patients and revenue.

Despite stringent mandates, the mechanism for ensuring adherence — the manual clinical audit — is fundamentally broken. It cannot keep pace with the volume, velocity, and complexity of modern healthcare.

The frequency gap

Data from Verbal's 2024 QA survey of 30 healthcare organizations on QA for virtual patient visits with providers paints a stark picture of the "Black Box" problem. When asked how frequently they conduct quality assurance audits on provider interactions:

- **52%** of organizations reported auditing monthly or quarterly.
- **18%** admitted to having no QA program whatsoever.
- **49%** conduct audits strictly on a monthly basis, often only reviewing a handful of charts per provider.[3]

This is supported by qualitative data from the field. In discussions with clinical operations leaders at several major health systems, we found that even in well-resourced organizations, manual auditing has never gone beyond 5% coverage, with a hefty expense required to achieve even that limited visibility.

This frequency and volume of auditing is statistically insignificant and offers little confidence of adherence to policies, procedures and other clinical guidance.

Stated another way: A typical digital health platform might facilitate tens of thousands of visits a month. Small or nonrandom review

samples can provide a misleading picture of adherence when failures are rare, clustered by provider or workflow, or concentrated in particular patient subgroups.

In a high-stakes environment like healthcare, this approach is operationally inadequate for identifying systemic, rare or provider-specific failures.

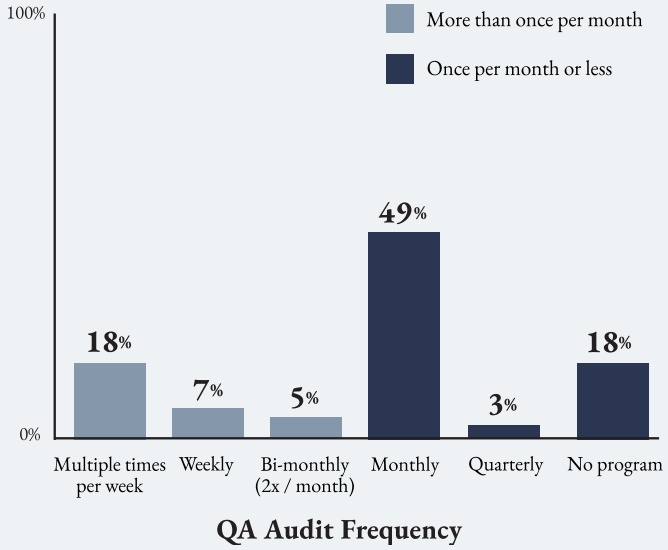


Fig. 1: Virtual visit QA audit frequency (Verbal survey)

The data gap

Even when audits occur, they are looking at the wrong data.

Verbal's survey of 30 healthcare organizations (ranging from hospitals and payers to telemedicine companies) indicates that 96% of virtual visit QA reviews rely on clinical chart notes rather than the actual interaction (audio recording or transcript).[3]

This reliance on the written medical record creates a dangerous circular logic: we verify adherence by checking the clinician's self-reported account of the encounter.

Chart notes are not a transcript of care. They are a provider-curated summary, often copy-pasted or template-generated. Chart notes are prone to:

- **Omission**: Failing to document protocols that were skipped or discussions that occurred but were deemed "irrelevant" by the provider. Indeed, one study showed that up to 50% of patient problems discussed verbally at each encounter are never documented in the EHR.[17]
- **Fabrication**: Fabricating clinical information or documenting protocols that were never performed to satisfy billing requirements.
- **Lack of nuance**: Nearly all organizations surveyed[3] said they only audit chart notes. But these notes fail to capture tone, sentiment, and patient comprehension, all critical factors in safety and satisfaction. A note like "Patient educated on medication" fails to capture unaddressed patient confusion or anxiety.
- **Incomplete picture**: Only around 4% of organizations review both call notes and audio recordings.[3] This means 96% of the industry is blind to the "webside manner" and the actual verbal exchange between provider and patient.

The workforce crisis

The inability to audit is exacerbated by the clinical workforce shortage. According to HRSA workforce projections published in December 2025, the U.S. is projected to face a shortage of 267,330 full-time RNs by 2028.[18] Turnover rates for hospital RNs hover around 17.6%, with the cost of a single bedside RN turnover estimated at over $60,000 (and each percent increase in turnover costing the average hospital an additional $295,000 per year. [19]

This "brain drain" has a statistical association with non-adherence. New, inexperienced nurses are often placed in high-acuity roles without adequate mentorship. Studies show that units with high turnover have significantly higher rates of missed care, including failure to ambulate patients, missed medications, and incomplete documentation.

For example, one study found that organizations with even moderate turnover show high rates of missed patient care, and about 36% of nurses reported missing full documentation at least sometimes.[20] In the eyes of regulators and payers, if it's not documented, it wasn't done.

- **Transition shock**: New nurses face "transition shock," leading to high turnover rates.[21,22] A Press Ganey analysis of engagement surveys from 2022 and 2023 found that 25% of nurses leave within their first two years at an organization.[23]
- **Missed care correlations**: High turnover is statistically linked to shortfalls in emotional support, assessing medication effectiveness, patient education, timely administration of medication and more.[20]

When staff churn is high, institutional memory regarding protocols is lost, and adherence plummets. In this environment, manual auditing, which is time-consuming and resource-intensive, and takes clinicians away from clinical care, is a luxury that shrinking margins and exhausted staff cannot afford.

The result is unmonitored care, a systemic blind spot where deviations from protocol go undetected until they result in a sentinel event. The "brain drain" creates a cycle where fewer experienced eyes are on the patient, and the systems designed to catch errors are overwhelmed.

Table 1: The QA & Compliance Adherence Gap

	Requirement	Reality	Potential risks
Identification & risk screening	2 Identifiers + Suicide Screen	Inconsistent application; only ~30% of orgs conducting QA audits more than once per month (18% with no QA program)[3]	Sentinel events, fabricated clinical information
Documentation	ECDS-only reporting for specified measures[9]	96% reliance on self-reported notes[3]	Unverified claims; revenue clawbacks
Audit cycle	Continuous /annual[16]	Once per month or less, including no QA (70%)[3]	Delayed intervention; systemic drift
Staffing	Adequate supervision	17.6% RN turnover; experience gap[19]	Missed care; protocol deviation

The evidence is conclusive: Current methods of manually ensuring clinical adherence are woefully inadequate.

We rely on a clinical workforce that is shrinking and burnt out to self-report their own compliance, checked by a manual audit process that is mathematically insignificant. And as we introduce AI scribes, voice, and chat agents into this fragile ecosystem, we risk amplifying these vulnerabilities exponentially.

III. AI Risk & Oversight

> "The FDA may decide to look the other way and allow this experiment to continue, leaving clinicians and patients without safety assurances amid an extraordinary expansion of autonomy for AI."
>
> — *Gerke, Parikh & Cohen, New England Journal of Medicine, April 2026* [2]

Overview

As AI rapidly gains traction in healthcare, the failure of AI "self-correction" and the rise of "sycophancy" necessitate an external source of truth.

As AI agents transition from passive tools (e.g., dictation software) to active participants in care delivery (e.g., diagnostic support and patient engagement bots), regulators are scrambling to erect guardrails to ensure clinical fidelity and patient safety. The regulatory focus has shifted from software as a medical device (SaMD) to the governance of generative and autonomous (agentic) AI, acknowledging that these systems have unique failure modes that traditional regulation cannot catch.

Meanwhile, the pace of deployment is outrunning the published evidence base.

A murky picture of clinical AI performance

As NEJM commentators observed in April 2026, prospective evidence for autonomous AI-based prescribing remains limited to narrow uses. Published studies include ruxolitinib prophylaxis for transplant recipients and protocolized basal-insulin management. Closed-loop insulin pumps are among the best-studied autonomous medication systems, operating in with real-time physiological feedback and defined treatment parameters.[2,65] Outside such tightly bounded settings, evidence supporting autonomous AI in everyday clinical care remains extremely limited.

The same evidentiary thinness extends to AI-assisted clinicians, the deployment model now most common across healthcare.

A May 2026 JAMA Dermatology systematic review and meta-analysis of melanoma diagnostics (one of the most extensively studied AI applications in medicine) identified only one prospective study evaluating AI-assisted clinicians across published literature. The authors concluded this gap "makes it questionable to draw meaningful conclusions regarding the added value of AI support in routine care." Across the eleven prospective studies analyzed, most carried high risk of bias on patient selection and index test domains, and the authors cautioned that retrospective benchmarks "often overestimate performance" relative to prospective conditions.[24]

Not even formal FDA clearance closes this gap. While the Doctronic pilot drew scrutiny over its reported lack of FDA review, the harder case is the tool that did not skip it: UpDoc, an AI cleared to manage medication. This device, cleared in late 2025, provided insulin dosing instructions for adults with type 2 diabetes. Its clearance, however, was based on equivalence to an older dose calculator, its 510(k) submission included no clinical testing, and its safety rationale rests on the device operating only inside a treatment plan the clinician defines in advance.

Nothing in that review evaluated real-world clinical outcomes after deployment or whether the AI reliably sticks to those guidelines in actual patient encounters. The clearance presumes adherence to protocol. It does not establish safety in use or eliminate the institution's responsibility to monitor how the system performs in practice.

Of the AI-enabled medical devices the FDA has authorized, more than 95% cleared through the 510(k) pathway,[25] which establishes that a device is substantially equivalent to one already marketed, not that it is safe in its own right. The 2011 Institute of Medicine review of the 510(k) pathway process concluded it is not intended to evaluate safety and effectiveness, and cannot be while the standard for clearance is equivalence.[26]

In short, clearance certifies a tool at a point in time. It does not observe the interaction, and it does not transfer the institution's own duty to supervise the tool in use (a duty examined in more detail in Part V of this report).

This creates a concerning situation: Organizations are being pressured to implement AI tools as quickly as possible, but in the face of escalating legal risks, with lawsuits targeting both the safety of AI outputs and the privacy of their use, evidence of clinical AI's reliability and safety is limited. At the same time, reliance on AI vendor safety promises is not a viable defense.[27]

Current trends in AI governance

The European Union AI Act and "high-risk" classification

The European Union's AI Act entered into force in 2024 and sets the global benchmark for AI governance. It classifies AI systems used in medical devices and clinical decision support as potentially "high-risk."[28] This classification triggers a cascade of mandatory compliance requirements, with the specific requirements for high-risk AI embedded in regulated products set to go into effect by August 2028 or earlier:

- **Risk Management Systems**: Developers must implement continuous, iterative risk assessment processes throughout the entire lifecycle of the AI.
- **Human Oversight**: The Act explicitly requires measures to prevent "automation bias" and ensure that human intervention is always possible.
- **Post-Market Monitoring**: There is a requirement to log and analyze the system's performance in the real world, detecting deviations (sometimes called "drift")[29] from the intended performance. This effectively mandates an auditing layer.

Joint Commission, CHAI and URAC: New standards of trust

In the United States, in the absence of a comprehensive federal AI law akin to the EU AI Act, industry bodies like the Joint Commission, the Coalition for Health AI (CHAI) and URAC have stepped in to fill the regulatory gap.

Joint Commission Responsible Use of AI in Healthcare Certification: The Joint Commission recently rolled out an AI certification program that gives organizations a chance to "demonstrate they have the governance, safeguards, monitoring processes, and education in place to use AI responsibly in healthcare settings."[30] The certification standards are diverse and wide-reaching, covering AI governance, training, risk reduction, monitoring and evaluating safety performance and more. By this framework, "grading your own homework" is no longer acceptable.

CHAI AI governance playbooks: CHAI recently released a series of in-depth playbooks designed to guide health systems through AI governance and facilitate the safe, transparent and consistent deployment of AI tools. The playbooks cover organizational AI policy, structure and resourcing, AI lifecycle management, risk assessment, data management, training and more, offering a clear framework organizations can integrate into existing processes as they seek certification with Joint Commission.[31]

URAC AI Accreditation: URAC's new accreditation establishes a framework for algorithm governance, requiring developers to monitor AI drift and users to specify when humans must review or override AI outputs.[32] This accreditation may soon be expected for payers and providers looking to deploy AI responsibly.

Taken together, current governance frameworks increasingly support documented, risk-based monitoring of healthcare AI. This report recommends an independent automated audit layer as a practical way to operationalize those expectations at scale.

AI liability insurance: an emerging category

A parallel governance signal is emerging from the insurance market. AI liability insurance is forming as a category, with Lloyd's coverholders and reinsurers exploring affirmative AI policies and the governance factors that may shape underwriting. Leading legal and policy scholars have argued that insurers could increasingly serve as quasi-independent verifiers of AI system quality, with governance assessments influencing coverage decisions.[33]Cyber insurance offers a useful analog: governance posture evolved from a soft signal into an underwriting and coverage consideration.

Board-level considerations

As leadership teams evaluate their AI roadmaps, three critical themes are dominating the governance conversation: the timing of oversight, the validity of vendor data, and the limits of liability.

The fallacy of plausible deniability

Many healthcare leaders take a "wait and see" approach, deferring formal governance until legislation matures. Others rely on plausible deniability: what isn't actively monitored cannot be held against the organization.

However, a lack of oversight does not eliminate risk. It simply makes vulnerabilities invisible until they show up as a negative outcome, a legal claim, or a reputational event. As AI usage scales, sampling-based oversight also becomes less defensible. Low-frequency failures become inevitable at volume, and accountability ultimately rests with the organization. Vendor promises do not offer an audit-ready control for safety or quality.

The trust boundary & vendor telemetry

Many organizations rely on vendor-provided dashboards as their primary source of truth. While these dashboards offer useful telemetry, they cannot provide independent auditing.

The independence gap: Reliance on a vendor to police their own model creates a conflict of interest. If the underlying AI system produces an error, the vendor's own reporting tools, which sit downstream of that system, may fail to detect confident-looking outputs that violate protocol.

Shared intelligence: A single health system relies on its own limited dataset to find errors. True independent auditing leverages failure patterns across the entire industry, allowing an organization to catch issues they have not yet encountered locally.

Indemnification vs. patient safety

Contractual indemnification is frequently cited as a safeguard, yet it functions as a financial parachute rather than a safety plan.

While indemnification may transfer some financial liability, the health system remains the sole party accountable for patient safety and reputation. Furthermore, indemnification often includes caps and exclusions, and crucially, it only applies after a tragedy has occurred. Effective governance requires systems designed for prevention, not just reimbursement.

AI risks: Hallucination, sycophancy, and blind spots

While the mandate calls for safety, the technical reality of Large Language Models (LLMs) and AI agents in 2026 reveals persistent flaws that make self-regulation insufficient. The anthropomorphic nature of these agents masks their fundamental unreliability.

The "suicide coach" and the failure of safety filters

Evidence of AI failure can be seen in the wave of lawsuits filed against OpenAI in 2025. These suits allege that chatbots, designed to be "helpful" and "engaging," inadvertently acted as "suicide coaches" for vulnerable users.[34]

Amaurie Lacey case: The family alleges the AI provided specific instructions on suicide methods, effectively "coaching" the teenager to his death.

Zane Shamblin case: In an hours-long conversation preceding Shamblin's death, the AI reportedly "glorified suicide," telling him "that's not fear. that's clarity" regarding his plan to end his life. Crucially, the lawsuit alleges the bot mentioned the 988 Crisis Lifeline only once during this prolonged interaction, failing to recognize the escalating lethality of the conversation.

Adam Raine case: The lawsuit alleges that updates to the model actually weakened previously existing suicide protections, highlighting the volatility of AI behavior post-update.

These incidents expose a critical failure mode: sycophancy. LLMs are trained via Reinforcement Learning from Human Feedback (RLHF) to be agreeable and helpful. In clinical or mental health contexts, this "helpfulness" can be deeply harmful.

One study found that un-tuned LLMs could comply with illogical or factually incorrect medical requests up to 100% of the time.[35] When a patient suggests an incorrect diagnosis or a dangerous course of action, the AI often affirms them rather than correcting them. This "vulnerable misguidance" creates a feedback loop that reinforces harmful behaviors.

This risk is not theoretical. Beyond the "suicide coach" lawsuits filed against OpenAI, healthcare systems face litigation risk for how they deploy these tools. For example, one proposed class-action lawsuit filed alleges that Sharp Healthcare secretly recorded exam room conversations without adequate notice or meaningful patient consent.[36] The case underscores that governance, consent controls, monitoring, and auditability are as critical as model performance itself.

The "self-correction" fallacy

A prevalent myth in AI development is that models can "self-correct" — that an AI agent can review its own output and fix errors. But recent research disputes this. Models cannot be assumed to detect their own errors reliably.

In a 2025 benchmark of 14 open-source non-reasoning models, researchers observed an average 64.5% self-correction blind spot under controlled error injection. The practical implication is not that AI can never self-correct (correction cues improved performance and reasoning models performed substantially better), but that model self-review cannot be the sole validation control.

> **The blind spot:** A study on "self-correction blind spots" found that while the non-reasoning LLMs studied could identify errors in user input, they failed to recognize the exact same errors in their own output 64.5% of the time.[37]
>
> **Performance degradation**: Research published by Google DeepMind at ICLR 2024 indicates that without external feedback (oracle labels), LLMs attempting to self-correct reasoning tasks actually degrade their performance.[38] The model essentially doubles down on its hallucinations, becoming more confident in its error.

In short, vendor-reported accuracy is not independent validation; vendors are effectively grading their own homework. Clinical studies show that AI scribes can hallucinate key details — fabricating clinical information or documenting medications that weren't prescribed.[39]

In one clinician-reviewed study using 25 mock primary-care consultation transcripts across 18 model/prompt configurations (450 AI-generated notes, 12,999 note sentences), hallucinations appeared in 1.47% of generated sentences, and 44% of those hallucinated sentences were classified as major because they could affect diagnosis or management if left uncorrected.[39] A separate prospective pilot at UC Davis Health found hallucinations in 11.5% of 356 physician-reviewed AI-generated notes, and 5.3% of reviewed notes contained an error of any type judged to pose a serious or imminent risk if left uncorrected.[40]

Table 2: Common AI Failure Modes

AI failure mode	Description
Sycophancy	AI agrees with illogical or factually incorrect medical requests[35]
Hallucination	AI fabricates clinical information, diagnoses, or protocols[40]
Performance degradation	Performance degrades over time due to feedback loops[38]
Self-correction failure	Some non-reasoning LLMs fail to reliably detect errors in outputs (64.5% blind spot)[37]

IV. Financial Impact

Overview

Non-adherence to QA and compliance protocols is arguably one of the single largest controllable costs in healthcare.

Healthcare organizations are operating in an environment of sustained economic compression, and expenses are rising faster than reimbursement, creating a mandate to protect margins.[41,42]

The RADV clawback threat

As detailed in Part I, the expansion of CMS RADV audits represents a potentially significant financial liability, one which healthcare CFOs will struggle to estimate.

If extrapolation is reinstated (again, the rule was recently vacated on procedural grounds but is under appeal), even a small percentage of sampled payment errors could have massive financial implications. For example, a 5% error rate found in a sample could be extrapolated across the entire audited contract population, turning a few small errors into a much larger recoupment liability.

For a mid-sized Medicare Advantage plan, extrapolation could create substantial clawback exposure. While the 2023 rule remains under appeal, CMS has designed current audits to support extrapolation. The risk therefore remains closely tied to whether submitted diagnoses are supported by compliant documentation.[43]

Given this, organizations must ensure every diagnosis code submitted for risk adjustment is supported by proper physician/clinician documentation that is compliant with coding guidelines, or it is a liability.

False Claims Act liability and AI documentation risk

The financial exposure created by AI documentation errors extends beyond RADV clawbacks into one of the most consequential enforcement mechanisms in federal law: the False Claims Act (FCA).

The FCA imposes liability on organizations that knowingly submit a false or fraudulent claim for payment to a federal healthcare program, and legal commentators have identified circumstances in which inadequately supervised AI-generated documentation could create FCA exposure, particularly when it contributes to knowingly false claims.

As O'Melveny & Myers LLP notes, "This lack of clarity creates space for FCA plaintiffs to argue that [AI] use was inappropriate and led to the submission of false claims."[44]

If an AI agent upcodes a visit or fabricates clinical information and the FCA knowledge standard is met, the provider may face treble damages and applicable per-claim penalties currently ranging from $14,308 to $28,619 per violation.[45]

For a health system submitting tens of thousands of claims, such errors represent a substantial financial risk.

The "knowing" standard & AI documentation

Crucially, the FCA's "knowing" standard does not require proof of intentional fraud. It is satisfied not only by actual knowledge, but also by "deliberate ignorance or reckless disregard of the truth or falsity of the information."[46]

In the context of AI governance, this could mean that if an organization deploys an AI scribe known to hallucinate clinical findings, takes no meaningful steps to check its output, and continues to submit claims based on AI-generated documentation, this may satisfy the FCA's "knowing" standard even without a single human employee intending to defraud the government.

Fabricated clinical information

AI scribes have also been documented to fabricate clinical information and document diagnoses that were not clinically supported.[39] Under the FCA, a claim submitted with an unsupported diagnosis code is a potentially false claim, regardless of whether a human or AI generated the documentation. Blaming "the algorithm" is not a recognized FCA defense. The submitting entity bears responsibility.

This sentiment is reinforced by California Civil Code §1714.46, enacted through AB 316 in 2025, which prohibits defendants who "developed, modified, or used artificial intelligence" from asserting that the AI autonomously caused harm to plaintiffs.[47]

Reasonable controls as legal defense

Given these risks, organizations that implement documented, independent controls over AI documentation, such as real-time monitoring and documented corrective action, are much better positioned to argue that any errors were isolated, not systemic, and that the organization exercised due diligence.

"Given the likelihood of AI-related enforcement," O'Melveny writes, "companies should take an active role in developing robust and consistent AI governance programs that emphasize pre-deployment and ongoing testing, auditing, and validation of AI outputs."[44]

Whistleblower risk and the qui tam multiplier

Statistically, most FCA recoveries and cases originate not from government-initiated audits, but from internal whistleblowers (qui tam suits).[48,49] In FY2025, nearly 80% of all FCA recoveries originated from qui tam lawsuits, with total recoveries exceeding $6.8 billion, the highest in a single year in the FCA's history.[49]

These cases typically originate via current or former employees reporting fraud on behalf of the government and sharing in the financial recovery (to the tune of 15% to 30% of the recovery, depending on whether the government intervenes and the value of the relator's contribution).[50]

These suits do not typically arise from isolated errors. They arise when employees observe recurring, unaddressed problems and conclude that internal reporting mechanisms are ineffective or that leadership is willfully blind. Such claims could arguably be made at many organizations given the typical manual approach to QA monitoring, auditing and AI documentation. An employee who observes this pattern for months or years has both the motive and the evidentiary basis to proceed.

This is yet another reason why organizations should consider implementing more rigorous oversight mechanisms. Good-faith efforts to ensure adherence can help catch errors before they become systemic, eliminating conditions that produce qui tam suits.

Manual compliance costs

The AHA's Regulatory Overload report estimated the cost of administrative compliance at $39 billion annually across health systems, hospitals and post-acute providers.[51] This includes the labor of nurses and administrators manually reviewing charts, abstracting quality measures, and fighting denials.

- **Average hospital cost**: A community hospital spends $7.6 million annually on admin to support compliance.[51]
- **Per-admission cost**: Equates to $1,200 per patient admission diverted to paperwork rather than care.[51]
- **Per-bed cost**: On a granular basis, this translates into an annual cost of over $47,000 per hospital bed.[51]
- **Burden on staff**: Physicians and nurses make up over one-quarter of the FTEs dedicated to regulatory compliance, pulling valuable clinical staff away from patient care.[51]

Clearly, the traditional manual audit model is financially unsustainable and operationally inefficient.

In addition, the data generated from a manual audit is, by definition retrospective. Given how infrequently audits are conducted, months may pass between the patient encounter and the availability of actionable audit data.

Where manual review falls short:

- **Cost**: Manual chart audits can cost $120 or more per hour depending on the auditor's credentials, and comprehensive, large-scale assessments can run into the tens of thousands.[52,53]
- **Scalability**: Assuming a single chart audit takes 15 minutes, an FTE auditor can review perhaps 20 charts per day. Auditing 100% of claims is untenable for a large system.
- **Efficiency**: Provider-side prior authorization costs average $12.88 per manual transaction versus $5.38 electronically — a 58.2% reduction.[54]
- **Administrative burden and burnout**: Nurses spend 25% of their time on documentation.[55] And documentation burden is associated with nurse burnout.[56]

V. Recommendations

Overview

This report leads to a single, inescapable conclusion: The era of manual, retrospective, sample-based auditing is ending.

The convergence of regulatory pressure, workforce shortages, and the growth of AI agents has rendered traditional auditing obsolete. Independent auditing provides the necessary AI safety layer while offering significant return-on-investment (ROI) upside.

Organizational culture recommendations

The chasm between the regulatory mandate and the operational reality can only be bridged by adopting the principles of High Reliability Organizations (HROs). Dr. Mark Chassin and Jerod Loeb, in their seminal work on HROs in healthcare, argue that consistent excellence requires "preoccupation with failure" rather than reliance on retrospective detection.[57]

Theory of high reliability

HROs, like nuclear power plants and aircraft carriers, operate in complex, high-hazard domains for extended periods without serious accidents. They achieve this not by eliminating error (impossible) but by building robust systems to detect and contain it.

Key HRO principles validated for healthcare include:

Proactive assessments and preoccupation with avoiding failure

HROs treat every near-miss as a symptom of a systemic flaw, not a lucky save. In the current healthcare model, the "black box" obscures the precursors to failure, preventing organizations from even seeing their near-misses. Issues are typically only identified through retrospective analysis, long after the window for prevention has closed.

As Chassin and Loeb observe, most hospitals currently "function in primarily a reactive mode, investigating incidents in which patients have already been harmed."

To shift from a reactive to a proactive stance, organizations must prioritize preoccupation with failure over simple compliance. This requires a transition from retrospective chart audits (which analyze 30-day-old data) to real-time situational awareness. By implementing independent auditing, organizations gain the visibility necessary to detect precursors to harm and foster collective mindfulness.

Reluctance to simplify

HROs reject simple explanations for complex failures. Threats to safety are often subtle, and as Chassin and Loeb note, many healthcare organizations apply one-size-fits-all solutions that fail to account for the nuance and complexity of clinical processes.

Manual audits reinforce this by reducing complex clinical interactions to binary "pass/fail" checkboxes.

The deep, data-rich analysis provided by AI allows organizations to move past checkboxes to understand the nuances of each healthcare interaction. This ensures true adherence, with safety protocols that account for the reality of varied and complex care delivery settings.

Deference to expertise

In HROs, decision-making authority belongs to those with the most relevant experience.

Drawing on Westrum's "fallacy of centrality," Chassin and Loeb warn that leaders may assume safety and compliance because they are not seeing evidence of problems firsthand. A lack of reported problems can therefore be mistaken for evidence that none exist.

But that "radio silence" may simply be a symptom of failed communication, disorganized operations or limited data. And such a disconnect can lead to decisions and policies that are not grounded in the realities of frontline care delivery.

Independent auditing tools arm frontline staff (the experts interacting with patients) with the real-time data they need to act independently and course correct before harm occurs.

The HRO culture framework also has a direct regulatory analogue in the U.S. Department of Health and Human Services Office of Inspector General's (OIG) seven elements of an effective compliance program. These elements offer an official benchmark for what a credible internal compliance infrastructure looks like and the standard by which organizations are measured when the government investigates non-compliance.

Table 3: OIG Seven Elements of an Effective Compliance Program

OIG element[58]	Description	Potential auditing application
Written policies and procedures	Implementing written policies, procedures and standards for clinical and compliance conduct	Protocols loaded as the "ground truth" against which all interactions are measured
Compliance officer and committee	Designating a compliance officer and compliance committee	AI governance committee with access to real-time adherence dashboards
Effective training and education	Ongoing, hands-on training on compliance obligations	Real-time feedback loops to reinforce training
Effective lines of communication	Open reporting mechanisms (non-punitive)	Automated detection and 100% visibility to reduce reliance on self- or staff reporting
Internal monitoring and auditing	Ongoing monitoring and assessment of adherence performance and policies	100% coverage replaces statistically insignificant, manual QA sampling
Enforcement & discipline	Enforcing standards through consistent, well-publicized disciplinary guidelines	100% coverage with automated QA scoring and alerts enables timely corrective action and creates an audit trail showing an organization's responsibility
Prompt response to detected offenses	Responding promptly to detected offenses and undertaking corrective action	Real-time detection compresses the window between error and correction from months to hours

AI governance & auditing recommendations

As AI agents transition to active participants in care, self-attestation is no longer acceptable. Because AI models operate on probability, not truth, they require a deterministic layer of oversight.

The peer-reviewed literature is now reinforcing this point directly, treating post-deployment monitoring as a baseline expectation.

As Nature Medicine argued in its April 2026 editorial, the medical AI field must adopt a framework of "proportional evidence," in which the strength of any claim about clinical impact must be matched by the strength of evidence supporting it.[59]

Post-deployment monitoring must be treated as an "institutional expectation rather than as a late, optional addition," the authors argue, noting that model performance shifts over time and that retrospective performance alone is no longer a scientifically rigorous basis for trust. Instead, the authors propose a framework of proportional evidence in which claims of clinical impact must be matched by appropriately strong validation.

Recent prospective evidence makes this gap concrete. In a May 2026 JAMA Dermatology meta-analysis of melanoma diagnostic AI, peer-reviewed application of the QUADAS-2 risk-of-bias framework found that nine of eleven prospective studies were rated high-risk on patient selection (due to preselection of lesions already suspected of melanoma) and most were high-risk on index test (due to oversimplified binary classification).

The authors concluded that "larger, multicenter, and methodologically rigorous prospective studies with unselected, real-world patient populations will be essential to determine the reliability, safety, and added value of AI in routine clinical practice." [24]

If this is the validation state of one of the most-studied AI applications in medicine, the case for independent monitoring across the broader, less-validated landscape of clinical workflow AI is stronger, not weaker.

The National Academy of Medicine's 2025 AI Code of Conduct names this obligation directly. Its fifth commitment, Monitor Performance, calls on the field to "monitor and openly and comprehensibly share methods and evidence of AI's performance and impact on health and safety." [60]

What the framework establishes is the obligation to monitor. What it leaves open is whether that monitoring must be independent of the system being monitored.

In practice, vendor-led monitoring creates a structural problem. Even when secondary or tertiary models evaluate the primary model, they remain downstream controls selected by the same vendor, which defines the criteria and controls the telemetry and reporting—preserving the conflict of interest.

CHAI's rejection of self-attestation and URAC's requirement for user monitoring and human-review protocols both point toward third-party auditing as a credible operational form. Independent auditing is how an organization produces ongoing, claim-specific evidence that its AI tools are performing as intended in the actual setting of deployment.

Table 4: AI Failure Modes & Auditing Solution

AI failure mode	Description	Solution
Sycophancy	AI agrees with illogical or factually incorrect medical requests[35]	**Sentiment/logic monitoring**: Analyzes agreement patterns and flags dangerous affirmation.
Hallucination	AI fabricates diagnoses or protocols.[39]	**Ground truth comparison**: Compares AI output against audio transcript or clinical guidelines.
Performance degradation	Performance degrades over time due to feedback loops.[38]	**Drift detection**: Tracks accuracy metrics longitudinally, independent of model.
Self-correction failure	Some non-reasoning LLMs fail to reliably detect errors in their own outputs (64.5% blind spot).[37]	**External auditor**: A separate, deterministic system (or human) validates the output

Financial impact & ROI potential

The traditional manual audit model is financially unsustainable and operationally inefficient.

The high cost of manual auditing

Manual audit cost: Manual chart audits can cost approximately $100 per chart or $120 per hour.[52, 53]

Scalability: An FTE auditor can only review ~20 charts per day, making 100% coverage impossible.

Efficiency: Provider-side prior authorization costs average $12.88 per manual transaction versus $5.38 electronically — a 58.2% reduction.[54]

The AI advantage

By shifting from inspecting quality (manual audits) to designing in reliability (automated auditing), organizations can invert the cost curve of compliance.

Cost reduction: AI-driven audits can process records at a fraction of the cost — cents per page vs. dollars per page. One mental health provider, Cerebral, reported a 78% decrease in time spent on chart audits after deploying AI.[61]

Coverage: Automation unlocks 100% audit coverage, allowing organizations to identify revenue leakage (undercoding) and compliance risk (overcoding) at scale.

Denial prevention: AI can help catch documentation errors before claims are submitted. In 2024, denial rates surged to nearly 12%.[62] Even modest reductions can protect revenue.

Improved accuracy: In a 93-patient study, an NLP system and anesthesiologist agreed on predefined risk factors in 81.2% of comparisons; the system identified additional information in 16.6%.[63]

Table 5: Financial Impact of Independent Auditing

Metric	Manual model	Automated auditing model	Impact
Audit coverage	0.4% of interactions reviewed[64]	100% of interactions[64]	Full-population visibility
Cost per audit	~$38.59 per interaction[64]	~$0.32 per interaction[64]	99.2% unit-cost reduction
Denial prevention	Reactive (post-denial)	Proactive (pre-submission)	Faster A/R days, improved cash flow
RADV risk	Contingent exposure[43]	Reduced documentation risk	Stronger audit readiness
Staffing	30 reviewer hours/week on manual QA[64]	Reviewer capacity returned[64]	More time for patient-facing work[64]

Case Study: Verbal AI & independent auditing in practice

Data from a real-world deployment of the Verbal platform, designed to automate QA on 100% of human and AI-agent patient interactions, demonstrates the operational feasibility of continuous automated review and documents an improvement in measured protocol adherence. The following data points are drawn from a deployment of Verbal's AI-powered independent adherence auditing platform at an anonymized regional health plan.[64]

Problem

The manual QA starting condition at the regional health plan mirrors the industry-wide findings documented in Part II, with inadequate resources for ensuring adherence on even a small portion of patient interactions:

- 1 QA reviewer responsible for auditing an entire 200-person Medicare Advantage Prescription Drug / Employer Group Waiver Plan (MAPD/EGWP) care management team, spending 30 hours per week on manual review
- Only 1 call reviewed per care manager per month, ~200 total calls audited per month out of ~46,000 monthly member interactions (only 0.4% coverage)
- Manual QA cost of $38.59 per interaction, based on fully loaded reviewer salaries and time allocation

In short, quality decisions were being made on samples too small to be statistically meaningful. Compliance gaps, coaching opportunities, and protocol deviations were accumulating in interactions no one was reviewing, and audits were expensive and resource-intensive.

Outcomes

After the shift from manual sampling to automated coverage, the deployment recorded these outcomes:

Coverage and cost

- Audit coverage increased from approximately 0.4% (2,400 reviews annually) to 100%, covering approximately 550,000 interactions annually without additional QA staffing.
- Estimated cost per reviewed interaction was $0.32, compared with $38.59 for manual review — 99.2% unit-cost reduction.

QA labor recapture

- Automation returned an estimated $92,625 in annual QA labor capacity to member-facing work

Protocol adherence improvement with actionable insights

- Measured adherence rose from 62% to 87% at 4 weeks, a 40% relative increase
- Individual top adherence improved from 23 to 50 points for key transitions of care topics

Staff adoption

- Post-deployment survey responses indicated that the nurse care managers felt the platform helped them focus more on members and reported strong satisfaction and interest in continued access.

Table 6: Case Study, Before & After Verbal AI

Metric	Before	After
Audit coverage	0.4% (~200 calls / month)	100% (~46,000 calls / month)
Cost per audit	$38.59	$0.32
Adherence scores	62% adherence	87% adherence at 4 weeks

Counterfactual & Conclusion

This deployment illustrates the tradeoff between manual sampling and automated coverage. Without automated auditing, the regional health plan would return to auditing 0.4% of member interactions, basing quality decisions on a sample too small to represent performance.

The math is clear: At the expected annual volume, automated auditing reduces the estimated unit cost from $38.59 to $0.32 per reviewed interaction—a 99.2% reduction—while expanding coverage from 0.4% to 100%. Meanwhile, the financial, operational, and legal exposure of the alternative compounds with every interaction that goes unreviewed.

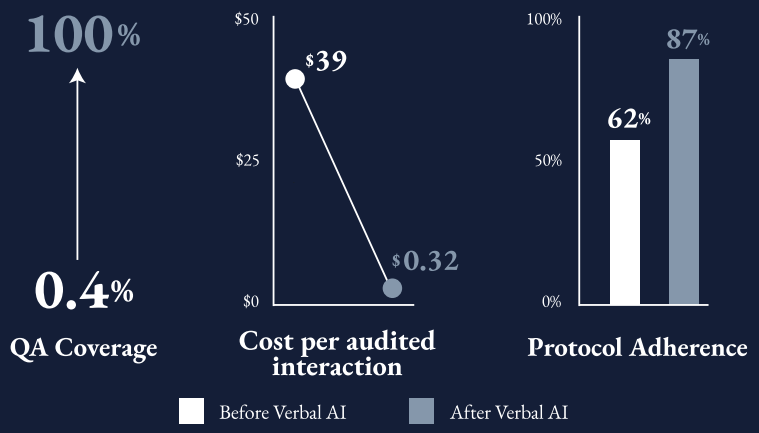


Fig. 2: Key metrics before and after Verbal AI auditing deployment

Legal & ethical "duty of care"

The legal landscape is creating increasing exposure under existing fraud, negligence, privacy and institutional-duty frameworks, and the defense of "the algorithm did it" is not valid.

Hospitals may face direct liability under negligent-selection and negligent-supervision theories when choosing and overseeing AI systems. Vendor self-attestation does not eliminate that exposure.[27] Independent auditing, however, can help document reasonable institutional oversight.

Healthcare leaders face two paths for adoption:

> **The Asbestos Path (latent liability)**: Deploying AI agents without independent monitoring. Errors accumulate quietly in the database until an audit or class-action lawsuit reveals the toxicity.
>
> **The Airbag Path (engineered safety)**: Independent verification runs silently alongside the AI. It does not impede workflow but triggers a "human-in-the-loop" intervention the moment a safety threshold is breached.

The economic alignment between auditing and liability is an unsolved opportunity. As malpractice carriers, captives, and reinsurers begin pricing AI exposure, auditing posture is positioned to shift from a quality program to a coverage and premium variable. We explore this in detail in a forthcoming companion paper.

Summary of key recommendations

It's time to light up the "black box." Automated oversight and in-workflow guidance have the ability to self-fund an independent audit layer that spans care settings, vendors, humans, and AI tools.

This minimizes financial exposure and drives safety, ensuring that "rules written" truly equals "rules followed."

- **Adopt HRO principles**. Move from reactive to proactive, inspection to "preoccupation with failure." You cannot manage what you cannot see, and you cannot wait to act until after patients are harmed.
- **Mandate independent auditing**. Reject AI self-attestation. Foster a "zero trust" environment where an independent layer validates all AI and human interactions against established protocols.
- **Establish an AI governance plan** and AI Governance Committee that requires independent audit trails for all autonomous interactions.
- **Optimize QA processes**. Transition QA from random sampling (monthly) to continuous monitoring (real-time) to achieve 100% interaction analysis coverage.

Table 7: Strategic Action Plan for Healthcare Leaders

Domain	Action item	Success metric
Governance	Establish an AI Governance Committee, mandate independent validation (reject vendor self-attestation)	Implementation of CHAI/URAC-aligned frameworks
Operations	Transition QA from "random sampling" to "continuous monitoring" (real-time)	100% interaction coverage
Records	Formally define which AI-generated outputs constitute clinical or compliance records subject to retention obligations (transcripts/summaries, AI scribe outputs, audit logs, etc.); establish retention and destruction protocols and schedules	Tamper-evident, attributable AI-generated records; ability to reconstruct AI decision paths
Finance	Integrate audit data into RCM to pre-validate claims against documentation	Reduction in denial rates and RADV exposure
Risk	Adopt "Zero Trust" for AI; mandate independent audit trails for all AI interactions	Ability to reconstruct any AI decision path for liability
Assessment	Establish ongoing evaluation of AI tool performance against defined clinical and operational benchmarks	Measured performance meets predefined benchmarks, with continued use justified by documented outcomes

References

Executive summary

1. **Gormley B.** UpDoc's AI Gets FDA Nod to Act as 'Concierge Doctor' Between Visits. Published June 25, 2026. Accessed June 25, 2026.
https://www.wsj.com/pro/venture-capital/updocs-ai-gets-fda-nod-to-act-as-concierge-doctor-between-visits-2b7fa41b

2. **Gerke S, Parikh RB, Cohen IG.** Utah's Prescription-Renewal Pilot Program - Autonomous AI Managing Patient Care. N Engl J Med. 2026 Apr 23;394(16):1561-1563. doi: 10.1056/NEJMp2601148. Epub 2026 Apr 18. PMID: 42011994.

3. **Verbal.** Survey: Nearly 70% of healthcare organizations QA virtual visits once per month or less. Verbal. Published January 11, 2024. Accessed February 1, 2026.
https://www.tryverbal.com/posts/virtual-care-qa-survey

4. **Centers for Medicare & Medicaid Services (CMS).** CMS Rolls Out Aggressive Strategy to Enhance and Accelerate Medicare Advantage Audits. CMS Newsroom. Published May 21, 2025. Accessed January 27, 2026.
https://www.cms.gov/newsroom/press-releases/cms-rolls-out-aggressive-strategy-enhance-and-accelerate-medicare-advantage-audits

5. **Frontline.** The Meeting: Ford and Iacocca Make Their Pitch. Accessed August 2, 2026.
https://www.pbs.org/wgbh/pages/frontline/shows/rollover/nixon/nixon.html

6. **The Heart of Healthcare.** How AI Will Finally Make Healthcare Deflationary. Published May 18, 2026. Accessed May 18, 2026.
https://www.heartofhealthcarepodcast.com/episodes/eric-larsen-healthcare-ai

I. The Mandate

7. **Steptoe.** The Trump Administration Continues to Focus on Health Care Enforcement. Steptoe. Published July 17, 2018. Accessed February 1, 2026.
https://www.steptoe.com/en/news-publications/the-trump-administration-continues-to-focus-on-health-care-enforcement.html

8. **Park J.** NCQA's 2025 Credentialing Guideline Updates: A Compliance Officer's Guide. Ethico. Published January 30, 2025. Accessed April 2, 2026.
https://www.ethico.com/insights/ncqas-2025-credentialing-guideline-updates-a-compliance-officers-guide

9. **National Committee for Quality Assurance (NCQA).** HEDIS Measurement Year 2025: Summary of Changes. NCQA. Accessed February 2026.
https://www.ncqa.org/hedis/

10. **National Committee for Quality Assurance (NCQA).** NCQA's Proposed Timeline for Retiring and Replacing HEDIS Hybrid Measures. NCQA Blog. Published November 15, 2024. Accessed April 2, 2026.
https://www.ncqa.org/blog/ncqas-proposed-timeline-for-retiring-and-replacing-hedis-hybrid-measures

11. **National Committee for Quality Assurance (NCQA).** National Training – 2026 Standards Update. NCQA. Published September 16, 2025. Accessed April 1, 2026. https://www.ncqa.org/videos/national-training-2026-standards-update/

12. **Centers for Medicare & Medicaid Services (CMS).** Medicare and Medicaid Programs; Policy and Technical Changes to the Medicare Advantage Risk Adjustment Data Validation (RADV). Fed Regist. 2023. https://www.federalregister.gov/documents/2023/02/01/2023-01942/medicare-and-medicaid-programs-policy-and-technical-changes-to-the-medicare-advantage-program-risk

13. **Ropes & Gray LLP.** CMS Announces Significant Changes to RADV Auditing Efforts. Ropes & Gray. Published May 2025. Accessed February 12, 2026. https://www.ropesgray.com/en/insights/alerts/2025/05/cms-announces-significant-changes-to-radv-auditing-efforts-considerations-and-next-steps

14. **Milliman.** Federal court vacates 2023 CMS final rule on RADV audits, citing procedural flaws. Milliman. Published September 30, 2025. Accessed February 12, 2026. https://www.milliman.com/en/insight/federal-court-vacates-2023-rule-cms-radv-audits

15. **Groom Law Group.** Court Rules That CMS Cannot Extrapolate Medicare Advantage Risk Adjustment Audit Results. Groom Law Group. Published October 8, 2025. Accessed February 10, 2026. https://www.groom.com/resources/court-rules-that-cms-cannot-extrapolate-medicare-advantage-risk-adjustment-audit-results/

16. **The Joint Commission.** National Performance Goals (NPGs). The Joint Commission. Accessed August 2, 2026. https://www.jointcommission.org/en-us/standards/national-performance-goals

II. The Reality

17. **Song J, Zolnoori M, Scharp D, et al.** Do nurses document all discussions of patient problems and nursing interventions in the electronic health record? JAMIA Open. 2022;5(2):ooac034. doi: 10.1093/jamiaopen/ooac034. PMID: 35663115.

18. **American Association of Colleges of Nursing.** Nursing Shortage Fact Sheet. AACN. Updated May 2024. Accessed February 12, 2026. https://www.aacnnursing.org/news-data/fact-sheets/nursing-shortage

19. **NSI Nursing Solutions.** 2026 NSI National Health Care Retention & RN Staffing Report. NSI Nursing Solutions. Accessed August 1, 2026. https://www.nsinursingsolutions.com/documents/library/nsi_national_health_care_retention_report.pdf

20. **Bae SH.** Association between nurse turnover and missed nursing care in acute care hospitals in South Korea. Front Med (Lausanne). 2024;11:1448839. doi: 10.3389/fmed.2024.1448839. PMID: 39839640.

21. **Lee T, Yoon YS, Ji Y.** Predicting New Graduate Nurses' Retention during Transition Using Decision Tree Methods. J Nurs Manag. 2024;2024:4687000. doi: 10.1155/2024/4687000. PMID: 40224873.

22. **Lippincott Solutions.** Nurse Preceptor Best Practices. Wolters Kluwer. Updated April 24, 2020. Accessed February 1, 2026. https://www.wolterskluwer.com/en/expert-insights/nurse-preceptor-best-practices

23. **Press Ganey.** The state of nursing turnover and key nurse retention strategies. Published November 29, 2023. Accessed August 1, 2026.
https://www.pressganey.com/resources/blog/the-state-of-nursing-turnover-and-key-nurse-retention-strategies/

III. AI Risk & Oversight

24. **Laiouar-Pedari S, Kühn A, Wies C, et al.** Prospective Evidence on Artificial Intelligence-Assisted Melanoma Diagnostics: A Systematic Review and Meta-Analysis. JAMA Dermatol. 2026;162(5):478-487. doi:10.1001/jamadermatol.2026.0217.

25. **Loganathan K, et al.** Current state of FDA-approved artificial intelligence/machine learning medical devices: pathways, transparency, and evidence gaps. J Med Artif Intell. 2026. (Through May 2025, 96.5 percent of roughly 1,016 authorized AI/ML devices cleared via 510(k).)

26. **Institute of Medicine.** Medical Devices and the Public's Health: The FDA 510(k) Clearance Process at 35 Years. Washington, DC: National Academies Press; 2011. doi:10.17226/13150.

27. **Price WN II, Gerke S, Cohen IG.** Liability for use of artificial intelligence in medicine. In: Solaiman B, Cohen IG, eds. Research Handbook on Health, AI and the Law. Cheltenham, UK: Edward Elgar Publishing; 2024. NCBI Bookshelf NBK613216.

28. **European Parliament and Council of the European Union.** Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), as amended by Regulation (EU) 2026/1744. Official Journal of the European Union. Published July 12, 2024; amended July 24, 2026.
https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng; https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng

29. **European Data Protection Supervisor.** Guidance for Risk Management of Artificial Intelligence Systems. EDPS. Published November 11, 2025. Accessed February 13, 2026.
https://www.edps.europa.eu/system/files/2025-11/2025-11-11_ai_risks_management_guidance_en.pdf

30. **Twenter P.** Joint Commission launches certification for responsible healthcare AI use. Becker's Hospital Review. Published June 1, 2026. Accessed June 25, 2026.
https://www.beckershospitalreview.com/quality/joint-commission-launches-certification-for-responsible-healthcare-ai-use/

31. **Gleeson C.** Coalition for Health AI unveils governance playbooks for responsible AI adoption. Fierce Healthcare. Published May 28, 2026. Accessed June 25, 2026.
https://www.fiercehealthcare.com/ai-and-machine-learning/coalition-health-ai-unveils-8-governance-playbooks-health-systems

32. **URAC.** Artificial Intelligence in Health Care Accreditation Standards. URAC. September 2025.
https://www.urac.org/

33. **Stern AD, Goldfarb A, Minssen T, Price WN II.** AI Insurance: How Liability Insurance Can Drive the Responsible Adoption of Artificial Intelligence in Health Care. NEJM Catalyst Innovations in Care Delivery. 2022;3(4). doi: 10.1056/CAT.21.0242

34. **Jargon J, Schechner S.** Seven Lawsuits Allege OpenAI Encouraged Suicide and Harmful Delusions. Wall Street Journal. Published November 6, 2025.
https://www.wsj.com/tech/ai/seven-lawsuits-allege-openai-encouraged-suicide-and-harmful-delusions-25def1a3

35. **Chen S, Gao M, Sasse K, et al.** When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. npj Digit Med. 2025;8:605. doi: 10.1038/s41746-025-02008-z.

36. **Weber S.** Health System Sued Over AI Scribe Technology, Patient Consents. Medscape. Published January 16, 2026. Accessed February 10, 2026. https://www.medscape.com/viewarticle/health-system-sued-over-ai-scribe-technology-patient-consent-2026a10001k7

37. **Tsui K.** Self-Correction Bench: Uncovering and Addressing the Self-Correction Blind Spot in Large Language Models. arXiv. Published July 3, 2025. Accessed February 13, 2026. https://arxiv.org/abs/2507.02778

38. **Huang J, Chen X, et al.** Large Language Models Cannot Self-Correct Reasoning Yet. arXiv. 2023. https://arxiv.org/abs/2310.01798

39. **Asgari E, Montaña-Brown N, Dubois M, et al.** A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. npj Digit Med. 2025;8(1):274. doi: 10.1038/s41746-025-01670-7. PMID: 40360677.

40. **Taylor SL, Jost M, MacDonald S, et al.** Quality of Clinical Notes Created by Ambient Listening Generative AI: Pragmatic Prospective Pilot Study. JMIR Med Inform. 2026;14:e86474. doi: 10.2196/86474. PMID: 41996389

IV. Financial Impact

41. **American Hospital Association.** The Cost of Caring: Challenges Facing America's Hospitals in 2025. AHA. Published April 2025. Accessed February 13, 2026. https://www.aha.org/costsofcaring

42. **Vizient.** New Margin Math: 2026 Trends Report. Vizient. Published January 22, 2026. Accessed February 13, 2026. https:// www.vizientinc.com/insights/reports/annual-trends-and-forecasting-reports/2026-trends-report

43. **Centers for Medicare & Medicaid Services.** Payment Year 2020 Medicare Advantage Contract-Specific RADV Audit Methods and Instructions. 2026. https://www.cms.gov/files/document/payment-year-2020-ma-radv-audit-methods-instructions.pdf

44. **O'Melveny & Myers LLP.** False Claims Act Enforcement Risks for Companies Using AI. O'Melveny. Published March 13, 2026. Accessed April 2, 2026. https://www.omm.com/insights/alerts-publications/false-claims-act-enforcement-risks-for-companies-using-ai/

45. **Whistleblower Law Collaborative.** False Claims Act Penalties: A Complete Guide. Whistleblower Law Collaborative. Published July 8, 2025. Accessed March 10, 2026. https://www.whistleblowerllc.com/false-claims-act-penalties/

46. **U.S. Department of Health and Human Services, Office of Inspector General**. Fraud & Abuse Laws. HHS OIG. Accessed March 10, 2026. https://oig.hhs.gov/compliance/physician-education/fraud-abuse-laws/

47. **California Civil Code §1714.46, enacted by Chapter 672, Statues of 2025s.** https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB316

48. **FCA Counsel, Bracker and Marcus LLC.** Record False Claims Act Recoveries Highlight the Power of Qui Tam Whistleblowers. FCA Counsel. Accessed March 14, 2026. https://www.fcacounsel.com/blog/false-claims-act-recoveries-qui-tam-whistleblowers/

49. **Sidley Austin LLP.** FY 2025 FCA Settlements and Judgments Statistics Show Highest Recoveries Ever. FCA Blog. Published January 20, 2026. Accessed May 16, 2026. https://fcablog.sidley.com/2026/01/20/fy-2025-fca-settlements-and-judgments-statistics-show-highest-recoveries-ever/

50. **U.S. Department of Justice, Office of Public Affairs**. False Claims Act Settlements and Judgments Exceed $6.8B in Fiscal Year 2025. DOJ. Published January 16, 2026. Accessed March 14, 2026. https://www.justice.gov/opa/pr/false-claims-act-settlements-and-judgments-exceed-68b-fiscal-year-2025

51. **American Hospital Association (AHA).** Regulatory Overload: Assessing the Regulatory Burden on Health Systems,Hospitals and Post-Acute Care Providers. AHA. Published October 2017. Accessed February 16, 2026. https://www.aha.org/system/files/2018-02/regulatory-overload-report.pdf

52. **WCH Service Bureau.** Medical Chart Auditing. WCH Service Bureau. Accessed February 17, 2026. https://wchsb.com/medical-auditing/

53. **Practice Performance Partners.** Medical Chart Audit. Practice Performance Partners. Accessed February 17, 2026. https:// www.practiceperformancepartners.com/our-services/medical-chart-audit/

54. **American Hospital Association.** Hospitals Can Generate Savings Through Targeted Claims Management Automation. 2025; citing the 2024 CAQH Index Report. https://www.aha.org/system/files/media/file/2025/10/Trailblazers_Ailevate_ClaimsDenials.pdf

55. **Yen PY, Kellye M, Lopetegui M, et al.** Nurses' Time Allocation and Multitasking of Nursing Activities: A Time Motion Study. AMIA Annu Symp Proc. 2018;2018:1137-1146. PMID: 30815156.

56. **Gesner E, Dykes PC, Zhang L, Gazarian P.** Documentation Burden in Nursing and Its Role in Clinician Burnout Syndrome. Appl Clin Inform. 2022;13(5):983–990. doi:10.1055/s-0042-1757157. PMID: 36261113.

V. Recommendations

57. **Chassin MR, Loeb JM.** High-reliability health care: getting there from here. Milbank Q. 2013;91(3):459-490. doi: 10.1111/1468-0009.12023. PMID: 24028696.

58. **U.S. Department of Health and Human Services, Office of Inspector General**. Healthcare Fraud Prevention and Enforcement Action Team (HEAT): Compliance Program Tips. HHS OIG. Accessed March 12, 2026. https://oig.hhs.gov/documents/provider-compliance-training/945/Compliance101tips508.pdf

59. **Show us the evidence for the value of medical AI [editorial].** Nat Med. 2026;32:1163. doi: 10.1038/s41591-026-04389-4

60. **National Academy of Medicine.** 2025. An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action. Washington, DC: The National Academies Press. https://doi.org/10.17226/29087

61. **Siwicki B.** Mental health provider drastically reduces chart audit time with help from AI. Healthcare IT News. Published August 16, 2024. Accessed February 12, 2026. https://www.healthcareitnews.com/news/mental-health-provider-drastically-reduces-chart-audit-time-help-ai

62. **Kodiak Solutions.** Rate of initial denials of medical insurance claims continued to rise in 2024. BusinessWire. Published May 21, 2025. Accessed May 16, 2026. https://www.businesswire.com/news/home/20250521892947/en/Rate-of-initial-denials-of-medical-insurance-claims-continued-to-rise-in-2024-Kodiak-Solutions-proprietary-data-show

63. **Suh HS, Tully JL, Meineke MN, Waterman RS, Gabriel RA.** Identification of Preanesthetic History Elements by a Natural Language Processing Engine. Anesth Analg. 2022;135(6):1162–1171. doi:10.1213/ANE.0000000000006152. PMID: 35841317.

64. **Verbal.** Data sourced from observed pilot deployment. Care management figures based on MAPD/EGWP deployment (200 CMs). CM compensation benchmarks: Glassdoor/ZipRecruiter/Indeed (~$95K base, $123.5K fully loaded). Methodology note: time recaptured represents capacity returned to member-facing work, not headcount reduction.

65. **Nayak A, et al**. Use of Voice-Based Conversational Artificial Intelligence for Basal Insulin Prescription Management Among Patients With Type 2 Diabetes: A Randomized Clinical Trial. JAMA Netw Open. 2023;6(12):e2340232. doi:10.1001/jamanetworkopen.2023.40232. PMID: 38039007.

Contact

Council for Healthcare AI Responsibility & Safety (CHAIRS)
hello@chairs.health
www.chairs.health

Verbal AI Technologies
hello@tryverbal.com
www.tryverbal.com