IDEM LABS

Activated, Not Audited: AI Summarization in the Legal Medical Record.

The Risks of Clinical AI Summaries and the Case for 3rd-Party Evaluation

Executive Summary

By the end of 2025, roughly 50% of US non-federal acute care hospitals were using Electronic Healthcare Record (EHR)-integrated generative AI in clinical care.[1] Epic software's chart summarization feature alone is invoked over **16 million times per month** across its customer base, a near-3x increase since November 2025.[2] Two-thirds of Epic providers have used a generative AI feature. In contrast to the previous era of healthcare technology acquisition, **70 to 80 percent** of AI deployments arrive as default activations of features that ship with the EHR contract, not as discrete procurements that passed through formal vendor selection.[3]

Systematic evaluation of these tools for hallucination rate, omission rate, accuracy, and demographic equity are essentially absent across the five most-deployed vendors in this category (Epic, Oracle Health, Meditech, Abridge Inside, Microsoft Dragon Copilot). Where evaluation does exist, it is benchmarked to the wrong standard. The dominant metrics were built to score news summarization, where the cost of an error is a clumsy sentence; they now serve as the safety check for tools that move medications and diagnoses through the legal medical record millions of times a month. A tool is validated against its highest-stakes use, not its most forgiving one, and by that measure **the metrics in common use were never built to detect the failures that matter in clinical care.**

Providers are beginning to see the risk these systems pose. In May 2026, Ontario's Auditor General audited 20 AI scribe vendors approved for use by Ontario doctors. **All 20 had inaccuracies**: 9 hallucinated content, 12 captured a different drug than the one prescribed, and 17 missed important mental-health details.[4]

The combined picture is one of systematic under-evaluation. The tools with the broadest reach into patient care are the least rigorously assessed; the vendors with the most rigorous evaluation posture operate at far smaller scales; and the measurement instruments in common use were not designed to capture clinically meaningful hallucinations and omissions across heterogeneous real-world clinical environments. Current benchmarks remain too narrow, too curated, and too disconnected from production deployment to support confident claims about safety, reliability, or equity at national scale. As a result, healthcare organizations are deploying generative systems into high-stakes workflows under conditions where **neither vendors nor**

health systems can reliably quantify how these models fail across diverse patient populations and clinical settings.

We assemble the evidence behind this gap and lay out five underestimated risks health systems are carrying on AI summarization tools, finding that **real regulatory and malpractice liability falls on the health system rather than the vendor.** We then propose three sequential actions built around independent auditing: inventory every summarization feature in production, commission a principled and scalable independent evaluation of how each performs, and close the legal and governance loop that evaluation opens. Idem Labs, the independent clinical-AI validation and audit firm behind this paper, specializes in these solutions for health systems.

1. Why This Paper Exists

The pre-deployment validation conversation in clinical AI has been thorough. The Coalition for Health AI's Applied Model Card framework, the Joint Commission and CHAI's seven-element guidance from September 2025,[5] Mayo Clinic Platform's Validate program, Stanford HAI's MedHELM framework, Duke-Margolis's governance work, and Kaiser Permanente's AIM-HI grant program have collectively built out what "validate before you deploy" should look like for a clinical AI tool.

Most AI summarization tools today are **activated rather than deployed** and do not follow this guidance**.** Tools ship with the EHR contract, get turned on at build or upgrade, and run inside the legal medical record without passing through the procurement gate that triggers the pre-deployment validation workflow. The validation conversation does not address them because it assumes the tool was procured, a condition that no longer holds for the majority of in-production AI summarization.

The post-deployment audit conversation has been thinner, and **no independent commercial offering today audits the AI summarization features a health system has already turned on.** The same activation that bypasses the procurement gate bypasses the scrutiny that gate carried, and nothing downstream replaces it. Whatever evaluation happens cannot come from the vendor, which has no commercial interest in surfacing the failure modes of a feature it sells; a rigorous evaluation that finds problems is a liability to the product, not an investment in it. Independence is therefore a requirement rather than a preference: only a third party with no stake in the outcome produces findings that carry weight in malpractice discovery, with carriers, and at the board level. The party that profits from the tool being trusted cannot be the party that certifies it is trustworthy.

2. The Deployment Landscape

2.1 The Two In-Scope Use Cases

This paper covers two applications of AI to medical-record summarization. The first is **chart summarization for visit prep and in-workflow review:** pre-visit chart summaries, problem-list synthesis, hover-summaries inside the EHR, in-basket message summaries, and encounter-context summaries surfaced during the visit. The second is **inpatient handoff and discharge summaries:** auto-drafted sign-out notes, ICU-to-floor and ED-to-floor handoffs, discharge summaries, and transition-of-care documents.

2.2 Where the Tools Sit

Adoption concentrates in providers using Epic. KLAS Research reports Epic is used in **43.7% of acute hospitals** and **56.9% of beds** at the start of 2026.[11] Within Epic's customer base, about two-thirds of providers have used at least one generative AI feature, and Epic's chart-prep summarizer is invoked over 16 million times per month across that base. Epic's generative AI tool ART runs in 150+ health systems and generates about 1 million message drafts per month.[12] Oracle Health, Meditech, and athenahealth each have in-scope summarization features in production.

2.3 The Procurement-Versus-Default Split

The American Hospital Association's 2024 IT Supplement reports that **80% of US hospitals using AI use EHR-vendor-supplied models.**[18] Most of those features did not pass through a discrete vendor selection. athenahealth distributes its AI features at no incremental cost via routine updates. Meditech bundles generative AI into the Expanse subscription. Epic ships Art Insights, In Basket ART, and AI Charting inside standard Epic contracts, activated via the system's build-and-governance process rather than a separate purchase decision.

When KLAS asks where organizations are with AI, most are piloting but not yet enterprise-scaling, with governance committees, return on investment (ROI) validation, and workflow integration as the gating factors.[19] **The implication for health-system leadership is that the effective "buyer" for these features is increasingly the AI governance committee** that decides which already-paid-for features to activate, in what order, for which specialties. The pre-deployment validation literature was built for a procurement gate. **The governance committee is operating under a different motion entirely, and the literature has not caught up.**

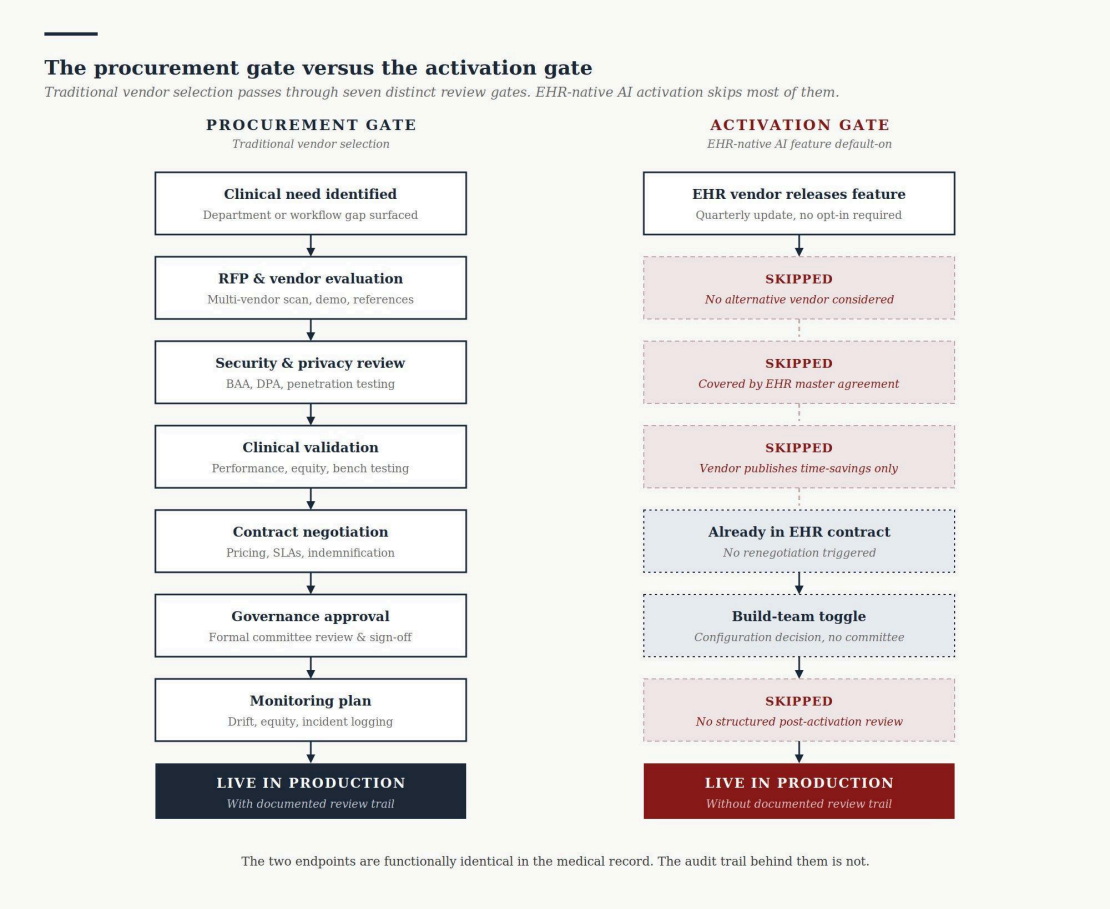

Figure 1. *The procurement gate versus the activation gate. Traditional vendor selection passes through seven distinct review gates. Clinical AI activation skips most of them, leading to a sparser validation process.*

3. The Evaluation Gap

3.1 How the Most-Deployed Vendors Evaluate Their Tools

Epic has not conducted or disclosed independent evaluation of hallucination rate, omission rate, accuracy, or demographic equity for its generative AI summary features. The performance data available in the public literature comes from third-party academic studies: one ambulatory implementation study reported a per-visit time savings of 18 seconds on Epic Signal.[20] (A randomized trial protocol is registered but data are not yet published.)[21] Neither represents a vendor-conducted evaluation of the summarization features themselves. Epic's open-source AI Trust and Assurance suite is the strongest infrastructure-transparency offering in the market, enabling health systems to run local validation,[22] though it shifts the burden of evaluation to the customer rather than reflecting any internal assessment Epic has run. Across the other

most-deployed vendors, Oracle Health,[23] Meditech, and athenahealth each report **workflow efficiency metrics with no evidence of systematic evaluation of output quality, error rates, or equity**.

3.2 How Smaller Clinical-Academic Vendors Evaluate Their Tools

Smaller clinical-academic vendors have invested more in evaluation. Layer Health evaluates against manual abstraction benchmarks in academic-style work.[24] Carta Healthcare uses inter-rater reliability as an audit metric for its AI clinical data abstraction product.[25] Glass Health evaluates against peer-reviewed differential diagnosis benchmarks, and Regard Health has published peer-reviewed evaluation work on diagnosis-finding accuracy. Relative to the EHR-native vendors, this reflects a genuine evaluation posture. Yet, even concerted efforts to benchmark these AI summarization tools run into critical methodological limitations.

3.3 Existing Evaluation Methodologies are Incompatible with EHR AI

The uncomfortable reality is that the traditional Machine Learning evaluation approaches are not fit for EHR AI summarization. The dominant automated metrics for summarization quality, ROUGE and BLEU, were developed primarily for **news article summarization** and measure surface-level n-gram overlap rather than clinical meaning and are entirely insufficient according to *Frontiers in Digital Health*.[37] A systematic review in *npj Health Systems* reached the same conclusion: evaluation of clinical LLM summarization must address hallucinations, omissions, and factual accuracy in ways that standard natural language processing (NLP) metrics do not capture.[38] Human evaluation closes address some of these issues, but the studies that exist rely on curated datasets and small annotator panels that do not generalize to production EHR environments as they cannot scale to millions of summaries a month and cannot be implemented with every model update. **A benchmark built on a few hundred samples from a single academic center does not necessarily generalize to systems deployed nationally that are invoked millions of times a month** spanning community hospitals, safety-net systems, and tertiary care centers with structurally different patient populations and documentation cultures.

The combined picture is one of systematic under-evaluation. The tools with the broadest reach into patient care are the least rigorously assessed; the vendors with the most rigorous evaluation posture operate at far smaller scales; and the measurement instruments in common use were not designed to capture clinically meaningful hallucinations and omissions across heterogeneous real-world environments. Current benchmarks remain too narrow, too curated, and too disconnected from production deployment to support confident claims about safety, reliability, or equity at national scale. As a result, healthcare organizations are deploying generative systems into high-stakes workflows under conditions where **neither vendors nor health systems can reliably quantify how these models fail** across diverse patient populations and clinical settings.

4. The Ontario Audit: Under-Evaluation as a Patient Safety Failure

In May 2026, Ontario's Auditor General published an audit of the 20 AI scribe vendors approved for use by Ontario doctors.[4] **All 20 vendors tools had inaccuracies:** nine hallucinated content (one fabricated a referral to therapy, another ordered blood tests the physician did not request), 12 captured a different drug than the one prescribed, and 17 missed important mental-health details.

The audit is the closest publicly available approximation of what an independent third-party evaluator finds when it audits AI in the medical record. The vendors named had been approved for clinical use by provider systems (i.e., they were not flagged as fringe or unvetted) and were part of the recommended product list by the government.

The Ontario findings are not a problem unique to Canadian healthcare. Walk the same failure mode into a US hospital running a clinical AI summary feature for chart prep or discharge summary, and the problems are identical: a discharge summary auto-drafted by the AI surfaces a referral the physician did not order or transposes a discharge medication; the summary is authenticated by an exhausted hospitalist who does not catch the substitution; the legal medical record now contains AI-generated content that does not match the clinical intent; and the audit trail does not distinguish AI-drafted content from clinician-edited content with enough granularity to reconstruct what happened.

The risk here is that discovery in a malpractice suit would surfaces the absence of Model Card documentation for the feature, no documented governance committee minutes, no audit-trail granularity, no equity testing, and no formal HTI-1 predictive DSI compliance review that charts where in the EHR data summarization facts came from.

In February 2026, OIG's Medicare Advantage guidance explicitly named AI-generated coding prompts as a risk-adjustment abuse vector,[26] and Kaiser Permanente's **$556M settlement in January 2026** — the largest Medicare Advantage risk-adjustment False Claims Act settlement in history — was specifically related to the use of automated mechanisms for finding unsubmitted diagnoses.[27] The inference chain from automated chart synthesis to billing integrity is the pattern OIG and Department of Justice are now policing. The direction is clear, **unaudited AI summarization represents clear liability risk.**

5. The Regulatory Environment

Three regulatory threads bear directly on clinical AI summarization: Office of the National Coordinator for Health Information Technology's (ONC) Health Data, Technology, and Interoperability (HTI)-1 predictive DSI rule, the FDA's CDS exemption framework; and a growing patchwork of state law is becoming the operative governance floor regardless of federal action. **Taken together, the picture is of unsettled and weakening federal oversight against a rising state and accreditation baseline with meaningful liability exposure to providers and carriers at each layer.**

5.1 ONC HTI-1 Predictive DSI Applicability

ONC's HTI-1 Final Rule (effective January 1, 2025) requires developers of certified Health IT Modules to publish **31 source attributes** for any Predictive Decision Support Intervention.[28] A chart summary that selects, condenses, and synthesizes patient data into clinically actionable content is, on the most defensible reading, a predictive DSI; ONC's own presentation and multiple law-firm analyses converge on that position,[29] and no EHR vendor has publicly argued otherwise. Epic, Oracle Health, and Meditech all face the obligation but compliance in the public record is uneven. The rule is in flux: HTI-5 (proposed December 2025 by ONC) would eliminate the predictive DSI source attributes and model card requirements entirely.[8] Until HTI-5 is finalized, HTI-1 still applies. Even if it is enacted, documented vendor compliance will remain useful evidence of due diligence.

5.2 FDA CDS Exemption Defensibility

The FDA's four-criterion test for whether a summarization tool is a regulated device turns on **Criterion 4:** whether the clinician can independently review the basis of the output.[30] A summary that synthesizes across notes without showing its sources fails that criterion. The January 2026 revised guidance sharpens the requirement, specifically naming transparency of logic, traceability, and disclosure of validation and limitations.[31] Most clinical AI summarization features do not meet that bar. The FDA has not yet taken enforcement action against a clinical AI summarization product, though its April 2026 warning letter on AI in drug manufacturing signals that **human review and validation are non-delegable** in regulated contexts.[32]

5.3 State Law and the Joint Commission

State law is the binding constraint today. California AB 3030 mandates clinician review or disclosure for generative AI patient communications, with penalties up to **$25,000 per violation.**[33] Colorado SB 24-205 reaches systems that are a "substantial factor" in consequential healthcare decisions.[34] Texas, Utah, and Tennessee have enacted or proposed analogous requirements to Colorado. Alongside these, Joint Commission and CHAI's September 2025 seven-element framework covering governance, validation, equity, and incident reporting is becoming the operational bar most US hospitals will adopt, regardless of whether accreditation eventually requires it.[5]

6. Five Underestimated Risks

The combined deployment, evaluation, and regulatory evidence assembles into a structured risk inventory. Each risk below applies to any clinical AI summarization feature running in a US hospital today.

1. Evaluation capability is absent. The most-deployed summarization features have not been evaluated for hallucination rate, omission rate, or demographic equity at the feature level, and the tools to do so at production scale do not exist in most health system environments. The Ontario audit found inaccuracies in all 20 vendors it assessed.[4] Furthermore, existing methodologies are categorically ill-suited for evaluating EHR AI summaries.

2. Audit trail integrity is undocumented. AI-generated content reviewed and signed by a clinician inherits clinician authorship, which is the mechanism by which an undetected AI error becomes a clinician-attributed documentation failure. Unedited insertion without review is a documentation integrity violation under 42 CFR 482.24. Most health systems have not confirmed which of these conditions applies to each in-scope feature in their environment nor do they provide documentation tooling.

3. Regulatory exposure is real and immediate. HTI-1's 31 predictive DSI source attributes still apply through 2026 enforcement-discretion windows,[28] and most health systems have not confirmed in writing what their EHR vendor has published for each in-scope feature. HTI-5 may rescind the requirement,[8] but the documentation remains useful evidence of due diligence in the context of malpractice discovery.

4. AI summarization is becoming a malpractice target. A malpractice claim tied to an AI-generated chart summary is no longer a theoretical risk, and the exposure falls almost entirely on health systems. AI-generated content sits in the legal medical record with clinician authentication is the only documented step between an AI error and patient harm. Audit trail granularity in most EHR environments is insufficient to reconstruct what the AI produced versus what the clinician modified. Evaluation documentation such as model cards, governance committee minutes, equity testing, HTI-1 compliance review is absent for most in-scope features, so plaintiff discovery finds not just the error but the absence of any systematic effort to prevent it. Health systems that cannot produce evaluation documentation of hallucination rates and AI usage at discovery are carrying this exposure alone.

5. Incentive structures work against rigorous evaluation. Vendors have no commercial interest in surfacing failure modes in tools they sell, as a rigorous independent evaluation that finds problems is a business liability. Clinicians face a parallel misalignment: authentication workflows reward throughput, and there is no direct feedback loop when an AI error propagates through a signed note into the downstream record. The result is inertia in systematic under-scrutiny at both feature development and implementation, with the evaluation burden falling on health systems that have neither the methodology nor the incentive to pick it up.

7. What to Do About It

The risks above resolve into three sequential actions for a health system's AI governance committee. First, take inventory of every AI summarization feature running in the environment. Second, commission an independent evaluation of how those features perform, which is only credible once that inventory is complete. Third, close the legal and governance loop, which is only actionable once the evaluation exists.

7.1 Know What You Are Running

The first step is establishing the baseline picture, which most health systems do not have. For every AI summarization feature to live in the environment, the governance committee needs to document the use case, vendor and product version, activation date and decision path, specialties and departments in production, and monthly invocation volume. Against each feature, the committee should document the vendor's **HTI-1 source-attribute disclosure** mapped to the 31 attributes with gaps identified, the vendor's FDA CDS positioning on Criterion 4, and whether AI-generated content carries a model identifier and version preserved over the record retention window. Where the vendor has not addressed any of these, a written question routed through GC establishes the record.

Most health systems will find at this step that they have more in-scope features running than the governance committee has formally reviewed, and that vendor documentation on the questions that matter is thinner than expected. That is the baseline finding and the foundation for everything that follows.

7.2 Commission a Principled, Scalable, Independent Evaluation

Independent third-party evaluation is a prerequisite for credible governance**.** A third party with no stake in the findings and specific expertise in clinical AI summarization evaluation is the only evaluator whose conclusions carry weight in malpractice discovery, carrier conversations, and board-level review.

A rigorous evaluation centers on systematic estimation of **hallucination and omission rates** in the clinical AI summarization domain, using a clinician-validated reference panel drawn from the health system's own patient population rather than a curated benchmark dataset. **Demographic subgroup performance** should be tested systematically across age, sex, race, ethnicity, and preferred language, with sample sizes sufficient to detect clinically meaningful differences rather than provide token equity coverage. Audit trail integrity and contract and indemnification terms should be reviewed as part of the same engagement, with findings tied directly to the risk profile the performance evaluation surfaces.

A key advantage of designing the evaluation for scalability from the outset is that it can be re-run on a defined cadence without rebuilding from scratch, which transforms a one-time governance exercise into an ongoing monitoring capability. Evaluations designed for a single

snapshot and not for periodic re-execution tend to atrophy into documentation artifacts rather than active risk management instruments.

7.3 Close the Legal and Governance Loop

Once the evaluation exists, three things should happen in parallel. The documentation should be routed to the **malpractice carrier and General Counsel (GC):** carrier knowledge of AI use in the legal record unlocks coverage clarity at renewal, and GC review unlocks the vendor contract conversation at the next anniversary. Ongoing monitoring should be stood up against the same methodology used in the evaluation, with random-sample quality review, drift detection, periodic equity re-test, and incident logging reviewed by the governance committee on a defined cadence. Additionally, the GC should define the **activation gate,** which is the process by which the governance committee reviews and approves each new AI feature before it goes live. The activation gate should require a written vendor disclosure before launch, governance committee sign-off, a clinical specialty review for first-deploy cohorts, and an explicit retirement pathway if monitoring surfaces issues. **The activation gate replaces the procurement gate for the era of default-on AI, shifting the system's center of gravity from passive receipt to active authorization.**

---

8. About Idem Labs

Idem Labs is a specialty consultancy focused on independent validation and audit of clinical AI tools founded in 2026 by Matt Allison and Josh Ward, PhD. Idem provides pre-deployment validation for AI tool procurements, post-deployment audit for AI features already in production (including the clinical AI summarization features this paper covers), and governance framework support for health systems.

Idem operates in alignment with the Coalition for Health AI's Assurance Standards Guide, the Joint Commission and CHAI's September 2025 guidance, and the NIST AI Risk Management Framework. We are not a standards body and do not claim to set the bar; we operate alongside emerging standards. Three reference engagements are open at a discount in exchange for case-study rights. For a scoping conversation, contact matt@idemlabs.ai.

Methodology Note

This paper synthesizes research and competitive scanning conducted in May 2026. Sources include peer-reviewed literature, regulatory filings, vendor disclosures, trade press, and analyst reports. Primary research (buyer-side interviews) is in process and will inform a second-version paper later in 2026.

No EHR vendor was a paying party for this research. No health system was named in any specific risk finding without public attribution. Anonymized aggregate findings from Idem's engagement work may inform future versions of this paper subject to client review and approval.

References

1. Apathy NC, Holmgren AJ, Cross DA, et al. Uptake of Generative AI Integrated With Electronic Health Records in US Hospitals. JAMA Network Open. 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC12701511/
2. Epic Systems. AI Charting Rolls Out Alongside an Expanding Set of Built-in AI Capabilities. Epic, 2025. https://www.epic.com/epic/post/epic-ai-charting-rolls-out-alongside-an-expanding-set-of-built-in-ai-capabilities/
3. Idem Labs analysis of AHA 2024 IT Supplement (Apathy et al. 2025), Menlo Ventures 2025 State of AI in Healthcare, and KLAS Healthcare AI Update 2025. See methodology note.
4. Office of the Auditor General of Ontario. Audit of Medical AI Transcription Vendors. May 2026. https://www.cbc.ca/news/canada/toronto/ai-scribe-system-hallucinations-9.7197049
5. The Joint Commission and Coalition for Health AI. Responsible Use of AI in Healthcare: Initial Guidance. September 17, 2025. https://www.jointcommission.org/en-us/knowledge-library/news/2025-09-jc-and-chai-release-initial-guidance-to-support-responsible-ai-adoption
6. US Food and Drug Administration. Clinical Decision Support Software: Guidance for Industry and FDA Staff. September 2022. https://www.fda.gov/media/162345/download
7. Office of the National Coordinator for Health IT. HTI-1 Final Rule. January 9, 2024. https://healthit.gov/regulations/hti-rules/hti-1-final-rule/
8. Office of the National Coordinator for Health IT. HTI-5 Proposed Rule Fact Sheet. December 22, 2025. https://www.healthit.gov/topic/laws-regulation-and-policy/hti-5-proposed-rule-fact-sheet
9. Bruce G. CHAI's AI oversight ambitions falter with scrapped AI labs. Fierce Healthcare. 2025. https://www.fiercehealthcare.com/ai-and-machine-learning/inside-chais-failed-assurance-labs
10. California Legislature. Assembly Bill 3030. Effective January 1, 2025. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB3030
11. KLAS Research. Acute Care EHR Market Share 2026, as reported via healthsystemcio.com, May 14, 2026. https://healthsystemcio.com/2026/05/14/acute-care-ehr-market-share-2026/

12. Epic Systems. Epic User Group Meeting 2025 deployment statistics, as reported via Healthcare IT News. https://www.healthcareitnews.com/news/epic-unveils-ai-agents-showcases-new-foundational-models
18. American Hospital Association. 2024 IT Supplement Survey of Hospitals. Published 2025.
19. KLAS Research. Healthcare AI Update 2025, as reported via Becker's Hospital Review. https://www.beckershospitalreview.com/healthcare-information-technology/innovation/where-healthcare-is-really-using-ai-in-2025-klas/
20. AI-generated encounter summaries: early insights. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC12403210/
21. medRxiv registered protocol for Epic chart summary randomized trial. https://www.medrxiv.org/content/10.64898/2026.02.20.26346503v1.full
22. Bruce G. Epic releases AI validation software for health systems. Fierce Healthcare. https://www.fiercehealthcare.com/ai-and-machine-learning/epic-releases-ai-validation-software-health-systems
23. Oracle Health press release on Clinical AI Agent documentation time reduction. March 2025. https://www.oracle.com/news/announcement/physicians-reduce-documentation-time-with-oracle-health-clinical-ai-agent-2025-03-04/
24. Layer Health Series A announcement. March 2025. https://www.layerhealth.com/resources/layer-health-raises-series-a
25. Carta Healthcare. AI Platform with Anthropic Claude. https://www.carta.healthcare/clinical-data-management/ai-platform/
26. Office of Inspector General. Medicare Advantage Risk Adjustment Guidance. February 2026.
27. Kaiser Permanente Medicare Advantage False Claims Act Settlement, January 2026, as reported via White & Case. https://www.whitecase.com/insight-our-thinking/healthcare-fraud-enforcement-2025-year-aggressive-action-expanding-risk
28. Office of the National Coordinator for Health IT. HTI-1 DSI Final Rule Presentation. January 2024. https://www.healthit.gov/sites/default/files/page/2024-01/DSI_HTI1%20Final%20Rule%20Presentation_508.pdf
29. Mintz Levin. HHS-ONC HTI-1 Final Rule Introduces New Transparency. January 2024. https://www.mintz.com/insights-center/viewpoints/2146/2024-01-08-hhs-onc-hti-1-final-rule-introduces-new-transparency
30. Covington. Five Key Takeaways from FDA's Revised CDS Guidance. January 2026. https://www.cov.com/news-and-insights/insights/2026/01/5-key-takeaways-from-fdas-revised-clinical-decision-support-software-guidance
31. FDA. Revised Clinical Decision Support Guidance. January 2026.
32. DLA Piper. FDA Warning Letter on AI in Drug Manufacturing. April 2026. https://www.dlapiper.com/en-us/insights/publications/2026/04/fda-warning-letter-highlights-risks-of-using-ai-in-drug-manufacturing
33. Morgan Lewis. California Law Requiring Disclaimers by Healthcare Providers Using GenAI. https://www.morganlewis.com/pubs/2024/12/california-law-requiring-disclaimers-by-healthcare-providers-using-genai-will-affect-providers-and-genai-developers
34. Colorado General Assembly. SB 24-205. https://leg.colorado.gov/bills/sb24-205
35. The Doctors Company. Public statements on AI and professional liability coverage, as reported via ABA Journal. https://www.americanbar.org/groups/journal/articles/2025/does-your-professional-liability-insurance-cover-ai-mistakes-dont-be-so-sure/
36. ProAssurance. ProVisions May 2025 issue. https://agents.proassurance.com/provisions/may25

37. Frontiers in Digital Health. MEDAI-LLM-SUMM: A reporting checklist for medical text summarization studies using large language models. January 31, 2026. https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2026.1761601/full

38. npj Health Systems. Current and future state of evaluation of large language models for medical summarization tasks. February 3, 2025. https://www.nature.com/articles/s44401-024-00011-2